University of Birmingham
Research at Birmingham

Removing speech artifacts from electroencephalographic recordings during overt picture naming

Porcaro, Camillo; Medaglia, Maria Teresa; Krott, Andrea

DOI:
10.1016/j.neuroimage.2014.10.049

License:
Other (please specify with Rights Statement)

Document Version
Peer reviewed version

Citation for published version (Harvard):
Porcaro, C, Medaglia, MT & Krott, A 2015, 'Removing speech artifacts from electroencephalographic recordings during overt picture naming', *NeuroImage*, vol. 105, pp. 171-180.
https://doi.org/10.1016/j.neuroimage.2014.10.049

Link to publication on Research at Birmingham portal

Publisher Rights Statement:
NOTICE: this is the author's version of a work that was accepted for publication. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published as Porcaro, Camillo, Medaglia, Maria Teresa, Krott, Andrea, Removing speech artifacts from electroencephalographic recordings during overt picture naming, NeuroImage (2014), doi: 10.1016/j.neuroimage.2014.10.049

General rights
Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

• Users may freely distribute the URL that is used to identify this publication.
• Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
• User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
• Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy
While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Download date: 01. Mar. 2021

Accepted Manuscript

Removing speech artifacts from electroencephalographic recordings during overt picture naming

Camillo Porcaro, Maria Teresa Medaglia, Andrea Krott

PII: S1053-8119(14)00875-1
DOI: doi: 10.1016/j.neuroimage.2014.10.049
Reference: YNIMG 11746

To appear in: *NeuroImage*

Accepted date: 17 October 2014

Please cite this article as: Porcaro, Camillo, Medaglia, Maria Teresa, Krott, Andrea, Removing speech artifacts from electroencephalographic recordings during overt picture naming, *NeuroImage* (2014), doi: 10.1016/j.neuroimage.2014.10.049

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Removing speech artifacts from electroencephalographic recordings during overt picture naming

Camillo Porcaro[1,3,4], Maria Teresa Medaglia[1,2], Andrea Krott[2]

[1]LET'S - ISTC - CNR, Ospedale Fatebenefratelli, Isola Tiberina, Rome, Italy
[2]School of Psychology and BUIC, University of Birmingham, UK
[3]Institute of Neuroscience, Newcastle University, UK
[4]Neural Control of Movement Lab, Department of Health Sciences and Technology ETH Zurich, Switzerland

Keywords: Overt Speech Production, Electroencephalography (EEG), Independent Component Analysis (ICA), Speech Artifacts

Corresponding author:
Dr. Camillo Porcaro
LET'S – ISTC – CNR,
Ospedale Fatebenefratelli, Isola Tiberina,
00186 Rome, Italy
Email: camillo.porcaro@istc.cnr.it
Phone: +39 066837382
Fax: +39 066837382

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Abstract

A number of electroencephalography (EEG) studies have investigated the time course of brain activation during overt word production. The interpretation of their results is complicated by the fact that articulatory movements may mask the cognitive components of interest. The first aim of the present study was to investigate when speech artifacts occur during word production planning and what effects they have on the spatio-temporal neural activation pattern. The second aim was to propose a new method that strongly attenuates speech artifacts during overt picture naming and to compare it with existing methods. EEG and surface electromyograms (EMG) of the lips were recorded while participants overtly named pictures in a picture-word interference paradigm. The comparison of the raw data with lip EMG and the comparison of source localizations of raw and corrected EEG data showed that speech artifacts occurred mainly from ~400 ms post stimulus onset, but some earlier artifacts mean that they occur much earlier than hitherto assumed. We compared previously used methods of speech artifacts removal (SAR) with a new method, which is based on Independent Component Analysis (SAR-ICA). Our new method clearly outperformed other methods. In contrast to other methods, there was only a weak correlation between the lip EMG and the corrected data by SAR-ICA. Also, only the data corrected with our method showed activation of cerebral sources consistent with meta-analyses of word production.

ACCEPTED MANUSCRIPT

1. Introduction

The act of speaking is complex. Even the production of the simplest utterances, for instance saying 'dog' as a response to a picture of a dog, involves a diverse set of cognitive processes. Though current models of word production characterize these processes somewhat differently (Caramazza, 1997; Goldrick & Rapp, 2002; Indefrey, 2011; Levelt, Roelofs, & Meyer, 1999), there is good consensus that word production involves several steps, from conceptual processing over selection of a lexical representation, to phonological and phonetic encoding processes.

A powerful research tool in current psycholinguistics is electroencephalography (EEG), which allows researchers to gain insight into the precise time course of the cognitive processes involved. However, EEG research of word production has been hampered by the fact that speech artifacts might contaminate the cognitive components of interest (Brooker & Donald, 1980; Grözinger, Kornhuber, & Kriebel, 1975; Wohlert, 1993). For instance, prevocalization potentials have been found to be severely affected by artifacts from the temporalis and the masseter muscles. Both are used for closing the lower jaw. The temporalis spreads widely over the frontal/temporal/parietal junction of the brain (Brooker & Donald, 1980), and the master is located between the cheekbone and the jaw.

Many event-related potential (ERP) studies investigating the time course of word planning have therefore relied on covert speech planning (i.e., speakers were asked to plan words without producing them), metalinguistic tasks (e.g. deciding whether a word includes a target sound), or delayed naming (i.e., naming once a cue is given) rather than immediate overt speech production (e.g., Hauk, Rockstroh, & Eulitz, 2001; Jescheniak, Schriefers, Garrett, & Friederici, 2002; Laganaro, Morand, Michel, Spinelli, & Schnider, 2011; Laganaro et al., 2009; Schiller, Bles, & Jansma, 2003; Schmitt, Münte, & Kutas, 2000; Van Turennout, Hagoort, & Brown, 1997, 1998). Though these studies have led to important insights, they

ACCEPTED MANUSCRIPT

have their limitations. For instance, in covert speech experiments it is not possible to guarantee that participants actually follow the instructions, and these tasks either do not include all cognitive steps involved in overt speech production or the timing of the steps is altered. Thus, in many contexts the use of immediate overt production tasks is preferable.

Studies using overt production tasks have generally focused on early processes of word planning that occur within 400-600 ms after stimulus presentation, with the assumption that those processes are artifact-free because they occur before actual speech (for an overview see Christoffels, Firk, & Schiller, 2007; Costa, Strijkers, Martin, & Thierry, 2009; Ganushchak, Christoffels, & Schiller, 2011; Strijkers, Costa, & Thierry, 2010; Strijkers, Holcomb, & Costa, 2011; Verhoef, Roelofs, & Chwilla, 2009). But no study to date has investigated the exact timing of speech artifacts during picture naming, and we know of only two studies that have given some indication of how speech artifacts have affected ERPs (Laganaro & Perret, 2011; Riès, Janssen, Burle, & Alario, 2013).

The first aim of the present study was therefore to investigate when speech artifacts actually occur during word production planning and what effects they have on the spatio-temporal neural activation pattern. To answer the former question, we examined electromyograms (EMG) of the lips during a word production study. To answer the latter question, we compared the relationship of the lip EMG with the raw ERP data recorded during the same experiment. A strong correlation between the lip EMG and the raw data would indicate that ERP data during word production studies are highly affected by speech artifacts. Furthermore, we conducted source-localization analyses of the ERP data, before and after speech artifact attenuation. The source localization analysis of the ERPs of the raw data can reveal the extent to which the spatio-temporal pattern of the ERPs is a reflection of cognitive processes or speech movements. Indefrey and Levelt (2004) conducted a comprehensive meta-analysis of behavioural word production studies to estimate the time-

ACCEPTED MANUSCRIPT

course of naming a picture and combined the results with brain imaging findings within the word production literature (82 experiments). This resulted in a map of the involved brain areas and the time course of their activation (see also the update by Indefrey, 2011). Their estimates for simple picture naming predicts a progression from early occipital and ventral-temporal activation during conceptual preparation (0 and 175 ms), via activation at the left middle temporal gyrus during lemma retrieval and lemma selection (150-250 ms) and posterior temporal lobe during phonological code retrieval (including Wernicke's area, 250-330 ms) to frontal activation during syllabification and articulation (400-600 ms, including Broca's area) (see also Hultén, Vihla, Laine, & Salmelin, 2009; Levelt, Praamstra, Meyer, Helenius, & Salmelin, 1998; Salmelin, Hari, Lounasmaa, & Sams, 1994; Sörös, Cornelissen, Laine, & Salmelin, 2003; Vihla, Laine, & Salmelin, 2006). Localizing ERPs should show such a progression if the ERPs truly reflect cognitive processes. Source localization is therefore an excellent tool to reveal in how far and during which processing time step the speech artifacts contaminate the raw data.

The recording of EMG from orofacial speech muscles is not new even if it has not typically been used in combination with EEG recordings. Instead, it has been used to study, for instance, silent recitation (Livesay et al., 1996), speech in stutterers (Choo et al., 2010), covert verbal hallucinations in schizophrenia (Rapin et al., 2013), and articulation by aging participants (Rastatter et al., 1987b) and by articulatory disordered children (Rastatter et al. 1987a). The muscle complex that has typically been focused on is the orbicularis oris (OO), which is situated in the lips and controls lip posture during overt speech. It consists of four sub-muscles, the left and right orbicularis oris superior (OOS) in the upper lip and the left and right orbicularis oris inferior (OOI) in the lower lip. The fibers of the lip muscles are not clearly separated from surrounding muscles. It is therefore not possible to obtain a signal that stems exclusively from the OO (Blair & Smith, 1986). However, Blair and Smith (1986)

ACCEPTED MANUSCRIPT

argue, and subsequent studies have shown, that useful data can be obtained if electrodes are placed at the same place across participants. We therefore chose to record EMG from the OO in our study.

The second aim of this study was to propose a new method that strongly attenuates speech artifacts during overt picture naming and to compare it with existing methods. While the majority of overt speech production studies have not attempted to remove speech artifacts, there are exceptions. Some studies have removed the strongest articulation-driven potentials by excluding responses whose acoustic onsets fell within the time-window of interest (e.g., Strijkers et al., 2011; Costa et al., 2009). However, as evident in the present investigation, articulators move much earlier than the acoustic onset. Such a procedure therefore does not necessarily lead to artifact-free data. Others applied filters: a low-pass filter of 12 Hz (Ganushchak & Schiller, 2008) or a band-pass filter of 0.2-20 Hz (Laganaro & Perret, 2011). Laganaro & Perret (2011) report that their filter did not make much of a difference to their results. However, the frequency content of facial muscle artifacts can overlap to a large extent with that of brain signals. Filtering therefore means that brain signals might be partly filtered out and results might nevertheless change (for a discussion see De Vos et al., 2010; Friedman & Thayer, 1991). To overcome this problem, De Vos and colleagues (De Vos et al., 2010) proposed a Blind Source Separation method based on Canonical Correlation Analysis (BSS-CCA) to separate cortical sources from electromyographic (EMG) responses. Their method appears to be the most promising approach and has been implemented, for instance, in studies by Riès and colleagues (Riès et al., 2013; Riès, Janssen, Dufau, Alario, & Burle, 2011). In the current study we introduced a new method of speech artifact removal, based on an Independent Component Analysis (ICA) procedure (Barbati, Porcaro, Zappasodi, Rossini, & Tecchio, 2004), appropriately modified to

ACCEPTED MANUSCRIPT

fit the problem under investigation. For simplicity, we will refer to this approach as Speech Artifact Removal by ICA (SAR-ICA).

We compared the artifact attenuation performance of our new SAR-ICA method with that of previously applied methods, namely BSS-CCA and the two filters 0.1-12 Hz (Ganushchak & Schiller, 2008) and 0.2-20 Hz (Laganaro & Perret, 2011). For validation purposes we conducted the following two analyses.

First, we compared the different methods (0.1-12 Hz, 0.2-20Hz, BSS-CCA, and SAR-ICA) with regard to the relationship of the lip EMG with the ERP data. A large reduction in the correlation strength between the lip EMG and the ERP data due to artifact attenuation suggests that a method is well suited to remove artifactual ERP components that are strongly related to articulatory muscle movements during word planning. As an additional validation of our SAR-ICA method, we investigated the relationship of the lip EMG with the ICA component of speech artifacts. A strong correlation between the artifact ICA component and the lip EMG together with a weak correlation between the lip EMG and the corrected data would confirm that our ICA method indeed separated and identified the components that are strongly related to speech artifacts during word planning.

Second, as another validation method for the artifact attenuation procedures we used the results of the aforementioned source localisation analyses of the raw data and compared them with source localisations of the corrected data. Results are reported for our SAR-ICA procedure and the most promising other artifact attenuation method, the BSS-CCA procedure. As explained above, for picture naming one expects a progression from early occipital and ventral-temporal activation via activation at the left middle temporal gyrus and posterior temporal lobe to frontal activation (Indefrey, 2011; 2004). Consequently, a successful artifact attenuation procedure should reveal such a progression.

ACCEPTED MANUSCRIPT

The speech production data that we used for all these analyses were the responses in a paradigm that is very typical for the study of word production planning and that has been used in several ERP and brain imaging studies with overt responses, namely the picture-word interference paradigm. In this paradigm participants name pictures with superimposed distracter words (Schriefers, Meyer, & Levelt, 1990). We used visually distracters that were either semantically related (e.g. orange) or unrelated (e.g. arrow) to the name of the picture (e.g. banana). Analyses of the distracter manipulation, including effects of distracter attenuation on the results, are not focus of this study and will be reported elsewhere.

2. Materials and Methods

Ethical approval for the research was obtained from the Ethics Board of the School of Psychology at Birmingham University.

2.1 Participants

Eighteen monolingual native English speakers (mean age 23.3, SD 3.7, 10 males) took part in the experiment and received either course credit or £20 for their participation. All were right-handed determined by the Edinburgh handedness inventory (Oldfield, 1971) and all had normal or corrected-to-normal vision.

2.2 Materials

We selected 24 line-drawings of common objects (Snodgrass & Vanderwart, 1980) and paired them with distracter words from the same semantic category (e.g., banana (target picture) – orange (distracter word)). We created unrelated target-distracter pairs by re-pairing targets and distracter words. This ensured that related and unrelated distracters as a group were perfectly matched for all possible variables (length, frequency etc.).

ACCEPTED MANUSCRIPT

2.3 Procedure

Participants were seated in a quiet and normally illuminated test room 1 m away from a 17" monitor with a resolution of 800 x 600 pixels. Before the experiment, participants studied a picture booklet to familiarize themselves with the names that they were asked to use when naming the pictures. In the experiment, participants saw the pictures with superimposed written distracter words as shown in Figure 1. They were asked to name the pictures as fast and as accurately as possible, while ignoring the distracter words. E-Prime (Psychological Software Tools, Inc.) was used to control stimulus presentation and data collection. Responses were recorded for off-line error analysis and response times were measured using a voicekey (PST SRBox). Each trial began with the presentation of a fixation cross for 800 ms. Then followed a stimulus, which disappeared when the voice key was triggered or after 400 ms, whatever occurred first. The short presentation time ensured that participants responded quickly and that re-reading of the distracters was minimized. Each trial lasted up to 3000 ms. The maximum response time was 2200 ms. We instructed participants to keep movements to a minimum (apart from moving their articulatory muscles in order to speak), to blink only after providing a response.

We created two stimulus lists of picture-distracter pairs. Each list contained all pictures, but half of the pictures occurred with related distracters, the other half with unrelated distracters (and vice versa for the other list). The order of the items within the lists was random. Each participant saw each list four times in alternation. The order of the lists was counterbalanced across participants. There was a short pause after the completion of every second list. The trials up to the first break were considered practice trials and excluded from the analyses. The experiment lasted about 75 minutes (including breaks).

3. EEG, EMG AND EOG RECORDING PARAMETERS

ACCEPTED MANUSCRIPT

Electroencephalograms (EEG) were acquired using a 128 channel BioSemi Active Two EEG system, with electrodes placed in a nylon cap according to the 10–5 system (Oostenveld & Praamstra, 2001). Horizontal and vertical electrooculograms (EOG) and upper and lower lip electromyograms (EMG) were monitored by bipolar derivations. For the EMG we placed surface electrodes at the left orbicularis oris superior (OOS) and left orbicularis oris inferior (OOI), half way between the center and the corner of the mouth. This position has previously been found to be the optimal place for recordings of the OOI (Lapatki et al., 2010). Both EEG and EMG were sampled at 512 Hz. They were off-line referenced to an average of the left and right mastoids, and filtered with a band-pass of 0.1-30 Hz.

4. Data Analysis

For all analyses, we excluded trials with missed or incorrect responses, with reaction times below 250 ms or above 1800 ms, and with disfluencies or self-repair.

4.1 Speech Artifact Removal using an automatic ICA approach (SAR-ICA)

An independent Component Analysis (ICA) procedure, like many other blind source separation (BSS) techniques, decomposes the EEG data into sources with independent time course on the basis of the statistical properties of the generated signal (Makeig, Debener, Onton, & Delorme, 2004; Medaglia et al., 2009; Porcaro et al., 2009; Porcaro, Ostwald, Hadjipapas, Barnes, & Bagshaw, 2011; Porcaro et al., 2006). Following ICA model application, introduced for example in the context of fMRI (Beckmann & Smith, 2004; Porcaro, Ostwald, & Bagshaw, 2010) and Foetal Magnetoencephalography (Porcaro et al., 2006), we applied an automatic ICA procedure (an appropriately modified version of Barbati et al., 2004) to the Raw Data to identify and classify artifactual non-cerebral activities, i.e. eye movements, speech artifacts, environmental and channels noise, without rejecting the

ACCEPTED MANUSCRIPT

contaminated epochs. This method is based on statistical and spectral characteristics of Independent Components (ICs). Briefly, it consists of three main steps: (1) applicattion of ICA for blind source separation (here we used fastICA (Hyvarinen, 1999)). (2) automatic detection of artifactual components, based on statistical characteristics (percentage of kurtosis-outlier segments to detect cardiac artifacts, global kurtosis coefficient to detect environmental noise and percentage of entropy outlier segments to detect ocular artifacts) and spectral IC characteristics (significant correlation between PSD EOG/EMG and ICs, $p<0.01$). More specifically, for the statistical indexes (kurtosis and entropy), these measure distributions were normalized with respect to all ICs (mean 0 and standard deviation 1). In this way, thresholds in terms of number of standard deviations from the mean were applied (standard threshold set at ± 1.64) and, if a certain percentage of segments (in our applications more than 33%) exceeded rejection thresholds, the corresponding IC was marked for rejection. (3) The final step was a control cycle on the 'discrepancy', i.e. on the difference between the original data and those reconstructed using only ICs retained after automatic artifact detection. The aim of the control cycle was to give a quick visual feedback on the quality of the automatic artifact identification. This step is based on visualizing PSD and ERP of the discrepancy. The feedback was positive when discrepancy contained only artifacts and noise. In case of negative feedback (i.e. presence of any brain activity in terms of brain rhythms or evoked activity in the discrepancy), the index thresholds used for artifactual IC detection was reduced and step (2) was repeated on these new thresholds.

In addition to the criteria used in step (2) by Barbati et al. (2004), we used the following indices: a) a correlation index between lip EMG and ICs to identify which ICs are contaminated by speech artifacts ($p<0.01$), b) topographic distributions to identify ICs contaminated by ocular artifacts (see Figure 2 - first row eye blinking and second row horizontal eye movement) and noise (see Figure 2 - fourth row electrode noise and fifth row

ACCEPTED MANUSCRIPT

environmental noise) through visual inspection. c) IC source localization was used to identify noise components such as environmental noise. For the source localization we used the equivalent current dipole (ECD) model with four concentric conductive spheres. See routine DIPFIT2 (Oostenveld & Oostendorp, 2002) of EEGLAB v5.0 (available at http:// www.sccn.ucsd.edu/eeglab (Delorme & Makeig, 2004)). EEGLAB expresses ECD positions in Talairach coordinates and projects them onto the template brain of the Montreal Neurological Institute (MNI). ICs with dipole localization outside the brain were considered artifacts. In general, the detection system rejected an IC when at least one of the criteria described above was satisfied.

On the basis of the indexes described for step (2), we classified the ICs in the selected IC subsets described below. To this end, each IC dynamic was averaged within a time window of 2800 ms, i.e. from 500 ms before the picture presentation onset till 2300 ms afterwards. We then used averaged trials, single trials, topographical distributions and the localizations of the components to manually classify all ICs (Delorme & Makeig, 2004; Makeig, Debener, et al., 2004; Makeig, Delorme, et al., 2004; Medaglia et al., 2009; Porcaro et al., 2009) into the following 5 clusters (see Figure 3): Cluster1 – Evoked Responses (ER) of the Corrected Data (i.e. data without artifactual ICs), Cluster 2 – ER of Speech artifacts, Cluster 3 – ER of Environmental Noise, Cluster 4 – ER of Vertical Ocular artifacts (Blinking) and Cluster 5 – ER of Horizontal Ocular artifacts (Horizontal Eye Movements).

After the identification of ICs and the cluster membership, the data at the scalp electrodes for each cluster were obtained by retro-projecting the selected ICs, $\mathbf{EEG}\_Cluster_k^i = \mathbf{A}_k\mathbf{IC}_k$ where $\mathbf{A}_k$ is the estimated mixing vector for the source $\mathbf{IC}_k$ and $\mathbf{EEG}\_Cluster_k^i$ is the resulting $\mathbf{IC}_k$ retro-projection on the channel space. The five ER clusters were obtained by averaging each cluster dataset ($\mathbf{EEG}\_Cluster_k^i$) time-logged to the picture presentation onset (see Figure 3).

ACCEPTED MANUSCRIPT

4.2 Artifact attenuation by filtering and by BSS-CCA

Results obtained by SAR-ICA were compared with previously used methods of speech artifact removal, i.e. BSS-CCA (De Vos et al., 2010), a low-pass filter (12 Hz; Ganushchak & Schiller, 2008) and a band-pass filter (0.2-20 Hz in Laganaro & Perret, 2011). For the filtering, signals were forward–backward filtered using a second order Butterworth filter in bands 0.1-12 Hz and 0.2-20 Hz.

The artifact attenuation by BSS-CCA (De Vos et al., 2010) was performed on the same data as the one used for SAR-ICA. While both BSS-CCA and SAR-ICA are both blind source separation methods, only BSS-CCA decomposes the EEG data into sources that are sorted in decreasing order of autocorrelation (highly autocorrelated sources ranked first, weakly autocorrelated ones ranked last). A criterion has to be used to select the sources related with movement artifacts. We used the automatized version of the BSS-CCA method that we obtained by sending an email to the author (as suggested in De Vos et al., 2010). In this automatized version, the criterion for movement component selection is based on Power Spectral Density (PSD). Components are considered to be movement artifact activity if their average power in the EMG band (approximated by 15–30 Hz) is at least 1/n (with n being by default set to 7) of the average power in the EEG band (approximated by 0.1–15 Hz). The parameter *n* is usually empirically determined. In case of our particular dataset and in order to use BSS-CCA in an optimal way, we asked the author of the BSS-CCA method to set parameters for us, who recommended the default setting as this was a good compromise also in our case. Moreover, De Vos et al. (2010) showed that results of this method do not critically depend on the parameter settings.

ACCEPTED MANUSCRIPT

4.3 The Effect of Speech Artifacts on the ERPs and Evaluation of Speech Artifact Attenuation Methods

To investigate the effect of the speech artifacts on the ERPs, we compared the time-course of the lip EMG with the time-course of the Raw Data. Because it is common practice in psycholinguistic ERP research to remove eye artifacts, we were interested in differences between Corrected Data and the Raw Data with eye artifacts removed. Therefore, we used our ICA procedure to create a version of the Raw Data with eye artifacts (Clusters 4 and 5) removed. Note that whenever we refer to the Raw Data we mean the Raw Data without eye artifacts. Whenever we refer to the Corrected Data using the SAR-ICA method we mean Cluster 1 described above.

To validate that Cluster 2 (ER of Speech artifacts) identified by the ICA analysis indeed captured speech artifacts, we correlated it with the separately recorded lip EMG. As a validation procedure of all methods, we correlated the lip EMG with the ERP of the Raw Data and the ERPs of the Corrected Data.

Finally, to investigate the effect of the speech artifacts on spatio-temporal activation patterns as well as to conduct an additional validation of the artifact attenuation procedures, we submitted the Raw Data and the Corrected Data to source localization analyses. We present the results using SAR-ICA as well as BSS-CCA, as the best alternative procedure. Localizations for the filtered data (0.1-12 Hz and 0.2-20 Hz) were very similar to those of the original raw data. We applied the sLORETA algorithm (Pascual-Marqui, 2002) as implemented in CURRY 6 (Neuroscan, Hamburg, Germany, http://www.neuroscan.com/) using a regular grid with a spacing of 4 mm throughout the brain region and a four shell spherical head model. The results were projected onto the template brain of the Montreal Neurological Institute (MNI) within CURRY software. Source localizations were carried out

ACCEPTED MANUSCRIPT

for all ERP peaks up to speech onset (100 ms, 140 ms, 290 ms, 360 ms, 470 ms, 540 ms, 615 ms, and 660 ms).

5. RESULTS

5.1 Lip EMG and the Raw Data

Figure 4 shows the time course of lip EMG (black line) across 2000 ms post stimulus onset. Lip movements were especially pronounced from just after 400 ms onwards. The average speech onset measured by the voice key was 850 ms. Thus, lips started moving at least 400 ms earlier than the trigger of the voice key. We compared the Global Field Power (GFP) (Lehmann & Skrandies, 1984) of the lip EMG with the Raw Data. Figure 4 suggests a strong contamination of the Raw Data by speech artifacts. The GFP of the Raw Data followed a similar time-course as the lip EMG, especially after ~400 ms (compare black line = Lip EMG and dark blue line = Raw Data), with a very strong correlation between the two data sets (see Table 1).

5.2 Validation of SAR-ICA's Speech Artifact Component

We compared the evoked response of the lip EMG with the Speech Artifact component (Cluster 2) of our SAR-ICA analysis to confirm that the identified component indeed captured speech motor movements. Focusing on lip movements, Figure 5 shows again strong movements after ~400 ms. More surprisingly, an additional smaller burst is visible at 160 ms. Given that the Speech Artifact component was categorized mostly on the basis of its correlation with the lip EMG, the Speech Artifact component should follow the same time course as the lip EMG and should highly correlate with it. Table 1 shows that this was indeed the case. The single subject correlations in Table 2 show very similar results across participants.

ACCEPTED MANUSCRIPT

5.3 Validation of Artifact Attenuation Methods

To validate the quality of the removal, we compared the GFP of the lip EMG with the Corrected Data for the four different methods of artifact removal (0.2-20 Hz filter, 0.1-12 Hz filter, BSS-CCA and SAR-ICA). As Figure 4 shows, the 0.1-12 Hz filter (light blue line) led to data almost identical to the Raw Data apart from the high frequency present in the Raw Data (dark blue line). It therefore turned out to be the worst artifact attenuation method. The BSS-CCA approach (green line) and the 0.2-20 Hz filter (pink line) were able to somewhat reduce the speech artifact. However, the reduction after 400 ms post stimulus onset is not enough to warrant a safe investigation of brain activity in this later time-window. It also should be noted that the BSS-CCA procedure reduced the activity during the first 400 ms, while the 0.2-20 Hz filter first reduced and then considerably increased activity before 400 ms. Finally, the SAR-ICA (red line) reduced the presence of the speech artifact after 400 ms remarkably well and at the same time was able to keep the brain activity at the level of the Raw Data in the first 400 ms. To reinforce this qualitative impression, correlation analyses were performed between Lip EMG and Corrected Data for the different artifact attenuation methods (see Table 1). In relation to the correlation for the Raw Data, the correlation for Corrected Data applying either the BSS-CCA method or the two filters remained very strong. Within this set of methods, the BSS-CCA led to the weakest correlation, even though the correlation coefficients are very similar. In contrast, the correlation with the lip EMG was extremely reduced after applying SAR-ICA. Table 2 lists the single subject correlations for the SAR-ICA method and, as an example for the remaining methods, for the BSS-CCA method. These individual correlations confirm the overall results.

5.4 Source Localizations of Raw Data and Corrected Data for BSS-CCA and SAR-ICA *methods*

ACCEPTED MANUSCRIPT

We identified sources for all ERP components before speech onset, i.e. at 100 ms, 140 ms, 290 ms, 360 ms, 470 ms, 540 ms, 615 ms, and 660 ms post picture onset. Figure 6 shows the spatio-temporal activation sequence across these components for the grand average across all subjects for both Raw Data (left) and Corrected Data (BSS-CCA middle and SAR-ICA right). Both the Raw Data and the Corrected Data using BSS-CCA showed only the activation of an occipito-temporal network (plus cerebellum), with very similar activation patterns (apart from 360ms and 470ms).

In contrast, the localization of the Corrected Data using SAR-ICA showed the expected involvement of an occipito-temporo-frontal network (plus cerebellum), with the following dynamics (for details see Table 3): bilateral occipito-temporal areas (BA 18, 19, 20, 37) were activated from 100 ms to 290 ms. Predominant left temporo(-frontal) areas were activated from 360 ms. Wernicke's area (BA 22) was activated from 360 ms to 540 ms, while Broca's areas (BA 45) was identified at 360 ms as well as at a late 660 ms. Activation from 540 ms involved also other left-frontal areas (BA 10, 11, 46), and the final localization at 660 ms included the pre-central gyrus (BA 4 and 6).

6. Discussion

Almost all existing ERP studies into the time-course of speech production planning have either avoided immediate overt speech production because of speech artifacts, or they have considered neural correlates in early time windows only, usually until about 400-600 ms after picture onset (but see Laganaro & Perret, 2011; Laganaro, Valente, & Perret, 2012; Riès et al., 2013), assuming that this way contamination of artifacts are avoided. But no study to date has thoroughly investigated when speech artifacts actually occur and what effects they have on the spatio-temporal neural activation pattern. The first aim of the present study was

ACCEPTED MANUSCRIPT

therefore to fill this gap. For that we conducted a picture-word interference experiment and recorded not only ERPs, but also lip EMG.

The onset of lip movements in our data was earlier than expected. The ICA analysis of the ERPs and the separately recorded lip EMG identified large speech motor effects after about 400 ms, as one would expect. However, there was also a small burst of lip movements around 160 ms after stimulus presentation. As mentioned above, finding motor effects closer to speech onset (i.e. ~400 ms post stimulus onset and ~400 ms pre speech onset) is not surprising because articulators such as lips and tongue have to be brought into place before the articulation can begin. Earlier lip movements probably arose because participants moved their lips as a response to the picture presentation, in order to get ready to speak. Importantly, these results show that speech artifacts in a picture naming study occur much earlier than at voice onset.

The effects of speech artifacts on the ERPs were very strong, especially after about 400 ms post stimulus onset, evidenced by a very strong correlation between the lip EMG and the raw data and the spatio-temporal activation patterns revealed by source localization analyses of the ERPs. Removing speech artifacts (using our new SAR-ICA method) greatly improved the data quality in terms of source localizations of the ERPs. Source localization analyses of the raw ERP data detected reliable sources only at occipital brain areas and the cerebellum. In contrast, for data corrected with respect to speech artifacts (using our new SAR-ICA method), sources followed the expected progression from occipital to temporal and frontal areas, including the middle temporal gyrus, Wernicke's and Broca's areas, as found in previous imaging studies and brain lesion studies (Hultén et al., 2009; Indefrey, 2011; Indefrey & Levelt, 2004; Levelt et al., 1998; Salmelin et al., 1994; Sörös et al., 2003; Vihla et al., 2006). Future studies should therefore make the effort of attenuating speech artifacts from the data, especially if effects closer to speech onset are investigated.

ACCEPTED MANUSCRIPT

The second aim was to propose a new method that strongly attenuates speech artifacts during overt picture naming *and to compare it with existing methods*. A number of artifact attenuation methods had been used in the literature, with the most promising method being the BSS-CCA method by Vos et al. (De Vos et al., 2010), but also two filters (12 Hz in Ganushchak & Schiller, 2008; 0.2-20 Hz in Laganaro & Perret, 2011). Evidenced by the relationship of corrected data with the lip EMG, only our new SAR-ICA method succeeded in removing the artifacts. Furthermore, the almost perfect correlation between the speech artifact component identified by our ICA procedure and lip EMG confirmed that the speech artifact categorization by our ICA method was successful. In contrast to the performance of our SAR-ICA method, corrected data resulting from any of the other methods still showed a very strong correlation with the lip EMG. In terms of reduction of the increased GFP due to speech artifacts after about 400 ms, the 0.2-20 Hz filter and the BSS-CCA method somewhat succeeded, but did not reach the same level as the SAR-ICA procedure.

A comparison of the sources of the raw data with those of the corrected data confirmed the excellent artifact attenuation performance of the SAR-ICA method. While the raw data did not show any expected temporal or frontal activity, the corrected data using our SAR-ICA method did show the expected progression of activity. In contrast, corrected data using the BSS-CCA showed activity similar to the raw data with no activity of expected temporal or frontal areas.

One might wonder whether removing speech artifacts from ERPs actually changes the results when investigating effects of experimental manipulations. We know of only two studies that give some indication of how speech artifacts affect the data. Laganaro and Perret (2011) used the 0.2-20 Hz filter and reported that cleaning did not impact on the results. Riès and colleagues (2013) implemented the BSS-CCA method and reported that some of their components were affected by the artifact attenuation. However, as we have seen in our

ACCEPTED MANUSCRIPT

analysis, the methods that were applied in these two studies are only somewhat affective in attenuating artifacts from data. Our SAR-ICA method removed artifacts much more successfully and should therefore have a bigger impact on differences between experimental conditions.

The excellent results obtained in attenuating speech artifacts using SAR-ICA notwithstanding, one has to be aware that ICA approaches have a weakness with respect to components that are temporally correlated with each other. In our specific case, the assumption that the ICs are temporally independent might be violated by a temporal correlation between the primary motor activity during articulation and the movement of articulators. An ICA approach would not be able to separate such correlated activity (Calhoun et al., 2001). However, we have some evidence that we do not need to be overly concerned about such a correlation. If the correlation was strong enough then one would expect that motor activities would be part of the speech artifact cluster (or any other artifact cluster) and not in the data corrected using SAR-ICA. We therefore conducted a localization analysis on the artifactual clusters (and in particular on the speech artifact cluster) and found no brain activities in the sensory-motor areas. In contrast, those activations were found in the SAR-ICA corrected data close to speech onset (around 660 ms), as one would expect.

Due to speech artifacts, processes of word production planning right before or after speech onset have been studied less often with overt speech paradigms than processes close to stimulus onset. A few studies have locked ERPs to overt responses. This has first been done in order to study effects that occur after the start of vocalization, namely error-related negativity (ERN) (Ganushchak & Schiller, 2008; Masaki, Tanaka, Takasawa, & Yamazaki, 2001) and positivity error (PE) (Riès et al., 2011). Recently studies have also reported response-locked analyses to investigate processes leading up to overt speech responses (Laganaro & Perret, 2011; Laganaro et al., 2012; Riès et al., 2013). As one would expect, our

ACCEPTED MANUSCRIPT

study showed that the time-window around speech onset is the one where the strongest articulatory movements occur. It is therefore important to clean the data from these artifacts when investigating processes close to voice-key onset. This had paid off in the study by Riès and colleagues (Riès et al., 2011). They found an ERN response for correct trials using the BSS-CCA artifact attenuation method, which had not been found by other studies.

Given our results, one might conclude that major lip movements in overt speech production studies start around 400 ms post stimulus onset and therefore any effect earlier than 400 ms can be safely investigated. But the timing of lip movements cannot be generalized to other studies. We used a picture-word interference paradigm, which results in relatively long response times, especially when compared to picture naming in the absence of distracters. Articulators might move earlier or later than 400 ms depending on the stimuli and the complexity of the task. One should also not ignore the small, but consistent lip movement around 160 ms in our study, which suggest that earlier articulatory artifacts than the major movements directly linked to the overt production of the response cannot be completely ruled out.

To conclude, we have shown the strong impact that speech artifacts have on both ERPs of word production processes and their sources, especially closer to speech onset (from about 400 ms post stimulus onset and about 400 ms pre voicekey onset in our experiment), and to a small degree also during an early time-window. In addition, we demonstrated that our SAR-ICA procedure successfully removed speech artifacts from overt speech production ERP data, and greatly outperformed alternative methods. In future ERP studies of overt speech production, care should be taken to identify and remove speech artifacts. Our procedure provides a safer investigation of word production planning processes, especially close to speech onset.

ACCEPTED MANUSCRIPT

7. Acknowledgements

We are grateful to Ellen Seiss who helped with setting up and testing the experiment, and Francesca Stregapede, who helped with conducting the experiment. We thank the Experimental Psychology Society (UK) as well as the College of Life and Environmental Sciences of the University of Birmingham for financial support. Furthermore, we would like to thank Antje S. Meyer and Peter Praamstra for their valuable input into this study and the report.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

References

Aksan, N., & Kochanska, G. (2004). Links between systems of inhibition from infancy to preschool years. *Child Development, 75*(5), 1477-1490. doi: 10.1111/j.1467-8624.2004.00752.x

Barbati, G., Porcaro, C., Zappasodi, F., Rossini, P. M., & Tecchio, F. (2004). Optimization of an independent component analysis approach for artifact identification and removal in magnetoencephalographic signals. *Clinical Neurophysiology, 115*, 1220-1232.

Beckmann, C. F., & Smith, S. M. (2004). Probabilistic independent component analsysis for functional magnetic resonance imaging. *IEEETransactions on Medical Imaging, 23*(2), 137-152.

Blair, C., & Smith, A. (1986). EMG recording in human lip muscles - Can single muscles be isolated. *Journal of Speech and Hearing Research, 29*(2), 256-266.

Brooker, B. H., & Donald, M. W. (1980). Contribution of the speech musculature to apparent human EEG asymmetries prior to vocalization. *Brain and Language, 9*(2), 226-245. doi: 10.1016/0093-934x(80)90143-1

Calhoun VD, Adali T, Pearlson GD, Pekar JJ. (2001). Spatial and temporal independent component analysis of functional MRI data containing a pair of task-related waveforms. *Hum Brain Mapp*, 3:43-53.

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology, 14*(1), 177-208.

Choo, A. L., Robb, M. P., Dalrymple-Alford, J. C., Huckabee, M.-L., & O'Beirne, G. A. (2010). Different Lip Asymmetry in Adults Who Stutter: Electromyographic Evidence during Speech and Non-Speech. *Folia Phoniatrica Et Logopaedica, 62*(3), 143-147. doi: 10.1159/000287213

Christoffels, I. K., Firk, C., & Schiller, N. O. (2007). Bilingual language control: An event-related brain potential study. *Brain Research, 1147*, 192-208.

Costa, A., Strijkers, K., Martin, C., & Thierry, G. (2009). The time course of word retrieval revealed by event-related brain potentials during overt speech. *Proceedings of the National Academy of Sciences of the United States of America, 106*(50), 21442-21446. doi: 10.1073/pnas.0908921106

De Vos, M., Riès, S., Vanderperren, K., Vanrumste, B., Alario, F. X., Van Huffel, S., & Burle, B. (2010). Removal of muscle artifacts from EEG recordings of spoken language production. *Neuroinformatics, 8*, 135-150.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of singletrial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(19-21).

Friedman, B. H., & Thayer, J. F. (1991). Facial muscle-activity and EEG recordings - redundancy analysis. *Electroencephalography and Clinical Neurophysiology, 79*(5), 358-360. doi: 10.1016/0013-4694(91)90200-n

Ganushchak, L. Y., Christoffels, I. K., & Schiller, N. O. (2011). The use of electroencephalography in language production research: a review. *Frontiers in psychology, 2*, 208.

Ganushchak, L. Y., & Schiller, N. O. (2008). Motivation and semantic context affect brain error-monitoring activity: An event-related brain potentials study. *NeuroImage, 39*(1), 395-405.

Goldrick, M., & Rapp, B. (2002). A restricted interaction account (RIA) of spoken word production: The best of both worlds. *Aphasiology, 16*(1-2), 20-55.

ACCEPTED MANUSCRIPT

Grözinger, B., Kornhuber, H. H., & Kriebel, J. (1975). Methodological problems in the investigation of cerebral potentials preceding speech - determining the onset and suppressing artefacts caused by speech. *Neuropsychologia, 13*(3), 263-270.

Hauk, O., Rockstroh, B., & Eulitz, C. (2001). Grapheme monitoring in picture naming: An electrophysiological study of language production. *Brain Topography, 14*(1), 3-13.

Hyvarinen A. (1999). Fast and robust fixed point algorithms for independent component analysis. *IEEE Trans Neural Netw* 10:126–634.

Hultén, A., Vihla, M., Laine, M., & Salmelin, R. (2009). Accessing Newly Learned Names and Meanings in the Native Language. *Human Brain Mapping, 30*(3), 976-989. doi: 10.1002/hbm.20561

Indefrey, P. (2011). The spatial and temporal signatures of word production components: a critical update. *Frontiers in psychology, 2*, 255.

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition, 92*(1-2), 101-144.

Jescheniak, J. R. D., Schriefers, H., Garrett, M. F., & Friederici, A. D. (2002). Exploring the activation of semantic and phonological codes during speech planning with event-related brain potentials. *Journal of Cognitive Neuroscience, 14*(6), 951-964.

Laganaro, M., Morand, S., Michel, C. M., Spinelli, L., & Schnider, A. (2011). ERP Correlates of Word Production before and after Stroke in an Aphasic Patient. *Journal of Cognitive Neuroscience, 23*(2), 374-381. doi: 10.1162/jocn.2010.21412

Laganaro, M., Morand, S., Schwitter, V., Zimmermann, C., Camen, C., & Schnider, A. (2009). Electrophysiological correlates of different anomic patterns in comparison with normal word production. *Cortex, 45*(6), 697-707. doi: 10.1016/j.cortex.2008.09.007

Laganaro, M., & Perret, C. (2011). Comparing Electrophysiological Correlates of Word Production in Immediate and Delayed Naming Through the Analysis of Word Age of Acquisition Effects. *Brain Topography, 24*(1), 19-29. doi: 10.1007/s10548-010-0162-x

Laganaro, M., Valente, A., & Perret, C. (2012). Time course of word production in fast and slow speakers: A high density ERP topographic study. *NeuroImage, 59*(4), 3881-3888. doi: 10.1016/j.neuroimage.2011.10.082

Lapatki, B. G., Oostenveld, R., Van Dijk, J. P., Jonas, I. E., Zwarts, M. J., & Stegeman, D. F. (2010). Optimal placement of bipolar surface EMG electrodes in the face based on single motor unit analysis. *Psychophysiology, 47*(2), 299-314. doi: 10.1111/j.1469-8986.2009.00935.x

Lehmann, D., & Skrandies, W. (1984). Spatial analysis of evoked potentials in man - a review. *Progress in Neurobiology, 23*, 227-250.

Levelt, W. J. M., Praamstra, P., Meyer, A. S., Helenius, P. I., & Salmelin, R. (1998). An MEG study of picture naming. *Journal of Cognitive Neuroscience, 10*(5), 553-567. doi: 10.1162/089892998562960

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*(1), 1-75.

Livesay, J., Liebke, A., Samaras, M., & Stanley, A. (1996). Covert speech behavior during a silent langauge recitation task. *Perception and Motor Skills, 83*, 1355-1362.

Makeig, S., Debener, S., Onton, J., & Delorme, A. (2004). Mining event-related brain dynamics. *Trends in Cognitive Sciences, 8*(5), 204-210.

Makeig, S., Delorme, A., Westerfield, M., Jung, T. P., Townsend, J., Courchesne, E., & Sejnowski, T. J. (2004). Electroencephalographic brain dynamics following manually responded visual targets. *PLoS Biology, 2*(6), 0747-0762.

ACCEPTED MANUSCRIPT

Masaki, H., Tanaka, H., Takasawa, N., & Yamazaki, K. (2001). Error-related brain potentials elicited by vocal errors. *Neuroreport, 12*(9), 1851-1855.

Medaglia, M. T., Tecchio, F., Seri, S., Di Lorenzo, G., Rossini, P. M., & Porcaro, C. (2009). Contradition in universal and particular resasoning. *Human Brain Mapping, 30*(1), 4187-4197.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia, 9,* 97-113.

Oostenveld, R., & Oostendorp, T. F. (2002). Validating the boundary element method for forward and inverse EEG computations in the presence of a hole in the skull. *Human Brain Mapping, 17*(3), 179-192. doi: 10.1002/hbm.10061

Oostenveld, R., & Praamstra, P. (2001). The five percent electrode system for high-resolution EEG and ERP measurements. *Clinical Neurophysiology, 112*(4), 713-719.

Pascual-Marqui, R. D. (2002). Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. *Methods and Findings in Experimental & Clinical Pharmacology, 24D*, 5-12.

Porcaro, C., Coppola, M., Di Lorenzo, G., Zappasodi, F., SIracusano, A., Pierelli, F., . . . Seri, S. (2009). Hand somatosensory subcortical and scortical sources assessed by functional source separation: An EEG study. *Human Brain Mapping, 30*, 660-674.

Porcaro, C., Ostwald, D., & Bagshaw, A. P. (2010). Functional source separation improves the quality of single trial visual evoked potentials recorded during concurrent EEG-fMRI. *NeuroImage, 50*(1), 112-123.

Porcaro, C., Ostwald, D., Hadjipapas, A., Barnes, G. R., & Bagshaw, A. P. (2011). The relationship between the visual evoked potential and the gamma band investigated by blind and semi-blind methods. *NeuroImage, 56*(3), 1059-1071.

Porcaro, C., Zappasodi, F., Barbati, G., Salustri, C., Pizzella, V., Rossini, P. M., & Tecchio, F. (2006). Fetal auditory respones to external sounds and mother's heart beat: detection improved by Independent Component Analysis. *Brain Research, 1101*(1), 51-58.

Rapin, L., Dohen, M., Polosan, M., Perrier, P., & Loevenbruck, H. (2013). An EMG study of the lip muscles during covert auditory verbal hallucinations in schizophrenia. *Journal of speech, language, and hearing research : JSLHR, 56*(6), S1882-1893. doi: 10.1044/1092-4388(2013/12-0210)

Rastatter, M., McGuire, R., & Blair, B. (1987a). EMG activity of orbicularis oris superior, orbicularis oris inferior, and masseter mucles of mild and moderate articulatory disordered children. *Perceptual and Motor Skills, 64*(3), 725-726.

Rastatter, M. P., McGuire, R. A., Bushong, L., & Loposky, M. (1987b). Speech-motor equivalence in aging subjects. *Perceptual and Motor Skills, 64*(2), 635-638.

Riès, S., Janssen, N., Burle, B., & Alario, F. X. (2013). Response-Locked Brain Dynamics of Word Production. *PLoS One, 8*(3). doi: e5819710.1371/journal.pone.0058197

Riès, S., Janssen, N., Dufau, S., Alario, F. X., & Burle, B. (2011). General-purpose monitoring during speech production. *Journal of Cognitive Neuroscience, 23*(6), 1419-1536.

Salmelin, R., Hari, R., Lounasmaa, O. V., & Sams, M. (1994). Dynamics of brain activation during picture naming. *Nature, 368*(6470), 463-465. doi: 10.1038/368463a0

Schiller, N. O., Bles, M., & Jansma, B. M. (2003). Tracking the time course of phonological encoding in speech production: an event-related brain potential study. *Cognitive Brain Research, 17*(3), 819-831. doi: 10.1016/s0926-6410(03)00204-0

Schmitt, B. M., Münte, T. F., & Kutas, M. (2000). Electrophysiological estimates of the time course of semantic and phonological encoding during implicit picture naming. *Psychophysiology, 37*(4), 473-484.

ACCEPTED MANUSCRIPT

Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production - picture-word interference studies. *Journal Of Memory And Language, 29*(1), 86-102.

Snodgrass, J. G., & Vanderwart, M. (1980). Standardized set of 260 pictures - norms for name agreement, image agreement, familiarity, and visual complexity. *Journal Of Experimental Psychology-Human Learning And Memory, 6*(2), 174-215.

Sörös, P., Cornelissen, K., Laine, M., & Salmelin, R. (2003). Naming actions and objects: cortical dynamics in healthy adults and in an anomic patient with a dissociation in action/object naming. *NeuroImage, 19*(4), 1787-1801. doi: 10.1016/s1053-8119(03)00217-9

Strijkers, K., Costa, A., & Thierry, G. (2010). Tracking Lexical Access in Speech Production: Electrophysiological Correlates of Word Frequency and Cognate Effects. *Cerebral Cortex, 20*(4), 912-928. doi: 10.1093/cercor/bhp153

Strijkers, K., Holcomb, P. J., & Costa, A. (2011). Conscious intention to speak proactively facilitates lexical access during overt object naming. *Journal Of Memory And Language, 65*(4), 345-362. doi: 10.1016/j.jml.2011.06.002

Van Turennout, M., Hagoort, P., & Brown, C. M. (1997). Electrophysiological evidence on the time course of semantic and phonological processes in speech production. *Journal Of Experimental Psychology-Learning Memory And Cognition, 23*(4), 787-806.

Van Turennout, M., Hagoort, P., & Brown, C. M. (1998). Brain activity during speaking: from syntax to phonology in 40 milliseconds. *Science, 280*(5363), 572-574.

Verhoef, K., Roelofs, A., & Chwilla, D. J. (2009). Role of inhibition in language switching: Evidence from event-related brain potentials in overt picture naming. *Cognition, 110*(1), 84-99. doi: 10.1016/j.cognition.2008.10.013

Vihla, M., Laine, M., & Salmelin, R. (2006). Cortical dynamics of visual/semantic vs. phonological analysis in picture confrontation. *NeuroImage, 33*(2), 732-738. doi: 10.1016/j.neuroimage.2006.06.040

Wohlert, A. B. (1993). Event-Related Brain Potentials Preceding Speech and Nonspeech Oral Movements of Varying Complexity. *Journal of Speech and Hearing Research, 36*(5), 897-905.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 1 – Correlation coefficients R for the relationship of Lip EMG with Raw Data, with Corrected Data using the different artifact attenuation methods, and with the Speech Artifact Component (Cluster 2) of the SAR-ICA method

	R	
	0-850 ms	0-1350 ms
Cluster 2 by SAR-ICA	0.99[a]	0.98[a]
Raw Data	0.86[a]	0.91[a]
0.1-12 Hz filter	0.86[a]	0.91[a]
0.2-20 Hz filter	0.85[a]	0.87[a]
Corrected data by BSS-CCA	0.84[a]	0.86[a]
Corrected data by SAR-ICA	0.22[b]	0.24[b]

[a]$p < 0.001$, [b]$p < 0.01$, two-tailed

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 2 – Correlation coefficients R for the relationship of Lip EMG with Speech Artifacts (Cluster 2 by SAR-ICA), Raw Data and Corrected Data using BSS-CCA and SAR-ICA methods - Single Subject Analysis

	R							
	0-850 ms				0-1350 ms			
Subject	Cluster 2	Raw Data	Corrected data BSS-CCA	Corrected data SAR-ICA	Cluster 2	Raw Data	Corrected data BSS-CCA	Corrected data SAR-ICA
S1	0.98	0.56	0.52	0.14	0.98	0.95	0.95	0.27
S2	0.95	0.85	0.85	0.32	0.90	0.84	0.84	0.33
S3	0.81	0.65	0.64	0.16	0.80	0.40	0.39	0.12
S4	0.95	0.80	0.80	0.11	0.94	0.72	0.72	0.10
S5	0.94	0.90	0.90	0.13	0.97	0.91	0.91	0.24
S6	0.47	0.46	0.31	0.17	0.75	0.63	0.63	0.15
S7	0.88	0.80	0.80	0.13	0.91	0.69	0.69	0.11
S8	0.97	0.67	0.67	0.16	0.96	0.84	0.84	0.23
S9	0.71	0.36	0.35	0.10	0.44	0.72	0.71	0.35
S10	0.99	0.86	0.85	0.34	0.97	0.84	0.84	0.22
S11	0.98	0.59	0.58	0.13	0.87	0.85	0.84	0.13
S12	0.87	0.71	0.70	0.10	0.81	0.73	0.72	0.10
S13	0.83	0.62	0.62	0.12	0.82	0.70	0.70	0.05
S14	0.99	0.89	0.89	0.27	0.99	0.85	0.85	0.20
S15	0.81	0.81	0.81	0.34	0.84	0.61	0.61	0.19
S16	0.88	0.79	0.79	0.21	0.85	0.73	0.72	0.16
S17	0.99	0.95	0.45	0.28	0.93	0.86	0.85	0.17
S18	0.94	0.84	0.84	0.11	0.97	0.90	0.90	0.31

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 3 – Brain Areas activated in overt speech production (Corrected Data using SAR-ICA)

Networks Timing (ms)	Location	BA (Left)	BA (Right)
Occipito-Temporal (100)	Fusiform Gyrus (FG)	19;37	19;37
	Inferior Occipital Gyrus (IOG)	18;19	18;19
	Middle Occipital Gyrus (MOG)	19; 37	19; 37
	Inferior Temporal Gyrus (ITG)	20; 37	20; 37
	Middle Temporal Gyrus (MTG)	37	37
Occipito-Temporo (140)	Fusiform Gyrus (FG)	19; 37	19; 37
	Inferior Occipital Gyrus (IOG)	18;19	18;19
	Middle Occipital Gyrus (MOG)	19; 37	19; 37
	Inferior Temporal Gyrus (ITG)	20; 37	20; 37
	Middle Temporal Gyrus (MTG)	20; 37	20; 37
Occipito-Temporal (290)	Fusiform Gyrus (FG)	19	19
	Inferior Occipital Gyrus (IOG)	18; 19	18; 19
	Inferior Temporal Gyrus (ITG)	20; 37	20; 37
Temporo-Frontal (360)	Fusiform Gyrus (FG)	20; 37	
	Inferior Temporal Gyrus (ITG)	20;37	20
	Middle Temporal Gyrus (MTG)	21; 37	21
	Superior Temporal Gyrus (**Wernicke's area**)	**22**	
	Inferior Frontal Gyrus (**Broca's area**)	**45**; 46	
Occipito-Temporal (470)	Inferior Occipital Gyrus (IOG)	18	
	Inferior Temporal Gyrus (ITG)	20	
	Middle Temporal Gyrus (MTG)	21	
	Superior Temporal Gyrus (**Wernicke's area**)	**22**	
Occipito-Temporal (540)	Inferior Occipital Gyrus (IOG)	18	18
	Inferior Temporal Gyrus (ITG)	20	
	Middle Temporal Gyrus (MTG)	21	
	Superior Temporal Gyrus (**Wernicke's area**)	**22**	
	Middle Frontal Gyrus (MFG)	10; 11	
Frontal (615)	Middle Frontal Gyrus (MFG)	10; 11	
Fronto-Temporal and Sensory-Motor (660)	Inferior Temporal Gyrus (ITG)		20;21;37
	Middle Temporal Gyrus (MTG)		21
	Middle Frontal Gyrus (MFG)	10; 11	
	Inferior Frontal Gyrus (**Broca's area**)	**45**; 46	
	Post-central Gyrus		1;2;3
	Pre-central Gyrus		4;6

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Figure legends

Figure 1 – Example of Experimental Stimuli
The Figure shows an example of a stimulus with a semantically related distracter and an example of a stimulus with an unrelated distracter.

Figure 2 – Artifact identification
First column shows the Event Related Potentials (ERP) of the ICs in phase with the picture presentation onset: ocular artifact (vertical and horizontal movements) in first and second row respectively; Speech artifact in third row; Electrode and Environmental noise in forth and fifth row respectively. Second column shows for each IC the assial, coronal and sagittal view of the equivalent current dipole (ECD) model superimposed onto the template brain of the Montreal Neurological Institute (MNI). Third column shows the spatial distributions of each IC obtained by representing the corresponding IC weights (Topographic Map). Fourth column shows the probability density function to describe the distribution probability (pdf) of each IC. The red line indicates the normal probability density, the values of kurtosis (K), and skewness (S) are also provided. As ICA procedure performs whitened preprocessing, each quantity is expressed in arbitrary unit (a.u.).

Figure 3 – SAR-ICA Cluster Identification
Evoked Responses (ER) in the -500 – 2200 ms time window time-logged to the picture presentation onset (vertical dashed line) are shown for the 5 clusters identified by Independent Component Analysis (ICA). Blue rectangle: ER Raw Data not corrected by ocular artifacts. For each cluster and for the Raw Data the potential distribution on the scalp (topographic map) is shown at different latencies corresponding to the major peaks of the ER (100ms, 140ms, 290ms, 360ms, 470ms and 660ms). Red rectangle (Cluster 1): ER Corrected Data (i.e. data without artifactual Independent Components), Cluster 2: ER Speech artifacts, Cluster 3: Environmental Noise, Cluster 4: ER Ocular artifacts - Blinking, and Cluster 5: ER Ocular artifacts - Horizontal Eye Movements.

Figure 4 – Global Field Power comparison of Lip EMG, Raw Data, and Corrected Data using the different artifact attenuation methods (0.2-20 Hz filter, 0.1-12 Hz filter, BSS-CCA and SAR-ICA)
Grand-average Global Field Power (GFP) of Lip EMG (black line), Raw Data (dark blue line), corrected data using 0.2-20 Hz filter (pink line), 0.1-12 Hz filter (light blue line), the BSS-CCA method (green line), and the SAR-ICA method (red line). Dashed line indicates picture onset. The Lip EMG was scaled to simplify the visualization.

Figure 5 – Comparison of normalized Lip EMG and Speech Artifacts
Grand-averages of Lip EMG (blue line) and Speech Artifacts (Cluster 2 identified in the ICA analysis, red line) plus correlation values of the two data sets are shown. Dashed line indicates picture onset. To simplify the representation in the figure the data was normalized.

Figure 6 – Dynamics of Brain Activation of Raw and Corrected Data (BSS-CCA, SAR-ICA), determined by source localization
Spatiotemporal activation of the Raw Data (left) and Corrected Data (BSS-CCA middle and SAR-ICA right) during the task, applying the localization procedure described in the method section to the data averaged across participants.

ACCEPTED MANUSCRIPT

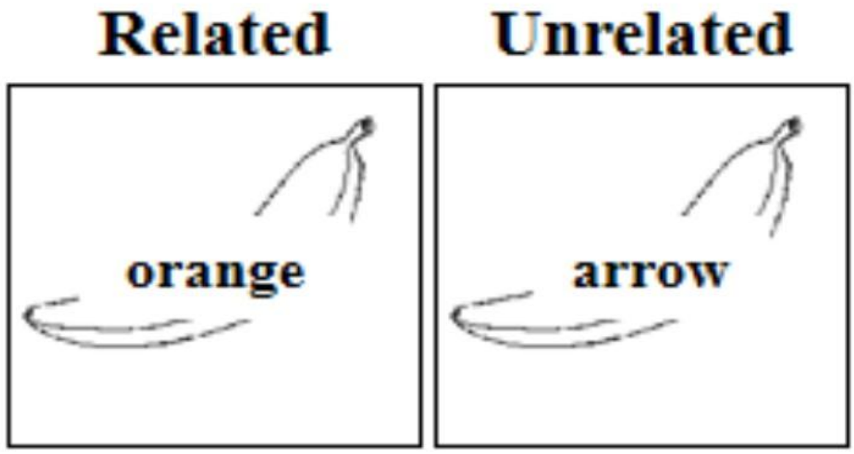

Figure 1

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

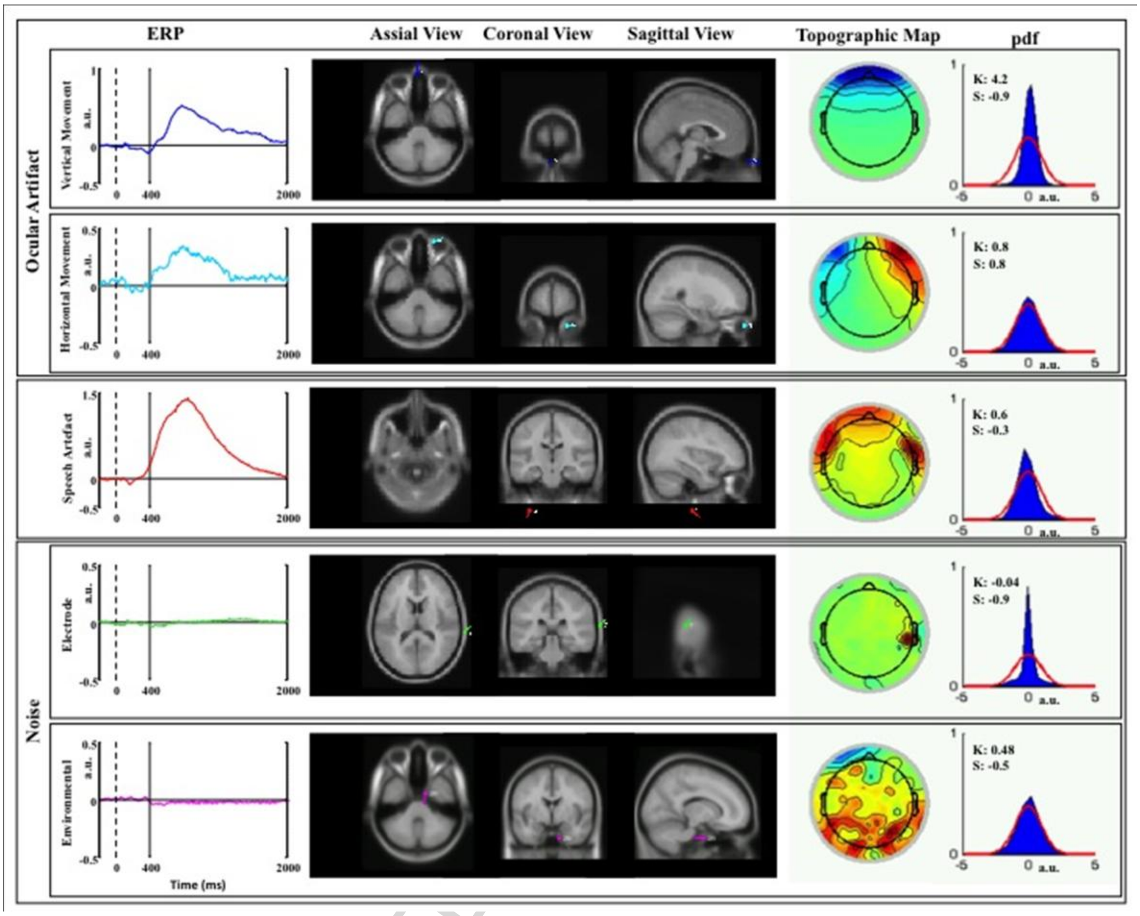

Figure 2

ACCEPTED MANUSCRIPT

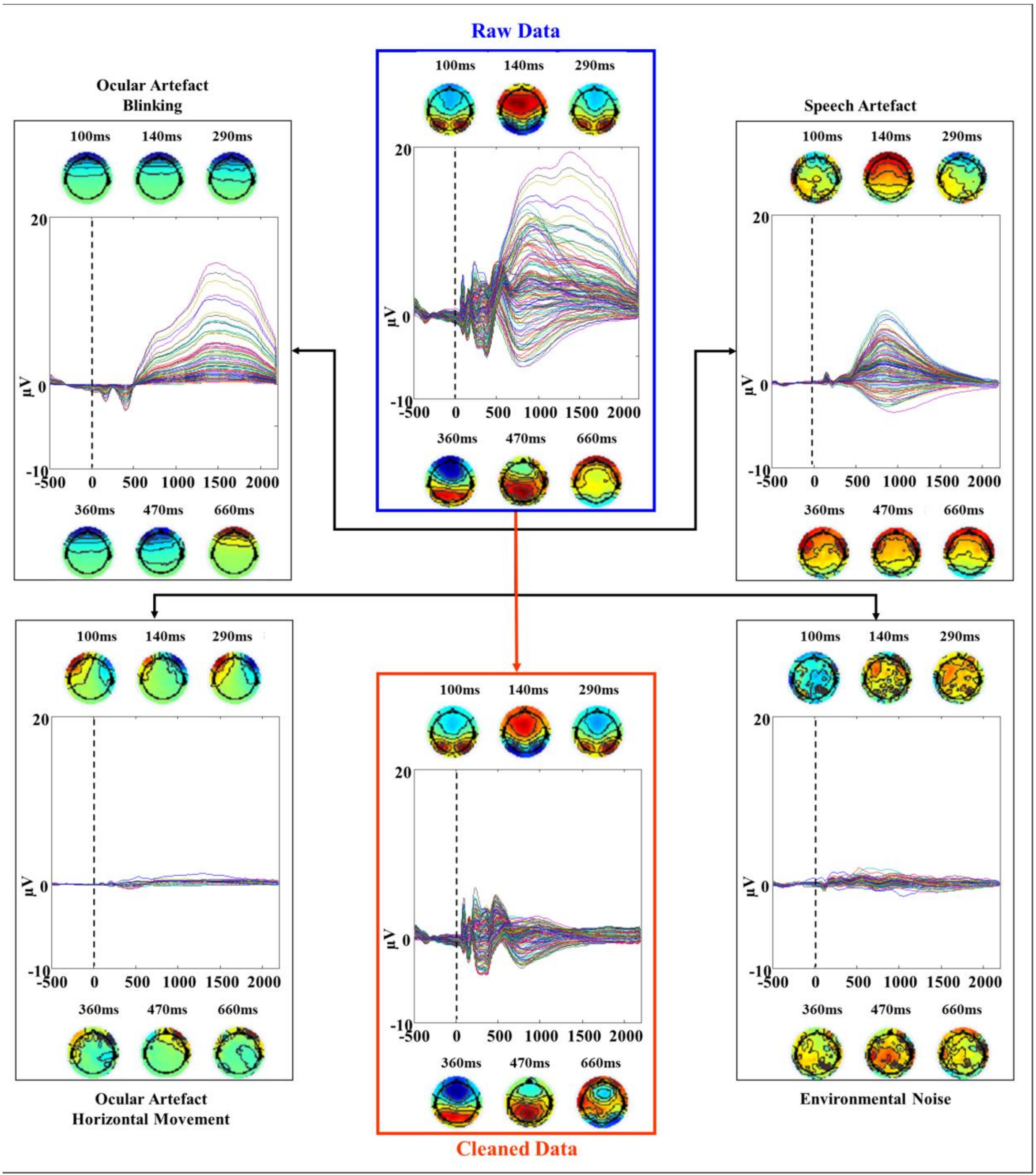

Figure 3

ACCEPTED MANUSCRIPT

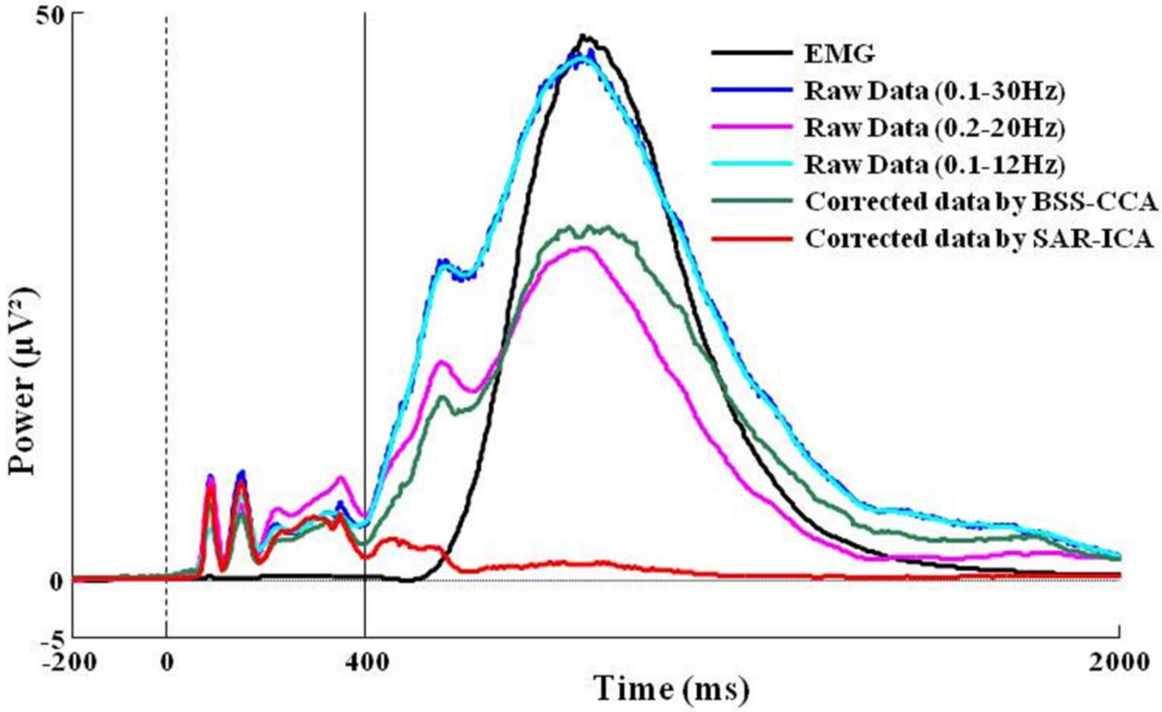

Figure 4

ACCEPTED

ACCEPTED MANUSCRIPT

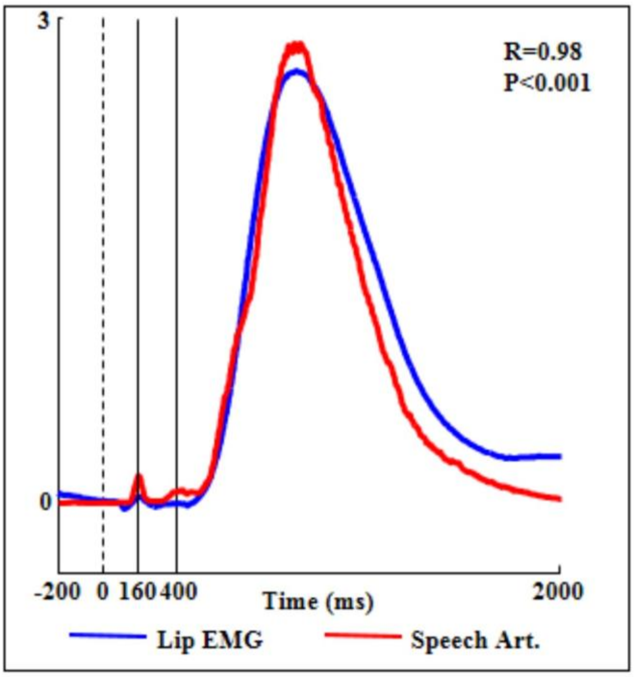

Figure 5

ACCEPTED MANUSCRIPT

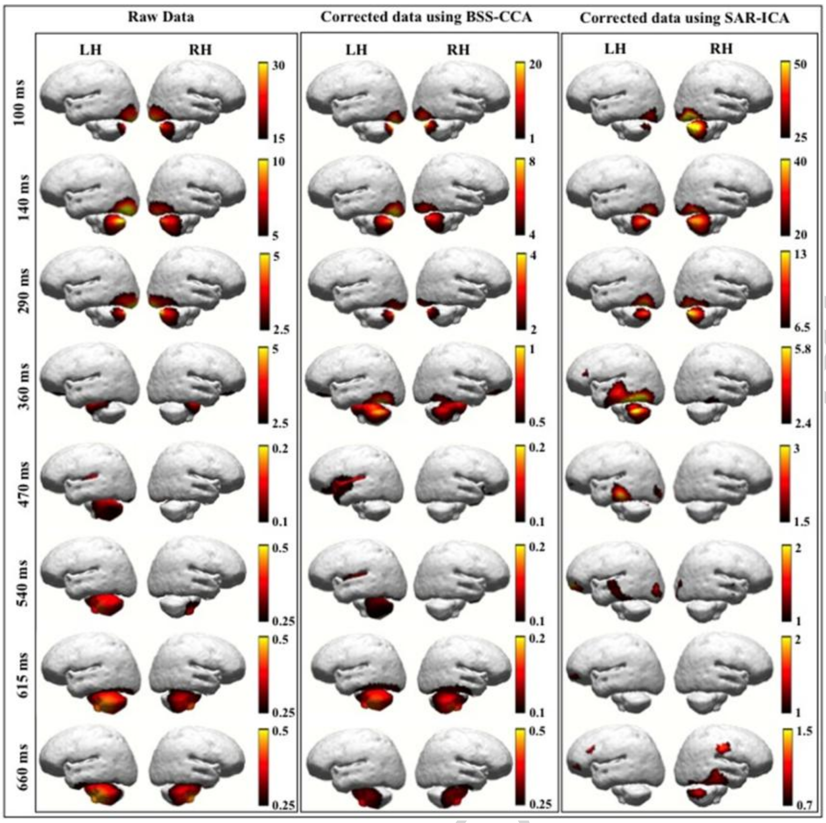

Figure 6

ACCEPTED MANUSCRIPT

Highlights

- SAR-ICA is a useful tool for speech artifact attenuation in overt speech naming
- SAR-ICA out-performs previous artifact attenuation methods
- Speech artifacts affect neural pattern during both early and late word planning
- SAR-ICA allows investigation of late ERPs during overt speech production planning

ACCEPTED MANUSCRIPT