Reconstructing Auditory Evoked Cortical Response with Fewer Trials

Zoey Y. Zhang
San Jose State University

Follow this and additional works at: https://scholarworks.sjsu.edu/etd_theses

Recommended Citation
https://scholarworks.sjsu.edu/etd_theses/5488

This Thesis is brought to you for free and open access by the Master's Theses and Graduate Research at SJSU ScholarWorks. It has been accepted for inclusion in Master's Theses by an authorized administrator of SJSU ScholarWorks. For more information, please contact scholarworks@sjsu.edu.
RECONSTRUCTING AUDITORY EVOKED CORTICAL RESPONSES WITH FEWER TRIALS

A Thesis
Presented to
The Faculty of the Department of Biomedical Engineering
San José State University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science

by
Zoey Y. Zhang
December 2023
The Designated Thesis Committee Approves the Thesis Titled

RECONSTRUCTING AUDITORY EVOKED CORTICAL RESPONSES WITH FEWER TRIALS

by

Zoey Y. Zhang

APPROVED FOR THE DEPARTMENT OF BIOMEDICAL ENGINEERING

SAN JOSÉ STATE UNIVERSITY

December 2023

Folarin Erogbogbo, Ph.D.  Department of Biomedical Engineering

Carly Demopoulos, Ph.D.  Department of Radiology and Biomedical Sciences, Department of Psychiatry and Behavioral Sciences
                        University of California, San Francisco

Srikantan Nagarajan, Ph.D.  Department of Radiology and Biomedical Imaging
                           University of California, San Francisco

Alessandro Bellofiore, Ph.D.  Department of Biomedical Engineering
ABSTRACT

RECONSTRUCTING AUDITORY EVOKED CORTICAL RESPONSES WITH FEWER TRIALS

by Zoey Y. Zhang

Auditory evoked fields (AEF) paradigm is a common research tool to study human auditory responses by using Magnetoencephalography (MEG), a neuroimaging tool. AEF paradigm requires repetition over many trials to achieve adequate signal-to-noise ratio (SNR), but Autistic children and some other populations cannot tolerate prolonged exam time due to various reasons. To address these challenges, this project uses a novel machine learning algorithm, Champagne with baseline noise learning, to reconstruct AEF data with fewer trials (80%, 60%, 40%, 20%) but produce the same results as using all AEF trials (100%). The results show that this novel machine learning algorithm can produce reliable latency results with only 60% of all AEF trials.
DEDICATION

This thesis is dedicated to my grandma Yulan, who passed away on June 26, 2006, and my previous colleague, Dr. Irwin Feinberg, who passed away on August 25, 2022.

Thank you for sharing your glorious adventures with me.
ACKNOWLEDGEMENTS

Foremost, I would want to thank my parents who have spent all their resources to raise me and support me along the way. I am always a difficult child with various complicated health conditions, but my parents never thought about giving me up. Without my parents, I probably would not be alive today. I also appreciate the efforts of my health care team and any health care workers who had participated in treating/rescuing me. Thank you for keeping me alive and functional.

I would like to thank my mentors: Dr. Neal Fleming and Dr. Ian Campbell. Both Neal and Ian spent tremendous time and resources train me to be an expert in electrophysiology. They also try their best to support my passion and some absurd research ideas unconditionally along the way. Whenever I have had any concerns or questions about my health, my life, and my career, Neal and Ian always try their best to help and support me. I would also want to thank many other researchers who had trained me along the way. These people helped me build such strong work ethic. Without these skills, I would never have these many achievements in the academia.

I am also deeply grateful to my current boss, Dr. Carly Demopoulos and her coworker, Dr. Srikantan (Sri) Nagarajan. Sri was the one who provided the fundamental idea and necessary resources for my master thesis. Although he is always busy, he squeezed enough time to help me with my master thesis. Carly has also never stopped helping me with all the times she could provide. I am grateful that I have the chance to work in Carly’s lab. She gives me lots of freedom and data, which are the two essential requirements to make a
biomedical engineer happy and thrive. In addition, I want to thank my coworkers and several lab members from Sri’s team who helped my master thesis.

Lastly, to my goofy, and loving cat, Simba, who is currently lying right next to my laptop: my deepest gratitude. You always stay right next to me quietly whenever I am working on this thesis. When I felt exhausted and frustrated, you always tried your best to make me happy. You provide such a great comfort and relief along my journey of drafting this master thesis.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Tables</td>
<td>ix</td>
</tr>
<tr>
<td>List of Figures</td>
<td>x</td>
</tr>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>2. Literature Review</td>
<td>6</td>
</tr>
<tr>
<td>2.1 Auditory Evoked Potential Paradigm in ASD</td>
<td>6</td>
</tr>
<tr>
<td>2.2 Introduction to Source Localization for MEG</td>
<td>10</td>
</tr>
<tr>
<td>3. Research Objective</td>
<td>15</td>
</tr>
<tr>
<td>4. Methods</td>
<td>16</td>
</tr>
<tr>
<td>4.1 Subjects and Experimental Procedure</td>
<td>16</td>
</tr>
<tr>
<td>4.2 AEF Data Processing</td>
<td>16</td>
</tr>
<tr>
<td>4.3 M50, M100, M200 Peak Selection and Analysis</td>
<td>17</td>
</tr>
<tr>
<td>5. Results</td>
<td>20</td>
</tr>
<tr>
<td>5.1 AEF Data Overview</td>
<td>20</td>
</tr>
<tr>
<td>5.2 M50, M100 and M200 Amplitude and Latency</td>
<td>21</td>
</tr>
<tr>
<td>5.3 Linear Regression of All Trial Paradigm vs Fewer Trial Paradigm</td>
<td>26</td>
</tr>
<tr>
<td>6. Discussion</td>
<td>32</td>
</tr>
<tr>
<td>References</td>
<td>34</td>
</tr>
<tr>
<td>Appendix: Supplemental Figures</td>
<td>38</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1. Percentages of Missing Functional Voxels and M50, M100, or M200 Peaks for the LH and RH ................................................................. 20
LIST OF FIGURES

Figure 1. Individual Example of Reversed Magnetic Flux Photography .......... 18
Figure 2. Individual Example of Non-reversed Magnetic Flux Photography ...... 19
Figure 3. Individual Example of Source Space Data .............................................. 21
Figure 4. Individual Example of Sensor Space Data .............................................. 22
Figure 5. Violin Plot of LH and RH M50 Amplitude ........................................... 23
Figure 6. Violin Plot of LH and RH M50 Latency ................................................... 24
Figure 7. Violin Plot of LH and RH M100 Amplitude ........................................... 24
Figure 8. Violin Plot of LH and RH M100 Latency ................................................ 25
Figure 9. Violin Plot of LH and RH M200 Amplitude ........................................... 25
Figure 10. Violin Plot of LH and RH M200 Latency ............................................. 26
Figure 11. Linear Regression of LH M50 Latency for All Trial Paradigm vs Fewer Trial Paradigm ................................................................. 27
Figure 12. Linear Regression of LH M100 Latency for All Trial Paradigm vs Fewer Trial Paradigm ................................................................. 28
Figure 13. Linear Regression of RH M100 Latency for All Trial Paradigm vs Fewer Trial Paradigm ................................................................. 28
Figure 14. Linear Regression of LH M200 Latency for All Trial Paradigm vs Fewer Trial Paradigm ................................................................. 29
Figure 15. Linear Regression of RH M50 Latency for All Trial Paradigm vs Fewer Trial Paradigm ................................................................. 30
Figure 16. Linear Regression of RH M200 Latency for All Trial Paradigm vs Fewer Trial Paradigm ................................................................. 30
Figure 17. Linear Regression of RH M100 Amplitude for All Trial Paradigm vs Fewer Trial Paradigm ............................................................... 31
Figure 18. Linear Regression of LH M50 Amplitude for All Trial Paradigm vs Fewer Trial Paradigm ............................................................... 38
Figure 19. Linear Regression of LH M100 Amplitude for All Trial Paradigm vs Fewer Trial Paradigm. ................................................................. 38

Figure 20. Linear Regression of LH M200 Amplitude for All Trial Paradigm vs Fewer Trial Paradigm. ................................................................. 39

Figure 21. Linear Regression of RH M50 Amplitude for Full Trial Paradigm vs Less Trial Paradigm. ................................................................. 39

Figure 22. Linear Regression of RH M200 Amplitude for All Trial Paradigm vs Fewer Trial Paradigm. ................................................................. 40
1. Introduction

Magnetoencephalography (MEG) is a neuroimaging technique used to study the magnetic fields generated by the brain’s electrical activity. It is a non-invasive method that measures the magnetic fields produced by the flow of ionic currents within the brain’s neurons. MEG is particularly useful for investigating the dynamic and temporal aspects of brain function, such as the processing of sensory information, perception, attention, and motor control. The magnetic signals measured by MEG are very weak but highly specific to the brain’s electrical activity, making MEG an important tool for understanding the workings of the human brain. Additionally, MEG data can be combined with structural brain images acquired from other neuroimaging technique, such as magnetic resonance imaging (MRI). This integration can provide a more comprehensive view of brain function through source reconstruction and imaging the electrophysiological data captured by MEG.

MEG can be used to aid in diagnosis and monitoring for diverse neurological conditions, such as epilepsy and brain tumors. MEG’s current clinical usage is brain mapping for brain tumors or determining epileptic foci. In addition, researchers use MEG extensively in basic and translational research to study brain function in healthy patients and patients with atypical brain function, such as autism spectrum disorders. Since MEG provides unique aspects of dynamic and temporal brain function, studies have used it to understand the human brain and its role in cognition, emotion, and behavior for various neurological diseases [1].

MEG offers several key advantages over other common clinical neuroimaging techniques, such as MRI. Firstly, MEG is non-invasive, meaning that it does not require any insertion of electrodes into the brain or exposure to ionizing radiation. Therefore, it is a safe
and highly tolerated procedure, especially for patients who cannot tolerate MRI due to noise or claustrophobia. Secondly, MEG provides a highly sensitive and specific measurement of the brain’s electrical activity, allowing subtle changes in brain function that may not be visible using other imaging techniques to be identified.

However, unlike MRI, X-ray, and positron emission tomography, MEG does not provide a direct image of the brain’s structure [1]. In MEG, data is collected based on the position of the MEG machine’s sensors rather than the subject. Therefore, modern MEG systems include a sub-system to determine the subject’s head position relative to the MEG sensors. To accurately identify the source of the brain’s electrical activity, it is necessary to know the head’s size, orientation, and movement parameters during acquisition. Additionally, MEG cannot provide the direct structural details of the brain as an image, as provided by MRI.

This limitation can be overcome by reconstructing the MEG data to MRI. This process involves registering the MEG data with the structural information obtained from MRI using the 3D head model generated during MEG. The registration process involves transforming the MEG data into the same coordinate system as the MRI data so that the neural activity measured by MEG can be localized within the structural framework provided by MRI.

Once reconstructed in source space, MEG data can be used for source localization. It is typically achieved using mathematical algorithms, such as beamforming or other inverse solutions, that analyze the magnetic field patterns recorded by the MEG sensors. More recently, some labs have developed tomographic inversion methods that use Bayesian theories to perform MEG source localization. Specifically, this project will use the
Champagne algorithm developed by the Biomagnetic Imaging Lab (BIL) at the University of California, San Francisco (UCSF) [2].

As a collaborator of the BIL, we (the Neurodevelopmental Assessment and Imaging Lab at UCSF) also use MEG, MRI, and other psychological behavior and cognition assessments to study the effects of sensorimotor functioning in children with neurodevelopmental conditions, with a particular emphasis on autism spectrum disorders (ASDs). ASD is a neurodevelopmental disorder characterized by difficulties in social interaction, communication, restrictive, repetitive patterns of behavior or interest, and sensory processing differences. It is a spectrum disorder, meaning that its symptoms and severity can vary greatly between individuals. Individuals with autism may struggle with social cues, such as eye contact and understanding nonverbal communication, and may have difficulty in social interactions and developing, maintaining and understanding social relationships. They may also have restricted patterns of interest or repetitive behaviors, such as hand flapping or rocking, and sensitivities to sensory stimuli, such as loud noises or certain textures.

Specifically, one of our ongoing projects aims to gain insight into the neural mechanisms underlying auditory processing in the ASD pediatric population by examining Auditory Evoked Fields (AEF) via MEG. The AEF paradigm allow for recording the magnetic fields generated by the electrical activities in the brain in response to auditory stimuli, such as pure tone sounds. The AEF paradigm measures the neural activity in the auditory cortex and can provide valuable information about the processing of auditory information in the brain. Specifically, it provides important information about the timing and organization of the auditory processing pathways in the brain.
However, conducting such research in autism, or pediatric populations in general, can be challenging because children may have difficulty cooperating with the assessment procedures. For example, children may have limited attention or interest in participating in the assessment, difficulty remaining still, etc. Children with autism also may have sensitivities to stimuli, and difficulties with communication may make it difficult for them to understand instructions and expectations. All these challenges can make it difficult to obtain accurate and sufficient AEF test data from children, in particular children with autism.

To address these challenges, the long-term goal of this project is to evaluate if shorter AEF test is achievable with using novel machine learning algorithm. By reducing the number of trials, AEF assessment would be more accessible to pediatric populations and others who struggle to tolerate long scans. This would also result in reduced research costs and still allow researchers to understand cortical auditory processing in these clinical populations. To achieve the long-term goal, we performed a study to evaluate the degree to which the baseline learning Champagne algorithm can be used to capture reliable results with fewer trials.

This machine learning algorithm is based on the original Champagne algorithm, a MEG source localization method published in 2009 [2]. The Champagne algorithm is based on a Bayesian framework and uses prior knowledge of the brain’s anatomy and electromagnetic properties to estimate the location of neural activity with high spatial precision. On top of this Champagne algorithm, Cai et al. [3] proposed an additional noise learning mechanism that enables the Champagne algorithm with convex bounding (CB_NL) to effectively estimate and incorporate the noise into the analysis, resulting in more robust and accurate results.
Consequently, CB_NL has been demonstrated to accurately locate brain activity with just a few trials or even one, while other algorithms fail to do so with <12 trials [3].

Using a similar Champagne algorithm, that learns noise from pre-stimulatory phase and fixed for post-stimulatory period, we propose that shorter AEF paradigms can produce results comparable to the longer paradigm that is standard (e.g., 120 trials). To test this hypothesis, we will apply this Champagne with baseline noise learning algorithm to the AEF data of both ASD and healthy-control populations. We aim to use 100%, 80%, 60%, 40%, and 20% of the total trials to perform source localization with Champagne with baseline noise learning. We will then evaluate the reconstruction results’ latency, normalized magnitude, and absolute power and compare them across different trial percentages. Through this process, we aim to determine the minimum percentage of all artifact-free AEF trials that can produce results comparable to those obtained using all trials. Based on our study results, we hope to identify a shortened AEF test that retains anatomical significance while minimizing the testing time. This shortened AEF test will help to alleviate any potential frustration caused by prolonged testing times in pediatric populations and increase scan time efficiency. This thesis uses AEF paradigm in ASD, which is reviewed in literature review 2.1, and using machine learning algorithm to reconstruct AEF results, which is reviewed in literature review 2.2.
2. Literature Review

2.1 Auditory Evoked Potential Paradigm in ASD

The American Psychiatric Association’s Diagnostic and Statistical Manual (Fifth Edition) [4] indicates that a child must exhibit persistent deficits in social communication and social interaction and restricted, repetitive patterns of behavior, interests, or activities, or differences in sensory processing to receive an ASD diagnosis. ASD is a neurodevelopmental disorder, as symptoms present in the early developmental period. According to CDC’s Autism and Developmental Disabilities Monitoring (ADDM) network, the prevalence of ASD across 11 sites within the network is 27.6 per 100 children, meaning that 1 in 36 children is diagnosed with ASD [5]. Boys are more likely to be diagnosed than girls. Given the relatively high prevalence, it is important to understand the potential causes of ASD to propose effective treatments.

Auditory dysfunction is fairly common in children with ASD [6-8] and have been shown to have both direct and indirect correlations with language impairments because it is necessary for children to hear language before their brains can process it [6, 7, 9, 10]. Therefore, I have chosen to specifically focus on finding an efficient measurement of auditory processing in children with autism for my master’s project.

The AEF paradigm is a commonly used tool to study the auditory cortex in MEG scans. To better understand the general trends of auditory processing in both typically developing (TD) and ASD populations, Cardy et al. studied the auditory field potential responses of adults, adolescents, and children to 1000 Hz sinusoidal tones [11]. Their study included 8 TD subjects and 10 subjects with ASD. The authors found that children in the TD group had a
significant delay in M100 latency and smaller M100 amplitude than adults. In contrast, the M50 amplitude and latencies were significantly larger in ASD and TD children than in adults. The M50/M100 amplitude ratio was found to have a significant positive linear relationship with age. Based on these findings, Cardy et al. proposed that M50 amplitude progressively decreases while M100 amplitude progressively increases with age. Therefore, M50 is more prominent in children, and M100 is more prominent in adults.

Matsuzaki et al. conducted a study similar to Cardy et al., testing AEF responses to 200, 300, 500, and 1000 Hz sinusoidal tones in a much larger sample population comprising 132 subjects aged 6 to 42 years (55 TD and 77 ASD) [11, 12]. They reported that adults and children in the ASD group had statistically significant delays in M50 and M100 latencies compared to the TD group. These delays were found to persist throughout the lifespan without converging or increasing. Similar findings were also reported in young children by Stephen et al. [13] Furthermore, Matsuzaki et al. [12] reported that both the ASD and TD groups showed a significant negative association between M50 latency and age. However, a similar significant association between M100 latency and age was observed in the TD group but not in the ASD group.

The significant negative association between M100 latency and age in the TD group is also supported by Gage et al. [14], who conducted similar AEF experiments only with children. Their study included 15 children with ASD and 17 TD children who were asked to listen to sinusoidal tones at 200 and 1000 Hz frequencies. They used simple linear regression analyses on the M100 latency for both groups, finding that the TD group but not the ASD group showed a significant negative relationship between M100 latency and age.
In addition to their linear regression findings, Gage et al. found that the ASD group exhibited longer M100 latencies than the TD group [14]. Additionally, in another experiment by Gage et al. [15] on the same population of 15 ASD and 17 TD children, they found that the TD group exhibited significant M100 latency difference for all four tones (200, 500, 100, and 1000 Hz) in both the right and left hemispheres. In contrast, the ASD group showed similar results to the TD group for the left hemisphere only. For the right hemisphere, the ASD group showed significant differences in M100 latency between 200 and 1000 Hz but not between the other tones.

In response to the intriguing hemisphere differences reported by Gage et al. [15], more studies have explicitly investigated the AEF responses of the left and right hemispheres in ASD and TD children. A study by Roberts et al. [16] that specifically examined M50 and M100 in different hemispheres of the superior temporal gyrus found that ASD children had delayed M100 in the right hemisphere across four different sinusoidal tone frequencies (200, 500, 100, and 1000 Hz). However, neither the ASD nor TD children showed left or right hemisphere differences in M50 source strength. In addition, neither group showed significant left-hemisphere M100. Port et al. and Williams et al. also found similar results [17, 18]. In addition to the M50 and M100 findings mentioned in the studies above, Demopoulos et al. reported a significant difference in M200 latency in the left hemisphere between the ASD and TD groups [6].

All AEF studies mentioned above focus on auditory responses to sinusoidal tones. One might wonder about the auditory responses to vowels or even words in children with ASD. Cardy et al. [19] conducted an AEF study of sinusoidal tones and vowels with 9 TD and 7
ASD children. Surprisingly, no difference was found in M50 and M100 for different vowels between the ASD and TD groups. However, when measuring latency shift, calculated using a magnetic mismatch field minus M100 latency, children with ASD showed a statistically significant prolonged shift compared to TD children. Yoshimura et al. [8] conducted a similar study on young children. In a sample population comprising 33 ASD and 30 TD young children, they played a Japanese syllable commonly used by Japanese mothers to their infants during the AEF test. They found that shorter M50 latency in both the right and left hemispheres was significantly correlated with higher language-related performance in young TD children but not in young ASD children. Brennan et al. [20] conducted an AEF experiment that presented legal and illegal phonotactic sequences to children with ASD and TD. They found that only children with ASD showed significant sensitivity to legal and illegal phonotactic sequences. Moreover, this sensitivity emerged relatively late to the stimulus.

As mentioned at the beginning of this section, it can be difficult for children with autism to process language if they cannot hear the word. Therefore, research groups have studied the correlation between auditory response and language impairment in children with autism, given the above studies showing abnormal auditory responses to words or phonotactic sequences in ASD children. Cardy et al. [7] conducted an AEF experiment using a 1000 Hz tone in both ASD and TD children, finding that the M50 latency in the right hemisphere could predict oral language ability with 71% accuracy. Demopoulos et al. [6] also found a significant association between left M200 latency and communication measurements.
2.2 Introduction to Source Localization for MEG

The previous section discussed important findings on the AEF paradigm in the ASD population. There have been numerous MEG source localization techniques to interpret AEF results correctly. This section starts with a brief overview of the MEG mechanism and common source localization methods.

MEG measures the magnetic fields produced by the electrical activity of neurons in the brain using superconducting sensors called SQUIDs, which are located in a helmet-shaped device that is placed over the head. When neurons fire, they create tiny electrical currents that flow in their axon’s direction, creating a magnetic field perpendicular to the current’s flow. The SQUID sensors detect the magnetic fields from the brain and convert them into electrical signals that can be recorded and analyzed, providing information about the timing and location of neuronal activity with millisecond and millimeter-level precision [20, 21].

Besides SQUID, the other common MEG device is optically pumped magnetometers (OPM) developed by Kominis et al. [22]. Unlike the superconducting quantum interference devices (SQUIDs) used in conventional MEG, OPMs do not require cryogenic cooling and are based on a different physical principle. OPMs use laser light to excite atoms in a small glass cell, causing them to emit light at a specific wavelength when exposed to a magnetic field. By detecting the changes in the light emitted by the atoms, OPMs can measure the magnetic fields produced by the electrical activity of neurons in the brain. While OPMs can potentially make MEG more portable and accessible, they currently have lower sensitivity and signal-to-noise ratio than SQUIDs.
The forward and inverse models are commonly used to interpret MEG signals [21]. The forward model predicts how the electrical activity of neurons in the brain generates the magnetic fields measured by MEG sensors. In contrast, the inverse model estimates the underlying neuronal sources of the measured magnetic fields. Since the BIL lab specifically focuses on applying the inverse models to MEG data, this review will not further describe the forward model.

Common inverse models include parametric inversion on dipole fitting, spatial filtering, and topographic inversion. Parametric inversion on dipole fitting is a typical and classic method proposed by Hämäläinen et al. [23] They estimated the location and orientation of known dipole sources of MEG signals using a parametric inversion approach. Spatial filtering is also a common inverse model. Common methods within this model include MNE [24, 25], sLORETA [26], dSPM [27], and other beamforming methods.

Tomographic inversion is a relatively recent development compared to dipole fitting and spatial filtering methods. It is based on Bayesian principles and involves two algorithms: MSP [28] and Champagne [2, 29]. The Champagne method was proposed and developed by the BIL at UCSF, and it is the algorithm that will be used in this project. Champagne is a method for estimating multiple correlated neural sources’ location, orientation, and time course using MEG. As mentioned above, this method is based on Bayesian inference, which enables incorporating prior knowledge and uncertainty into the estimation process.

Specifically, Champagne uses a hierarchical Bayesian model to estimate the neural sources’ parameters [2]. The model assumes that each neural source is represented by a dipole, which has a location, orientation, and strength. Additionally, the model assumes that
each dipole’s time course is described by a basis function that captures the temporal dynamics of the neural activity. The model’s estimation process involves optimizing a posterior probability distribution using MCMC techniques. The posterior distribution combines the prior knowledge about the parameters with the likelihood of the MEG data given the parameters. The Champagne model’s main strength is its robustness to noise and other sources of uncertainty in the MEG data.

2.2.1 Novel Robust Algorithm for Reconstructing AEF

While the traditional Champagne algorithm demonstrates an excellent ability for source reconstruction [2, 29], such algorithms’ accuracy and computational efficiency can be limited by the presence of noise in the data. Several modifications of the traditional Champagne algorithms have been proposed to overcome this limitation [3, 30-32, 33].

In early 2020, Cai et al. [30] proposed a robust empirical Bayesian algorithm for reconstructing the brain’s electrical activity using electromagnetic brain imaging. The proposed algorithm, Smooth Champagne, incorporates a prior Champagne algorithm that accounts for the spatial correlation of the brain sources and a noise model that accounts for the noise in the data. The authors showed the effectiveness of the proposed algorithm using simulated and real datasets, including one real MEG dataset of AEF for four subjects. Their results showed that the MxNE, Champagne, and Smooth Champagne algorithms were more effective in localizing bilateral auditory activity than the other tested algorithms. Moreover, they showed that the Smooth Champagne algorithm performed better with noise and was more robust than existing methods using the other real datasets. They concluded that the Smooth Champagne algorithm has the potential to advance the electromagnetic brain
imaging field by providing a more accurate and robust method for identifying neural activity patterns in the brain.

Cai et al. [3] later proposed a similar noise learning algorithm that aims to reconstruct with fewer trials but achieve the same significant result as reconstructing with all trials. This new algorithm, Champagne with CB_NL, is based on the principle of robust statistical estimation and is designed to provide an accurate estimate of the data’s noise level while being insensitive to outliers and deviations from normality. The Champagne with CB_NL algorithm first estimates each sensor’s noise level using a robust standard deviation estimator. Then, it computes the average noise level across all sensors, which serves as the final estimate of the data’s noise level. It is designed to be robust to outliers and non-normality by using a weighted median standard deviation estimator, which down-weights the influence of outliers in the data.

Using Champagne with CB_NL, Cai et al. [3] performed source localization on a real MEG AEF dataset with five subjects. Specifically, they used 63, 12, 2, and 1 AEF trials to reconstruct the M100 auditory response using Champagne with CB_NL. Their results showed that Champagne with CB_NL could accurately identify bilateral brain activation with focused reconstructions, even with limited or just a single trial. The number of trials used did not appear to impact the accuracy of the reconstructions.

Besides Smooth Champagne and Champagne with CB_NL [3, 30], Hashemi et al. [33] proposed a majorization minimization (MM) framework. Unlike Smooth Champagne and Champagne with CB_NL, which are one type of modified Champagne algorithm, the MM framework uses and unifies various sparse Bayesian learning algorithm methods. Therefore,
this proposed framework combines the advantages of different sparse Bayesian learning algorithms, such as sparse Bayesian inference, empirical Bayesian inference, and hierarchical Bayesian inference, to improve the accuracy and efficiency of electromagnetic brain imaging. However, like the Smooth Champagne and Champagne with CB_NL algorithms, the MM framework aims to derive one efficient optimization algorithm that can handle different sparse Bayesian learning algorithms. In addition, the MM framework was also tested on both simulated and real data, including one real MEG dataset of AEF for one subject. They reconstructed AEF responses with 10, 20, 40, 60, and 100 trials and showed improved accuracy and efficiency in reconstructing the auditory response sources in the brain.

The other related algorithms are three Bayesian inference algorithms and time-frequency Champagne (TFC) by Cai et al. [31, 32]. The three Bayesian algorithms are jointly used to estimate brain activity and noise from the measured signals [32]. Such algorithms showed improved performance over existing methods when tested on simulated and real data. Similarly, TFC [31] is an empirical Bayesian method that can localize the time-frequency dynamics of neural activity from electromagnetic signals. This method is based on a hierarchical Bayesian model that incorporates prior knowledge about the brain activity distribution and the sparsity of the neural sources. TFC also showed improved accuracy and reliability over existing methods. Since the three Bayesian inference methods and the TFC method are not as pertinent as other methods discussed in this section, this literature review will not elaborate on them extensively.
3. Research Objective

This project’s primary goal was to evaluate a condensed version of the standard AEF test that can still yield optimal results as the standard version. This study aimed to achieve this objective by comprehensively comparing the outcomes of AEF source localization with various percentages (20%, 40%, 60%, 80%, and 100%) of the all artifact-free AEF trials using the Champagne algorithm with baseline noise learning. The latency and normalized response amplitude produced by these different AEF source localization paradigms were compared using linear regression.
4. Methods

4.1 Subjects and Experimental Procedure

The primary research method used in this study is applying Champagne with baseline noise learning to an AEF dataset. This algorithm was designed to learn noise from the pre-stimulatory period of the AEF test and keeps this learning fixed for its post-stimulatory period. The AEF dataset used in this project comprised 83 children, 25 TD and 58 ASD, recruited through methods specified in the NIDCD-funded R01DC019167-01A1 parent project, K23DC016637-01A1 and Weill Award for Clinical Neuroscience Research funding. This project was approved by the Institutional Review Board (IRB) at UCSF. The AEF paradigm data were collected during MEG scans at the specified locations listed in the grants. During the AEF test, participants were exposed to a 1000 Hz tone burst for 100 ms through MEG-compatible headphones at 65 dB, delivered binaurally. Participants listened passively to these stimuli for 120 trials with an interstimulus interval of 2 seconds.

4.2 AEF Data Processing

Before applying the Champagne algorithm with baseline noise learning, the AEF data were visually inspected to identify and remove any artifacts, such as 77 and muscle, eye blink, and motion artifacts. Next, 20%, 40%, 60%, 80%, and 100% of artifact-free trials were averaged and inputted into the Champagne algorithm with baseline noise learning. Specifically, the trials were taken sequentially from the start of the paradigm after eliminating any trials with artifacts. For example, if a subject had only 2 artifact-free trials out of the first 24 trials, 20% would be theses first 2 trials.
After obtaining results from the Champagne algorithm with baseline noise learning, individual functional voxels in both the left (LH) and right (RH) hemispheres of the source localization results were visually identified using the Nutmeg toolbox [34] in MATLAB 2012b [35]. Next, three raw orientations of the functional voxels were reconstructed using the principal component analysis (PCA) function in MATLAB 2021a [36], to compress time series across three raw orientations. With retaining only the first principal component of PCA, the reconstructed PCA results were normalized using the pre-stim z-normalization method. The absolute value of this pre-stim z-normalization result was obtained and plotted to identify the M50, M100, and M200 peaks.

4.3 M50, M100, M200 Peak Selection and Analysis

The latency and normalized amplitude of the three peaks (M50, M100, and M200) in the z-normalized results were visually inspected and recorded. Specifically, we evaluated two different peak selection methods. The first method was the one used by Cardy et al. [11]. The ASD population is well known for delayed auditory responses [12, 15-17, 19]. Because latency is not a reliable criterion for AEF peak identification in this clinical population, Cardy et al. [11] proposed using magnetic flux photography since the M50/M100 and M100/M200 peaks typically show inverse magnetic flux patterns. For example, in Figure 1, Figure 1a shows a particular subject’s RH source with M50, M100, M200 labeled with red dash line. Part Figure 1b, 1c, 1d, show the magnetic flux correspond to M50, M100 and M200. This subject’s magnetic flux for RH M50 shows a positive over negative flux pattern (i.e., red over blue), and the RH M100 magnetic flux shows a negative over positive pattern.
(i.e., blue over red). The magnetic flux pattern of M100 is the inverse of the magnetic flux pattern of M50 for this subject.

However, this method did not suit our AEF dataset. The main issue was that the magnetic flux photography did not show an inverse magnetic flux pattern for M50/M100 and M100/M200 in over half of our subjects (56.9%) in all trial sensor space data. Figure 2 shows an example of this phenomenon. Figure 2a is the reconstructed LH space result from the Champagne algorithm with baseline noise learning for the 100% trials paradigm. Figure 2b shows the raw MEG sensor data for all channels in the LH and RH. The red lines represent the M50, M100, and M200 peaks. Figure 2c, 2d, 2e show the magnetic flux photography for the RH and LH M50, M100, and M200 peaks. It is evident from this figure that the magnetic flux of M50 and M100 shows the same pattern for both hemispheres.

Therefore, we developed a second peak selection method. First, sensor and source space data are temporally aligned to ensure that at least two common peaks can be identified. For
example, the red dash lines in Figure 2a represent the peaks in the reconstructed source space, and the red solid lines in Figure 2b represent the temporally aligned peaks in the sensor space. The source space’s peaks are temporally aligned with the peaks in the sensor space. The only exception was that we did not take any peaks before 40 ms since it would be too early to be considered an auditory response.

After obtaining the normalized response amplitude and latency of the three peaks (M50, M100, M200), bivariate linear regression was used to compare the all trial result with the fewer trial results (e.g., 100% vs. 80%, 100% vs. 60%, 100% vs. 40%) in MATLAB 2021a [36]. The bisquare weight function was used to minimize the effect of outliers on these linear regression results [36]. This analysis treated all subjects as a single group but separately examined each peak’s latency and amplitude in each hemisphere. The $R$-value and 95% confidential interval were also calculated for each linear regression result.
5. Results

5.1 AEF Data Overview

Out of the 83 subjects, 69 had usable AEF data, and 14 either did not have AEF data or their AEF data had <30% usable trials. Among the 69 subjects with usable AEF data, 22 were TD children, and 47 were children with ASD. All analyses in this study treated TD children and children with ASD as one group because it aimed to evaluate an algorithm rather than study the differences between the TD and ASD groups.

The percentages of missing LH and RH functional voxels for five paradigms (20%, 40%, 60%, 80%, and 100% trial reconstruction) are listed in Table 1. If subjects had no active LH functional voxel, they would not have any LH auditory response (i.e., M50, M100, and M200 peaks for LH). As expected, the 20% paradigm results had higher percentages of missing functional voxels for both the LH and RH than the other paradigms. The LH had much higher rates of missing functional voxels than the RH.

Table 1. Percentages of Missing Functional Voxels and M50, M100, or M200 Peaks for the LH and RH

<table>
<thead>
<tr>
<th>Missing Left Hemisphere Functional Voxels (%)</th>
<th>100% Trials Paradigm</th>
<th>80% Trials Paradigm</th>
<th>60% Trials Paradigm</th>
<th>40% Trials Paradigm</th>
<th>20% Trials Paradigm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Missing Right Hemisphere Functional Voxels (%)</td>
<td>2.90%</td>
<td>5.79%</td>
<td>10.14%</td>
<td>5.79%</td>
<td>11.59%</td>
</tr>
<tr>
<td>Missing Left Hemisphere M50, M100 or M200 Peaks for Subjects who have Functional Voxels (%)</td>
<td>1.45%</td>
<td>4.34%</td>
<td>1.45%</td>
<td>2.90%</td>
<td>2.90%</td>
</tr>
<tr>
<td>Missing Right Hemisphere M50, M100 or M200 peaks for Subjects who have Functional Voxels (%)</td>
<td>8.69%</td>
<td>5.79%</td>
<td>5.79%</td>
<td>1.45%</td>
<td>7.24%</td>
</tr>
</tbody>
</table>

For subjects with active functional voxels, the percentages of missing M50, M100, or M200 peaks for each hemisphere are listed in Table 1. Unlike the missing rate of functional
voxels, the 80% paradigm had the highest missingness for the three LH peaks, and the 100% paradigm had the highest missingness for the three RH peaks. Nevertheless, the missing percentages of all paradigms were <10% for both hemispheres.

5.2 M50, M100 and M200 Amplitude and Latency

The source reconstruction results with the Champagne algorithm with baseline noise learning are similar across the five paradigms. Specifically, Figure 3 shows the source space data of one subject. Figure 2a-e shows the results for the 100%, 80%, 60%, 40%, and 20% trial paradigms, respectively. The normalized response amplitude of these five paradigms decreased from 100% to 20%, where the 100% trial paradigm had the largest amplitude, and the 20% trial paradigm had the smallest amplitude for both the LH and RH. In contrast, the latencies of the M50, M100, and M200 peaks only changed slightly across paradigms, except for the RH 20% trial paradigm, potentially because it had the most noise.

Figure 3. Individual Example of Source Space Data.
The noise observed in the 20% trial paradigm’s source space could also be seen in the sensor space. Figure 4 shows the raw MEG sensor data for the same subject in Figure 3. Similarly, Figure 4a-e shows the sensor space for the 100%, 80%, 60%, 40%, and 20% trial paradigms, respectively. For each paradigm, manually identified M50, M100, M200 by using peak selection method 2 are listed in magenta for LH and in green for RH. The 100% paradigm’s sensor space shows three clear peaks. However, the sensor space became noisier as the number of trials decreased, and the three peaks became harder to identify. The 20% paradigm’s sensor space contained the most noise. In addition, the latencies of three peaks are relatively similar for all paradigms, except the 20% trials paradigm.

Figure 4. Individual Example of Sensor Space Data.
Similar results were observed at the group level. Violin plots [37] were created to visualize the distribution and average of amplitude and latency for the LH and RH M50, M100, and M200 peaks (Figures 5–10). In each figure, the top subplot shows the data for the LH, and the bottom subplot shows the data for the RH. Specifically, Figures 5 and 6 show the amplitude and latency for M50, Figures 7 and 8 show the amplitude and latency for M100, and Figures 9 and 10 show the amplitude and latency for M200. Notably, the average amplitudes differed across paradigms for each peak and hemisphere. In contrast, the average latencies were similar across paradigms for each peak and hemisphere. These group-level results are consistent with those shown in the above individual examples (Figures 3 and 4).

Figure 5. Violin Plot of LH and RH M50 Amplitude.
Figure 6. Violin Plot of LH and RH M50 Latency.

Figure 7. Violin Plot of LH and RH M100 Amplitude.
Figure 8. Violin Plot of LH and RH M100 Latency.

Figure 9. Violin Plot of LH and RH M200 Amplitude.
5.3 Linear Regression of All Trial Paradigm vs Fewer Trial Paradigm

The results of bivariate linear regressions comparing the all trial paradigm (100%) with the fewer trial paradigms (80%, 60%, 40%, and 20%) are consistent with the violin plots. The latency results generally had much higher $R$ values than the amplitude results, except for the 20% trial paradigm.

Figures in this section show the linear regression results for LH and RH M50, M100, and M200 latency and amplitude in which the TD subjects and subjects with ASD were treated as a single group. Each figure contains four subplots: 100% vs. 80%, 100% vs. 60%, 100% vs. 40%, and 100% vs. 20%. In these figures, each subject’s response is represented by a single dot, where each red dot represents a subject with ASD, and each blue dot represents a TD subject. In each figure, the y-axis shows the amplitude or latency for the all trial paradigm,
and the x-axis shows the amplitude or latency obtained for the fewer trial paradigms. In each subplot, the black solid line shows the linear regression fit, and the dashed line shows the 95% confidential interval of the linear regression. The $R$-value for each linear regression analysis is labeled at either the top left or right of each subplot.

### 5.3.1 Linear Regression Results of Latency

The 100% vs 80% linear regression for LH M50 latency showed a strong correlation ($R = 0.88$; Figure 11). The correlation between the all trial and fewer trial paradigms decreased as the number of trials decreased. As such, the $R$-value was only 0.39 for 100% vs. 20% linear regression. Similar results were also seen with LH M100 latency (Figure 12) and RH M100 latency (Figure 13).

![Figure 11. Linear Regression of LH M50 Latency for All Trial Paradigm vs Fewer Trial Paradigm.](image-url)
Figure 12. Linear Regression of LH M100 Latency for All Trial Paradigm vs Fewer Trial Paradigm.

Figure 13. Linear Regression of RH M100 Latency for All Trial Paradigm vs Fewer Trial Paradigm.
For LH M200 latency, linear regression only showed a strong correlation for 100% vs. 80% ($R = 0.85$) and 100% vs. 60% ($R = 0.77$) comparisons (Figure 14). In addition, the $R$-value for 100% vs. 20% comparison ($R = 0.61$) was slightly higher than that for 100% of 40% comparison ($R = 0.52$). Similarly, linear regressions for RH M50 latency (Figure 15) and RH M200 latency (Figure 16) showed strong correlations for the 100% vs. 80% and 100% vs. 60% comparisons but not for the 100% vs. 40% and 100% vs. 20% comparisons.

Among all linear regressions for LH and RH latency, LH M100 ($R = 0.96$), RH M100 ($R = 0.96$), and RH M200 ($R = 0.97$) had the highest $R$-values in the 100% vs. 80% comparison. In contrast, the correlation was minimal for RH M50 ($R = 0.12$) and RH M100 ($R = 0.28$) in the 100% vs. 20% comparison.
Figure 15. Linear Regression of RH M50 Latency for All Trial Paradigm vs Fewer Trial Paradigm.

Figure 16. Linear Regression of RH M200 Latency for All Trial Paradigm vs Fewer Trial Paradigm.
5.3.2 Linear Regression Results of Normalized Response Amplitude

The linear regression results for normalized response amplitude for the all trial vs. fewer trial paradigms showed weaker correlations than with latency. The RH M100 amplitude (Figure 17) showed the strongest correlation in the 100% vs. 80% comparison ($R = 0.84$). In contrast, the linear regression results did not show strong correlations for the other peaks (LH M50/M100/M200 and RH M50/M200; all $R < 0.8$). The figures of these linear regression results are provided in the appendix A (Figure 18 – Figure 22).

Figure 17. Linear Regression of RH M100 Amplitude for All Trial Paradigm vs Fewer Trial Paradigm.
6. Discussion

In this study, the average latencies of LH and RH M50, M100, and M200 were similar across the five paradigms (100%, 80%, 60%, 40%, and 20%), but the averages of normalized response amplitude were not. The bivariate linear regression results comparing the all trial paradigm (100%) with the fewer trial paradigms (80%, 60%, 40%, and 20%) also support these findings, where latency showed strong correlations, but amplitude did not. Based on these findings, future studies could use the latency produced by the Champagne algorithm with baseline noise learning with fewer trials as reliable AEF results. However, the normalized response amplitude produced by this algorithm with fewer trials is unreliable. We hypothesize that the greater inconsistency with amplitude might reflect the signal-to-noise ratio (SNR), explaining its decreases when the algorithm reconstructed source space with fewer trials in a single subject (Figure 4) and its weak correlations in the linear regression results.

The bivariate linear regression results for latency also showed that the 80% and 60% trial paradigms were strongly correlated with the 100% trial paradigm for LH and RH M50, M100, and M200. In other words, 80% and 60% trial paradigms could produce similar results to the 100% trial paradigm. In contrast, the 40% trial paradigm could produce similar results to the 100% trial paradigm for LH M50, LH M100, and RH M100 but not for RH M50, RH M100, and LH/RH M200. Therefore, the 40% trial paradigm could produce some reliable latency results. Finally, as expected, latency results only weakly correlated between the 20% and 100% trial paradigms since the 20% trial paradigm used the fewest trials to reconstruct the source space among all paradigms. Based on these linear regression results
for latency, the Champagne algorithm with baseline noise learning algorithm can produce reliable results with as few as 60% of trials, meaning that the total number of AEF trials could be reduced by 40%.

This study had several limitations. First, the artifact rejection of AEF data was performed manually. Human errors could occur during this process, where trials with noise might be included, or trials without noise might be rejected accidentally. Second, the Champagne algorithm with baseline noise learning was only tested on one dataset. While our dataset has relatively large numbers of subjects (69 in total), this algorithm should be tested on more AEF datasets to confirm that our findings are replicable. Third, the normalized response amplitudes of the Champagne algorithm with baseline noise learning might represent the SNR of the source reconstruction, not the actual magnitude of the peaks. Therefore, studies aiming to understand the magnitude of AEF peaks might find this algorithm less useful.

Future improvements to the Champagne algorithm with baseline noise learning may focus on capturing noise for paradigms with <40% trials and outliers since the latency and amplitude outliers in our results could be affected by noise. Future studies should test whether this algorithm could use fewer AEF trials to produce a similar clinical correlation as the all AEF trial paradigm. If any neuropsychological variable correlates significantly with AEF latency or amplitude in the all trial paradigm, assessing whether this significant clinical correlation is also observed with fewer trial paradigms would be necessary.
References


Appendix: Supplemental Figures

Figure 18. Linear Regression of LH M50 Amplitude for All Trial Paradigm vs Fewer Trial Paradigm.

Figure 19. Linear Regression of LH M100 Amplitude for All Trial Paradigm vs Fewer Trial Paradigm.
Figure 20. Linear Regression of LH M200 Amplitude for All Trial Paradigm vs Fewer Trial Paradigm.

Figure 21. Linear Regression of RH M50 Amplitude for Full Trial Paradigm vs Less Trial Paradigm.
Figure 22. Linear Regression of RH M200 Amplitude for All Trial Paradigm vs Fewer Trial Paradigm.