Title: Removal of EMG Artifact from EEG During Naturalistic Speech Production

Authors: Garret Kurteff, M.S.[1], Rosemary A. Lester-Smith, PhD[1], CCC-SLP, Liberty S. Hamilton, PhD[1, 2]*
[1] The University of Texas at Austin, Moody College of Communication, Department of Speech, Language and Hearing Sciences
[2] The University of Texas at Austin, Dell Medical School, Department of Neurology

Journal: NeuroImage

***Corresponding Author:** Liberty S Hamilton, liberty.hamilton@austin.utexas.edu

Abstract (282 words)

Speech production is under-studied compared to speech perception largely due to artifacts in data collection caused by articulation. In electroencephalography (EEG), these artifacts include electromyographic activity (EMG) originating from the muscles that control articulation. This is unfortunate because EEG is well-suited for studying the rapid temporal changes in speech production. In addition, the few EEG studies of speech production are limited to the single-word level, which limits the generalizability of studies to how speech is used in everyday contexts. Here we demonstrate the successful removal of EMG artifact from 64-channel EEG collected during a naturalistic speech production and perception experiment. Participants overtly produced sentences from the MOCHA-TIMIT corpus, then listened to playback of themselves producing the sentences. Perception and production trials were preceded and followed by a broadband click tone. Trials containing excessive artifact and bad channels were manually rejected and ICA was used to remove electrooculographic artifact. Next, canonical correlation analysis (CCA), a technique previously used in single-word production studies, was used to correct EMG artifact. To confirm EMG was successfully removed from the EEG, responses epoched to sentence onset were compared before and after CCA. CCA-corrected data demonstrated a significant decrease in EMG amplitude compared to uncorrected data. To confirm that EEG activity was not erroneously removed by CCA, responses to the inter-trial click tone were compared before and after CCA correction. The absence of a significant difference in click response amplitude before and after CCA confirms that the response removed by CCA is indeed EMG artifact. In this paper, the preprocessing steps necessary to isolate and remove EMG artifact are detailed so that any researchers wishing to study speech production beyond the word level using EEG may do so.

1. Introduction

Historically, the neuroscience of language has been studied in heavily constrained experiments that do not immediately resemble the language humans use in daily life (Hamilton & Huth, 2020; Matusz et al., 2019). In large part, the impetus for studying language this way has been limitations of imaging technologies. There is motivation to refine data collection techniques so that historically understudied aspects of language may be analyzed.

One such aspect is speech production: while overlaps in neural representation of speech production and perception exist, neurolinguistic studies often study speech perception alone (D'Ausilio et al., 2009; Meister et al., 2007; Watkins et al., 2003; Wilson et al., 2004). The emphasis on speech perception is in large part due to every widely used noninvasive neuroimaging technique being impacted by head movement during image acquisition (Burgess, 2020; Friston et al., 1996; Jiang et al., 2019). Because speech production involves the movement of speech articulators, head movement is a fundamental component of speech production. To combat this, researchers use methods that are additional degrees of abstraction away from natural speech, such as covert speech, where participants are instructed to *imagine*, instead of *execute*, the articulatory movements associated with speech (Okada et al., 2018; Shuster, 2003). Another technique for avoiding movement artifact is to acquire imaging data *before* articulation begins or *after* articulation has completed, as examining time windows where articulation is not actively taking place can prevent the influence of movement caused by articulation (Singh et al., 2018).

While methods that do not directly study articulated speech have greatly advanced our understanding of the neurobiological bases of speech production, all these methods do not directly observe the speech production that occurs in everyday scenarios. Language is a continuous and noisy signal that has many aspects that methods-constrained experiments cannot capture, which in recent years has motivated many speech researchers to favor naturalistic, unconstrained stimuli over the constrained stimuli used to build the foundational research of speech neuroscience (Huth et al., 2016; Lerner et al., 2011; Wehbe et al., 2014).

Motivated by the relative lack of speech production research and a movement towards naturalistic study of speech, here we examine methods for studying naturalistic, sentence-level speech production using electroencephalography (EEG). EEG is a relatively inexpensive and popular method for studying the neurobiology of language, but researchers have in large avoided the study of speech production due to widespread electromyographic (EMG) artifacts propagated across the scalp by facial musculature involved in articulation (Jiang et al., 2019). This is unfortunate because EEG, as an electrophysiological method, is well-suited to capture the rapid changes in articulatory and linguistic content associated with speech production. Phonemes often last for only tens of milliseconds, a timescale too rapid for hemodynamic imaging techniques like fMRI and fNIRS (Sejnowski et al., 2014). While invasive methods such as intracranial EEG (also called ECoG or sEEG) may have less contamination by EMG artifact, the surgical necessity of these procedures limits the availability of data to researchers, making EEG an excellent choice for studying speech production.

An *artifact* can be defined as any extraneous signal recorded by an EEG electrode that does not originate from the underlying neurons. This could include electrooculographic artifacts from eye movement, electromyographic artifacts from muscle contraction, or even physiological changes from sweat on the skin (Luck, n.d.). Techniques for removing artifact from EEG signal can be classified as either artifact *rejection* or *correction*. Artifact rejection is the complete exclusion of contaminated trials from analysis, and while effective in preserving the integrity of the signal, it has the potential drawback of substantially reducing the number of trials in an analysis, affecting statistical power. Simply collecting more trials to offset this is a potential solution, but often limited by the participants' comfort and focus, as many EEG experiments already involve multiple hours of simple, repetitive tasks. Artifact correction allows for contaminated trials to be retained in the analysis by subtracting voltage from the EEG signal to "zero out" artifacts. The drawbacks of artifact correction are that it can lead to Type I (false positive) or Type 2 (false negative) error in the "cleaned" signal. Figure 1 is a visualization of common types of EEG artifact from the EEG dataset reported here.

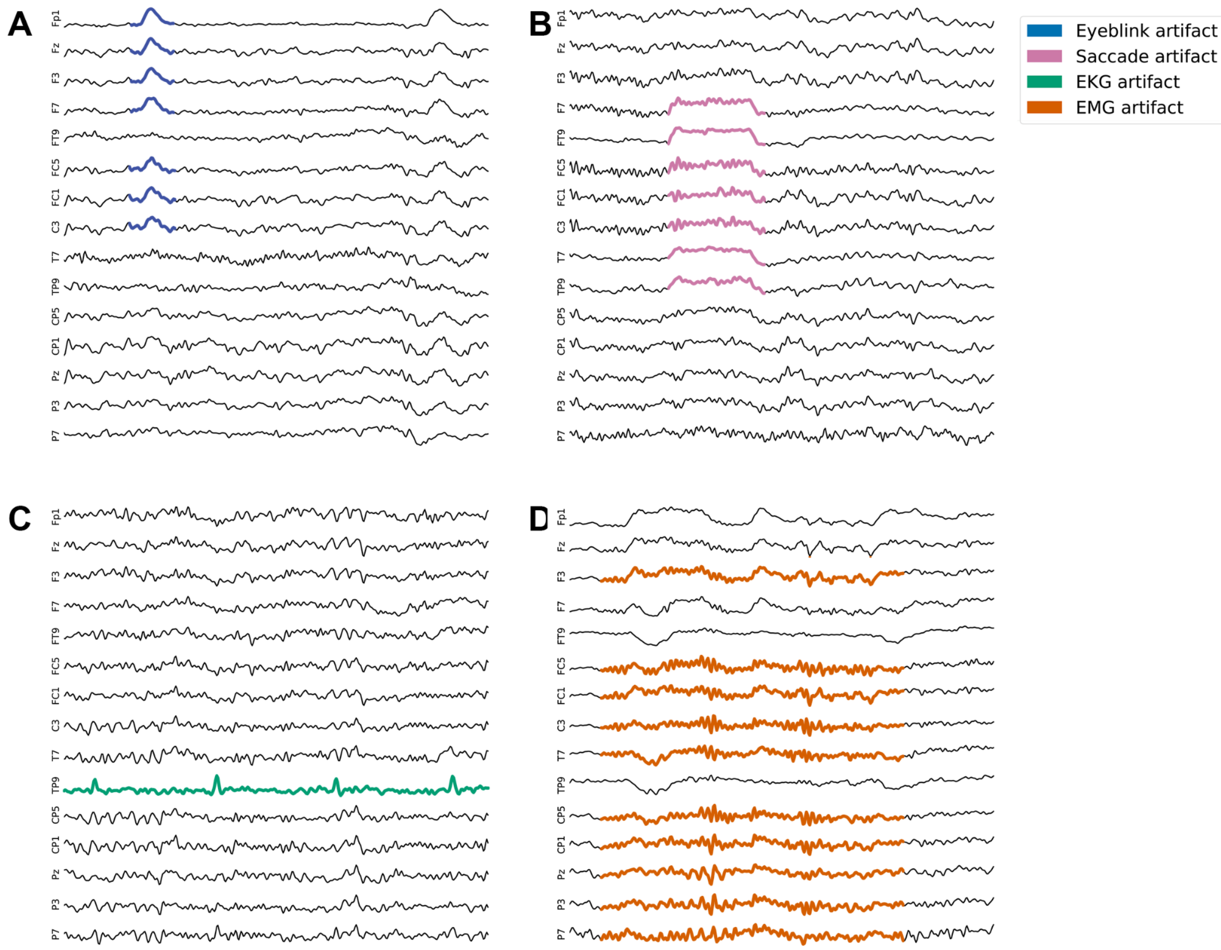

Figure 1. Common EEG artifacts visualized. Data in all panels is referenced to linked mastoids, notch filtered at 60Hz and bandpassed 1-30Hz. [A]: Vertical electrooculographic (EOG) artifact caused by blinking commonly manifests as a transient positive or negative deflections in bilateral frontal and temporal electrodes. [B]: Horizontal EOG artifact manifests as a dipole with opposite polarity for left and right lateral frontal and temporal electrodes.This artifact reflects horizontal saccade movements – quick leftward or rightward movements of the eyes. [C]: Electrocardiographic (EKG) artifact is generated by the heartbeat and is visible as rhythmic, transient spikes in lateral temporoparietal electrodes at a frequency that roughly corresponds with heart rate. [D]: Electromyographic (EMG) artifact is a broad category of artifact that can vary in timecourse and frequency depending on which articulatory muscles are being recorded. It is commonly seen in frontal, central and temporal electrodes, and may be seen especially with jaw clenching or movement.

With the exception of EMG, common types of EEG artifact are not a serious concern for EEG researchers, as there are well-documented methods for removing these artifacts via correction techniques (Dimigen, 2020; Gao et al., 2009; Gomez-Herrero et al., 2006; Gratton et al., 1983; Islam et al., 2016; Jiang et al., 2019; Keren et al., 2010; Luck, n.d.). Independent component analysis (ICA) is a popular example of such a technique and comes packaged with most EEG data analysis software including EEGLab (Delorme & Makeig, 2004), FieldTrip (Oostenveld et al., 2011), MNE-python (Gramfort et al., 2014), and others. ICA is a blind source separation technique which relies on similarities in topographic distribution, voltage change and timecourse to isolate artifacts as components that can then be subtracted from the signal.

Electrooculographic (EOG) and electrocardiographic (EKG) artifact (Figure 1) originate in large part from stereotyped muscle movement which results in high trial-to-trial consistency in response, making their correction via ICA trivial. However, EMG can originate from any of the muscles involved in speech production, of which there are many. This makes EMG much less predictable in its timecourse and topographic distribution, and ICA often fails to reliably correct for EMG artifact. The frequency range of EMG is also quite wide which only compounds the problem: some researchers categorize EMG within the alpha (8-13Hz) and beta (13-20Hz) bands (Friedman & Thayer, 1991), while others report broader frequency distributions in the 1-200Hz range (Goncharova et al., 2003).

There has been some success in isolating and correcting EMG through a blind source separation technique similar to ICA called Canonical Correlation Analysis (CCA, or sometimes BSSCCA). The primary difference between ICA and CCA is that CCA excels in identifying data that have low autocorrelation within a specified sliding time window. EMG activity is weakly autocorrelated compared to EOG, EKG, and EEG signal due to its variable timecourse. CCA has been successful at removing EMG from a synthetic data set (De Clercq et al., 2006), a single-word speech production task (Vos et al., 2010), and a phrase-level (four word) speech production task (Ries et al., 2021).

In this paper, we explore the application of CCA (in addition to standard EEG preprocessing techniques) to clean a dataset collected during a dual speaking and listening task with sentence stimuli. To verify the accuracy of these preprocessing techniques (i.e., no Type I or Type II error), comparisons are made between pre-corrected and post-corrected data and between speech production and speech perception data. The preprocessing steps are enumerated in detail sufficient for any researchers interested in applying similar techniques to the study of the neurobiological foundations of speech production.

2. Methods

2.1 Participants

21 participants (11 female, age 24.4±3.9) were recruited from flyers placed around the University of Texas at Austin campus. One participant was excluded from analysis due to a recording error, leaving N=20 participants in the final analysis. All participants were native English speakers with typical hearing as assessed through pure tone audiometry and a speech-in-noise hearing test (QuickSIN, Interacoustics). Participants provided written consent for participation in the study and were compensated at a rate of $15/hr. The sentence production and perception task lasted an average of one hour. All experimental procedures were approved by the Institutional Review Board at the University of Texas at Austin.

2.2 EEG Data Acquisition

64-channel scalp EEG and audio were recorded continuously using a BrainVision actiChamp amplifier (Brain Products, Gilching, Germany) with active electrodes at a sampling rate of 25kHz. This high sampling rate was used to acquire synchronized audio at high resolution, but is not required for EEG analysis and our system allows for only one universal sampling rate for all data. Audio signals were acquired using a StimTrak processor set on "pass through mode", such that high fidelity synchronized audio signals were acquired synchronously with neural data. Impedance level was kept below 15kΩ throughout recording using conductive gel (SuperVisc, EASYCAP) applied to the scalp at each electrode.

Four auxiliary electrodes were used to capture vEOG and EMG activity. vEOG electrodes were placed above and below the participant's left eye in line with the pupil, while EMG placement differed across participants due to the wide variety of muscles from which EMG can originate (Table 1). All placements were trialed on a participant who consented to additional time during pre-recording setup. In this participant, masseter placement allowed for the most reliable detection of EMG activity as assessed by real-time inspection of the auxiliary electrode activity during isolated muscle movement and speech production by two of the authors (G.K., R.A.L.S.); however, adhesion of masseter electrodes was unreliable from participant-to-participant due to facial hair, so placement on the orbicularis oris (upper lip) and mandible was also used for many participants (Figure 2B, Table 1). Auxiliary electrodes allowed for detection of EMG activity associated with the onset of articulation, as speech-onset articulation causes the largest artifacts in the temporal window of interest for event-related potential analysis. The masseter, orbicularis oris, mylohyoid and submentalis have all been implicated in jaw movement by previous facial EMG studies (Rastatter & De Jarnette, 1984; Stepp, 2012; Van Eijden et al., 1993).

Number of Subjects	Electrode 1 Placement	Electrode 2 Placement
11	Insertion of orbicularis oris (upper lip)	Mental protuberance of the mandible
6	Origin of masseter	Insertion of masseter
2	Origin of anterior belly of the digastric muscle	Insertion of anterior belly of the digastric muscle
2	No EMG electrode	No EMG electrode

Table 1. Summary of auxiliary electrode placement used for detecting EMG activity tallied for all subjects.

2.3 Task Design & Presentation

Stimuli were presented on an Apple iPad Air 2 and were controlled by the participant during data collection using custom interactive software developed in Swift (Apple). Stimuli were presented in a white font on a black background after a 1000ms fixation cross to minimize visual artifact in the EEG signal (Figure 2A). Accurate stimulus presentation timing was controlled by synchronizing events to the refresh rate of the screen. The iPad was placed on an overbed table to minimize arm travel needed to interact with the screen and ensure stimuli were presented at a comfortable reading distance. Participants were instructed to complete the task at a comfortable pace, and timing information on the task was collected by an automatically generated log file to assist in data preprocessing. Auditory stimuli were presented via foam-tipped insert earbuds (3M, E-A-Rtone Gold 10Ω, Minnesota, USA) and audio levels were tested prior to the start of the task.

The task was designed using a dual perception-production block paradigm, where trials consisted of a dyad of sentence production followed by sentence perception. Trials consisted of participants overtly reading a sentence then listening to a recording of themselves reading the produced sentence. Additionally, there was a manipulation of predictability in the perception trials such that blocks of 50 sentences would alternate between predictable, immediate playback of the production trial recording and unpredictable, shuffled playback of a randomly-selected production trial; however, this manipulation was not explored in the current study. Participants produced 50 unique sentences (100 for the first participant) in repetition for a total of 300 to 400 trials per participant to ensure a large number of trials for statistical analyses. Sentence stimuli were taken from the MultiCHannel Articulatory (MOCHA) database, a corpus of 460 sentences that includes a wide distribution of phonemes and phonological processes typically found in spoken English (Wrench, 1999). The 50 sentence subset used in this study was chosen at random; however, before random selection, 61 sentences were manually removed by the author (G.K.) for containing offensive semantic content (e.g., “Women may never become completely equal to men.”) or being difficult for an average reader to produce (e.g., “Many wealthy tycoons bought a yacht and a schooner.”) to reduce extraneous cognitive effects and error production, respectively.

2.4 EEG Signal Preprocessing

A visual summary of data acquisition and preprocessing is provided in Figure 2. All preprocessing was performed offline using custom Python scripts and functions from the MNE-python software package. EEG, EOG and EMG data were downsampled from 25kHz to 128Hz prior to analysis. A linked mastoid reference (Figure 2D) was applied followed by a 60Hz notch filter (Figure 2E) to remove electrical line noise artifact. For one subject, the reference electrode TP9 was a bad channel and was interpolated prior to referencing. Next, a 1-30Hz zero-phase, non-causal bandpass FIR filter (Hamming window, 0.0194 passband ripple with 53 dB stopband attenuation, -6 dB falloff) was applied to facilitate artifact rejection and independent component selection. Before manual annotation of bad segments, the power spectral density by channel was visually inspected to isolate and remove bad channels. The data were next inspected manually and bad channels and artifacts that were not removable during ICA or CCA were rejected (Figure 2F).

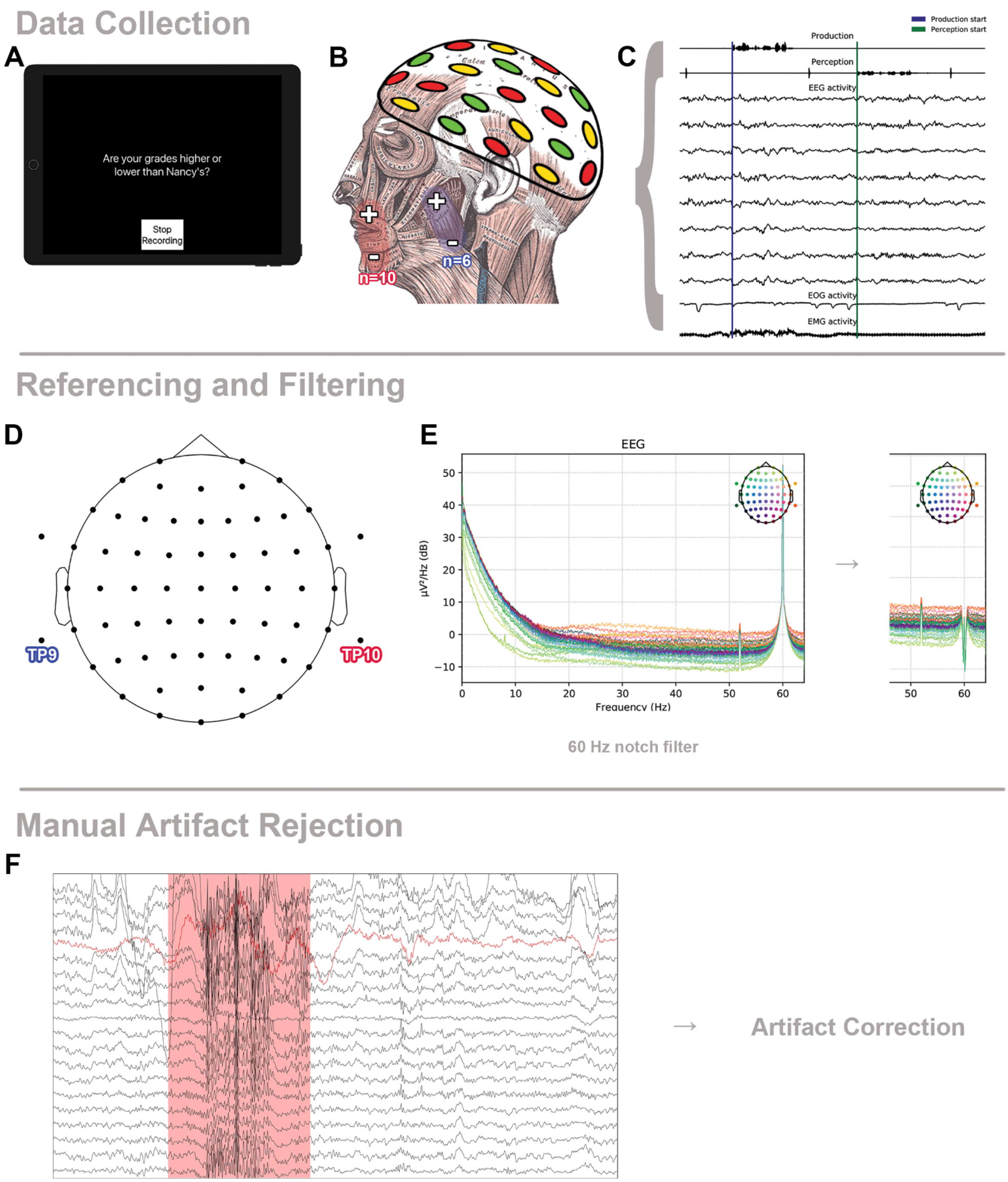
Data Collection
A
Are your grades higher or lower than Nancy's?
Stop Recording
B
n=10
n=6
C
Production start
Perception start
Production
Perception
EEG activity
EOG activity
EMG activity
Referencing and Filtering
D
TP9
TP10
E
EEG
µV²/Hz (dB)
Frequency (Hz)
60 Hz notch filter
Manual Artifact Rejection
F
Artifact Correction

Figure 2: Schematic of data collection and preprocessing. [A]: Stimulus presentation on an Apple iPad. White text presented on black screen is to minimize visual components of task. [B]: Schematic of EEG sensors and auxiliary EMG electrode placement on masseter and orbicularis oris/mandible. [C]: Example raw data showing synchronization of speech output (production), auditory input (perception), EEG, and auxiliary electrodes. Vertical lines indicate onset of production trial (blue) and perception trial (green). [D]: Sensor locations. Mastoid electrodes TP9 and TP10 are emphasized as they are used as reference. [E] Power spectral density of raw EEG before and after notch filter. [F] Example manual annotation of artifact before artifact correction. A bad channel is also visualized in this example (red line).

Artifact correction was performed after manual artifact rejection (Figure 3). Independent component analysis (ICA) was performed to remove EKG and EOG artifact (Figure 3B) . The number of ICA components selected was set to the number of non-bad channels. ICA was performed using an information-maximization approach, and manually rejected channels and segments were ignored during the fitting of ICA (Bell & Sejnowski, 1995). EOG epochs were generated by detecting peaks in the vEOG auxiliary electrodes using the create_eog_epochs function in MNE-python with a bandpass filter from 1-10 Hz. Peaks in the EOG were required to be at least ¼ of the trough to peak height of the data (the default peak threshold in mne-python). These were then correlated with the ICA components to further verify components related to EOG activity based on their scalp topography and concurrent aligned EOG peaks. Components related to EOG and EKG were visually inspected and rejected. While a 1-30Hz bandpass filter was applied to assist in component detection and inspection, we corrected the raw EEG data by removing the relevant ICA components from unfiltered EEG data. This is because CCA works best on minimally filtered data due to the wide range of frequencies in which EMG artifact can occur (Goncharova et al., 2003).

CCA was applied to data high pass filtered at 0.16Hz to remove low frequency EMG. CCA was performed using the AAR plugin for EEGLab in similar fashion to previous studies examining CCA-corrected speech production data (Ries et al., 2021; Vos et al., 2010). The AAR plugin estimated power spectral density of EEG and EMG components using a Welch spectrum estimator with a Hamming window. The approximate frequency separating the EEG and EMG was set to the default 15Hz. The options passed to the BSS algorithm included `‘eigratio’=1e6` and the default criterion options for optimization: `(‘ratio’=10, ‘range’=[0, 32])`. CCA was run in two passes: first a 30-second window to remove tonic muscle activity; second, a 2-second window to remove rapid bursts of EMG associated with speech production. For these analyses, both the window length and window shift were set to 30 or 2 seconds, respectively. Results of this process can be observed in Figure 3, which shows the raw data, data after ICA, and data after the two passes of CCA to remove EMG.

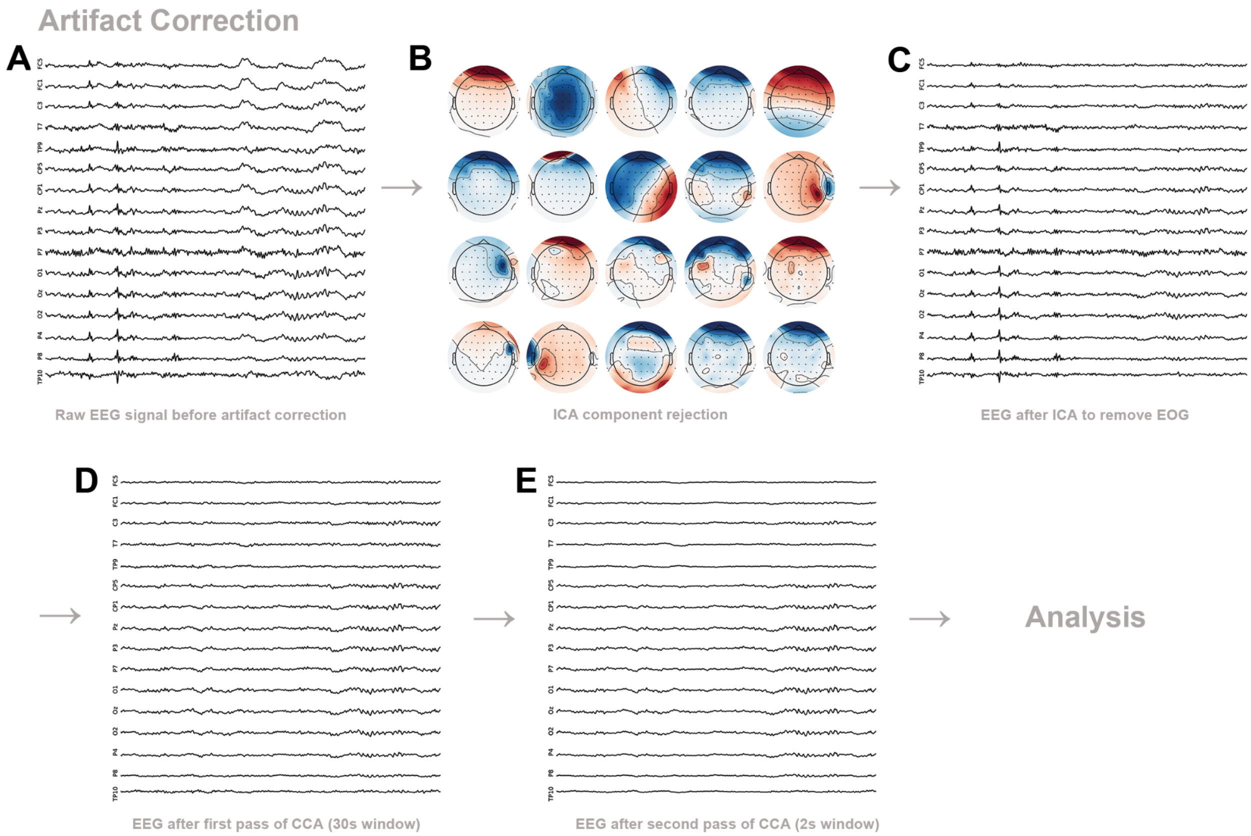

Figure 3. Order-of-operations schematic for artifact correction. Only 15 of 64 channels are displayed for visualization purposes. [A] Raw signal after steps enumerated in Figure 2. [B] ICA components. First 16 components are displayed but the number of components is set to the number of non-rejected channels. [C] EEG signal after rejected ICA components are subtracted from the waveform. [D] EEG after first pass of CCA using a 30 second window to remove tonic EMG activity. [E] EEG after second pass of CCA using a 2 second window to remove rapid EMG activity associated with articulatory movement. This signal is then bandpass filtered at 1-15 Hz before analysis.

2.5 Transcription and Epoching

Accurate timing information for words, phonemes, and sentences was generated to allow epoching of EEG data to levels of linguistic representation. A modified version of the Penn Phonetics Forced Aligner (P2FA) (Yuan & Liberman, 2008) was used to automatically generate Praat TextGrids (BOERSMA & P, 2001) using a transcript generated by the iPad log file that was manually checked and corrected for accuracy by undergraduate research volunteers. Automatically generated TextGrids were also checked for accuracy by undergraduate research volunteers. The author (G.K.) supervised the transcription process and checked the final TextGrids for accuracy before generating event files used in the analyses.

Event files containing start and stop times for each phoneme, word and sentence and information about trial type (perception versus production) were automatically generated using the iPad log files and TextGrids. These event files allowed for epoching of EEG activity to the onset of articulation at different levels of linguistic representation; however, analyses in this paper will be restricted to the sentence level. Another set of event files were automatically generated to allow for EEG alignment with each inter-trial click sound. Neural data was bandpass filtered 1-15Hz before ERP analysis.

2.6 Statistical Analysis

Linear-mixed effects (LME) models were created and assessed using the `lmertest` package in R to determine statistical significance of the ERP analyses (Kuznetsova et al., 2017). LME models are well-suited to analysis of large EEG datasets due to the minimal assumptions made about the structure of the data and the ability to examine behavioral effects across participants that account for the high degree of within-participant variation in EEG datasets. Topographic inspection of sentence-level ERP activity revealed a frontocentral ROI of nine channels that elicited the strongest response (F1, Fz, F2, FC1, FCz, FC2, C1, Cz and C2). To assess the effect of CCA correction on EEG time locked to EMG events, sentence events, and click events, we calculated the root mean square (RMS) difference wave of epoched responses averaged across these nine channels and used this as the response variable in the LME model. Difference waves were calculated by subtracting the channel-averaged response of the CCA-corrected data from the channel-averaged response of the raw data at each epoch. RMS values were used so response polarity would not influence model evaluation. Three sets of difference waves were created in this fashion corresponding to three sets of epochs: (1) peak EMG activity from facial electrodes obtained via the MNE function `mne.preprocessing.create_eog_epochs()` with a 1 - 30 Hz bandpass filter and threshold of at least ¼ the trough-to-peak threshold (2) sentence onsets and (3) inter-trial click tone responses. All three sets of epochs were obtained from -200ms to +500ms relative to the event. The LME model had a fixed effect of epoch type and a random effect of participant.

As an additional confirmation of CCA efficacy, RMS values were obtained from manually-annotated jaw EMG epochs in a single participant. A Wilcoxon signed-rank test was used to compare the mean of these epochs across channels between the raw and CCA-corrected datasets. Nonparametric statistics were used because the datasets' means did not follow a Gaussian distribution, which was confirmed via Kolmogorov-Smirnov test ($p<0.001$). Results were false discovery rate-corrected for multiple comparisons using the negative Benjamini/Yekutieli method (Benjamini & Yekutieli, 2001).

2.7 Data/code availability

Code for reproducing the analyses in this manuscript can be found at https://github.com/HamiltonLabUT/emg_removal/. The EEG dataset and corresponding event files can be downloaded at http://doi.org/10.17605/OSF.IO/FNRD9.

3. Results

Because studies of speech production using naturalistic stimuli above the word level are rare, it is important to confirm that EMG artifact correction techniques were successful. Because there is no "ground truth" in electroencephalography, there is no guaranteed method of confirming an artifact correction technique is both successful (absence of Type I false positive error) and accurate (absence of Type II false negative error). Thus, custom methods for confirming efficacy of artifact correction were developed.

3.1 Verification of specificity of removal of EMG artifact with minimal effect on EEG

In the context of this dataset, a lack of Type I error means that the neural signal is not falsely identified as EMG and removed from the dataset. To confirm signal integrity, the N1-P2 complex, a well-studied neural component related to auditory processing, was examined (Brumberg & Pitt, 2019; Hawco et al., 2009; Heinks-Maldonado et al., 2007; Lightfoot, 2016; Lijffijt et al., 2009; Martikainen et al., 2005). If EEG was falsely removed during CCA artifact correction, the N1-P2 complex would likely be degraded when CCA-corrected data are compared to pre-CCA data. In ERP data epoched to inter-trial broadband click tones, N1-P2 complex was preserved, suggesting that neural signal is not falsely removed from the dataset (Figure 4A, D). The integrity of task-related production ERP (Figure 4C, F) provides further corroboration that neural signal was not falsely removed from the dataset.

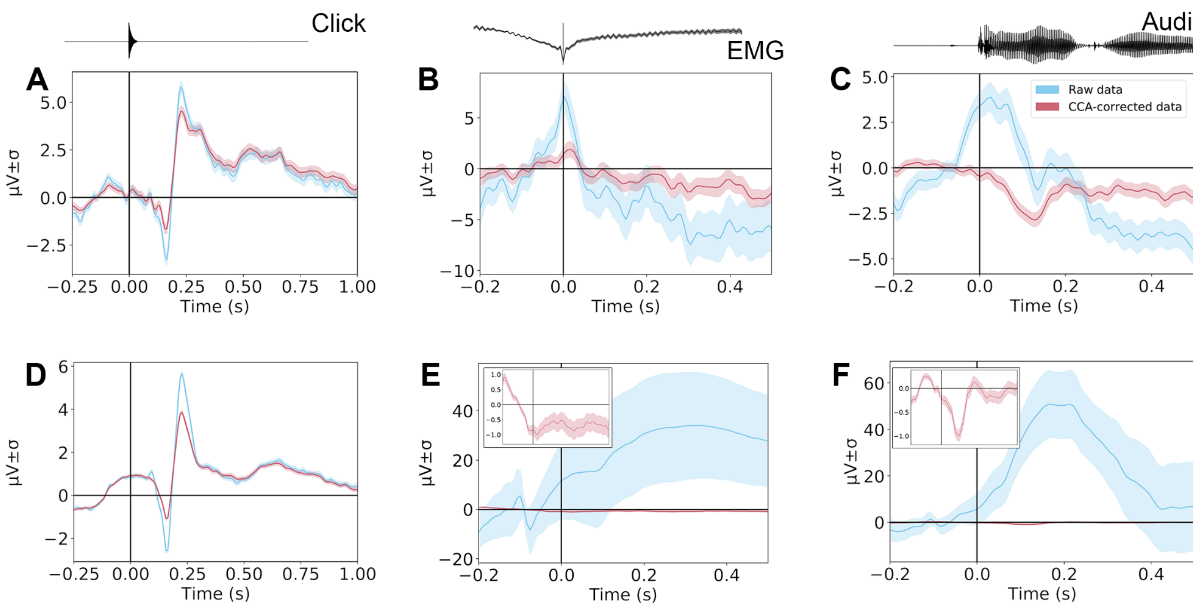

Figure 4. CCA correction removes EMG artifact without significantly affecting auditory responses, as shown by comparison of event-related potential activity between raw data (blue) and CCA-corrected data (red). The top row of panels [A,B,C] shows responses in a single subject (OP0008) while the bottom row of panels [D,E,F] shows grand average responses in 17 subjects. Left column [A,D]: ERP responses epoched to the inter-trial click tone. Middle column [B,E]: ERP responses epoched to EMG activity recorded from facial electrodes. Right column [C,F]: ERP responses epoched to the onset of sentence articulation. All panels include data averaged across nine electrodes: F1, Fz, F2, FC1, FCz, FC2, C1, Cz, and C2.

3.2 Verification of sensitivity of EMG artifact removal

A lack of Type II error in this dataset means that EMG activity is accurately removed from the dataset. EMG activity associated with articulation was identified automatically using activity from the facial electrodes (Figure 2B). A comparison of EMG-epoched responses between the raw dataset and the CCA-corrected dataset showed a large reduction in amplitude corresponding with the onset of EMG activity, which suggests that CCA was successful in removing EMG activity from the dataset (Figure 4B, E). A similar comparison using sentence-onset epoched instead of EMG-epoched responses showed that there was a large reduction in amplitude corresponding with sentence articulation (Figure 4C, F). Manual inspection of 109 jaw clench artifacts in an individual subject (OP0008) demonstrated that EMG activity not associated with peak activity from facial electrodes or sentence production was also removed from CCA-corrected data ($p < 0.0001$, Wilcoxon signed-rank test). Linear mixed effects models comparing sets of RMS difference waves are summarized in Figure 5. This result is consistent with the simultaneous removal of EMG artifact and preservation of neural responses revealed through visual inspection of the ERP data. Overall, there was close to zero difference in the amplitude of the click responses before and after artifact correction (mean RMS difference -2.99 µV). Because the click epochs take place during passive listening, this suggests that CCA is not removing non-EMG activity from the dataset. On the other hand, EMG and sentence epochs demonstrated a large difference wave before and after CCA correction (mean RMS difference 89.06 and 42.77 µV, respectively). This is because unlike the inter-trial click tone, sentence and EMG epochs were directly associated with overt articulation of speech.

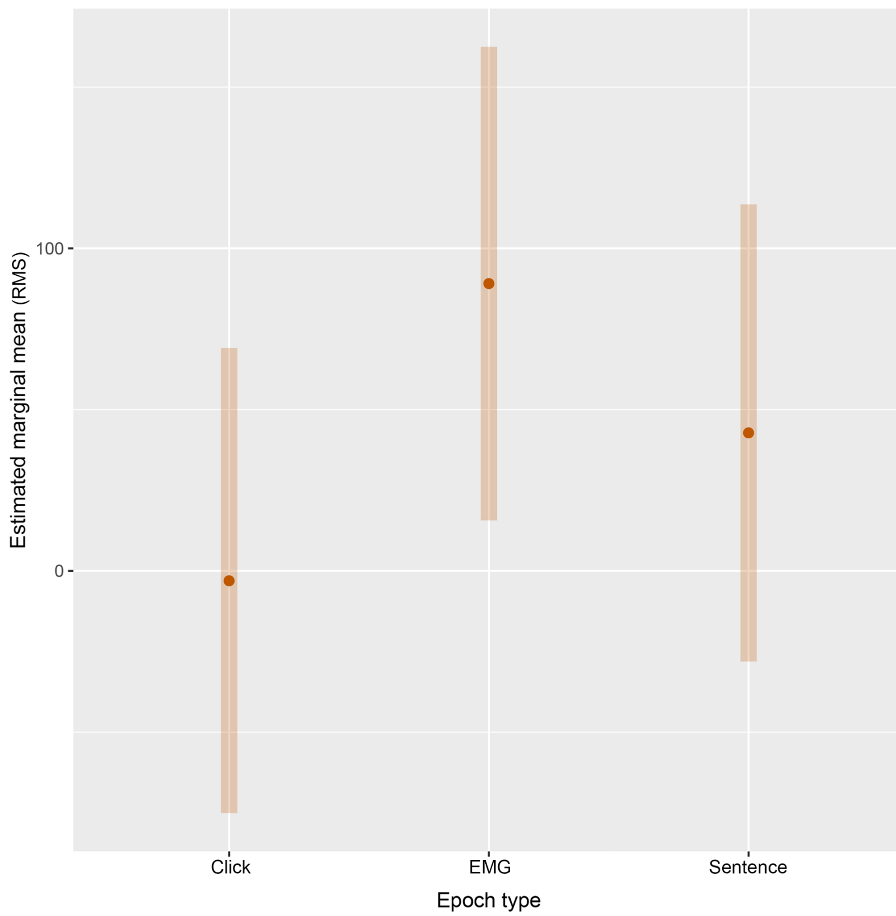

Figure 5. RMS difference waves before and after CCA artifact correction show a large difference for EMG and sentence epochs, but not for click epochs. Estimated marginal means of the difference waves from the linear mixed-effects model are split by epoch type. Shading represents the confidence interval for each epoch type calculated via Kenward-Roger approximation.

4. Discussion

EMG artifact was perceived by speech production researchers as an insurmountable hurdle to analyzing EEG data for a long period of time. While studies occasionally successfully analyze speech production data after correcting for EMG artifact, the procedure is far from standardized. Furthermore, analyses of these results are only performed at the single-word level with a handful of exceptions (Goregliad Fjaellingsdal et al., 2020; Pelayo-González et al., 2020; Ries et al., 2021). The problem of EMG artifact is further complicated by the lack of ground truth in artifact correction, meaning there is not a guaranteed method of confirming both the success and accuracy of an artifact correction technique. This study demonstrates the successful removal of EMG from a naturalistic, sentence-level speech production task using CCA. Additionally, we have provided some methods for confirming the effectiveness and accuracy of EMG artifact removal. Analysis of ERP data epoched to EMG-related activity recorded from facial electrodes and epoched to task-related activity reveals a significant reduction in EMG artifact for CCA-corrected EEG when compared to raw EEG. These results confirm that EMG activity is being reduced from the data. A nonsignificant reduction in response amplitude for inter-trial click epochs between pre-CCA and post-CCA EEG suggests that the integrity of neural signal is preserved after CCA correction.

These validations of the BSSCCA technique for continuous speech data are framed in terms of statistical error, such that a Type I (false positive) would represent neural activity being interpreted as EMG artifact by the algorithm and erroneously removed, leading to overcorrection, while Type II error would represent uncorrected EMG in the signal, or undercorrection. There is a degree of arbitration to the checks presented in this paper which comes from EEG activity being an inextricable mixture of electrical potential from many sources. In an effort to combat this, we have provided certification techniques for CCA from multiple events within the dataset: (1) EMG epochs automatically generated from auxiliary facial electrodes; (2) task-relevant sentence epochs; (3) task-independent inter-trial click epochs. We chose to include epochs from multiple sources in an effort to provide a more functional measure for the effectiveness of CCA given the lack of ground truth in EEG. However, other checks on the efficacy of CCA are possible. One such check is to examine task-relevant neural components before and after CCA. A comparison of peak-to-peak amplitudes may provide more information on the preservation of neural responses after CCA. Different time intervals within sentence production epochs may also provide insight on what signal is being removed through CCA: recall that historically, speech production researchers have analyzed pre-articulatory or post-articulatory epochs in order to avoid during-articulation EMG artifact. In theory, looking at these time windows before and after CCA should show no major differences, or, a preservation of neural signal.

It is also possible that there are additional ways to optimize EMG artifact correction outside the blind source separation algorithm used: many EEG studies perform manual artifact rejection and/or reject trials that fall outside an arbitrary threshold (e.g., ±10SD of the mean voltage for a channel). More deliberate setting of this threshold relative to the amplitude of EMG artifacts may provide an additional check for artifact removal in addition to artifact correction techniques. An alternative approach is to instead remove researcher degrees of freedom by handling artifact rejection through statistics-based methods: Bayesian linear mixed-effects models have been demonstrated as a successful artifact rejection technique that can increase the statistical power of a dataset without making any assumptions about the source of the EEG activity (Alday & van Paridon, 2021).

5. Conclusion

To our knowledge, this is the first paper to assess EMG artifact correction techniques in a naturalistic, sentence-level speech production dataset. EMG artifact has caused many speech researchers to assume that the study of above-word-level speech production using EEG is impractical due to the level of noise in the resulting dataset or overcorrection of the neural response by aggressive blind source separation techniques. Here we used canonical correlation analysis (CCA) to demonstrate both (1) removal of EMG artifact from the signal and (2) preservation of responses during passive listening in the signal through an examination of three sets of event-related potentials from task-related and non-task-related epochs. We hope that this study serves as a "proof-of-concept" for speech researchers who wish to study speech production with EEG due to its rapid temporal resolution, and we hope we inspire some groups to try such an experiment out for themselves.

6. Acknowledgements
The authors would like to thank Maansi Desai, Mary Lowery, and Ian Griffith for their assistance in data collection. In addition, the authors would also like to thank undergraduate research volunteers Nicole Currens, Jade Holder, Claire Huber, Amanda Martinez, Valerie Rae Mercado, Paranjaya Pokharel, Christopher Truong, and Cassandra Villareal for the assistance in phonetic transcription of the data. Lastly, the authors would like to thank Stéphanie Riès for her assistance in designing the preprocessing steps of the experiment. GK and LSH designed the study along with consultation from RLS. GK collected the data and performed the analysis. GK, RLS, and LSH contributed to manuscript preparation and editing. The authors have no conflicts of interest or relevant disclosures of competing interests.

References

Alday, P. M., & van Paridon, J. (2021). *Away from arbitrary thresholds: using robust statistics to improve artifact rejection in ERP*. https://doi.org/10.31234/osf.io/wqrb5

Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, *7*(6), 1129–1159.

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, *29*(4), 1165–1188.

BOERSMA, & P. (2001). Praat, a system for doing phonetics by computer. *Glot. Int.*, *5*(9), 341–345.

Brumberg, J. S., & Pitt, K. M. (2019). Motor-Induced Suppression of the N100 Event-Related Potential During Motor Imagery Control of a Speech Synthesizer Brain-Computer Interface. *Journal of Speech, Language, and Hearing Research: JSLHR*, *62*(7), 2133–2140.

Burgess, R. C. (2020). Recognizing and Correcting MEG Artifacts. *Journal of Clinical Neurophysiology: Official Publication of the American Electroencephalographic Society*, *37*(6), 508–517.

D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology: CB*, *19*(5), 381–385.

De Clercq, W., Vergult, A., Vanrumste, B., Van Paesschen, W., & Van Huffel, S. (2006). Canonical correlation analysis applied to remove muscle artifacts from the electroencephalogram. *IEEE Transactions on Bio-Medical Engineering*, *53*(12 Pt 1), 2583–2587.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21.

Dimigen, O. (2020). Optimizing the ICA-based removal of ocular EEG artifacts from free viewing experiments. *NeuroImage*, *207*, 116117.

Friedman, B. H., & Thayer, J. F. (1991). Facial muscle activity and EEG recordings: redundancy analysis. *Electroencephalography and Clinical Neurophysiology*, *79*(5), 358–360.

Friston, K. J., Williams, S., Howard, R., Frackowiak, R. S. J., & Turner, R. (1996). Movement-Related effects in fMRI time-series. In *Magnetic Resonance in Medicine* (Vol. 35, Issue 3, pp. 346–355). https://doi.org/10.1002/mrm.1910350312

Gao, J., Zheng, C., & Wang, P. (2009). Automatic Removal of Ocular Artifacts from EEG Signals. In *2009 2nd International Conference on Biomedical Engineering and Informatics*. https://doi.org/10.1109/bmei.2009.5305540

Gomez-Herrero, G., De Clercq, W., Anwar, H., Kara, O., Egiazarian, K., Van Huffel, S., & Van Paesschen, W. (2006). Automatic Removal of Ocular Artifacts in the EEG without an EOG Reference Channel. *Proceedings of the 7th Nordic Signal Processing Symposium - NORSIG 2006*, 130–133.

Goncharova, I. I., McFarland, D. J., Vaughan, T. M., & Wolpaw, J. R. (2003). EMG contamination of EEG: spectral and topographical characteristics. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *114*(9), 1580–1593.

Goregliad Fjaellingsdal, T., Schwenke, D., Scherbaum, S., Kuhlen, A. K., Bögels, S., Meekes, J., & Bleichner, M. G. (2020). Expectancy effects in the EEG during joint and spontaneous word-by-word sentence production in German. *Scientific Reports*, *10*(1), 5460.

Gramfort, A., 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.

Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, *55*(4), 468–484.

Hamilton, L. S., & Huth, A. G. (2020). The revolution will not be controlled: natural stimuli in speech neuroscience. In *Language, Cognition and Neuroscience* (Vol. 35, Issue 5, pp. 573–582). https://doi.org/10.1080/23273798.2018.1499946

Hawco, C. S., Jones, J. A., Ferretti, T. R., & Keough, D. (2009). ERP correlates of online monitoring of auditory feedback during vocalization. *Psychophysiology*, *46*(6), 1216–1225.

Heinks-Maldonado, T. H., Mathalon, D. H., Houde, J. F., Gray, M., Faustman, W. O., & Ford, J. M.

(2007). Relationship of imprecise corollary discharge in schizophrenia to auditory hallucinations. *Archives of General Psychiatry*, *64*(3), 286–296.

Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, *532*(7600), 453–458.

Islam, M. K., Rastegarnia, A., & Yang, Z. (2016). Methods for artifact detection and removal from scalp EEG: A review. *Neurophysiologie Clinique = Clinical Neurophysiology*, *46*(4-5), 287–305.

Jiang, X., Bian, G.-B., & Tian, Z. (2019). Removal of Artifacts from EEG Signals: A Review. *Sensors* , *19*(5). https://doi.org/10.3390/s19050987

Keren, A. S., Yuval-Greenberg, S., & Deouell, L. Y. (2010). Saccadic spike potentials in gamma-band EEG: characterization, detection and suppression. *NeuroImage*, *49*(3), 2248–2263.

Kuznetsova, A., Brockhoff, P., & Christensen, R. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, Articles*, *82*(13), 1–26.

Lerner, Y., Honey, C. J., Silbert, L. J., & Hasson, U. (2011). Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *31*(8), 2906–2915.

Lightfoot, G. (2016). Summary of the N1-P2 Cortical Auditory Evoked Potential to Estimate the Auditory Threshold in Adults. *Seminars in Hearing*, *37*(1), 1–8.

Lijffijt, M., Lane, S. D., Meier, S. L., Boutros, N. N., Burroughs, S., Steinberg, J. L., Moeller, F. G., & Swann, A. C. (2009). P50, N100, and P200 sensory gating: relationships with behavioral inhibition, attention, and working memory. *Psychophysiology*, *46*(5), 1059–1068.

Luck, S. J. (n.d.). *An Introduction to the Event-Related Potential Technique*.

Martikainen, M. H., Kaneko, K.-I., & Hari, R. (2005). Suppressed responses to self-triggered sounds in the human auditory cortex. *Cerebral Cortex* , *15*(3), 299–302.

Matusz, P. J., Dikker, S., Huth, A. G., & Perrodin, C. (2019). Are We Ready for Real-world Neuroscience? *Journal of Cognitive Neuroscience*, *31*(3), 327–338.

Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., & Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. *Current Biology: CB*, *17*(19), 1692–1696.

Okada, K., Matchin, W., & Hickok, G. (2018). Phonological Feature Repetition Suppression in the Left Inferior Frontal Gyrus. *Journal of Cognitive Neuroscience*, *30*(10), 1549–1557.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 156869.

Pelayo-González, H. J., Granados-Ramos, D., Alcaraz-Romero, V., & Reyes-Meza, V. (2020). Electroencephalographic activity during sentence production in healthy and schizophrenic men. *Revista Mexicana de Neurociencia*, *21*(2). https://doi.org/10.24875/rmn.19000132

Rastatter, M., & De Jarnette, G. (1984). EMG activity with the jaw fixed of orbicularis oris superior, orbicularis oris inferior and masseter muscles of articulatory disordered children. *Perceptual and Motor Skills*, *58*(1), 286.

Ries, S. K., Pinet, S., Nozari, N. B., & Knight, R. T. (2021). Characterizing multi-word speech production using event-related potentials. *Psychophysiology*, *58*(5), e13788.

Sejnowski, T. J., Churchland, P. S., & Movshon, J. A. (2014). Putting big data to good use in neuroscience. *Nature Neuroscience*, *17*(11), 1440–1441.

Shuster, L. I. (2003). *fMRI and Normal Speech Production*. https://doi.org/10.1044/nnsld13.3.16

Singh, T., Phillip, L., Behroozmand, R., Gleichgerrcht, E., Piai, V., Fridriksson, J., & Bonilha, L. (2018). Pre-articulatory electrical activity associated with correct naming in individuals with aphasia. *Brain and Language*, *177-178*, 1–6.

Stepp, C. E. (2012). Surface Electromyography for Speech and Swallowing Systems: Measurement, Analysis, and Interpretation. *Journal of Speech, Language, and Hearing Research*. https://doi.org/10.1044/1092-4388(2011/11-0214

Van Eijden, T. M., Blanksma, N. G., & Brugman, P. (1993). Amplitude and timing of EMG activity in the human masseter muscle during selected motor tasks. *Journal of Dental Research*, *72*(3), 599–606.

Vos, D. M., Riès, S., Vanderperren, K., Vanrumste, B., Alario, F.-X., Huffel, V. S., & Burle, B. (2010).

Removal of Muscle Artifacts from EEG Recordings of Spoken Language Production. *Neuroinformatics*, *8*(2), 135–150.

Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, *41*(8), 989–994.

Wehbe, L., Murphy, B., Talukdar, P., Fyshe, A., Ramdas, A., & Mitchell, T. (2014). Simultaneously Uncovering the Patterns of Brain Regions Involved in Different Story Reading Subprocesses. In *PLoS ONE* (Vol. 9, Issue 11, p. e112575). https://doi.org/10.1371/journal.pone.0112575

Wilson, S. M., Saygin, A. P., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, *7*(7), 701–702.

Wrench, A. (1999). *The MOCHA-TIMIT articulatory database*.

Yuan, J., & Liberman, M. (2008). Speaker identification on the SCOTUS corpus. *The Journal of the Acoustical Society of America*, *123*(5), 3878.