Correcting MEG Artifacts Caused by Overt Speech

Omid Abbasi[1]*, Nadine Steingräber[1] and Joachim Gross[1,2,3]

[1] *Institute for Biomagnetism and Biosignal Analysis, University of Münster, Münster, Germany,* [2] *Otto-Creutzfeldt-Center for Cognitive and Behavioral Neuroscience, University of Münster, Münster, Germany,* [3] *Centre for Cognitive Neuroimaging, University of Glasgow, Glasgow, United Kingdom*

Recording brain activity during speech production using magnetoencephalography (MEG) can help us to understand the dynamics of speech production. However, these measurements are challenging due to the induced artifacts coming from several sources such as facial muscle activity, lower jaw and head movements. Here, we aimed to characterize speech-related artifacts, focusing on head movements, and subsequently present an approach to remove these artifacts from MEG data. We recorded MEG from 11 healthy participants while they pronounced various syllables in different loudness. Head positions/orientations were extracted during speech production to investigate its role in MEG distortions. Finally, we present an artifact rejection approach using the combination of regression analysis and signal space projection (SSP) in order to correct the induced artifact from MEG data. Our results show that louder speech leads to stronger head movements and stronger MEG distortions. Our proposed artifact rejection approach could successfully remove the speech-related artifact and retrieve the underlying neurophysiological signals. As the presented artifact rejection approach was shown to remove artifacts arising from head movements, induced by overt speech in the MEG, it will facilitate research addressing the neural basis of speech production with MEG.

Keywords: speech production, MEG, movement artifact, head movement, signal space projection, regression analysis

INTRODUCTION

Our everyday life and especially social interactions are greatly affected by verbal communication and, consequently, the question of how our brain is able to produce and understand spoken language has been the focus of many studies in the field of cognitive neuroimaging. fMRI studies have the spatial resolution to potentially provide a spatially detailed account of brain areas involved in speech perception and production (Hickok, 2012; Hickok and Poeppel, 2016). However, scanner noise and sensitivity to head movements caused by overt speech make fMRI less than ideal – a limitation that is further aggravated by the limited temporal resolution of fMRI. In comparison, magnetoencephalography (MEG) and electroencephalography (EEG) provide excellent temporal resolution and are completely silent recording techniques (Gross, 2019). This makes them ideal candidates to non-invasively study the dynamics of speech perception and production. However, particularly speech production studies are challenging due to the induced artifacts coming from several sources such as facial muscle activity, lip and eye movements, and head movements (Ewald et al., 2012). Therefore, earlier studies have avoided recording neurophysiological activities during

speech production. Some tried to simply circumvent the induced artifacts by delaying continuous speech, using silent naming, or manual responses (Eulitz et al., 2000; Schmitt et al., 2000; Ewald et al., 2012). However, these approaches compromise our ability to directly investigate neural activity underlying speech production.

Recently, some studies have used MEG and EEG to study speech production (Masaki et al., 2001; De Vos et al., 2010; Ganushchak et al., 2011; Llorens et al., 2011; Ewald et al., 2012; Alexandrou et al., 2016; Munding et al., 2016; Meyer, 2018; Bourguignon et al., 2020). While EEG is less affected by head movement compared to MEG, source localization is generally thought to be more precise with MEG (Gross, 2019).

Therefore, here we focus on the use of MEG for studying speech production. While artifacts are an obvious problem for these studies, they have so far not been carefully described. In this study, we seek to comprehensively characterize speech artifacts in MEG and evaluate the performance of correction techniques with the aim to optimize source localization.

Several artifact correction approaches have been used to reduce speech artifacts in MEG signals. Temporal signal space separation (tSSS; Taulu and Simola, 2006) removes MEG artifacts originating from sources outside the brain as well as nearby sources. While tSSS can be used to compensate for head movements and suppress external artifacts, this is not suitable to remove facial muscle artifacts (which originate inside the sensor array).

In addition, independent component analysis (ICA) has been used to remove speech-related artifacts from the recorded MEG data (Alexandrou et al., 2017; Bourguignon et al., 2020). These studies used ICA to decompose data from the sensor level to the component level. Then, components related to artifacts were detected and removed from data. ICA is in principle also suitable for removing muscle artifacts but requires an accurate selection of artifactual components (Gross et al., 2013).

Besides head movements and muscle artifacts, movements of the lower jaw are another prominent source of artifacts during speech production. Most participants have tooth fillings, retainers, or other dental work that often has residual magnetic activity even after demagnetization. This can lead to strong artifacts that are independent of head movements. Since head movements are tracked by coils attached to the scalp, lower jaw movements are not corrected by head movement correction methods.

Here, we present an analysis pipeline that aims to correct these different types of artifacts using non-commercial methods that are applicable to data from any MEG system. We start by characterizing the dynamics of head movement and recorded facial electromyogram (EMG) activity relative to the produced speech envelope and the corresponding artifacts in the MEG signal. Finally, we present and evaluate a system-independent approach to remove the artifacts. This approach uses regression analysis for head movement correction and signal space projection (SSP) for remaining artifacts such as those arising from lower jaw movements.

To test this approach, we recorded MEG data from 11 participants who were instructed to pronounce different syllables in various loudnesses. We demonstrate that the head movement level is directly linked to the loudness level and leads to MEG distortion. Our proposed artifact rejection scheme was able to remove the movement-related artifact from MEG and retrieve neurophysiological signals.

MATERIALS AND METHODS

In this study, we aimed to comprehensively characterize movement-related artifacts induced during speech production and subsequently present an approach to remove these artifacts from MEG data. Therefore, we analyzed MEG data from two different experiments (*EXP1* and *EXP2*). *EXP1* was conducted in order to characterize the induced artifacts and to further investigate the feasibility of removing them from the recorded MEG data. Moreover, we also analyzed data from a larger experiment (*EXP2*) in order to evaluate if our artifact rejection approach removes neurophysiological information.

Participants

We recruited 11 healthy volunteers [6 males, mean age 25.0034 ± 2.7 years, median 25, range (22–31)] for *EXP1* and 25 healthy volunteers [13 males, mean age 24.9 ± 2.8 years, median 25, range (21–32)] for *EXP2* from a local participant pool. The study was approved by the local ethics committee and conducted in accordance with the Declaration of Helsinki. Prior written informed consent was obtained before the measurement and participants received monetary compensation after the experiment.

Recording

Magnetoencephalography, EMG, and speech signals were recorded simultaneously. A 275 whole-head sensor system (OMEGA 275, VSM Medtech Ltd., Vancouver, BC, Canada) was used for all of the recordings with a sampling frequency of 1200 Hz, except the speech recording which had a sampling rate of 44.1 kHz. Audio data were captured with a microphone, which was placed at a distance of 155 cm from the participants' mouth, in order not to cause any artifacts through the microphone itself. Three pairs of EMG surface electrodes were placed after tactile inspection to find the correct location to capture muscle activity from the *m. genioglossus, m. orbicularis oris*, and *m. zygomaticus major* (**Figure 1**). One pair of electrodes was used for each muscle with about 1 cm between electrodes. A low-pass online filter with a 300 Hz cut-off was applied to the recorded MEG and EMG data.

Paradigm

For both experiments participants were asked to sit relaxed while performing the given tasks and to keep their eyes focused on a white fixation cross. In *EXP1* they were instructed to pronounce three different syllables ("La," "Du," and "Pa") and each of them at three different volume levels: normal, loud, and without sound, meaning they should only perform the mouth movement. This makes nine conditions in total. A total of 50 trials were collected for each of them, consisting of 4 s in which participants pronounced the syllable separately 3 to 4 times and 2 s of rest in
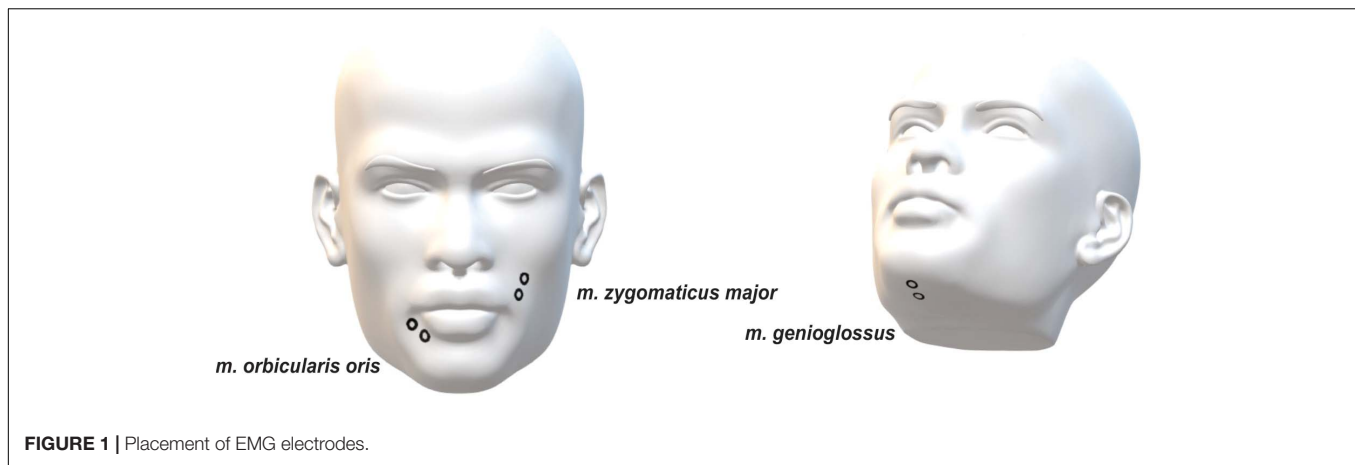
FIGURE 1 | Placement of EMG electrodes.

between trials. A color change from white to blue of the fixation cross indicated the beginning of the 4 s in which participants should speak. After these 4 s the color changed back to white, indicating that participants should stop speaking. A total of 2 s later the color would again change back to blue.

EXP 2 consisted of two separate recordings. The first recording contained speech generation tasks whereas the second one was completely about speech perception.

For the first recording, there were 14 trials for overt speech. Each trial consisted of a 60 s time period in which participants answered a given question. Three more trials were recorded in which subjects were told to speak as they normally would, using only the given syllable ("Du," "La," and "Pa") for three minutes each. As in *EXP1* a color change of the fixation cross from white to blue indicated the beginning of the time period in which participants should speak and the end was marked by a color change back to white.

In the second recording session participants listened to audio-recordings of their own voice, which were collected in the first recording.

Preprocessing and Data Analysis

Prior to data analysis, MEG data were visually inspected for detecting jump artifacts and bad channels. A discrete Fourier transform (DFT) filter was applied to eliminate 50 Hz line noise from the continuous MEG and EMG data. Moreover, EMG data was highpass-filtered at 20 Hz and rectified. Continuous head position and rotation were extracted from the fiducial coils placed at anatomical landmarks (nasion, left, and right ear canals) leading to six time-series representing the instantaneous location (X, Y, and Z) and orientation (Ox, Oy, and Oz) of the head. These time-series are in a 3D Cartesian coordinate system defined relative to the dewar in which X, Y, and Z are from the circumcenter of the three fiducial coils to anterior, left, and superior, respectively, and Ox, Oy, and Oz are the head rotation around X, Y, and Z-axis, respectively. The wideband amplitude envelope of the speech signal was computed using the method presented in Chandrasekaran et al. (2009). Nine logarithmically spaced frequency bands between 100 and 10,000 Hz were constructed by bandpass filtering (third-order, Butterworth
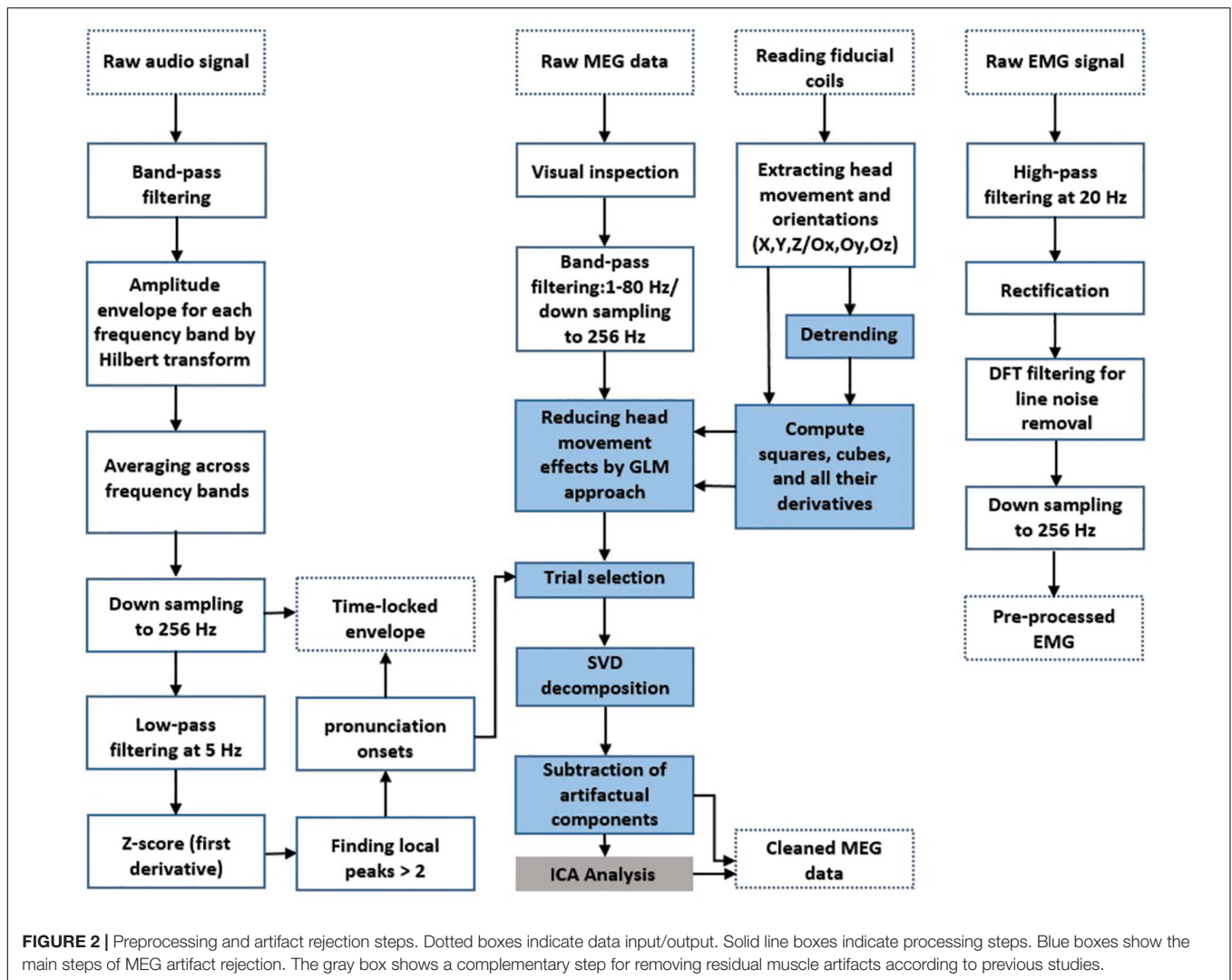
filters). Then, we computed the amplitude envelope for each frequency band as the absolute value of the Hilbert transform and downsampled them to 1200 Hz. Finally, we averaged them across bands and used the computed wideband amplitude envelope for all further analysis. We used speech envelope signals to extract syllable pronunciation onsets for further trial selection steps. First, the envelope signal was low-pass filtered at 5 Hz. Then, we computed the first derivative of the signal followed by z-scoring. Finally, we detected all the local peaks that have Z-value higher than 2. Our multi-step preprocessing procedures are depicted in **Figure 2**.

Magnetoencephalography, EMG, speech envelope, and head movement signals were downsampled to 256 Hz and were segmented from -1 s to 4 s time-locked to speech onset identified as peaks of the first derivative of the speech envelope signal. In the preprocessing and data analysis steps, custom-made scripts in Matlab R2020 (The Mathworks, Natick, MA, United States) in combination with the Matlab-based FieldTrip toolbox (Oostenveld et al., 2011) were used in accord with current MEG guidelines (Gross et al., 2013).

Artifact Rejection

For removing the speech-related artifacts we applied the following procedure:

(i) First, we reduced head movement-related artifacts by incorporating the head position time-series into a general linear model (GLM) using regression analysis (Stolk et al., 2013). We used the original head movement/orientation time-series (X, Y, Z, OX, OY, and OZ) and the same signals after removing slow drifts using a third-order polynomial fit. To account for more complex, non-linear dependencies of artifacts on movement we also calculated for each of these 12 time series their squares, cubes, and the temporal derivatives of all these signals in the model resulting in a total of 72 regressors. We then used the residuals of the GLM for further analysis. (ii) To further remove the residual artifact, we used the SSP method implemented in the Fieldtrip toolbox (Uusitalo and Ilmoniemi, 1997). We estimated the spatial subspace containing the speech-related artifact from the MEG data using SSP. The singular value decomposition (SVD) approach was applied to the MEG data to decompose

FIGURE 2 | Preprocessing and artifact rejection steps. Dotted boxes indicate data input/output. Solid line boxes indicate processing steps. Blue boxes show the main steps of MEG artifact rejection. The gray box shows a complementary step for removing residual muscle artifacts according to previous studies.

data into singular vectors (components). This method works well if it is applied to data where the artifact is maximized due to preprocessing or averaging. In our case, we averaged data time-locked to speech onset thereby enhancing the artifact in the averaged data. However, our results are virtually unchanged when SVD-components are computed from the single-trial data. In both cases, the data is dominated by speech-related artifacts. Therefore, the first SVD components are expected to reflect the artifact subspace. (iii) Artifactual components were detected via visual inspections and mutual information (MI) analysis and then removed from the single-trial data (Abbasi et al., 2016); (iv) finally, all remaining components were back-transformed to the sensor level. The blue boxes in **Figure 2** depict MEG artifact rejection steps.

Our conservative component selection consisted of three steps and once a component met these three criteria, we marked and removed it: First, visual inspection of components' time-series: artifactual components' time-series were affected by movement patterns and were detectable. Second, visual inspection of the components weight distribution on the topographic maps:

components with major power distribution on temporal areas were marked as artifactual components. Third, computing MI between SVD components and head movement/orientation signals: MI was calculated between SVD components and each of the extracted head movement/orientation signals (X, Y, Z, OX, OY, and OZ). These MI values were subsequently averaged across all movement signals for each pair. Components with high MI values were detected as artifactual components.

Artifact removal based on ICA approach: we also tried to remove the induced artifact using ICA approach and compare its performance with our proposed approach. Therefore, we used the above steps except that we replaced the SVD algorithm by the ICA algorithm in the second step. Throughout the analysis, we used the extended Infomax algorithm implemented in FieldTrip.

Artifact Level
In order to quantify the level of artifact and be able to compare the level of residual artifact between different artifact rejection steps, we constructed a spatial filter based on the detected artifactual components in EXP1 according to the following steps:

First, by performing SVD on data covariance matrix "M_{cov}," we obtained all the spatial patterns needed to fully explain the data (column vectors of U):

$$M_{cov} = USV^T \qquad (1)$$

Next, we constructed spatial filters "SF_a" by only selecting the first "K" spatial components identified as artifactual components from the artifact rejection step:

$$SF_A = [U_1, U_2, \ldots, U_k][U_1, U_2, \ldots, U_k]^T \qquad (2)$$

And finally we calculate the artifact level "AL" in every step by applying the constructed filter on the averaged data "M_{avg}"

$$AL = \sum \sum |SF_A M_{avg}| \qquad (3)$$

This approach projects the original data into the artifact subspace spanned by the K spatial components described above and thereby provides a representation of the artifact.

Additionally, we evaluated to what extent the proposed artifact rejection approach removes neurophysiological information. For this aim, we used the spatial filters constructed from artifactual SVD components in the speech production condition to clean the artifact-free syllable perception dataset in *EXP2*. The amount of reduction in MEG components in the syllable perception condition indicates the removal of neurophysiological information since there are no speech-artifacts in the perception condition.

Coherence Analysis

In order to quantify similarities between speech envelopes and EMG signals, we used coherence analysis. To this end, time-locked data were further segmented into short data segments (duration 2 s, overlap 0.5 s). Spectral power and cross-spectral density of these data were calculated using a multi-taper approach with a smoothing of 2 Hz (range 1–50 Hz). Coherence was computed for each participant between speech envelopes and EMG signals. These coherence values were subsequently averaged across all participants for each pair (EMG, Envelope) and each frequency. Finally, for each pair, the coherence values were averaged within the frequency range of 0–5 Hz.

Time-Frequency Analysis

We applied time–frequency analysis on the MEG data before and after artifact rejection in order to evaluate whether our proposed approach is able to retrieve motor activities in the speech production condition. Time–frequency analysis was performed on MEG data by a multitaper approach with a smoothing of 2 Hz on a 400 ms long sliding window. The step size of the sliding window was 25 ms and the spectral resolution was interpolated to 1 Hz.

Statistical Analysis

In our statistical analysis, we aimed to investigate any significant effect of the speech loudness and syllables on head movement/orientation. To meet this purpose, we used linear mixed effect modeling (LMEM). We computed the averaged

head movement/orientation in all directions (X, Y, and Z) and orientations (Ox, Oy, and Oz) for the first 4 s after the syllables pronunciation onset for each participant. All the average head movement/orientation values were entered into a fixed-effect model with loudness and syllable as factors. Moreover, we also investigated whether there are any significant effects of speech loudness, syllables, and EMG signal on the calculated coherence between the speech envelope and EMG signals using the LME method. In this case, the average coherence (1–5 Hz) between speech envelope and EMG signal was calculated for each experimental condition (loudness and syllables) and entered into a fixed-effect model with loudness, syllable, and EMG as factors.

Data Availability

The data used in the current study will be available from the corresponding author upon request subject to a formal data sharing agreement. The Matlab code of the proposed artifact rejection approach is publicly accessible through the Open Science Framework[1].
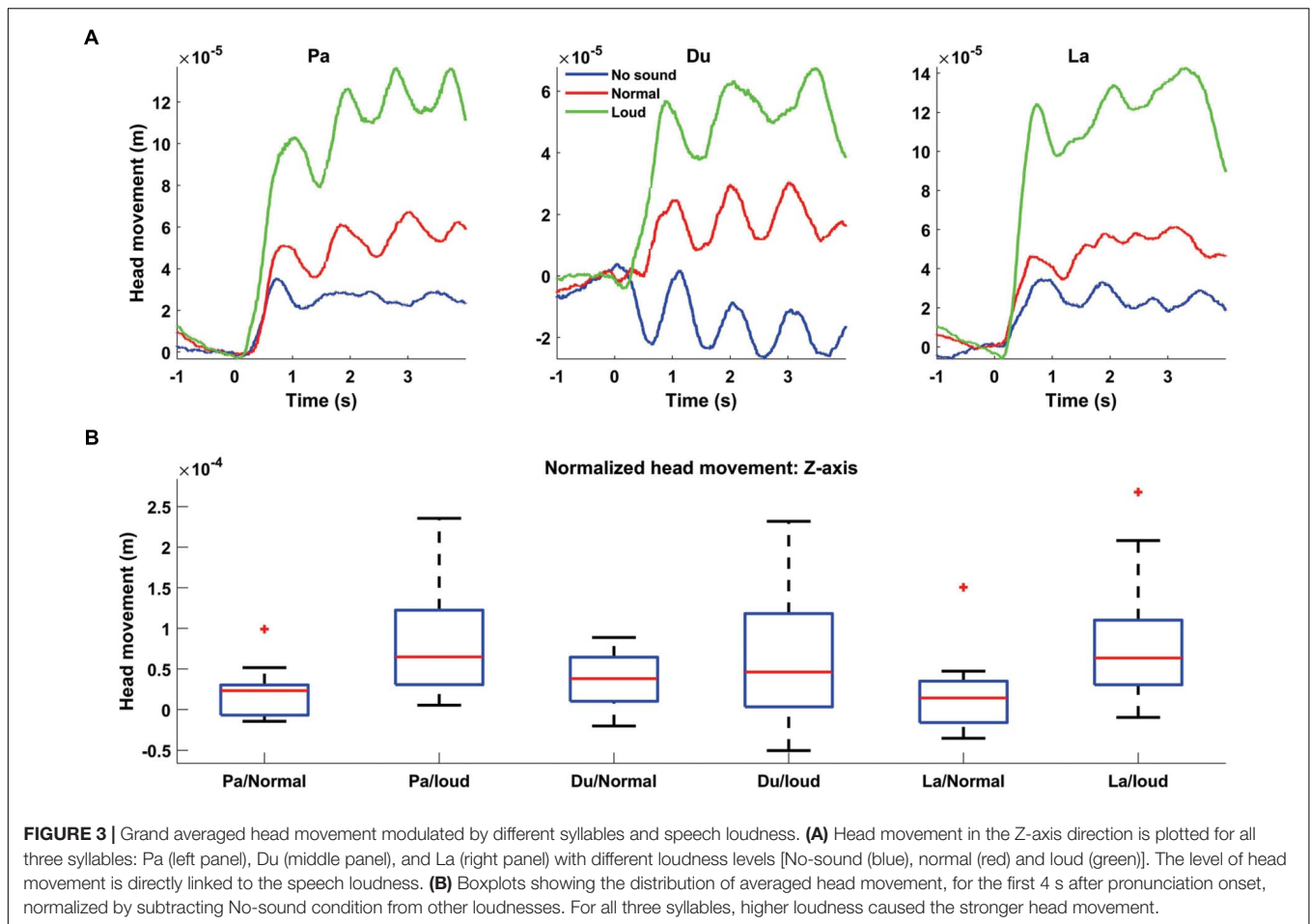
RESULTS

Since head-movement artifact characterization and its removal were two major aims of the current study, we report our results in two separate parts: artifact characterization and artifact rejection:

Artifact Characterization
EXP1-Head Movement Analysis During Speech Production

To investigate whether speech production causes head movement during MEG recording, we extracted continuous head localization information which was captured by fiducial coils. We extracted the head center movement information for different experimental conditions, i.e., different loudness: No-sound/normal/loud as well as different syllables: Pa/Du/La. **Figure 3A** illustrates the Z-axis grand averaged head movement during different syllables pronunciation. These results show that, independently for each syllable, the amount of head movements is directly linked to its loudness level. Next, we calculated the averaged head movement values for the first 4 s after the syllables pronunciation onset. We normalized the averaged value for every participant by subtracting the No-sound condition from other conditions. Boxplots in **Figure 3B** display that higher loudness caused larger Z-axis head movements. We tested for the effect of loudness and syllable on head movements/orientations by performing a linear mixed-effect model. We computed the averaged head movement/orientation in all directions (i.e., X/Y/Z/Ox/Oy/Oz) for the first 4 s after the syllables pronunciation onset. All the average head movement/orientation values were entered into a fixed-effect linear model with loudness and syllable as factors. Our results show that the main effects of loudness were highly significant on the head movement in Z direction [$t(98) = 5.9$, $p < 0.0001$]. Detailed results are presented in **Table 1**.

---
[1]https://osf.io/uc8st/

FIGURE 3 | Grand averaged head movement modulated by different syllables and speech loudness. (A) Head movement in the Z-axis direction is plotted for all three syllables: Pa (left panel), Du (middle panel), and La (right panel) with different loudness levels [No-sound (blue), normal (red) and loud (green)]. The level of head movement is directly linked to the speech loudness. (B) Boxplots showing the distribution of averaged head movement, for the first 4 s after pronunciation onset, normalized by subtracting No-sound condition from other loudnesses. For all three syllables, higher loudness caused the stronger head movement.
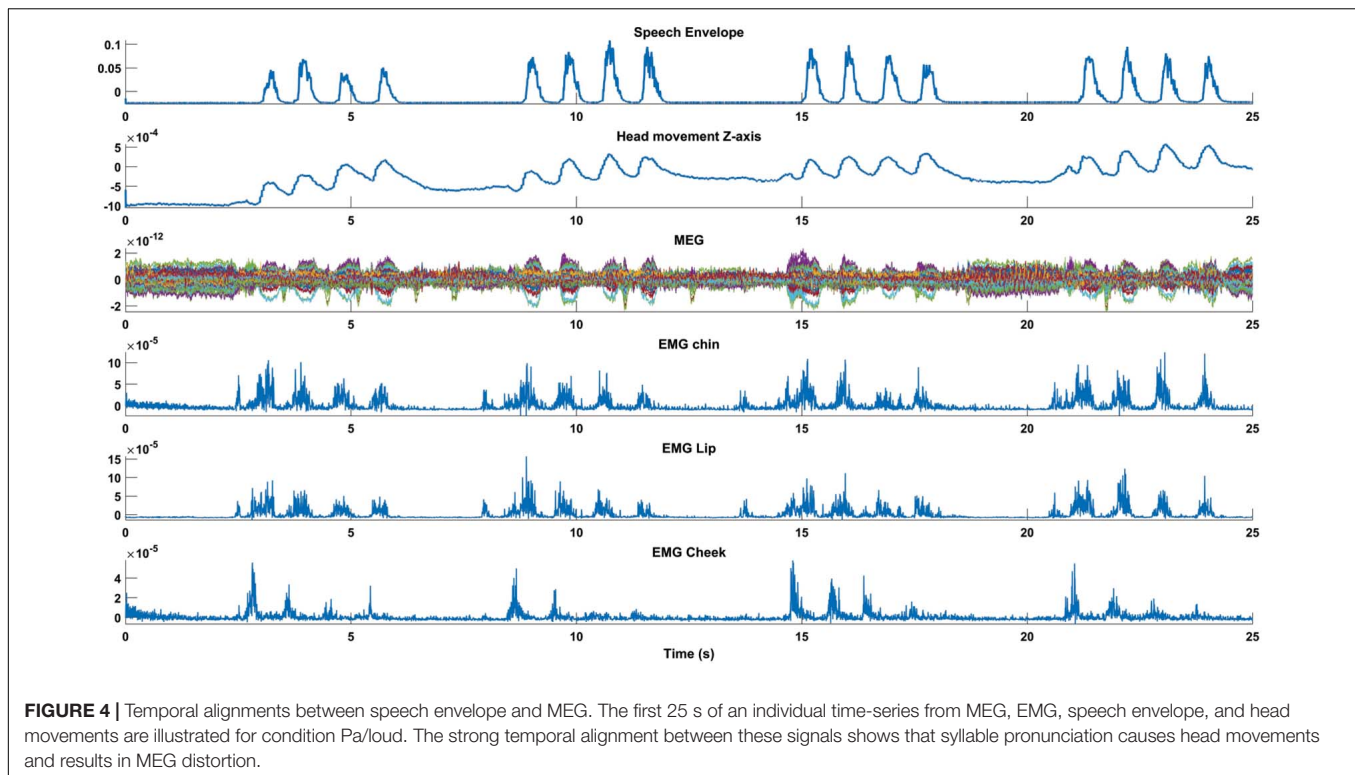
EXP1-Speech Envelope

For further analysis, we calculated the speech signal envelope. We observed a strong temporal alignment of the speech envelope with head position time-series, MEG, and EMG signals recorded from the chin, lip, and cheek during syllables pronunciations (**Figure 4**). The strong temporal alignment between these signals shows that syllable pronunciation could cause head movement and subsequently MEG distortion.

In the next step, we aimed to investigate the temporal relationship between the speech envelope and EMG signals

TABLE 1 | Statistical analysis results.

Head movements	X	Y	Z
Loudness	$t = 1.7$, $p = 0.07$	$t = -2.1$, $p = 0.03$	$t = 5.9$, $p = 4.632e{-}08$
Syllables	$t = 0.8$, $p = 0.39$	$t = 0.4$, $p = 0.63$	$t = 0.1$, $p = 0.9$
Head rotations	**Ox**	**Oy**	**Oz**
Loudness	$t = 3.3$, $p = 0.001$	$t = -0.02$, $p = 0.9$	$t = -3.3$, $p = 0.001$
Syllables	$t = 0.9$, $p = 0.35$	$t = 0.2$, $p = 0.8$	$t = -0.9$, $p = 0.35$

to better understand how they both reflect speech production artifacts in MEG. We therefore calculated coherence between both signals in order to objectively select an EMG channel that best reflects the speech envelope. We normalized the coherence value for every participant/syllable by subtracting the No-sound condition from other experimental conditions. Next, we averaged the normalized coherence values across syllables for every participant. We observed the strongest coherence between EMG Lip signal and speech envelope (**Figure 5A**). Hence, further analysis in this study is done by using EMG Lip signals. We also expected delays between the movement of articulators and onset of speech because auditory and visual speech are temporally coupled (Chandrasekaran et al., 2009; Park et al., 2016, 2018) but typically mouth movements precede auditory speech (Schwartz et al., 2004). Therefore, we computed coherence between EMG and time-shifted speech envelope signal to identify the optimal delay. We used coherence analysis because we are dealing with non-stationary signals where coherence can be optimized by shifting both time series to account for delays between both. **Figure 5B** illustrates the grand averaged coherence for delays between -333 and 333 ms. EMG and speech envelope showed the highest coherence at a negative delay of around -150 ms, indicating, as expected, that movement of articulators precedes auditory speech. Optimal temporal delays varied across

FIGURE 4 | Temporal alignments between speech envelope and MEG. The first 25 s of an individual time-series from MEG, EMG, speech envelope, and head movements are illustrated for condition Pa/loud. The strong temporal alignment between these signals shows that syllable pronunciation causes head movements and results in MEG distortion.

participants (**Figure 5C**). Therefore, for further analysis, the mean optimal temporal delay across the group (-150 ms) was taken into account and the coherence values were recalculated after shifting the speech envelope. The new coherence results illustrate high coherence in normal and loud speaking conditions (**Figure 5D**). We tested for the effect of loudness, syllable, and EMG signal on coherence between EMG and time-shifted speech envelope signal by performing a linear mixed-effect model. We computed the average coherence in the frequency range of 0–5 Hz. All the average coherence values were entered into a mixed-effect model with loudness, syllable, and EMG signal as factors. Our results show that the main effects of loudness and syllable were highly significant [loudness: $t(290) = 9.5$, $p < 0.001$; syllable: $t(290) = -4.3$, $p < 0.001$; EMG: $t(290) = -1.2$, p = 0.2].
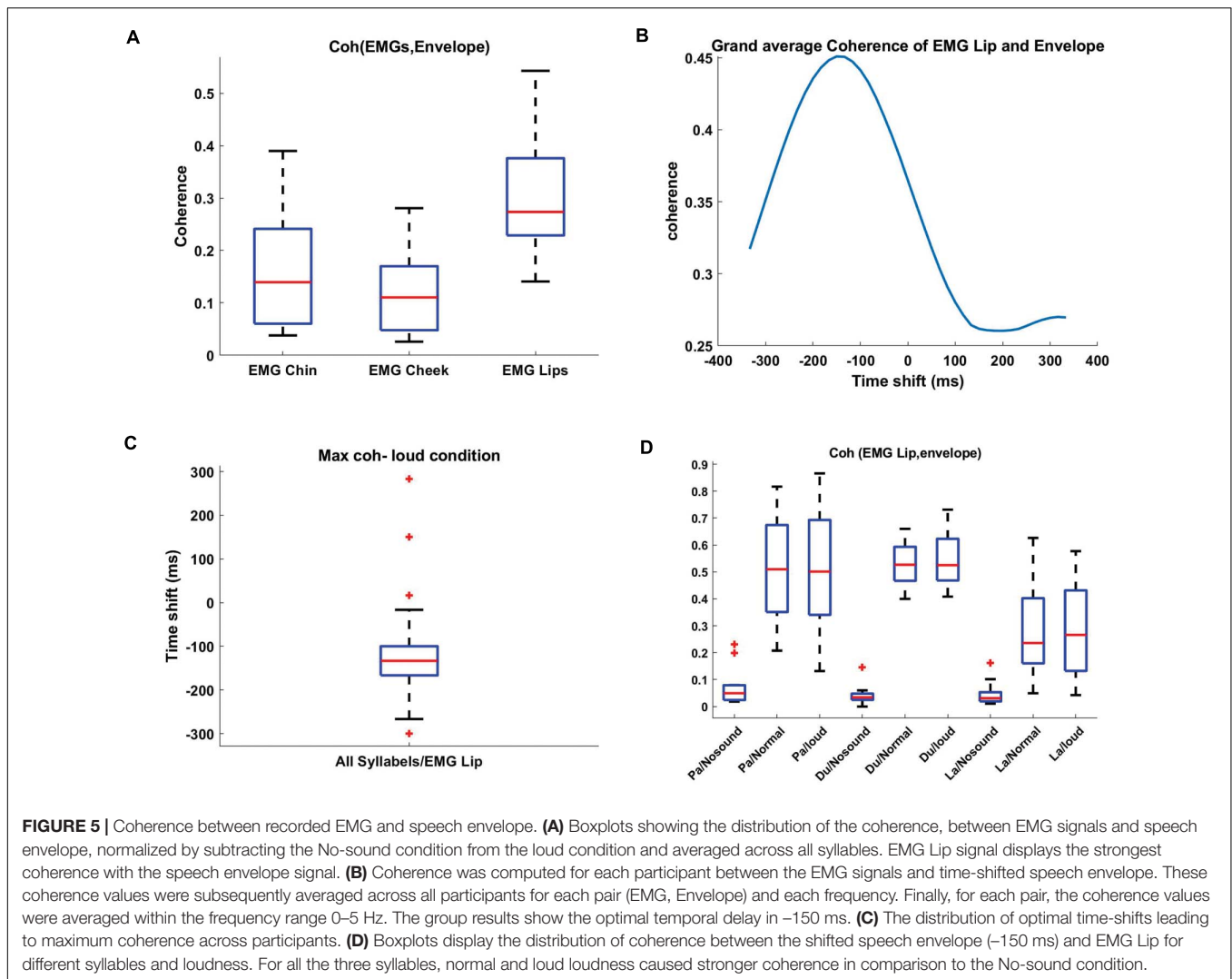
Artifact Rejection
EXP1-Recording MEG Data During Speech Production

Speech production caused head movement in all 11 participants. Subsequently, clear patterns of MEG distortion were observable for all the participants due to their head movement. We observed severe MEG distortion especially for participant #11 due to retainers (**Figure 6**, first panel). Therefore, here we first report the results from this participant to study to what extent we can remove the artifact. A successful artifact correction should ideally lead to data where auditory evoked responses to speech onset can be seen. This is particularly challenging in our case where self-produced speech leads to attenuation of auditory evoked responses (Houde et al., 2002).

Our initial artifact rejection step, the GLM-based approach, could reduce the confounding variance that was caused by head movement (**Figure 6**, second panel). We tried to further remove the induced artifact from the recorded MEG data using both ICA and our proposed method. After decomposing MEG data to independent components (ICs), we observed that ICA could not successfully decompose the data due to the non-stationary nature of the induced artifact. Artifact patterns were observed in most of the components' topographical maps and time-series. Therefore, we only removed the first four components (out of 20) containing the major part of the artifact by showing higher MI with the head movement/orientation signals. Although the neurophysiological pattern (auditory evoked response) was partly detectable shortly after the pronunciation onset, the residual artifact still significantly distorts the back-transformed MEG (**Figure 6**, third panel). Next, we used our proposed method to remove the speech-related artifacts from the recorded MEG data. Only four components were classified as artifactual components based on our criteria. After removing artifactual SVD components, we observed a clear auditory evoked response for data recorded from participant #11 (**Figure 6**, fourth panel).

We applied our proposed artifact rejection approach to clean all the recorded datasets in *EXP1* (participant #1 to #10). On average, 1.4 ± 1.1 components were classified as artifactual (MI mean = 0.16, MI SD = 0.04, range = 0.08–0.23) and 18.06 ± 1.17 components (MI mean = 0.04, MI SD = 0.02, range = 0.01–0.13) were classified as components containing signal of interest.

To evaluate if our artifact rejection approach could remove artifacts and retrieve neurophysiological information, we investigated if we could detect M100 activity before and after

FIGURE 5 | Coherence between recorded EMG and speech envelope. **(A)** Boxplots showing the distribution of the coherence, between EMG signals and speech envelope, normalized by subtracting the No-sound condition from the loud condition and averaged across all syllables. EMG Lip signal displays the strongest coherence with the speech envelope signal. **(B)** Coherence was computed for each participant between the EMG signals and time-shifted speech envelope. These coherence values were subsequently averaged across all participants for each pair (EMG, Envelope) and each frequency. Finally, for each pair, the coherence values were averaged within the frequency range 0–5 Hz. The group results show the optimal temporal delay in –150 ms. **(C)** The distribution of optimal time-shifts leading to maximum coherence across participants. **(D)** Boxplots display the distribution of coherence between the shifted speech envelope (–150 ms) and EMG Lip for different syllables and loudness. For all the three syllables, normal and loud loudness caused stronger coherence in comparison to the No-sound condition.
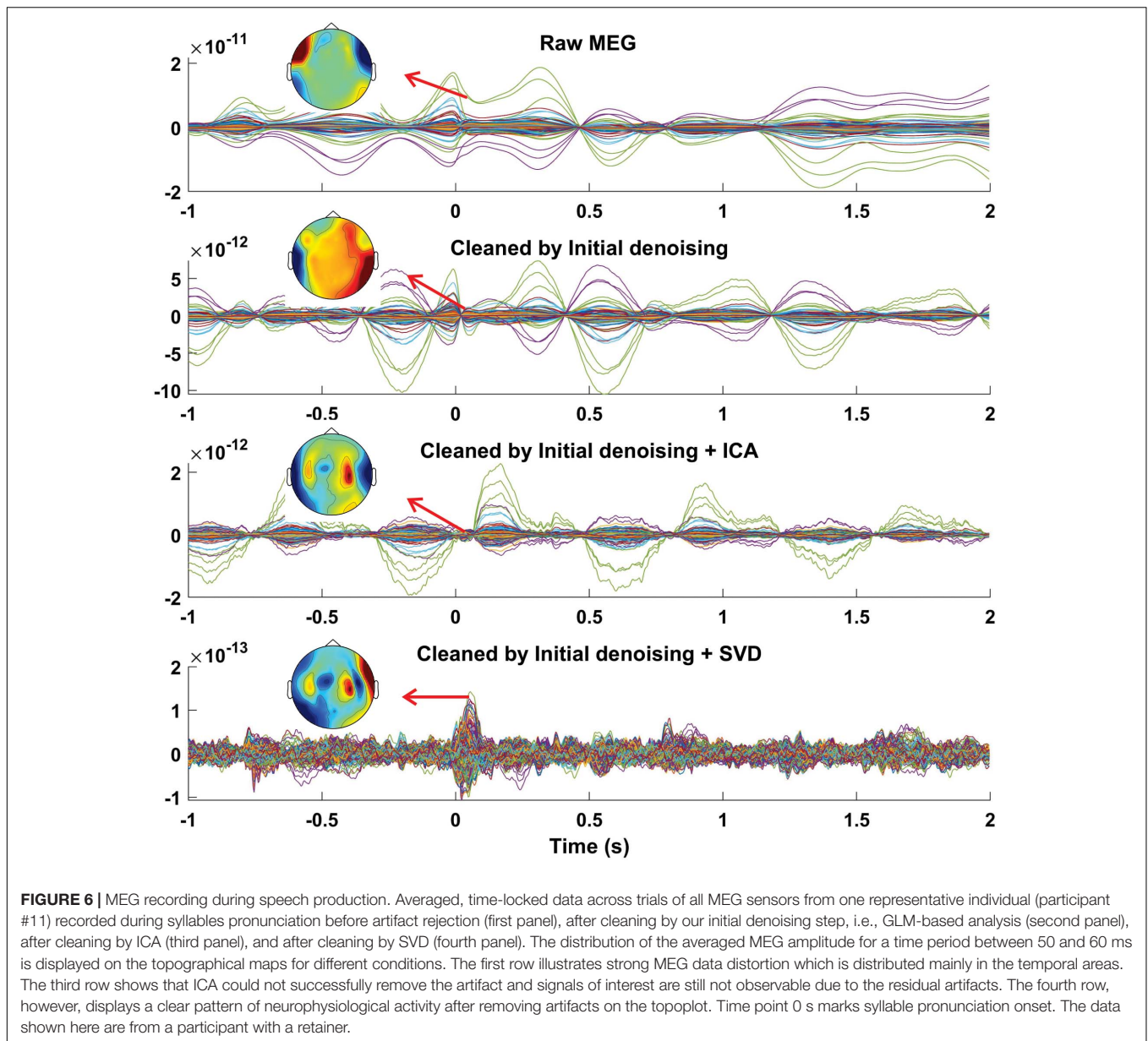
artifact rejection. Before any artifact rejection, it was not possible to detect M100 peak due to the data distortion around the syllables production onset (**Figure 7A**, left panel). However, after artifact removal, neurophysiological activities were detectable 50 ms after the syllables production onset on the grand average data (**Figure 7A**, right panel). The strongest MEG distortion was observed around the syllables pronunciation onset and our artifact rejection approach was capable of largely removing the artifact (**Figures 7B,C**). Residual artifact level was computed for the group data after different cleaning steps, i.e., Raw, GLM, and GLM + SVD. First, we tested for the effect of different artifact correction on the residual artifact level by performing a linear mixed-effect model. We computed the level of the artifact by applying the constructed spatial filter on data. All the artifact level values were entered into a fixed-effect model with method as factor. Our results show that the main effects of artifact corrections were highly significant [Methods: $t(28) = -4.4$, $p = 0.0001$]. Our results showed that our proposed approach could significantly better remove the artifact from the recorded data (**Figure 7C**, right panel).

Additionally, to examine if our proposed approach could retrieve modulated oscillatory neural activity in the motor area, we calculated the time–frequency spectrograms of the MEG sensors located above the left motor cortex before (**Figure 7D**, left panel) and after (**Figure 7D**, right panel) artifact rejection. As it can be seen, beta suppression and rebound were clearly observable before and after artifact rejection and our proposed method did not distort the recorded neurophysiological data.

EXP2-Recording MEG Data During Syllable Perception

We also evaluated if the proposed artifact rejection approach partly removes neurophysiological information. We applied the spatial filters constructed from the speech production dataset to the syllables perception dataset in *EXP2*. Since there is no speech artifact in the perception condition an ideal artifact correction would leave the perception data unchanged. We observed severe MEG distortion for participant #9 due to retainers in the speech production condition. Six components were classified as

FIGURE 6 | MEG recording during speech production. Averaged, time-locked data across trials of all MEG sensors from one representative individual (participant #11) recorded during syllables pronunciation before artifact rejection (first panel), after cleaning by our initial denoising step, i.e., GLM-based analysis (second panel), after cleaning by ICA (third panel), and after cleaning by SVD (fourth panel). The distribution of the averaged MEG amplitude for a time period between 50 and 60 ms is displayed on the topographical maps for different conditions. The first row illustrates strong MEG data distortion which is distributed mainly in the temporal areas. The third row shows that ICA could not successfully remove the artifact and signals of interest are still not observable due to the residual artifacts. The fourth row, however, displays a clear pattern of neurophysiological activity after removing artifacts on the topoplot. Time point 0 s marks syllable pronunciation onset. The data shown here are from a participant with a retainer.

artifactual for this dataset. We constructed a spatial filter from these six components and applied it to the syllables perception condition. We observed a clear M100 peak in response to syllable onset before (**Figure 8A**) and after (**Figure 8B**) artifact correction steps. Importantly, the auditory topography was not attenuated or distorted after artifact correction. Instead, small artifacts in temporal and frontal areas were reduced leading to an overall cleaner topography. Next, we quantified the performance of our artifact correction at the group level. We constructed the SSP filters for every participant in *EXP2* from the speech production dataset. After applying the constructed filter to syllable perception dataset, we observed that the proposed method leads to negligible changes of the M100 amplitude (median: 0%; 25th percentile: −5%; 75th percentile: 0.7%; **Figures 8C,D**). However, it should be noted that physiological activity might be reduced more by

artifact correction if the topographies of physiological and artifact signals show stronger collinearity.

In summary, this indicates that the SSP correction specifically targets the speech artifact and leaves the real neurophysiological signals of interest largely intact.

DISCUSSION

In this study, we carefully characterize the effects of head movement on the recorded MEG data during speech production. We demonstrate that MEG data is distorted due to head movement and the severity of the distortion is directly linked to the speech loudness level. We present an artifact rejection approach to remove artifacts caused by head movement from
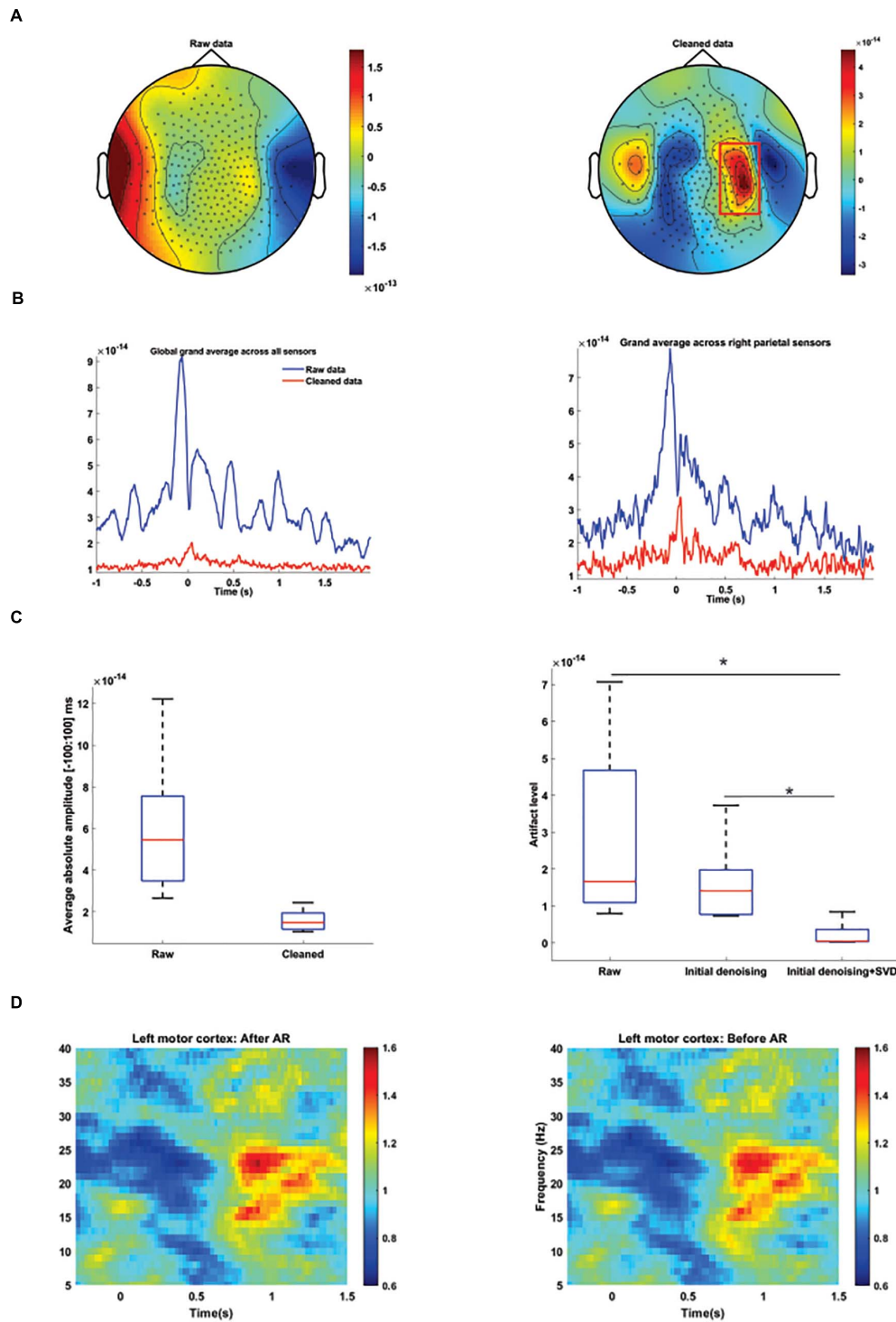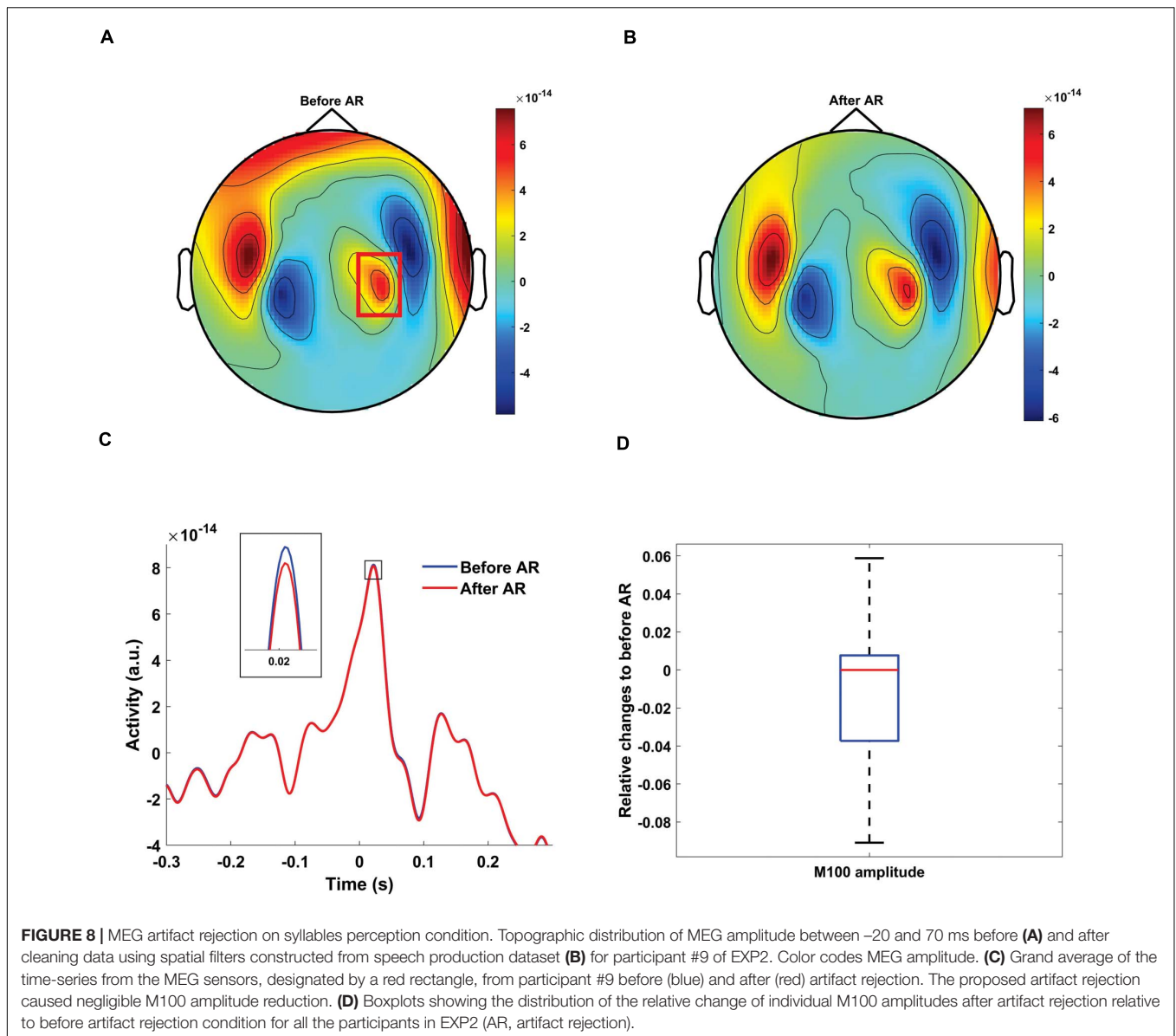
FIGURE 7 | MEG artifact rejection. **(A)** Average topographic distribution of MEG amplitude between 40 and 50 ms for raw data (left column) and after cleaning data by our proposed method (right column). M100 activity was masked before artifact rejection by speech related artifact. Color codes MEG amplitudes. **(B)** Absolute grand average of the time-series from all MEG sensors (left column) and sensors above the right parietal area (right column) from all the participants before (blue) and after (red) artifact rejection. Selected sensors are designated by a red rectangular. **(C)** Boxplots (left panel) showing the distribution of the average values of MEG amplitude between −100 and 100 ms across participants for raw and cleaned data. Right panel boxplots illustrate the distribution of artifact level after data cleaning by different approaches. Our proposed method could significantly remove artifacts from MEG data in comparison to other methods. Significant p-values (after t-test analysis) are indicated. $*p < 0.01$. **(D)** Time-frequency spectrograms of MEG sensors located above the left motor cortex before (left panel) and after (right panel) artifact rejection. Beta suppression and rebound were clearly observable before and after artifact rejection. Color codes relative changes. Time point 0 s marks syllable pronunciation onset.

FIGURE 8 | MEG artifact rejection on syllables perception condition. Topographic distribution of MEG amplitude between –20 and 70 ms before **(A)** and after cleaning data using spatial filters constructed from speech production dataset **(B)** for participant #9 of EXP2. Color codes MEG amplitude. **(C)** Grand average of the time-series from the MEG sensors, designated by a red rectangle, from participant #9 before (blue) and after (red) artifact rejection. The proposed artifact rejection caused negligible M100 amplitude reduction. **(D)** Boxplots showing the distribution of the relative change of individual M100 amplitudes after artifact rejection relative to before artifact rejection condition for all the participants in EXP2 (AR, artifact rejection).

MEG data. Our approach is based on a combination of regression analysis and SSP. The proposed method was successfully applied to the MEG data recorded during pronouncing syllables with different loudness.

Using the real-time head center positions extracted from the head localizer coils, we were able to accurately track head positions during different experimental conditions. We observed that the head movement level is directly linked to the loudness level of the pronunciation of the syllables. This loudness-level-dependent movement can easily distort the recorded MEG data during speech production studies. In addition to general head movement, the unique pattern of lower jaw movement and facial articulatory muscles, for producing every single word, add a complex artifact pattern to the recorded data especially when participants have some ferromagnetic implants in their mouth such as retainers. Moreover, we observed a strong connection

between speech envelopes and the facial articulatory muscles involved in speech production. Taken together, there are several different sources that contribute to MEG distortion during speech production studies. One of the major sources is the head and lower jaw movement. In this study, we mainly focus on removing this type of artifact.

A few earlier studies have investigated speech production using MEG. They used different strategies to either avoid or remove the speech-related artifacts. In the following, we discuss these approaches and compare them with our proposed artifact rejection approach.

Earlier studies tried to simply circumvent the induced artifacts by delaying continuous speech, silent naming, or manual responses (Eulitz et al., 2000; Schmitt et al., 2000; Liljeström et al., 2008; Sahin et al., 2009; Ewald et al., 2012). However, these approaches could prevent us from directly

investigating neural activities underlying speech production. In another study (Saarinen et al., 2006), reported the feasibility of investigating cortical oscillations around 20 Hz using MEG without artifact rejection. While they reported reliable observations, this approach might be only feasible for special cases such as extracting some features which are not affected by speech-related artifacts. Using EEG, other studies tried to investigate speech-related brain activity by removing artifactual epochs (Ewald et al., 2012) or by using heavy low-pass filtering (Masaki et al., 2001). While these approaches are transferable to MEG, they are not applicable when we record data during continuous speech tasks, which could be distorted by head movement continuously, especially in lower frequency bands.

Temporal signal space separation (Taulu and Simola, 2006) was developed to remove artifacts from MEG signals. This approach removes MEG artifacts based on a two-step algorithm: In the first step, it uses expansions of spherical harmonic functions to separate MEG signal components originating inside the sensor array from those originating elsewhere. In the second step of its algorithm, it recognizes and projects out the temporally correlated signal components between the internal and external spaces originating from nearby artifact sources not completely suppressed by the first step. Several studies have used tSSS approach to reduce speech artifacts in MEG data. They showed that after correcting head movement and suppressing external disturbances by tSSS they could investigate corticomuscular coherence, cortical neural network connectivities as well as the connection between evoked responses and rhythmic activities during speech production (Laaksonen et al., 2012; Ruspantini et al., 2012; Liljeström et al., 2015a,b). However, since most high amplitude speech-related artifacts come from sources such as facial muscle activity or lower jaw movement (which are inside the sensor array), tSSS could not completely clean the data. Therefore, several complementary strategies have been taken to further clean the data after applying tSSS. Ruspantini et al. (2012) trained their participants in order to produce small movements during speech production. Liljeström et al. (2015a) tried to avoid the residual artifact by selecting artifact-free time windows for studying cortical networks underlying preparation for speech production. Finally, in recent studies, the combination of tSSS and ICA has been used to remove speech-related artifacts from the recorded MEG data (Alexandrou et al., 2017; Bourguignon et al., 2020). In ICA, the spatial filters are derived by decomposing sensor data to the set of maximally temporally ICs (Delorme et al., 2007). Alexandrou et al. (2017) used ICA to remove artifacts induced by the movement of facial articulatory muscles during speech production. Their results showed that they observed artifacts separated into several components (Alexandrou et al., 2017). This is in line with our observation when we tried to clean data by means of ICA. For participant #11, for instance, topographical maps for all of the components showed a noisy pattern. There are several reasons why ICA might not be optimal for speech-related artifact rejection. The movement artifact induced in MEG signals is non-stationary due to the different articulation depending on the word to articulate and the location of the source of the artifact (e.g., jaw) changes over time (De Vos et al., 2010). This violates the

requirements for ICA and leads to a failure of ICA to clearly separate signals and artifacts.

Taken together, different strategies/approaches have been used to study speech production using MEG. However, there is still a demand to further improve speech-related artifact removal approaches. The unique characteristic of our proposed approach is that it makes use of head movement information to initially reduce movement artifacts and then uses the SSP approach to remove the residual artifacts.

In the first step of our proposed artifact rejection pipeline, we used regression analysis to reduce the artifact level. We could successfully reduce the confounding variance caused by head movement by incorporation of the head position/orientation time-series into a GLM. This approach has been previously applied on several MEG datasets recorded during visual attention, auditory expectation, as well as in somatosensory spatial attention tasks (Stolk et al., 2013). This step could play a vital role in improving the signal-to-noise ratio by removing a major part of the artifact. It should be noted that this step benefits from running twice. First, global head movements are corrected that occur across the duration of the measurement. Second, a third-order polynomial fit to the head movement traces is subtracted to optimize the representation of transient movements (such as movements associated with each syllable production).

For two out of 11 participants this step was sufficient to clean data and proceed with further analysis. The effectiveness of this step depends on the distortion level (level of confounding variance) and the quality of the extracted head position/orientation time-series.

To remove the residual artifact, we used the SSP approach in the second step. Using SSP, we can separate signals into a signal subspace and an artifact subspace (Uusitalo and Ilmoniemi, 1997). Data is then linearly projected onto the signal subspace. This is an effective approach for separating artifacts from signals especially when the artifact has a constant spatial pattern. This approach has been previously used to remove different types of artifacts such as ocular artifacts from MEG data (Pulvermüller and Shtyrov, 2009; Chang et al., 2018). During speech production, movement artifacts are represented in a distinct subspace (coming from either head movement or lower jaw movement) with varying amplitudes as a function of time. Therefore, SSP is an optimal approach for separating spatial subspaces containing movement artifacts. We used SVD to decompose MEG data into components. Movement-related components were selected by a conservative three-steps approach using components' topographical maps and time-series as well as their MI with head position/orientation time-series. The advantage of the latter component selection step is that it makes explicit use of the available information on artifact characteristics.

A potential limitation of the SSP-approach is the need to define the dimensionality of the artifact subspace. It is difficult to define clear criteria for this selection. In our data, MI values often suggested a specific number of artifact components. If this is not the case one could try different numbers of artifact components and assess the suppression of robust brain responses in analogy to what we do with the M100 response (e.g., **Figure 8**). However, future work is required to make a more automatic artifactual

component removal possibly using a step-wise approach with well-defined stopping criteria.

In this study, we demonstrated that our proposed approach was able to remove artifacts and retrieve neurophysiological information. Importantly, our approach could successfully clean MEG data strongly distorted due to implanted ferromagnetic dental retainers. Previous studies reported successful MEG measurements in the presence of implanted ferromagnetic stimulation hardware (Abbasi et al., 2016). Although the pre-measurement demagnetization step might reduce the artifact caused by retainers, any head/lower jaw movement can still induce strong artifacts in MEG sensors during measurement. Here we also showed the feasibility of recording participants with retainers. This could be very important for recruiting participants not only for studying speech production but also when planning any MEG experiment.

Another advantage of the proposed approach is that there is no need to perform any initial step such as defining epochs or frequency bands representing either artifact or neurophysiological signal. Moreover, this approach is system-independent and could be applied to any type of MEG data.

Additionally, facial muscle activities have been reported as another source of artifact in MEG speech studies. The power of induced muscle artifact lies in frequencies higher than 20 Hz (Muthukumaraswamy, 2013). In this study, we did not attempt to remove muscle artifacts from MEG data. Previous studies used different strategies such as the ICA approach to remove muscle artifacts (Shackman et al., 2009; McMenamin et al., 2011; Muthukumaraswamy, 2013). Generally, ICA shows better performance when data quality is higher. Therefore, in case muscle artifact removal is necessary, we recommend to improve the signal-to-noise ratio by means of our proposed approach and apply ICA as the final step.

CONCLUSION

Our newly presented artifact rejection approach could facilitate assessing speech production and perception mechanisms by MEG. The presented results revealed that it is also possible to remove complex patterns of artifact induced by implanted devices, while preserving essential neurophysiological information. Future applications might include MEG experiments during continuous speech production assessing the neural activity underlying speech production.

DATA AVAILABILITY STATEMENT

The data used in the current study will be available from the corresponding author upon request subject to a formal data sharing agreement.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the University of Münster. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

OA and JG contributed to conception and design of the study and developed the proposed method. NS collected the database. OA and NS performed the data analysis. OA, NS, and JG wrote the first draft of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

ACKNOWLEDGMENTS

REFERENCES

Abbasi, O., Hirschmann, J., Schmitz, G., Schnitzler, A., and Butz, M. (2016). Rejecting deep brain stimulation artefacts from MEG data using ICA and mutual information. *J. Neurosci. Methods* 268, 131–141. doi: 10.1016/j.jneumeth.2016.04.010

Alexandrou, A. M., Saarinen, T., Kujala, J., and Salmelin, R. (2016). A multimodal spectral approach to characterize rhythm in natural speech. *J. Acoust. Soc. Am.* 139, 215–226. doi: 10.1121/1.4939496

Alexandrou, A. M., Saarinen, T., Mäkelä, S., Kujala, J., and Salmelin, R. (2017). The right hemisphere is highlighted in connected natural speech production and perception. *Neuroimage* 152, 628–638. doi: 10.1016/j.neuroimage.2017.03.006

Bourguignon, M., Molinaro, N., Lizarazu, M., Taulu, S., Jousmäki, V., Lallier, M., et al. (2020). Neocortical activity tracks the hierarchical linguistic structures of self-produced speech during reading aloud. *Neuroimage* 216:116788. doi: 10.1016/j.neuroimage.2020.116788

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., and Ghazanfar, A. A. (2009). The natural statistics of audiovisual speech. *PLoS Comput. Biol.* 5:e1000436. doi: 10.1371/journal.pcbi.1000436

Chang, Y.-C. C., Khan, S., Taulu, S., Kuperberg, G., Brown, E. N., Hämäläinen, M. S., et al. (2018). Left-lateralized contributions of saccades to cortical activity during a one-back word recognition task. *Front. Neural Circuits* 12:38. doi: 10.3389/fncir.2018.00038

De Vos, M., Riès, S., Vanderperren, K., Vanrumste, B., Alario, F.-X., Van Huffel, S., et al. (2010). Removal of muscle artifacts from EEG recordings of spoken language production. *Neuroinformatics* 8, 135–150. doi: 10.1007/s12021-010-9071-0

Delorme, A., Sejnowski, T., and Makeig, S. (2007). Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *Neuroimage* 34, 1443–1449. doi: 10.1016/j.neuroimage.2006.11.004

Eulitz, C., Hauk, O., and Cohen, R. (2000). Electroencephalographic activity over temporal brain areas during phonological encoding in picture naming. *Clin. Neurophysiol.* 111, 2088–2097. doi: 10.1016/s1388-2457(00)00441-7

Ewald, A., Aristei, S., Nolte, G., and Abdel Rahman, R. (2012). Brain oscillations and functional connectivity during overt language production. *Front. Psychol.* 3:166. doi: 10.3389/fpsyg.2012.00166

Ganushchak, L. Y., Christoffels, I. K., and Schiller, N. O. (2011). The use of electroencephalography in language production research: a review. *Front. Psychol.* 2:208. doi: 10.3389/fpsyg.2011.00208

Gross, J. (2019). Magnetoencephalography in cognitive neuroscience: a primer. *Neuron* 104, 189–204. doi: 10.1016/j.neuron.2019.07.001

Gross, J., Baillet, S., Barnes, G. R., Henson, R. N., Hillebrand, A., Jensen, O., et al. (2013). Good practice for conducting and reporting MEG research. *Neuroimage* 65, 349–363. doi: 10.1016/j.neuroimage.2012.10.001

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nat. Rev. Neurosci.* 13, 135–145. doi: 10.1038/nrn3158

Hickok, G., and Poeppel, D. (2016). "Neural basis of speech perception," in *Neurobiology of Language*, eds M. J. Aminoff, F. Boller, and D. F. Swaab (Amsterdam: Elsevier), 299–310. doi: 10.1016/B978-0-12-407794-2.00025-0

Houde, J. F., Nagarajan, S. S., Sekihara, K., and Merzenich, M. M. (2002). Modulation of the auditory cortex during speech: an MEG study. *J. Cogn. Neurosci.* 14, 1125–1138. doi: 10.1162/089892902760807140

Laaksonen, H., Kujala, J., Hultén, A., Liljeström, M., and Salmelin, R. (2012). MEG evoked responses and rhythmic activity provide spatiotemporally complementary measures of neural activity in language production. *Neuroimage* 60, 29–36. doi: 10.1016/j.neuroimage.2011.11.087

Liljeström, M., Kujala, J., Stevenson, C., and Salmelin, R. (2015a). Dynamic reconfiguration of the language network preceding onset of speech in picture naming. *Hum. Brain Mapp.* 36, 1202–1216. doi: 10.1002/hbm.22697

Liljeström, M., Stevenson, C., Kujala, J., and Salmelin, R. (2015b). Task- and stimulus-related cortical networks in language production: exploring similarity of MEG- and fMRI-derived functional connectivity. *Neuroimage* 120, 75–87. doi: 10.1016/j.neuroimage.2015.07.017

Liljeström, M., Tarkiainen, A., Parviainen, T., Kujala, J., Numminen, J., Hiltunen, J., et al. (2008). Perceiving and naming actions and objects. *Neuroimage* 41, 1132–1141. doi: 10.1016/j.neuroimage.2008.03.016

Llorens, A., Trébuchon, A., Liégeois-Chauvel, C., and Alario, F.-X. (2011). Intra-cranial recordings of brain activity during language production. *Front. Psychol.* 2:375. doi: 10.3389/fpsyg.2011.00375

Masaki, H., Tanaka, H., Takasawa, N., and Yamazaki, K. (2001). Error-related brain potentials elicited by vocal errors. *Neuroreport* 12, 1851–1855. doi: 10.1097/00001756-200107030-00018

McMenamin, B. W., Shackman, A. J., Greischar, L. L., and Davidson, R. J. (2011). Electromyogenic artifacts and electroencephalographic inferences revisited. *Neuroimage* 54, 4–9. doi: 10.1016/j.neuroimage.2010.07.057

Meyer, L. (2018). The neural oscillations of speech processing and language comprehension: state of the art and emerging mechanisms. *Eur. J. Neurosci.* 48, 2609–2621. doi: 10.1111/ejn.13748

Munding, D., Dubarry, A.-S., and Alario, F.-X. (2016). On the cortical dynamics of word production: a review of the MEG evidence. *Lang. Cogn. Neurosci.* 31, 441–462. doi: 10.1080/23273798.2015.1071857

Muthukumaraswamy, S. D. (2013). High-frequency brain activity and muscle artifacts in MEG/EEG: a review and recommendations. *Front. Hum. Neurosci.* 7:138. doi: 10.3389/fnhum.2013.00138

Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* 2011:156869. doi: 10.1155/2011/156869

Park, H., Ince, R. A. A., Schyns, P. G., Thut, G., and Gross, J. (2018). Representational interactions during audiovisual speech entrainment: redundancy in left posterior superior temporal gyrus and synergy in left motor cortex. *PLoS Biol.* 16:e2006558. doi: 10.1371/journal.pbio.2006558

Park, H., Kayser, C., Thut, G., and Gross, J. (2016). Lip movements entrain the observers' low-frequency brain oscillations to facilitate speech intelligibility. *elife* 5:e14521. doi: 10.7554/eLife.14521

Pulvermüller, F., and Shtyrov, Y. (2009). Spatiotemporal signatures of large-scale synfire chains for speech processing as revealed by MEG. *Cereb. Cortex* 19, 79–88. doi: 10.1093/cercor/bhn060

Ruspantini, I., Saarinen, T., Belardinelli, P., Jalava, A., Parviainen, T., Kujala, J., et al. (2012). Corticomuscular coherence is tuned to the spontaneous rhythmicity of speech at 2-3 Hz. *J. Neurosci.* 32, 3786–3790. doi: 10.1523/JNEUROSCI.3191-11.2012

Saarinen, T., Laaksonen, H., Parviainen, T., and Salmelin, R. (2006). Motor cortex dynamics in visuomotor production of speech and non-speech mouth movements. *Cereb. Cortex* 16, 212–222. doi: 10.1093/cercor/bhi099

Sahin, N. T., Pinker, S., Cash, S. S., Schomer, D., and Halgren, E. (2009). Sequential processing of lexical, grammatical, and phonological information within Broca's area. *Science* 326, 445–449. doi: 10.1126/science.1174481

Schmitt, B. M., Munte, T. F., and Kutas, M. (2000). Electrophysiological estimates of the time course of semantic and phonological encoding during implicit picture naming. *Psychophysiology* 37, 473–484. doi: 10.1111/1469-8986.3740473

Schwartz, J.-L., Berthommier, F., and Savariaux, C. (2004). Seeing to hear better: evidence for early audio-visual interactions in speech identification. *Cognition* 93, B69–B78. doi: 10.1016/j.cognition.2004.01.006

Shackman, A. J., McMenamin, B. W., Slagter, H. A., Maxwell, J. S., Greischar, L. L., and Davidson, R. J. (2009). Electromyogenic artifacts and electroencephalographic inferences. *Brain Topogr.* 22, 7–12. doi: 10.1007/s10548-009-0079-4

Stolk, A., Todorovic, A., Schoffelen, J.-M., and Oostenveld, R. (2013). Online and offline tools for head movement compensation in MEG. *Neuroimage* 68, 39–48. doi: 10.1016/j.neuroimage.2012.11.047

Taulu, S., and Simola, J. (2006). Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. *Phys. Med. Biol.* 51, 1759–1768. doi: 10.1088/0031-9155/51/7/008

Uusitalo, M. A., and Ilmoniemi, R. J. (1997). Signal-space projection method for separating MEG or EEG into components. *Med. Biol. Eng. Comput.* 35, 135–140. doi: 10.1007/bf02534144