Brain Topogr
DOI 10.1007/s10548-016-0515-1

ORIGINAL PAPER

Articulation Artifacts During Overt Language Production in Event-Related Brain Potentials: Description and Correction

Guang Ouyang[2,3,6] · Werner Sommer[1] · Changsong Zhou[2,3,4,5] · Sabrina Aristei[1] · Thomas Pinkpank[1] · Rasha Abdel Rahman[1]

Received: 23 January 2016 / Accepted: 3 August 2016

Abstract Overt articulation produces strong artifacts in the electroencephalogram and in event-related potentials (ERPs), posing a serious problem for investigating language production with these variables. Here we describe the properties of articulation-related artifacts and propose a novel correction procedure. Experiment 1 co-recorded ERPs and trajectories of the articulators with an electromagnetic articulograph from a single participant. The generalization of the findings from the single participant to standard picture naming was investigated in Experiment 2. Both experiments provided evidence that articulation-induced artifacts may start up to 300 ms or more prior to voice onset or voice key onset—depending on the specific measure; they are highly similar in topography across many different phoneme patterns and differ mainly in their time course and amplitude. ERPs were separated from articulation-related artifacts with residue iteration decomposition (RIDE). After obtaining the artifact-free ERPs, their correlations with the articulatory trajectories dropped near to zero. Artifact removal with independent component analysis was less successful; while correlations with the articulatory movements remained substantial, early components prior to voice onset were attenuated in reconstructed ERPs. These findings offer new insights into the nature of articulation artifacts; together with RIDE as method for artifact removal the present report offers a fresh perspective for ERP studies requiring overt articulation.

Keywords Event-related potential · Speech artifact · Method · Artefact correction · Residue iteration decomposition · Language · Speech production · Electro-magneto articulography

Guang Ouyang and Werner Sommer have contributed equally to this study.

Electronic supplementary material The online version of this article (doi:10.1007/s10548-016-0515-1) contains supplementary material, which is available to authorized users.

✉ Changsong Zhou
cszhou@hkbu.edu.hk

✉ Rasha Abdel Rahman
rasha.abdel.rahman@hu-berlin.de

[1] Department of Psychology, Humboldt-Universität zu Berlin, Rudower Chaussee 18, 12489 Berlin, Germany

[2] Department of Physics, Hong Kong Baptist University, Kowloon Tong, Hong Kong

[3] Centre for Nonlinear Studies and The Beijing–Hong Kong–Singapore Joint Centre for Nonlinear and Complex Systems (Hong Kong), Institute of Computational and Theoretical Studies, Hong Kong Baptist University, Kowloon Tong, Hong Kong

[4] Research Centre, HKBU Institute of Research and Continuing Education, Virtual University Park Building, South Area Hi-tech Industrial Park, Shenzhen 518000, China

[5] Beijing Computational Science Research Center, Beijing 100084, China

[6] Department of Psychology, Ernst-Moritz-Arndt-Universität Greifswald, Greifswald, Germany

Published online: 10 August 2016

Introduction

The use of scalp-recorded electroencephalography (EEG) in investigating human cognition is well established because the EEG and event-related brain potentials (ERPs) provide valuable information about the temporal unfolding of cognitive processes and their organization. However, these techniques have long played only a minor role in language production research due to the contamination of the brain signals by articulation-induced artifacts. Overt articulatory activities—associated with electrical potentials due to facial muscle contractions and glosso-kinetic potentials due to tongue movements—can severely distort and even obscure the EEG signal (e.g., Brooker and Donald 1980; Grözinger et al. 1975; Wohlert 1993).

In order to avoid articulation-related artifacts, tasks like delayed naming, silent naming, or manual classifications have been applied as alternatives to overt articulation (e.g., Abdel Rahman et al. 2003; Eulitz et al. 2000; Ganushchak and Schiller 2008; Jescheniak et al. 2003; Schmitt et al. 2001; van Turennout et al. 1998). Although avoiding articulation-induced artifacts, these alternatives have disadvantages. Specifically, silent naming does not yield performance measures and button-press tasks may differ in crucial and unknown ways from natural language production (but see Abdel Rahman and Aristei 2010; Hutson et al. 2013).

Despite the unresolved artifact question, a quickly growing number of studies have acquired EEG data during overt naming tasks to investigate language production (e.g., Aristei et al. 2011; Costa et al. 2009; Dell'Acqua et al. 2010; Ewald et al. 2012; Hirschfeld et al. 2008; Janssen et al. 2014; Piai et al. 2012, 2014; Wirth et al. 2011; for evidence from magnetoencephalography see Maess et al. 2001; for reviews see Ganushchak et al. 2011; Indefrey 2011; Strijkers et al. 2010). One strategy used in these studies is to restrict the analyses in the EEG signals to time periods before articulation onset as measured by voice onset, assuming that these periods are not significantly contaminated by artifacts associated with overt articulation. However, articulation artifacts may start up to several hundred milliseconds prior to voice onset (Brooker and Donald 1980), in particular when the speech material starts with voiceless stops (e.g./p, t, k/), and their precise onset is unclear. Another strategy is to compare identical utterances in different experimental conditions, assuming that the artifacts associated with identical words do not differ between conditions. However, even if artifacts for identical words show similar topographical distributions, they may differentially affect the EEG in case of latency shifts in artifact onsets. In this case, using EEG-derived ERPs to understand reaction time differences between conditions may be problematic even when identical words are articulated and when only relatively early time periods before articulation are analyzed.

In order to take best advantage of EEG techniques in language production research a better understanding of articulation-related artifacts and a procedure for their removal is needed. Unfortunately, little is known about the nature of the artifacts and their contributions to the signals measured at scalp sites. Since the first reports few attempts have been made to characterize and eliminate articulation-related artifacts in the EEG (e.g., De Vos et al. 2010). One obstacle may be the difficulty to measure not only surface electromyogram (EMG) of lip, masseter or temporalis muscle activity but also articulatory movements within the vocal tract (e.g., tongue and jaw movements possibly reflected in temporal scalp regions). The opening of the jaw and displacement of the tongue (the latter investigated by Vanhatalo et al. 2003) most likely give rise to the large glossokinetic potentials seen at frontal and posterior scalp sites.

De Vos et al. (2010) used a blind source separation method based on Canonical Correlation Analysis to separate EMG artifacts from the EEG, focusing on muscle-related artifacts in the frequency range of 15–30 Hz. Most recently Porcaro et al. (2015) employed an independent component analysis (ICA) algorithm to separate articulation artifacts from the EEG recorded in a picture-word interference task. A novelty of this work is the use of the correlation between upper and lower lip EMG and independent components (ICs) as a criterion for successful removal of articulation artifacts. After discarding the ICs with high correlations with the EMG, previously very high correlations of the EEG and the EMG dropped to $r = .24$ and less.

The present study aims at providing a systematic description of articulation-related EEG artifacts, and suggesting—on this basis—an alternative method for their efficient removal. Previous studies have shown that articulation related artifacts consist in frontal-positive and posterior negative shifts of the EEG. However, it is not clear, how the parameters of this artifact, such as onset, scalp topography and amplitude depend on the phonological structure of the speech output.

Importantly, the characterization of articulation artifacts as well as the control of their successful removal crucially depends on how well the different sources of the artifact inside and outside the vocal tract can be captured. This poses a potential problem because overt speech involves complex sequences of muscle activities. While muscles give rise to EMG artifacts in the EEG, these artifacts are usually in the high-frequency range with a bandwidth of about 20–300 Hz (Muthukumaraswamy 2013). For example, the peak frequency of the masseter muscle, involved in

speaking (jaw closure) and affecting EEG activity at temporal scalp regions, is around 50–60 Hz (O'Donnell et al. 1974). High frequency artifacts can be greatly attenuated by low-pass filtering at 30 Hz, the usual range of interest for language research, but may distort the signal (De Vos et al. 2010). Another—and hitherto largely unattended—serious source of artifacts are the changes of the anatomical configuration of the jaw and tongue during articulation. Thus, Vanhatalo et al. (2003) showed that modest forward movements of the tongue induce slow and large electrical potential shifts of positive polarity in the face and anterior areas of the head, similar to what has been shown for the artifacts during speaking. Therefore one should ideally measure the movements of all articulators in order to comprehensively monitor the origins of speech related artifacts.

Consequently, as a first aim and complementing recent reports relating external muscle activity to EEG measures (e.g., Porcaro et al. 2015; De Vos et al. 2010), we used 3D electromagnetic articulography (EMA) to measure the movements of different articulators, including their directions in the vertical, lateral, and anterior-posterior direction; in addition we simultaneously co-registered the EEG (Exp. 1 with a single participant). This experiment also provided information on how the parameters of the artifact relate to different articulation patterns. Experiment 2 investigated the generalizability of the findings from Experiment 1 to a typical group study with overt picture naming and the production of existing words.

As a second aim we explored the feasibility of a new method to separate and remove the articulation artifact from the EEG data in both experiments, applying RIDE (e.g., Ouyang et al. 2015a, b) described in detail below. A comparison with ICA, as applied by De Vos et al. (2010) and Porcaro et al. (2015), was provided as well.

Experiment 1

Here we co-recorded articulator movements with EMA and the EEG while a single participant repeatedly articulated 15 different disyllabic words with specific articulatory patterns. The purpose of the recording was to (1) compare artifactual activity in the EEG for different articulatory patterns, and to (2) obtain information about the temporal relationships of both voice onset and artifact onset relative to articulatory movements, and (3) assess the success of two artifact correction approaches.

The problem of any artifact correction method without external criterion is to know the true—artifact-free—signal. This problem has beset methods for correcting eye-movement induced artifacts for a long time although an external criterion by recording eye-movements with eye-tracking is readily available (cf. Dimigen et al. 2011). In principle, by correlating EEG activity with an independent measure of the putative artifact, it can be assessed whether the EEG is influenced by the variable in question. If the relationship is diminished or—ideally—abolished after artifact correction it shows the success of the method.

Method

Participant

The participant was a female native speaker of German, aged 42 years.

Materials

A total of 15 disyllabic pseudowords were constructed covering a wide range of different articulators involved. We were interested in whether different articulatory patterns would affect the topographies of the artifact: Would there be one single topography as suggested by previous studies on overt articulation or would the topography differ as a function of the articulatory pattern? Here we put the question of specific versus unitary artifact topographies to a strong test. The disyllables used were: kata, taka, tata, kaka, kiti, tiki, titi, kiki, pama, pumu, pimi, pflami, lami, mipfla, and mila. According to the International Phonetic Alphabet of German (see Kohler 1999) the vowels /a, i, u/ are the most extreme vowels of the vowel space (also called corner vowels), with /a/ being the lowest vowel (low tongue position), /i/ being a front high vowel (tongue with the highest and most anterior position), and /u/ the most back high vowel in German (tongue with the highest and most posterior position). Furthermore, to produce an /a/ the jaw must be lowered, whereas /i/ and /u/ have a very high jaw configuration. Moreover, /u/ additionally involves lip protrusion. Using these three vowels in our speech material covers the primary phonological vowel features: high vs. low vowels, front vs. back vowels, and rounded vs. unrounded vowels. For the consonants we chose /p, m/ which are bilabials and are produced with the lips in coordination with the jaw, /t, l/ which are realized with the tongue tip, and /k/ which is realized with the tongue back. The consonants also differ with respect to an involvement of vocal fold oscillations—the key measure for the start of audible speech. /p, t, k/ are voiceless aspirated stops in German and start with an acoustically silent closure movement of the lips, tongue tip or tongue back respectively. They are then followed by an acoustically prominent burst with aspiration noise. We expected that these three consonants show particularly large differences between movement onset and voice onset. In contrast /m, l/ are all realized with phonation already during the

formation of closure and we therefore expected smaller onset differences for these sounds. Items like /mipfla/ and /pflami/ were used because they have a different phonotactic structure with complex consonant clusters /pfl/ (CCCV) in one syllable and simple consonant–vowel (CV) combinations in the other syllables. To sum, with this selection of disyllables we intended to cover a wide range of different articulatory patterns to test for general differences in the EEG artifact.

Procedure

Each trial started with a fixation cross presented for 500 ms, followed by the presentation of one of the disyllables on the screen for 1500 ms. After an interval of 1 s of blank screen the next trial started. The participant was to speak each word aloud. Each of the 15 disyllables was presented 40 times, yielding a total of 600 trials. The order of disyllables was randomized.

Data Recordings

The sound waves were recorded with a dynamic microphone (YOGA DM-302 600 Ohm) and fed into the headbox for EEG recordings where it was sampled with 1000 Hz, that is, with the same resolution as the EEG. Voice onsets were identified by applying a threshold of ±100 μV to the sound wave. Since the maximum level of the sound waves were between 2000 and 5000 μV, the threshold corresponded to 1–2 % of the maximum. According to visual inspection of the sound waves this threshold allowed to measure early onsets of the isolated and experimentally specified single words spoken in this experiment. As verified by visual inspection, there were no non-speech noises within the relevant epochs. Movements of the articulators were recorded at a sampling rate of 200 Hz with a 3D-Articulograph (AG 500, Carstens Medizinelektronik). EMA tracks midsagittal fleshpoint movements by measuring induced current from receiver sensors moving in a magnetic field. The magnetic field is generated by transmitter coils with frequencies from 9 to 16 kHz. Five sensor coils were affixed to the articulators in the mid-sagittal plane: one sensor at about 1 cm from the tongue tip (TT), one sensor for the tongue dorsum (TD) at about 3 cm from the tip and one at the tongue back (TB) at approximately 5 cm from the tip, one beneath the vermillion border of the lower lip (LL) and one just beneath the lower incisors (JAW). Three additional sensors were glued above the upper incisors (UI), the nose bridge (N), and the left mastoid process below the right ear (E), which served to compensate for head movements. The movement of each coil was recorded in the vertical (up–down), lateral (left–right), and anterior-posterior (front–back) direction. The system was calibrated and pre-processed using the AG 500's standard program.

The EEG was recorded with Brain Amp amplifiers (Brain Products) from 61 electrodes (A1 AF3 AF4 AF7 AF8 AFz C1 C2 C3 C4 C5 C6 CP1 CP2 CP3 CP4 CP5 CP6 CPz Cz F10 F3 F4 F5 F6 F7 F8 F9 FC1 FC2 FC3 FC4 FC5 FC6 Fp1 Fp2 Fpz FT7 FT8 Fz O1 O2 Oz P3 P4 P5 P6 P7 P8 PO3 PO4 PO7 PO8 PO9 POz Pz T7 TP10 TP7 TP8 TP9) where A1 served as initial reference. Sampling rate was 1000 Hz. Offline, the bandpass was set between 0.0159 Hz (12 dB/oct) and 70 Hz (48 dB/oct; zero phase shift) with a notch filter at 50 Hz. Blinks were corrected with the method of Gratton et al. (1983). Six trials with remaining artifacts were rejected according to the following criteria: Voltage steps <20 μV/ms; maximal allowed difference of values in intervals of 200 ms: 200 μV; minimal and maximal allowed amplitudes: ±200 μV. Finally, the EEG was re-calculated to a common average reference.

As verified by frequency analysis of the unfiltered EEG signal (time constant: 10 s, high cutoff: 1000 Hz, no notch filter), the high-frequency magnetic EMA did not disturb the EEG signals. There was no difference in FFT of EEG between times with EMA being active and inactive because the oscillations of the magnetic field of EMA are between 9 and 16 kHz, whereas EEG activity is typically below 50 Hz. In addition, the EEG equipment (electrodes and cable) did not disturb EMA data acquisition. EMA sensors were brought as close to the EEG electrodes as possible and we checked for error reports, unusual velocities or amplitudes and extreme tilt values in software from the manufacturer. Errors occurred only when an EMA coil touched the EEG electrode. During data recording this was never the case.

Data Preprocessing and Analysis

Residue Iteration Decomposition

ERP consists of several components with different trial-to-trial latency variabilities. Some components are time-locked to stimulus onset, others tend to co-vary with response time while still others are positioned somewhere in between. A disadvantage of stimulus-locked averaged ERPs is the blurring of non-stimulus-locked components. In order to solve this problem, RIDE was developed (Ouyang et al. 2011, 2015a, b). RIDE decomposes ERPs into separate component clusters with different trial-to-trial variabilites, based on external time markers or estimated single trial latencies. A general framework consists in separating ERPs into stimulus-locked, response-locked, and intermediate, latency-variable component clusters.

$$\text{ERP} = \text{S} + \text{C} + \text{R} \quad (1)$$

Components time-locked to the stimulus or to the response, respectively, are placed into S- and R-component clusters, respectively. Components that are synchronized neither to stimulus nor to response are classified into the (central) C-component cluster. RIDE has been shown to sharpen the distinction between different ERP components (e.g., Ouyang et al. 2011, 2013; Stürmer et al. 2013; Verleger et al. 2014; Wang et al. 2015).

The rational of applying RIDE to remove articulation artifacts is based on the fact that articulation artifacts are usually large in amplitude and highly variable in latency from trial to trial as indicated by the variance of voice onset latencies. Therefore the neurocognitive ERP (denoted as ERP) and articulation artifact (denoted as AA) can be treated as two different component clusters with different latency variability—time-locked to stimulus onset and speech onset, respectively—jointly constituting the Speech ERP:

$$\text{Speech ERP} = \text{ERP} + \text{AA} \quad (2)$$

The separation of these two component clusters fully conforms to the framework of RIDE as a method for separating components based on different latency variabilities. If the trial-to-trial latency variability of neurocognitive ERP components (e.g., P3b, N400) is of concern, the ERP separated from Eq. (2) can be further subdivided into S and C (and R if motor activity is of interest). In fact, the development of the RIDE algorithm aimed to deal with the variability of ERP sub-components (Ouyang et al. 2011). However, in the present work, the focus is on the removal of articulation artifacts from speech ERP data, following scheme (2). The separation of ERPs into further sub-components is not further considered in the present work.

We should point out that RIDE is not a general method for artifact rejection or correction, for example with respect to ocular or technical artifacts. However, we propose that RIDE can capture and remove articulation artifacts because they show significant trial-to-trial latency variability.

The complete RIDE algorithms and an introduction to a toolbox can be found in Ouyang et al. (2015a, b, http://cns.hkbu.edu.hk/RIDE.htm). Here, we will give a brief, self-contained description of how RIDE was applied to the present data. After the standard procedure of artifact rejection (see above), the data was prepared as a three-dimensional matrix (sampling points, electrodes, trials) in Matlab and was directly fed into the RIDE toolbox. Note that the standard procedure of artifact rejection (removing eye blinks, etc.) is unable to remove the articulation artifact. A full script for setting up the parameters and applying RIDE toolbox is included in the Appendix. Concisely, RIDE consists of a decomposition module as an inner iteration loop, that is, decomposing the data into different sub-components based on the single trial latency information, and a latency estimation module as an outer iteration loop, that is, iteratively improving the estimation of single trial latency of component(s) without external time marker.

Decomposition Module Suppose we already know the latency information of each component for each single trial, this inner module separates ERP into different component clusters. Initially all component clusters are set to be zero. To estimate the first component cluster, RIDE subtracts all the other components from each single trial, aligns the residual of all trials to the latency of the first component cluster and obtains the waveform as the *median waveform* over all time points. The same procedure is applied to obtain the second component cluster and so on. After the first round over all component clusters the algorithm starts again from the first one. The whole procedure is iterated till convergence.

Latency Estimation Module In the Decomposition Module the single trial latencies of all component clusters are supposed to be known. In reality, some ERP components (e.g., P3b, N400, P600) are not coupled to an external time marker. However, RIDE is not limited to separate components based on external time markers. For instance, in the typical scheme of separating ERPs into stimulus-locked, response-locked and intermediate component clusters S, R, and C, only the latencies of S and R clusters are known, that is, stimulus onsets (L_S) and response times (L_R), but the latencies of the assumed central component cluster C are unknown. RIDE uses a self-optimized iteration scheme for latency estimation for the marker-less component cluster, for example, the C component cluster, starting with an approximate initial estimation of L_C from the raw data. Woody's method is used to initially estimate L_C: cross-correlation time courses between the ERP and single trials are calculated for each single electrode and averaged across all electrodes. The lag of the maximum in the scalp-averaged cross-correlation time course for each single trial is taken as the single trial latency (L_C). Starting with this L_C estimate, S, C, and R can be obtained using the *Decomposition Module* till convergence. The single trial latencies of C are further improved by removing S and R from single trials and applying the cross-correlation method. The improvement of C latency and component wave-shapes of S, C and R forms an outer iteration (see the step-by-step procedure for the present application below).

The term C-component cluster refers to the component cluster that is latency-variable and has no overt latency information corresponding to external time markers (like stimulus onsets and response times). Likewise, the terms S-component cluster and R-component cluster refer to

stimulus-locked and response-locked component clusters, respectively. The RIDE method is extendable to other schemes, for example, dropping the R-component cluster when there is no response trigger, or allowing for more than one C-component cluster.

In the present data, there are two sets of time markers, stimulus onsets and voice onsets. We followed scheme (2) to separate the speech ERP into (pure) ERP and AA. Without considering further internal sub-components, the ERP is presumably locked to stimulus onset, and can thus be regarded like an S component cluster. The question is, whether AA is strictly locked to voice onset. If so, AA can be regarded like an R component cluster. Therefore, the scheme becomes: Speech ERP $\rightarrow$ S + R, which does not require any latency estimation since R is voice onset-triggered. If AA is not locked to voice onset, that is, if voice onset is a poor indicator of the latency of AA, AA can be treated like a C component cluster, the latency of which can be estimated from single trials. The scheme therefore becomes: Speech ERP $\rightarrow$ S + C. We followed both schemes and compared their performances in the result section.

A step-by-step procedure of decomposing ERP into S and C for the first dataset can be described as follows:

(1) Use the template matching (from the latency estimation module mentioned above) scheme to estimate the single trial latency of C;
(2) Average all single trials to stimulus onset and obtain an estimation of S;
(3) Remove S from all single trials, synchronize all single trials to the latency of C and obtain an estimation of C;
(4) Remove C from all single trials. Synchronize all single trials to stimulus onset and obtain an improved estimation of S;
(5) Iterate (3) and (4) to improve the estimation of S and C until convergence;
(6) Improve the estimation of the single trial latency of C from single trials after removal of S using the template matching scheme from the latency estimation module.
(7) Iterate (3)–(6) until convergence of the latency estimation and obtain the final S and C.

A step-by-step procedure of decomposing ERP into S and R for the first dataset would be as follows:

(1) Average all single trials time-locked to stimulus onset and obtain an estimate of S;
(2) Remove S from all single trials, synchronize all single trials to the voice onset and obtain an estimate of R;
(3) Remove R from all single trials. Synchronize all single trials to stimulus onset and obtain an improved estimate of S;
(4) Iterate (2) and (3) to improve the estimates of S and R until convergence.

A graphical scheme of the scenario is shown in Fig. 1.

Independent Component Analysis (ICA)

ICA (Delorme and Makeig 2004) was applied to the data after correction for eye-movements and after standard artifact rejection (see above). The purpose of applying ICA to 'cleaned' EEG data was to check whether ICA is efficient to remove the—likely artifactual—frontal positive/posterior negative activity in the EEG related to articulation that is comparable to a regular ERP component in amplitude and waveform (see also below). For the ICA on cleaned data, the default setting of the *runica* function in the EEGLAB version 13 (non-extended) was used. PCA reduction was not applied. Learning rate was heuristically determined by the toolbox based on the data and was $\ll 1$; block size was determined in the same way and was $\ll$ data length. The training was terminated when weight change was less than 1×10^{-6} and maximal number of training steps was 512.

In the context of ICA, 'component' refers to an independent source that contributes to the scalp activities with a specific weight distribution. In contrast, in the context of RIDE, 'components' are defined as parts of an ERP waveform that are temporally coupled and show similar trial-to-trial latency variability (Ouyang et al. 2011, 2015a, b).

After application of ICA, five ICs with topography patterns most similar to frontal positive/posterior negative artifacts were removed from the original data. The similarity between the ICs and frontal positive/posterior negative artifacts was calculated according to the following procedure. Firstly the topography from the grand average ERP was obtained from the time window during which the frontal positive/posterior negative artifact was most prominent, [600, 800 ms] for the first dataset by visual inspection. The Pearson correlation coefficient between this topography and the topography of each IC (projected to the scalp) was calculated. An artifact IC is expected to resemble the articulation artifact in both topography (frontal positive/posterior negative) and waveform (inverted u-shape). Note that the polarity of an IC is meaningless by definition, so an IC with frontal positive/posterior negative topography and inverted u-shaped waveform is equivalent to one with frontal negative/posterior positive topography and u-shaped waveform. So, ICs having similar topography with articulation artifact but opposite waveforms (one being u-shaped and the other one being inverted u-shaped) or the other way around, were not classified as reflecting the artifact. ICs

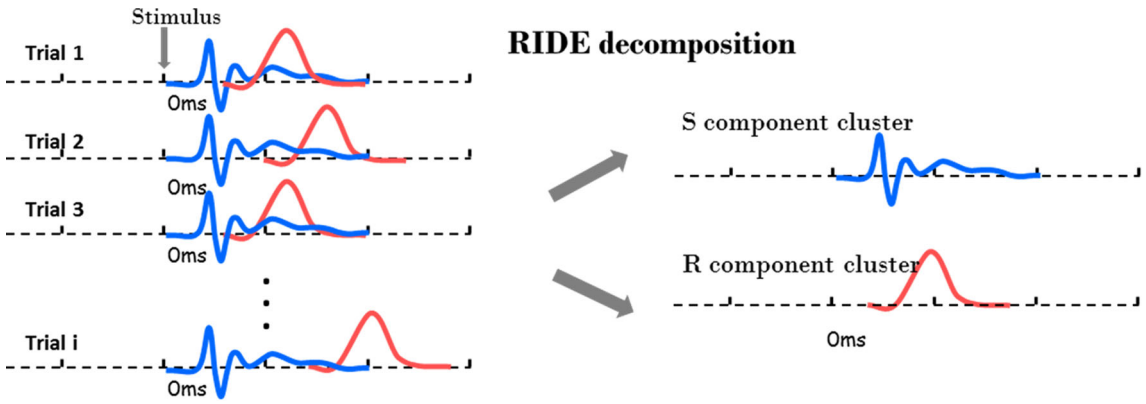

Fig. 1 Schematic illustration of the RIDE decomposition at one electrode site for two component clusters (S and R) with different trial-to-trial latency variability (Color figure online)

having opposite topographies to the articulation artifact and also opposite waveforms were classified as representing the artifact. Therefore, the sign of the correlations was corrected by multiplying it with the overall polarity (+1 or −1) of IC activation (stimulus-locked-averaged across trials and baselined at [−200, 0 ms]). Please note that, at variance with Porcaro et al. (2015), we did not employ the spectrum pattern of ICs as an additional criterion for selecting artifact ICs because, unlike high-frequency artifacts, articulation artifacts show predominantly low-frequency features (similar to a 1/f signal). Almost all the ICs showed 1/f-like patterns and the external signals from tongue, lip, and jaw movements were also low-frequent. Hence, all Pearson correlations between the spectra of the tongue movement activity and the spectra of ICs were very close to 1.0 ($ps < .01$). Therefore, the spectral information of ICs could not serve as a selection criterion in the present case.

Results

We will first describe the relationship between the ERPs and the articulation data, that is, articulator movements as recorded by the EMA and voice onsets. Then we will apply and evaluate two methods to eliminate the contribution of the articulation artifact to the ERP, namely, RIDE and ICA. We will compare the performances of RIDE with and without using external time markers (voice onsets) in separating the articulation artifact.

Relation Between ERPs and Articulation Data

On average, the interval between stimulus onset and voice onset was 649 ms ($SD = 44$ ms) ranging from 563 to 705 ms ("mila" to "titi"). Figure 2a shows the average ERP to each disyllable. Highly conspicuous is the very strong activity starting at about 500 ms after visual word presentation onset. Although the amplitudes differ across words, the late activity is always very strong and at around 100 μV across all electrodes.

Next we applied RIDE to the ERPs. As mentioned above, we applied two schemes for separating ERPs and artifacts, (1) by treating the artifact as time-locked to voice onset and (2) by estimating the single trial latencies of the artifacts. In order to compare the efficiency of these schemes, we used the following two criteria.

(1) If the voice onset time marker is more precise than latency estimation, the average coil activities will show clearer waveform pattern, e.g., less blurred, when time locked to the former as compared to the latter. In Fig. 3, left panel, we show the averaged coil activity (from the up-down movement) locked to both voice onset time and estimated latency by RIDE (as a C component), for one disyllable. It is clear that the waveform locked to voice onset is less blurred in terms of high frequency (∼6 Hz) content, though the low frequency parts are comparable. It is plausible because the latency estimated from single trials is based on the low-frequency content (the cross-correlation curve was low-pass filtered at 4 Hz in a default setting of the RIDE toolbox, Ouyang et al. 2015a) to avoid distortions by alpha EEG oscillations.

(2) Likewise, the artifact component separated by RIDE would show larger amplitudes or finer waveform details if the time marker is more precise. Figure 3, right panel, shows again the superimposed artifact waveform separated by RIDE based on the two schemes. It again supports the superiority of voice onset over estimated latency as the waveform is finer and displays larger amplitudes in high frequency oscillations (e.g., around measured voice onset). Such fine structure is more strongly smeared towards smaller amplitudes when the estimated latency has

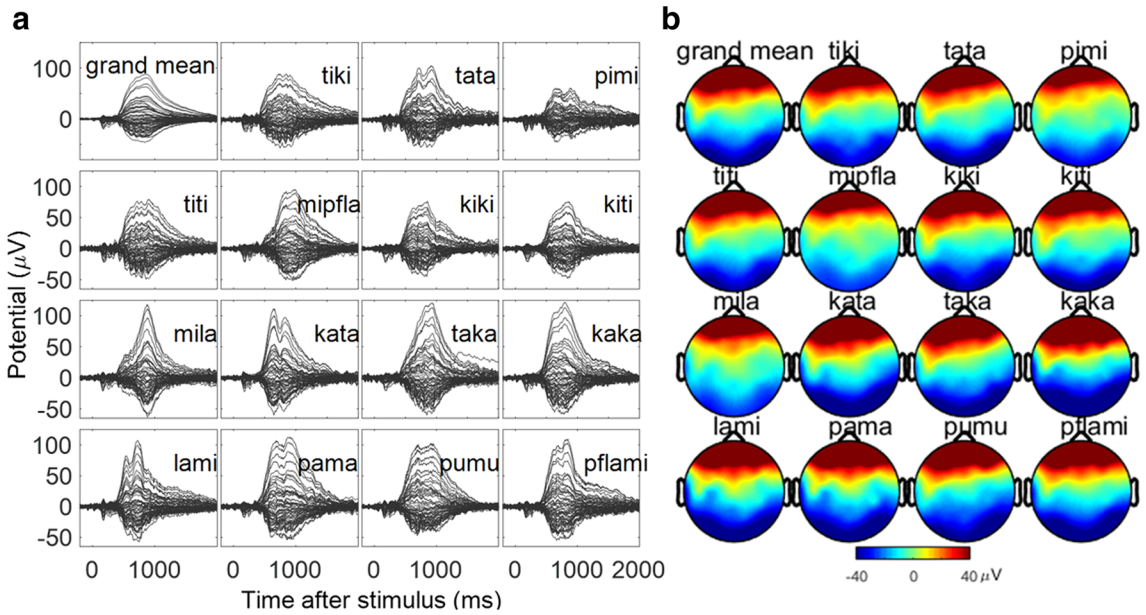

Fig. 2 **a** Grand mean ERP (*top left*) and individual ERPs for each disyllable. Waveforms for all electrodes are superimposed. **b** Topographies of articulation artifacts (R-component clusters, separated by RIDE using voice onsets) for each disyllable averaged for the time window where it was strongest [600, 800 ms] (Color figure online)

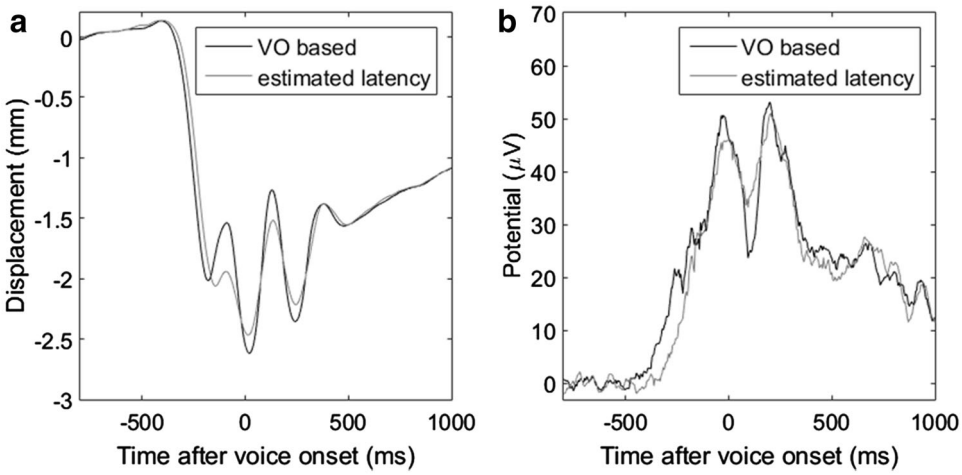

Fig. 3 Comparison of time markers used to extract artifact. **a** The average coil activity in up–down direction synchronized to the median of voice onsets (VO, *black*) and to the estimated C latency (*gray*). **b** The separated artifact component based on the voice onset (*black*) and on the estimated C latency (*gray*), from the Fpz electrode

additional jitter compared to the voice onset. The result from Fig. 3 is from a single disyllable, but the pattern (i.e., finer structure based on voice onset) is highly consistent across all disyllables. Because averaging the waveform across all disyllables again smears the high frequency pattern, only a single disyllable is shown rather than an average across all disyllables.

Despite an apparent advantage of voice onset based separation, the estimated C latency is precise enough to retain the main low-frequency part of artifact.

Figure 4 shows the decomposition of the overall ERP into S- and R-component clusters. In line with our expectation, the S component resembles a typical neurocognitive waveform pattern within the normal amplitude range whereas R captures the huge artifact. Note that there is a significant amount of temporal overlap between S and R.

Figure 2b depicts the scalp distribution of the R-components between 600 and 800 ms after stimulus presentation for each disyllable, where the R-component amplitudes were maximal. Please note the much larger color scale as compared to Fig. 4. The topographies are highly similar across the different words. Principle component analysis showed that one principle component explained 94.3 % of the variance in the RIDE-derived

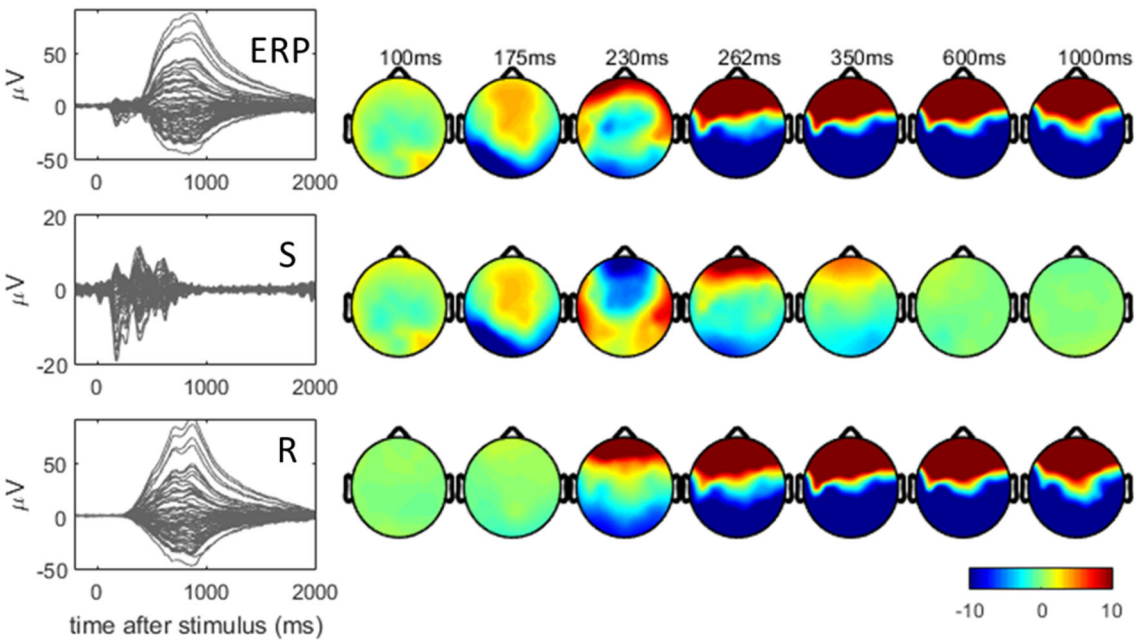

Fig. 4 RIDE separation of component clusters in Experiment 1. *Left* Grand mean ERP and RIDE-separated component clusters S and R. The scale of the *vertical axis* for S is enlarged to highlight the details of S, which basically captures all the early components of the ERP. *Right* The topography evolution of ERP, S, and R. The *color scale* is constrained to a small range (±10 μV) in order to better show the patterns in the early time window. Note the different *color scale* as compared to Fig. 2 (Color figure online)

R-component cluster; therefore, it seems that the R-component cluster, which putatively captures the articulation artifacts, is unitary and its topography does not notably depend on the particular articulation pattern.

To test the similarity of the topographies across disyllables, we applied the topographical dissimilarity measure (DISS) described by Murray et al. (2008). According to Murray et al. (2008), the DISS score between two scalp maps is defined as:

$$DISS = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{u_i-\bar{u}}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(u_i-\bar{u})^2}}-\frac{v_i-\bar{v}}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(v_i-\bar{v})^2}}\right)^2}.$$

where u_i (v_i) is the voltage of electrode i in the map u (v), $\bar{u}$ ($\bar{v}$) is the average voltage of all electrodes in the map u (v), and N is the number of the electrodes in each map. DISS values range from 0 to 2, reflecting identical and completely opposite topographies, respectively. Because there was only one participant in this experiment a permutation test was applied to single trials. In each permutation, all single trial ERPs for a pair of disyllables were mixed and each single trial was randomly assigned to one of the disyllables; afterwards the DISS value between the averaged topographies of the two newly assigned conditions was calculated. In this way 10,000 permutations were performed, yielding the same number of DISS values. The associated p value is defined by the proportion of DISS values of the permutations exceeding the DISS value between the two average scalp maps to the pair of disyllables in question.

The pairwise comparison among the 15 disyllables (105 pairs) generated mean DISS values of 0.143 ($SD = 0.044$), indicating that the topographies are highly similar across disyllables. Though mean DISS had small values, the measure is sensitive to subtle differences when submitted to permutation testing. The Bonferroni corrected p values generated by DISS analysis revealed topographical differences between 26 disyllable pairs with distinct articulation patterns (e.g., 'pama' and 'kiti'; $p < .05$), whereas 79 pairs with similar patterns had indistinguishable topographies (e.g., 'kiki' and 'kiti'; $p > .05$).

This single participant study allowed us to directly relate the EEG signal to articulatory movements. Figure 5a shows the global field power (GFP) of the activity captured in the R-component cluster in the interval [−800, 1000 ms] relative to voice onset, averaged over all stimuli. Superimposed we show the movements averaged across all five coils in each of the three spatial dimensions. This figure shows that GFP starts together with the movements of the articulators (at least in the vertical and lateral dimensions); this joint onset of movement and R-component cluster activity occurs already 300–400 ms prior to the audible voice onset. The vertical (UD) direction seems to give the clearest signals overall. Therefore, we used this signal to make comparisons for movement onset and R-component cluster onset for each disyllable (Fig. 5b).

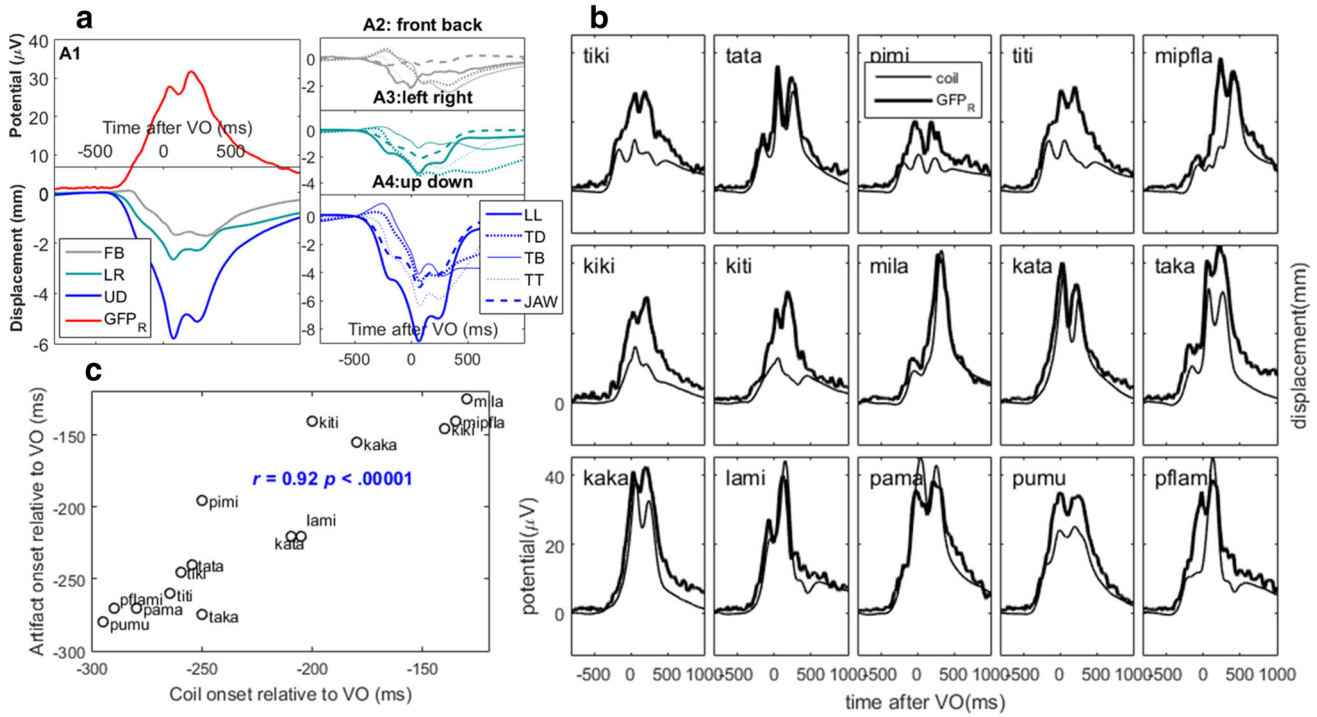

Fig. 5 **a** Relationship between global field power (GFP) of the RIDE-derived R-component cluster and coil activities synchronized to voice onset (VO). A1: GFP_R: GFP of R-component cluster (grand mean over all disyllables) and the averaged activities for each movement direction (*FB front–back*, *LR left–right*, *UD up–down*) from five coils and across all disyllables. A2-A4: Coil activities for each movement direction and coil (*LL* lower lip, *TD* tongue dorsum, *TB* tongue back, *TT* tongue tip, *JAW* lower incisors). **b** Superposition of the global field power of the R-component clusters (*black*, *left axis*) and the vertical movement signals (*gray*, *right axis*) for all 5 coils, synchronized to voice onset (time zero). Polarity of displacement has been inverted for ease of comparison with GFP of EEG. **c** Onsets of global field power of the R-component cluster versus onsets of articulation for each disyllable (Color figure online)

In Fig. 5b one can see that for each disyllable GFP and vertical movement onset are largely co-occurring. This impression was verified by measuring the onset of each signal type for each disyllable. The onsets were measured with thresholds of 5.0 μV for GFP_R and −1.0 mm for vertical coil displacements (baseline at [−800, −500 ms] relative to voice onset). Figure 5c shows the scatter plot for these measurements together with the label of each disyllable, demonstrating the very high correlation of these onsets (Pearson's $r = .92$; $p < .0001$). Although this correlation value should depend on the thresholds chosen for measuring the onsets of GFP_R and coil displacement, the present result is sufficient for showing the association between EEG artifact and articulation movement. It is worthwhile to note that the correlation value was .94 when the artifact was separated by the S + C scheme. This is plausible since the smoother artifact waveform was obtained due to the low-pass setting in C latency estimation (see Fig. 3b) makes the onset measurement less affected by noise. Figure 5c also shows that these onsets are quite variable and range from −100 to −300 ms depending on the word, given that the onset threshold was set to be relatively high for reliable detection.

Closer inspection of Fig. 5b shows that not only the onsets of the GFP in the EEG and the EMA are similar; also the time courses are sometimes very similar for at least some disyllables, although the shapes of these curves appear to be specific for the articulation patterns of the disyllables. To quantify the similarity of the time courses we calculated the Pearson correlations of the grand-averaged vertical coil signals with the grand-averaged GFP in the R-component cluster over time in the interval between −500 and +1000 ms relative to voice onset for each word. The Pearson correlations are lowest for the anterior-posterior direction ($r = -0.58 \pm 0.54$) and of similar high magnitudes for the vertical ($r = -0.88 \pm 0.1$) and lateral ($r = -0.90 \pm 0.08$) directions.

Artifact Correction

If RIDE is effective in eliminating articulation artifacts by taking out the R-component cluster, the similarity between the time courses of articulatory movements and the residual ERP should be much smaller than when the artifact is still present. We calculated the Pearson correlation between the mean waveforms from the five coils and the ERP time

course for every single trial and every electrode, which was then averaged for each disyllable. Figure 6, left column in each panel shows the distribution of the correlations for all disyllable before removal of the R-component cluster. After removing the R-component cluster from the ERP, correlations fell to $r < .1$ (Fig. 6a–c, Column 2 in each panel).

To see how well the artifact removal works even without utilizing the voice onset information we also showed the correlation value based on the S + C scheme, i.e., removal articulation artifact derived by C component based on single trial latency estimation assuming that voice onset is not available. Surprisingly, it seems that the S + C scheme actually removes the artifact more cleanly as shown by the slightly smaller correlation value (Fig. 6a–c, Column 3 in each panel). This strongly indicates that, although the artifact waveform extracted as C is not finer than voice onset-based in terms of waveform details (Fig. 3), the overall performance is not worse than voice onset-based (also see the comparison between the correlation values from Fig. 5c). Therefore the application of speech artifact removal on voice-trigger-free data is feasible.

As a comparison we also show the performance of ICA on articulation artifact rejection (Fig. 6a–c, Column 4). Although diminished, the correlations remained substantial after removal of the artifact-related ICs, suggesting that the artifacts are not completely removed. A paired t test showed significant differences in the correlations of the ERP and coil activities between the ICA and RIDE (VO-based) collapsed across movement directions: $t(14) = 3.7$, $p = .002$. To further test whether the remaining relatively high correlations after ICA correction are due to not removing enough ICs, we calculated the mean absolute correlations (for the forward–backward dimension; Fig. 6a) as a function of the number of ICs removed, which most strongly resemble the articulation artifact topography (Fig. 6d). We also plotted the change in N1 amplitude for each set of ICs removed (Fig. 6e).

Discussion

The single participant Experiment 1 aimed at relating the putative articulation artifact component in the ERP while pronouncing different disyllabic pseudowords to the activity of the articulators as measured with EMA. We obtained the following information about the properties of the observed anterior-positive/posterior-negative component extracted by means of RIDE.

First, the component showed very consistent scalp topography across rather different articulation patterns, though subtle differences can still be detected by sensitive DISS measures. This is a promising finding for any attempt to deal with the artifact. If the topography were in fact strongly different for different phonemes, the situation would become much more complex.

Second, the onset of what we consider the artifact component in the ERP was very tightly coupled to the onset of articulator movements with a high correlation close to unity, confirming the intimate relationship between

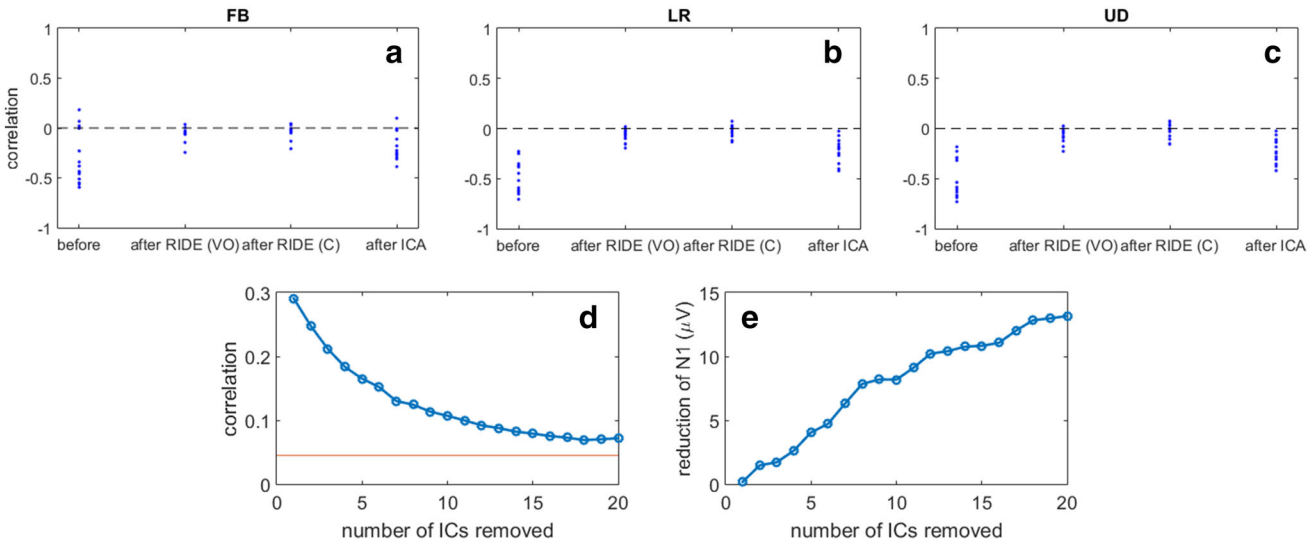

Fig. 6 **a–c** Correlation between time courses of coil movements and ERPs before (Column 1) and after removal of the articulation artifact based on voice onset (S + R scheme, Column 2) and estimated C latency (S + C scheme, Column 3) and after removal of the first 5 artifact-related ICs from ICA (Column 4). Each *dot* corresponds to one disyllable. **d** The reduction of mean correlations from A by ICA as a function of number of ICs removed in the order of topographical similarity with the artifact. For reference, the *horizontal orange line* shows the corresponding value for RIDE ($r = .045$). **e** The change in N1 amplitude from PO9, averaged within [160, 200 ms]) as a function of number of artifact ICs removed (Color figure online)

these activities. Both onsets were far earlier than that of the voice, sometimes—depending on the initial phonemes—preceding the voice trigger by 300 ms (Fig. 5a, c). This poses an important limitation to the use of ERPs preceding the voice onset in overt articulation tasks without appropriate artifact removal or a strict limitation of the speech material. Please note that the precise onset asynchrony between artifact onset and measured speech onset should strongly depend on the accuracy of the method used to determine speech onsets. We will come back to this issue in the "General Discussion" section. Third, although the artifact topographies did not differ strongly between articulation patterns, their time courses did. Specifically, the transmission of articulator displacement into ERP amplitude of the artifacts, rather than the topographies, is modulated. This was to be expected because different phoneme sequences may cause different artifact waveforms over time. Although more detailed analyses would exceed the scope of the present study and intentions, the specific coil information and how precisely the coils relate to the EEG artifact at different moments in time may reveal valuable additional information. For practical purposes of pure EEG recordings during speech production without EMA information, knowledge of such modulations is of little value. As we demonstrate here, the more fine grained EMA information is not necessary in order to deal with the artifact.

Finally, and of great importance for the question of artifact removal is the temporal relationship between the EEG and the articulator movements. In the present experiment we could use the articulator movements as an external criterion for the presence and strength of the artifact over time and its attenuation by artifact correction. In short, any EEG activity that shows a relationship to this movement measure over time may be caused by the articulator movement. Although muscle activity and hence movements are caused by the brain, the relationship between brain activity and movement or force is not a simple linear relationship (e.g., Sommer et al. 1994), in the sense that muscle activity is a direct function of the scalp recorded EEG signal. Indeed there were several articulation patterns (reflected in our measure lumping together all vertical movements of the coils) that showed correlations with the ERP over time of $r > .5$. Together with the very high correlation of the onsets of the R-component cluster in the GFP and the mean vertical coil movements, this strongly suggests a causal relationship between articulator movements and the frontal positive/posterior negative ERP. When only the S-component cluster, that is, the ERP without the R-component cluster, was correlated with the vertical coil movements, this correlation was strongly reduced even for those disyllables where the correlation was strongest. This is an indication that removal of the R-component cluster from the ERP is an efficient procedure to eliminate the articulation artifact from the ERP.

In contrast to RIDE, ICA could not remove the artifacts effectively, since the correlations with the coil movements were still considerable after removal of the major artifact-related ICs. It is interesting to note that the correlations of the ICA-corrected ERP with the articulator movements in the present experiment were similar in size to those reported by Porcaro et al. (2015) with lip-EMG. Moreover, ICA seems to distort early, articulation-ICs. This is illustrated in Fig. 6d, e. The distortion of early component can also be seen in the paper of Porcaro et al. (2015; Fig. 4). Although in principle the correlation of the ICA-corrected ERP and the articulator activity can decrease to values almost as low as those obtained after RIDE when more ICs are removed this comes at the cost of increasingly and seriously distorting the N1 (Fig. 6e) and also the P1 component (see Exp. 2).

Experiment 1 had mainly aimed at demonstrating that the R-component cluster as identified here is indeed related to movements of the articulators. This was only a single case study, requiring high motivation on the part of the participant, and quite unlike the typical language production study. Therefore, Experiment 2 analyzed ERP data from a typical picture naming task taken from a group of participants in order to assess the generalization of the findings of Experiment 1. In particular, we were interested whether in a more complex task the same component structure would be observed, and furthermore, we wanted to establish its relationship to voice onset. Finally we revisited the question of artifact removal with RIDE and compared it with the more common approach of ICA.

Experiment 2

Methods

Participants

24 right-handed native German speakers (14 women; mean age = 22.3 years; range 18–30) were paid for participation or received partial fulfillment of a curriculum requirement. All participants reported normal or corrected-to-normal visual acuity and normal color vision. Informed consent was obtained before the experiment. One participant was replaced because of high error rates.

Materials

We used 125 color photographs of common objects from five broad categories (animals, clothing, food and beverages, furniture, and tools, please see Appendix II for

details). The size of the photographs was 3.5 × 3.5 cm at an approximate viewing distance of 90 cm from the monitor.

Procedure and Design

Each trial began with a fixation cross, displayed in the centre of a light-grey computer monitor. After 500 ms the fixation cross was replaced by a picture, which remained on the screen until vocal response onset with a maximal duration of 2 s. The stimulus was followed by a blank screen for 1 s. Participants were instructed to name each picture as fast and accurately as possible. Naming times were recorded with a microphone and latencies were measured by voice key. Voice keys are a still frequently used method to determine naming latencies in current studies combining overt naming with EEG recordings, although many researchers use manual re-alignment of vocal onsets on basis of the spectrograms (e.g., Protopapas 2007). Naming accuracy and voice key functioning were monitored online by the experimenter. Trials in which the voice key triggered too early and trials with wrong naming responses were coded online by the experimenter and excluded from further analyses.

The experiment consisted of different conditions. However, for the present purpose we selected a baseline condition in which the pictures were presented in a random sequence. Each picture was presented and named 5 times, resulting in 625 trials. Prior to the experiment participants were familiarized with the pictures and their names as follows: all photographs were presented in random order on the screen and participants were asked to name each picture. If necessary, they were corrected or the picture name was provided by the experimenter. After this procedure was repeated once, participants were given a color sheet with all pictures and their names.

Data recording

The electroencephalogram (EEG) was recorded with Ag/AgCl electrodes from 64 sites according to the extended 10–20 system, referenced to the left mastoid, and at a sampling rate of 500 Hz (bandpass 0.032–70 Hz). Electrooculograms were recorded from the left and right outer canthi and beneath and above the left eye. Electrode impedance was kept below 5 kΩ. Prototypical eye movements for later artifact correction were obtained in a calibration procedure.

RIDE

Figure 7a shows the grand average ERP across all participants (mean voice key latency = 647 ms). The ERP was characterized by the usual early visual components followed by a pronounced parietal positive component, resembling the P3. Starting at around 650 ms a frontal positive/posterior negative component appeared, which was similar to what had been seen in Experiment 1 and identified as articulation artifact. Consistent with Experiment 1, we separated the speech ERP into ERP (denoted as S) and articulation artifact (denoted as R). The only difference to Experiment 1 was that, here R (the artifact) was extracted based on the voice key information since voice onset was not available.

Results and Discussion

RIDE Decomposition

Figure 7 shows the stimulus-locked time courses and topography evolutions of the ERP and each RIDE component cluster (S and R). ERP shows mixing of component clusters with dynamically changing topography over time. The scenario in the RIDE components is in line with expectations based on Experiment 1: The S-component cluster captured the early ERP components and the P3b-like activity, and the R component emerged at a later time stage and completely captured the frontal-positive/posterior-negative artifact activity.

Figure 8 depicts individual time courses of the GFP of the R component for all participants. It shows that the onsets of R can be as early as around −400 ms prior to voice key latency, consistent with Experiment 1. The topography of R (averaged from −300 ms to 500 ms relative to voice key latency) was somewhat variable across participants, with most of them showing the frontal-positive/parietal-negative pattern already observed in Experiment 1. Also, the amplitude was quite variable in the order of about one magnitude across participants.

ERP After Removal of Articulation Artifact

The topography and time course of S (grand averaged, here S represents the clean ERP) is shown in Fig. 9. The removal of the R-component cluster, the presumable artifact, significantly affected the ERP amplitude within the window as early as [300, 500 ms]. A paired *t* test revealed a significant amplitude difference between original and S between 300 and 500 ms is significant ($ps < .05$) for 42 out of 63 electrodes indicated by the red squares in the head cartoon inserted in Fig. 9. Though not adjusted for multiple comparisons, it strongly indicates the distortion of ERP amplitude by articulation artifact in a relatively early time window.

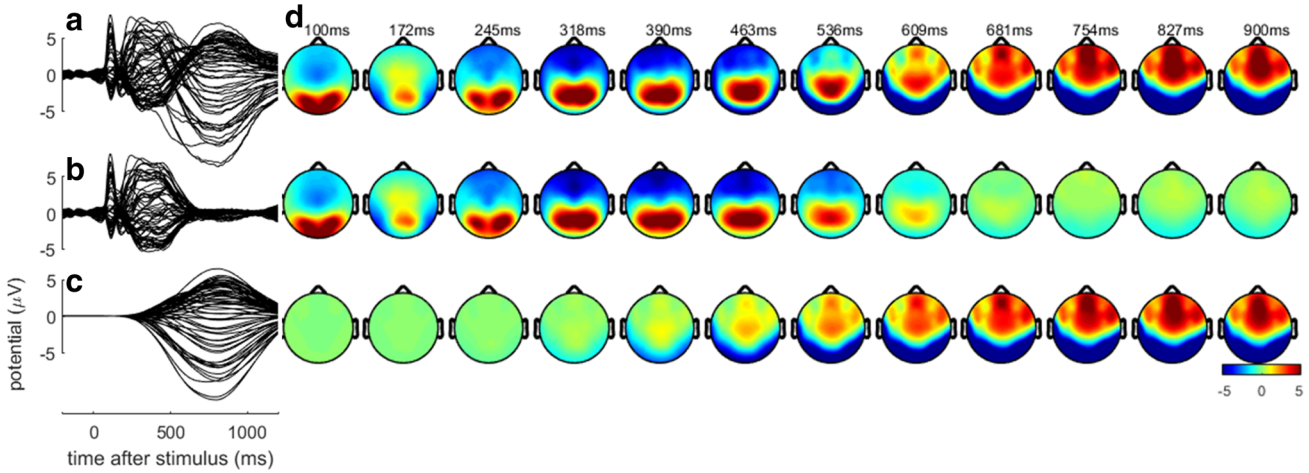

Fig. 7 *Left* **a–c** time courses superimposed for all electrodes for ERP (*up*), S (*middle*) and R (*down*). *Right* **d** corresponding topography evolutions for the wave shapes depicted on the *left* (Color figure online)

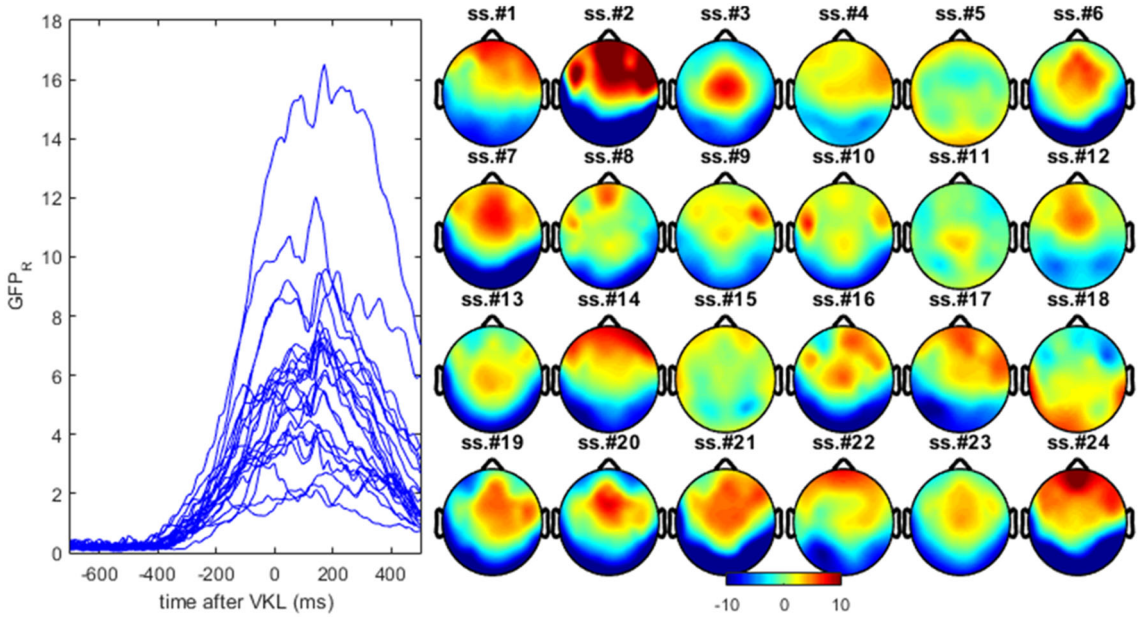

Fig. 8 The global field power of R component cluster (the artifact) locked to voice key latency (VKL, time 0) and its topography averaged from −300 ms to 500 ms for each participant (Color figure online)

Artifact Removal with ICA

We applied ICA on the single trial ERP data of each participant and decomposed it into 63 ICs, each of which has a topography pattern (namely weight distribution across scalp). We then calculated the correlations between the topography pattern of ICs and that of the articulation artifact derived from the grand average ERP in the time window 700–1000 ms. The sorted correlations are shown for two selected participants in Fig. 10 (ss. #21 and ss. #24 with conspicuous articulation artifact, see Fig. 7). A number of ICs were highly correlated with the articulation artifact in terms of topography. However, the correlations distributed rather linearly from −1 to +1, offering no clear cut transition between artifactual and non-artifactual ICs. Here the correlation of −1 is not related to the articulation artifact because we have corrected the sign of correlation by constraining both the topography and waveform to the same polarity of artifact (see procedures above). We show the topographies of the five ICs (red) with highest correlations in Fig. 10 and the variance explained in the ERP for each IC. For removal of the artifactual ICs, the activations of these five ICs were back-projected to the scalp and subtracted from the original ERP. Since there is no clear cutoff it is hard to provide a justification for the number of ICs to be artifactual. Here we picked the five ICs with the

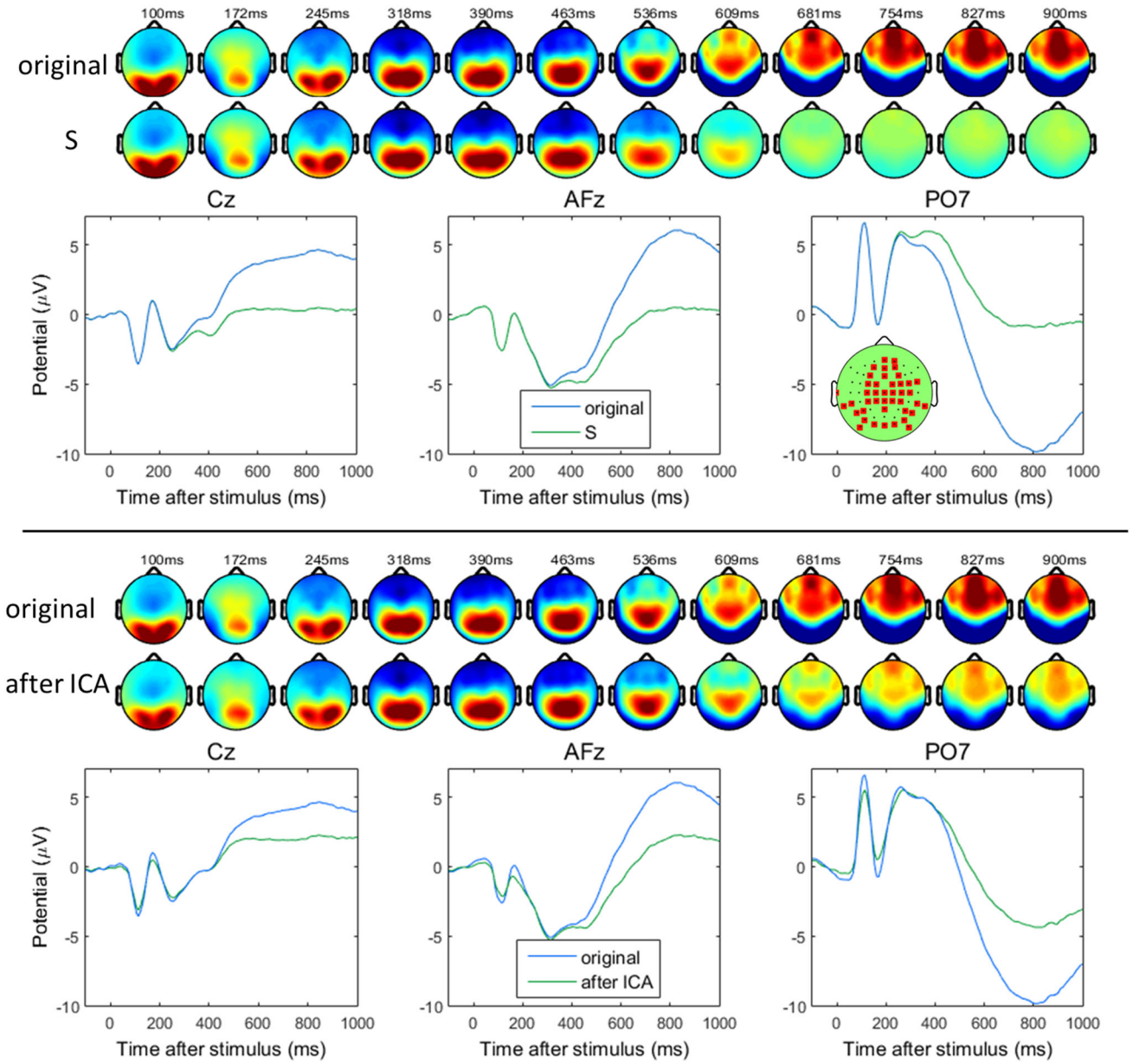

Fig. 9 *Top* Removal of the articulation artifact from the original ERP by RIDE. The topography array shows the evolution of (stimulus-locked) grand mean ERP and S. The three plots at the *bottom* are the time courses of ERP and S from different electrodes, for comparison. The *green insert* in the *right panel* shows the electrodes (*red squares*) where the difference between original ERP and S is significant already in [300, 500 ms]. *Bottom* Removal of articulation artifacts from the original ERP by ICA. Note the change in the early components in the plots of the waveform especially at posterior electrodes (Color figure online)

highest correlations, accounting for a substantial amount of variance in the data (Fig. 10).

The original ERP and the ICA-reconstructed ERP after removal of the articulation artifact is shown in Fig. 9 for the comparison with the results from RIDE. Removing the five most relevant ICs did not cleanly remove the artifact components—a significant residue remained in the reconstructed ERP (see the time window from 700 ms to 900 ms). Paired t tests revealed significant differences in amplitude reductions in the time window 600–900 ms between RIDE and ICA across different electrodes (AFz: $t(23) = 2.1$, $p = .04$; Cz: $t(23) = 5.2$, $p < .001$; PO7: $t(23) = 3.71$, $p < .01$; 38 out of 63 electrodes showed $p < .05$). The other problem is that, after removing the five ICs most closely resembling the articulation artifact, the components in the early time window (P1/N1) were also significantly altered, $t(23) = 2.4$, $p < 0.05$, at PO8 [100, 120 ms]. That means, the more ICs supposedly related to the frontal artifacts are removed, the less artifact will be left, but with the consequence to increasingly affect the

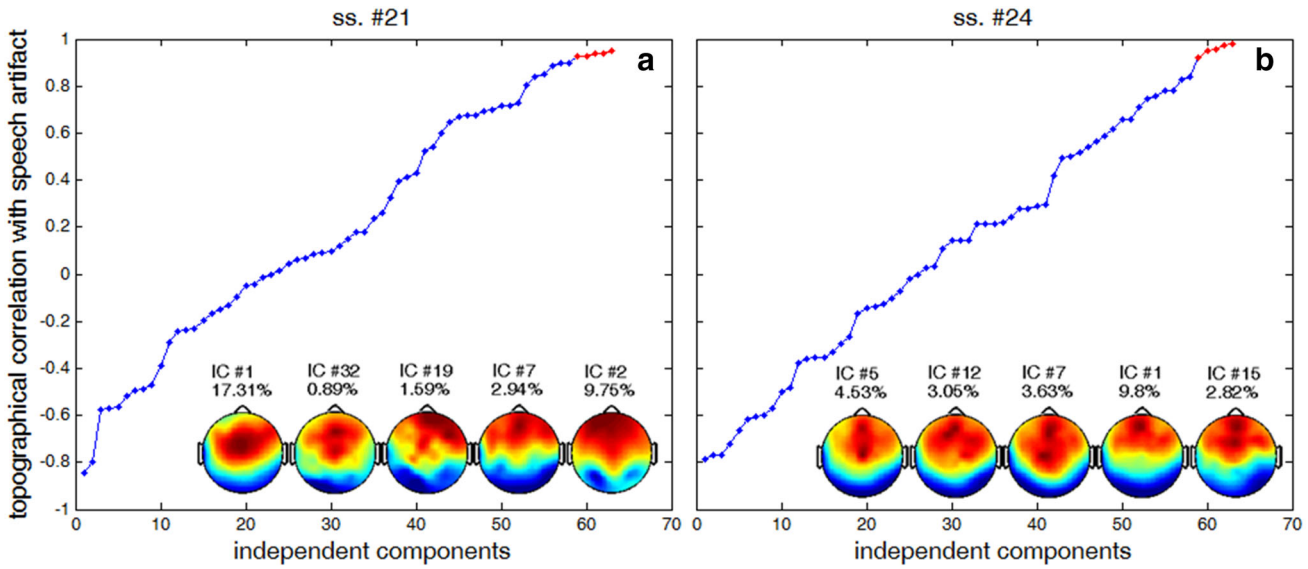

Fig. 10 Sorted correlations between the topographies of ICs and of the articulation artifact. The five ICs most closely resembling the frontal positive/posterior negative artifact are shown along with the variance explained. Data are from two selected participants with conspicuous articulation artifacts (Color figure online)

earlier ERP components. This is probably because ICA decomposition did not find ICs exclusively capturing the articulation artifact.

General Discussion

The aim of the present study was to describe the temporal and topographical properties of articulation-related artifacts during overt speech production in the EEG and to suggest a method to remove the artifacts from the EEG. To this end, we co-registered activity of the inner and outer vocal tract, measured with an electromagnetic articulograph, and an EEG during overt articulation from a single participant. The information derived from this study was used to describe the artifacts and their relation to articulatory movements. Furthermore, we used RIDE to identify and eliminate articulation artifacts from the EEG. In a second data set from a group experiment we applied RIDE to the data of a typical picture naming task in order to isolate and remove the articulation artifacts from the EEG.

Description of Articulation-Related ERP Activity

In both experiments, about 200–300 ms prior to voice onset or voice key the frontal positive/posterior negative activity appeared, which is typical when overt naming or speaking is required. For Experiment 2 the average range of this articulation-related activity across the scalp (max–min) was about 15 microvolt (at its peak latency) but varied across participants by about one order of magnitude. In the single participant of Experiment 1 the articulation-related activity was quite large but within the range of Experiment 2.

Crucially, the topographical distributions of the articulation-related activation pattern were very similar across disyllables in Experiment 1; for all disyllables the pattern consisted in a frontal positivity accompanied by a posterior negativity. In Experiment 2 the grand mean topography of the articulation-related activity was also characterized by a posterior negativity but the frontal positivity extended further to central sites. There was some inter-individual variability in the distribution of the fronto-central positivity, but overall the bipolar distribution was consistent.

Please note that the articulation artifact described here is different from (residual) blink artifacts. Blink topographies show a frontal focus and quickly fall off towards posterior sites (e.g., Lins et al. 1993). However our artifact topographies never showed such a distribution—they were invariably bipolar with negativities at posterior sites that were similarly focused as their positive counterparts at frontal sites.

Consistency of the topography of the articulation-related activity in the ERP can also be demonstrated by an analysis of widely different words in Experiment 2 (Fig. 11). In this analysis we averaged the ERPs with names of pictures that differed in many respects. However in contrast to the present Experiment 1 with only one participant, averaging was done across a group of 24 participants and per word. Here, although differing in onset, the later part of the ERP for each word was always characterized by the typical frontal positivity, accompanied by a posterior negativity.

The DISS analysis (Murray et al. 2008) done on the 7550 pairs among 125 different words showed very low DISS values ($M = 0.26$; $SD = 0.10$), suggesting very similar artifact topographies. Albeit this overall similarity, artifact topographies showed subtle variations between words when the permutation test was applied. 550 of the pairwise comparisons were statistically significant ($ps < .05$, after Bonferroni correction). Hence, we may conclude that the articulation-related activity has a rather consistent bipolar frontal (or fronto-central) positive/posterior negative pattern that nevertheless subtly differs as a function of the articulation pattern.

Experiment 1 showed that this articulation related pattern in the EEG is strongly correlated over time with the movement of the articulators (tongue, jaw, and lip). Indeed, not only facial EMG in the form of lip movements, but clearly also activities of the main articulators such as tongue and jaw movements significantly contribute to the artifact measured in the EEG, and different movement directions—in particular vertical movements—contribute to the artifact. Crucially, the onset of the frontal-positive/posterior negative activity is correlated almost with unity to the onset of articulator movements. Therefore, there is little room for doubt that the articulation-related EEG pattern is indeed largely non-neural in origin and can therefore be legitimately considered as artifactual contribution to the EEG.

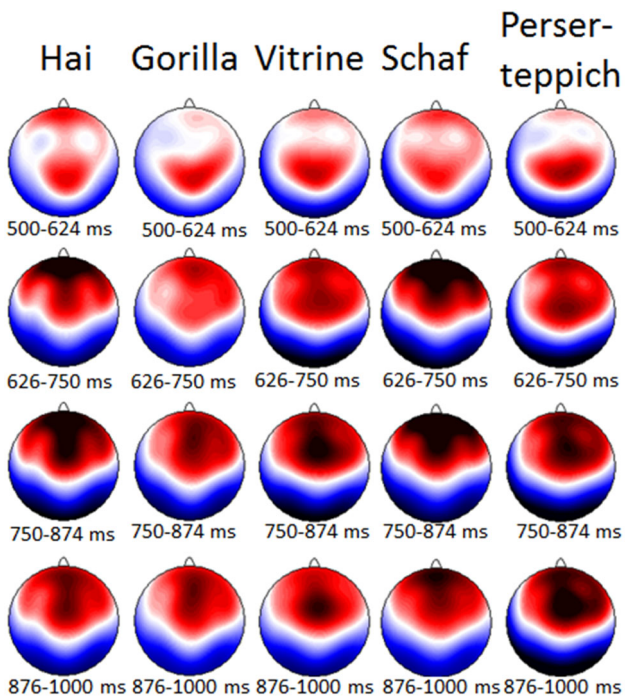

Fig. 11 Examples for the articulation-related artifact for several different words, averaged across 24 participants. Shown are the scalp distributions of the ERPs for time segments where the artifact dominates (Color figure online)

Concerning temporal properties, we found the articulation-related activity to start prior to voice onset in both experiments. As shown in Experiment 1, also articulatory movements—as more valid measures of early speech-related activity—precede voice onset. The decomposition of the ERP with RIDE allows a more precise look into the components. The frontal positivity appears to emerge about 200–300 ms before voice onset. This is of course not surprising because articulators have to be adjusted before any phonemes can be produced. However, the present findings point out that the threshold setting of measuring the voice onset is a critical issue when one wants to record ERPs in an experiment including overt articulation.

Removing Articulation-Related Activity

Given that the articulation-related artifactual part of the ERP in a language production task can be unambiguously identified and allocated to a separable ERP component, methods for separating the artifact-free contribution to the ERP from the artifactual part can be applied.

We assessed two methods to eliminate the articulation-related artifacts, RIDE and ICA. ICA has been frequently used for correcting artifacts of ocular origin and—especially if guided by eye-tracking—is an excellent method for this purpose. It has also been employed to correct artifactual EEG activity induced by articulation (Porcaro et al. 2015; De Vos et al. 2010). The question remains how well the ICA is able to correct the typically slow articulation-related activities. In speech production this can be judged firstly by the correlation with an external measure of the artifact. Although ICA was able to strongly reduce this correlation in both the present Experiment 1 and in the study of Porcaro et al. (2015), the remaining correlations of >.2 compared unfavorably with the correlations of <.1 obtained by RIDE. A second criterion is the effect of the correction methods on ERP components occurring before a speech artifact is likely to be present. In the very recent work by Porcaro et al. (2015) the speech artifact derived by ICA was shown to contain a visual P1/N1 complex (Porcaro et al. 2015; Fig. 3). If these components are removed as artifacts, also the early ERP signal will be affected. A similar effect was seen in the present experiments where the early components were diminished after ICA. In contrast, the reconstructed ERP after RIDE did not distort the early ERP components but cleanly removed only the frontal-positive/posterior negative artifact activity.

As often the case when one has to select from the numerous and person-specific ICs it is to some extent a subjective decision, which IC to drop and which to include in the reconstructed signal. In the present case, we saw a trade-off between eliminating the artifact and distorting early ERP components that are most likely unaffected by

articulation or leaving the early components unaffected but (likely) undercorrecting for articulation artifacts. This is due to the reason that many ICs do not exclusively contain the artifact-related component but also contain essential ERP components like P1/N1. ICA is based on the assumption that ICs are generated by independent sources, which are supposed to be manifested in the statistically independent signals. This assumption of independence in measured signals may not be satisfied even if they could be indeed generated by unrelated sources, which appears to be the case for the slow articulation-related artifacts and processing-generated ERP components (e.g., signal spectra could not be employed here for the selection of ICs). RIDE works with the different assumption that component clusters in the ERP have relatively independent latency variability, but does not require the independence in the signals of the components. A detailed comparison of the ICA and RIDE for simulated data was performed by Ouyang et al. (2015a, b) for different assumptions. In the case of speech production as shown for the present study, RIDE appears to be a better option to deal with articulation artifacts. Nevertheless, the performance of these methods may vary depending on the specific set-up of experiments. Specifically, the ICA may perform better when an even greater number of electrodes is used, especially when located at regions that are highly related to the artifacts. Further evaluation of the methods should be done.

Limitations

In the following, two limitations are discussed concerning the choice of speech onset measures. First, recordings of the actual speech signal are more precise measures of the actual voice onset than the voice key (for a discussion see Kessler et al. 2002), and a direct comparison with speech onset estimates from the speech signal would be desirable. On the other hand, as discussed above, voice keys are the most frequently employed method in current studies combining EEG measures with overt articulation. One reason is that most of these studies compare identical words in different conditions, in which case the precise onsets are not so crucial and the use of voice keys is an economically valid choice. Given this situation, the use of voice key triggers in Experiment 2 reflects typical—albeit possibly questionable—methodology in the area of language production using overt articulation and EEG measures. The present paper provides valuable information about the temporal relation between articulation artifacts and speech onsets as measured by both voice key triggers (Experiment 2) and by determination from voice recordings (Experiment 1).

The precise timing between artifact onset and speech onset depends on the method used to determine speech onsets, and should differ between these methods. For a test and critical discussion of biases of voice key measurements and the pros and cons of this and other methods, please see Kessler et al. (2002). Thus, our estimates of how early articulation-induced EEG artifacts will appear relative to speech onset should depend on how the latter is measured. However, in our measurements, the temporal relationship between the voice onset/key latency and the onset of articulation is remarkably similar across the two experiments. Table 1 gives the means and standard deviations (SD) of the voice onsets for all disyllables in Experiment 1 and the voice key latencies for all participants in Experiment 2. The onsets of the artifacts were measured with a 5 μV threshold to the GFP of the artifact component (R) in Experiment 1 and 1 μV for Experiment 2 since the overall amplitude was much smaller. However, both thresholds are approximately 10 % of the maximum. Interestingly, in both experiments, mean artifact onsets preceded the mean voice markers by almost 300 ms. This shows that (1) the kind of voice onset measure used is not decisive and that (2) the "artifact safe" period preceding such triggers ends already about 300 ms prior to the trigger.

With more accurate onset estimates the relative timing between the beginning of the articulation artifacts and the beginning of the articulation onset should also vary, and it would be desirable in future studies to directly compare speech onset times, especially when comparing different words across conditions, but also when comparing the same words with different onsets in different conditions. However, even though speech onsets are more precise measures, the principle problem remains that all measures need some minimal auditory input, while silent and undetected (pre)articulatory processes (and the corresponding EEG artifacts) may have already started (e.g., tongue or jaw movements), as demonstrated in Experiment 1. In this sense more precise speech onset measures are of limited help.

As to the usefulness or even necessity of voice markers for the extraction and separation of articulation artifact, our results showed satisfactory performance for all options. In the present study we used voice onset for Experiment 1 and voice key latency for Experiment 2. Taking the two experiments together, we showed that both voice onset and voice key are good candidates for representing the articulation latency, in the sense that the result of articulation removal is successful in a very high degree. If these time markers had been imprecise, the artifact waveform could not be cleanly removed or would even remain in the data, which is opposite to our results (Figs. 4, 5, and 7). The precision of voice onset is also shown to be slightly better than direct latency estimation from single trial ERP. Notably, however, the latency estimation scheme without any voice triggers is surprisingly good, given the noisy single trial ERP data. Hence, artifact removal without any voice markers seems to be a feasible option.

Table 1 The temporal relationship between voice onset/key latency and measured onset of artifact

	Experiment 1		Experiment 2	
	Mean (ms)	SD (ms)	Mean (ms)	SD (ms)
Voice onset/key latency	649	44	647	54
Onset of artifact	372	21	353	41
Difference	277		294	

Finally, as a minor point, RIDE is a secondary processing step applied on ERPs after preprocessing. That means, as long as the preprocessing steps affect the outcome of ERP, they affect the outcomes of RIDE (as well as other decomposition methods). Different reference montages can strongly affect the morphology of ERP (Nunez et al. 1997). And different reference montages have different advantages in certain circumstances. In principle, we advocate the optimal usage of reference. But in the present case, the performance of the artifact removal does not seem to be compromised by average referencing. As the topography showed (Figs. 4 and 7), the artifact were cleanly removed even when it was spread across the scalp.

Conclusions

From the present analyses we can draw the following conclusions. (1) During overt speaking there is a fronto(-central) positive/posterior negative EEG activity consistent in topography across words, phonemes, and participants although its amplitude and time course may vary greatly. Hitherto no such comparison between the artifacts elicited by different articulations has been available. (2) The onset and time course of this ERP activity are strongly related to the onset and time course of articulatory movements across different articulated disyllables, making it highly likely to be of non-neural, artifactual origin. (3) The artifact may start up to three hundred milliseconds before voice onset, indicating that the "safe" interval without articulation artifacts in the ERP may be shorter than has been thought. (4) RIDE allows to remove articulation artifacts without distorting pre-articulation ERP components, which is efficient for different types of voice onset markers and works even without any markers. Hence, RIDE provides a novel tool for removing the speech artifact from the EEG in unprecedented ways, opening new doors for EEG research using overt articulation.

Acknowledgments This work was partially supported by Hong Kong Baptist University (HKBU) Strategic Development Fund, the HKBU Faculty Research Grant (FRG2/13-14/022), the HKBU Matching Proof-of-Concept Fund and Germany-Hong Kong Joint Research Scheme (G-HK012/12), the National Natural Science Foundation of China (Grant No. 11275027) to G.O. and C.Z., the Germany-Hong Kong Joint Research Scheme (PPP 56062391) to W.S, and grant AB 277 4 from the German Research Council to RAR. This research was conducted using the resources of the High Performance Cluster Computing Centre, Hong Kong Baptist University, which receives funding from RGC, University Grant Committee of the HKSAR and HKBU.

We thank Susanne Fuchs for generously supporting us with her expertise on phonetics, Joerg Dreyer and Guido Kiecker for their technical support during the EMA-EEG recording session, and Leonardo Lancia for preprocessing the EMA data and post-synchronizing the EMA-EEG data.

Appendix

Appendix I. Matlab Script for Experiment 1

For Experiment 1 the epoched single trial ERP data after standard artifact rejection was prepared as a three-dimensional matrix for each participant and each condition (named *data* in Matlab workspace). The first dimension is the data sampling points containing points ranging from −200 to 2000 ms. The data was down-sampled to 200 Hz to increase computational speed. The second dimension is electrodes and the third is single trials. The script for parameter setup and calling RIDE toolbox for the two different schemes (S + C, S + R) are as follows:

Separating the speech ERP data into S + C:

```
cfg = [];
cfg.samp_interval = 5;
cfg.epoch_twd = [-200,2000];
cfg.comp.name = {'s','c'};
cfg.comp.twd = {[0,800],[300,1500]};
cfg.comp.latency = {0,'unknown'};
cfg = RIDE_cfg(cfg);

results = RIDE_call(data,cfg);
```

Separating the speech ERP data into S + R:

```
cfg = [];
cfg.samp_interval = 5;
cfg.epoch_twd = [-200,2000];
cfg.comp.name = {'s','r'};
cfg.comp.twd = {[0,800],[-500,800]};
cfg.comp.latency = {0,rt};
cfg = RIDE_cfg(cfg);
results = RIDE_call(data,cfg);
```

The time windows for each component were chosen to safely cover each component by visual inspection of the original ERP waveform. The time window is always relative to stimulus onset but for the R component (when the component name was named 'r' in the program), the time window is relative to response time (here the voice onset). That is why there is a negative value for the R time window [−500, 800 ms]. The variable 'rt' is the one dimension vector for response times. More details can be found in the manual from http://cns.hkbu.edu.hk/RIDE.htm.

Matlab Script for Experiment 2

For Experiment 2, the data was also down-sampled to 200 Hz to increase computational speed. The script for parameter setup and calling RIDE toolbox is as follows:

```
cfg = [];
cfg.samp_interval = 5;
cfg.epoch_twd = [-200,1200];
cfg.comp.name = {'s','rt'};
cfg.comp.twd = {[0,800], [-500,800]};
cfg.comp.latency = {0,rt};
cfg = RIDE_cfg(cfg);

results = RIDE_call(data,cfg);
```

Appendix II

See Table 2.

Table 2 Target pictures presented in Experiment 2

Sitzen (seating furniture)	Couch (couch)
	Hocker (stool)
	Ohrensessel (wing chair)
	Eckbank (corner seat)
	Bürostuhl (office chair)
Liegemöbel (lounge furniture)	Bett (bed)
	Futon (futon)
	Liege (divan bed)
	Hängematte (hammock)
	Schlafsofa (sofabed)
Aufbewahrung (storage)	Regal (shelf)
	Kleiderschrank (wardrobe)
	Vitrine (cabinet)
	Truhe (footlocker)
	Sideboard (sideboard)
Sanitär (sanitation)	Badewanne (bathtub)
	Pissoir (urinal)
	Waschbecken (basin)
	Dusche (shower)
	Bidet (bidet)
Textil (textile)	Perserteppich (Persian carpet)
	Vorhang (curtain)
	Rollo (roller blind)
	Badvorleger (bath mat)
	Tischdecke (tablecloth)
Vögel (birds)	Adler (eagle)
	Kolibri (hummingbird)
	Papagei (parrot)
	Geier (vulture)
	Eule (owl)
Fische (fishes)	Hai (shark)
	Aal (eel)
	Forelle (trout)
	Rochen (ray)
	Lachs (salmon)
Insekten (insects)	Fliege (fly)
	Biene (bee)
	Schmetterling (butterfly)
	Hirschkäfer (stag beetle)
	Ameise (ant)
Huftiere (hoofed animals)	Kamel (camel)
	Reh (deer)
	Pferd (horse)
	Esel (donkey)
	Schaf (sheep)
Affen (apes)	Schimpanse (chimpanzee)
	Pavian (baboon)
	Gorilla (gorilla)
	Orang-Utan (orang-utan)
	Mandrill (mandrill)

Kopfbedeckung (headdress)	Turban (turban)
	Hut (hat)
	Mütze (cap)
	Cappy (baseball cap)
	Zylinder (top hat)
Mäntel (coats)	Mantel (coat)
	Jacke (jacket)
	Poncho (poncho)
	Anorak (parka)
	Sakko (sportcoat)
Schmuck (jewelry)	Armreif (bracelet)
	Kette (necklace)
	Ohrring (earring)
	Brosche (brooch)
	Diadem (tiara)
Schuhe (footwear)	Stiefel (boots)
	Pumps (pumps)
	Turnschuh (sneaker)
	Clogs (clog)
	Mokkasin (moccasin)
Unterwäsche (underwear)	BH (bra)
	Tanga (thong)
	Socke (sock)
	Korsett (corset)
	Unterhemd (sleeveless shirt)
Obst (fruit)	Apfel (apple)
	Birne (pear)
	Kirsche (cherry)
	Trauben (grapes)
	Mandarine (mandarin)
Getränke (beverages)	Tee (tea)
	Milch (milk)
	Bier (beer)
	Wein (wine)
	Cocktail (cocktail)
Pilze (funguses)	Pfifferling (chanterelle)
	Steinpilz (porcini)
	Morchel (morel)
	Champignon (champignon)
	Bovist (puffball)
Kräuter (herbs)	Basilikum (basil)
	Petersilie (parsley)
	Dill (dill)
	Schnittlauch (chives)
	Rosmarin (rosemary)
Süßigkeiten (confectionery)	Kuchen (cake)
	Eis (icecream)
	Praline (praline)
	Bonbon (candy)
	Lakritze (licorice)
Küche (kitchen)	Schneebesen (egg whip)
	Kochlöffel (wooden spoon)
	Nudelholz (rollingpin)
	Messer (knife)
	Dosenöffner (canopener)
Bauernhof (farm)	Pflug (plow)
	Rechen (rake)
	Heugabel (hayfork)
	Sense (scythe)
	Axt (ax)
Friseur (coiffeur)	Kamm (comb)
	Lockenstab (curling iron)
	Schere (scissors)
	Bürste (brush)
	Haarnadel (hairpin)
Arzt (doctor)	Reflexhammer (reflexhammer)
	Spritze (syringe)
	Pinzette (tweezers)
	Akupunkturnadel (acupunture needle)
	Thermometer (thermometer)
Büro (office)	Edding (permanent marker)
	Tacker (stapler)
	Bleistift (pencil)
	Lineal (ruler)
	Klammer (paperclip)

References

Abdel Rahman R, Aristei S (2010) Now you see it… and now again: semantic interference reflects lexical competition in speech production with and without articulation. Psych Bull Rev 17(5):657–661. doi:10.3758/pbr.17.5.657

Abdel Rahman R, van Turennout M, Levelt WJM (2003) Phonological encoding is not contingent on semantic feature retrieval: an electrophysiological study on object naming. J Exp Psychol-Learn Mem Cogn 29(5):850–860. doi:10.1037/0278-7393.29.5.850

Aristei S, Melinger A, Abdel Rahman R (2011) Electrophysiological chronometry of semantic context effects in language production. J Cogn Neurosci 23(7):1567–1586. doi:10.1162/jocn.2010.21474

Brooker BH, Donald MW (1980) Contribution of the speech musculature to apparent human EEG asymmetries prior to vocalization. Brain Lang 9(2):226–245. doi:10.1016/0093-934X(80)90143-1

Costa A, Strijkers K, Martin C, Thierry G (2009) The time course of word retrieval revealed by event-related brain potentials during overt speech. Proc Natl Acad Sci USA 106(50):21442–21446. doi:10.1073/pnas.0908921106

De Vos M, Ries S, Vanderperren K, Vanrumste B, Alario FX, Huffel VS, Burle B (2010) Removal of muscle artifacts from EEG

recordings of spoken language production. Neuroinformatics 8(2):135–150. doi:10.1007/s12021-010-9071-0

Dell'Acqua R, Sessa P, Peressotti F, Mulatti C, Navarrete E, Grainger J (2010) ERP evidence for ultra-fast semantic processing in the picture-word interference paradigm. Front Psychol. doi:10.3389/fpsyg.2010.00177

Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 134(1):9–21. doi:10.1016/j.jneumeth.2003.10.009

Dimigen O, Sommer W, Hohlfeld A, Jacobs AM, Kliegl R (2011) Co-Registration of eye movements and EEG in natural reading: analyses and review. J Exp Psychol Gen 140(4):552–572. doi:10.1037/a0023885

Eulitz C, Hauk O, Cohen R (2000) Electroencephalographic activity over temporal brain areas during phonological encoding in picture naming. Clin Neurophysiol 111(11):2088–2097. doi:10.1016/S1388-2457(00)00441-7

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

Ganushchak LY, Schiller NO (2008) Brain error-monitoring activity is affected by semantic relatedness: an event-related brain potentials study. J Cogn Neurosci 20(5):927–940. doi:10.1162/jocn.2008.20514

Ganushchak L, Christoffels I, Schiller N (2011) The use of Electroencephalography in language production research: a review. Front Psychol. doi:10.3389/fpsyg.2011.00208

Gratton G, Coles MGH, Donchin E (1983) A new method for off-line removal of ocular artifact. Electroencephalogr Clin Neurophysiol 55(4):468–484. doi:10.1016/0013-4694(83)90135-9

Grözinger B, Kornhuber HH, Kriebel J (1975) Methodological problems in the investigation of cerebral potentials preceding speech: determining the onset and suppressing artefacts caused by speech. Neuropsychologia 13:263–270. doi:10.1016/0028-3932(75)90002-0

Hirschfeld G, Jansma B, Bolte J, Zwitserlood P (2008) Interference and facilitation in overt speech production investigated with event-related potentials. NeuroReport 19(12):1227–1230

Hutson J, Damian MF, Spalek K (2013) Distractor frequency effects in picture-word interference tasks with vocal and manual responses. Lang Cogn Process 28(5):615–632. doi:10.1080/01690965.2011.605599

Indefrey P (2011) The spatial and temporal signatures of word production components: a critical update. Front Psychol. doi:10.3389/fpsyg.2011.00255

Janssen N, Hernández-Cabrera JA, van der Meij M, Barber HA (2014) Tracking the time course of competition during word production: evidence for a post-retrieval mechanism of conflict resolution. Cerebral Cortex 25(9):2960–2969. doi:10.1093/cercor/bhu092

Jescheniak JD, Hahne A, Schriefers H (2003) Information flow in the mental lexicon during speech planning: evidence from event-related brain potentials. Cogn Brain Res 15(3):261–276. doi:10.1016/S0926-6410(02)00198-2

Kessler B, Treiman R, Mullennix J (2002) Phonetic biases in voice key response time measurements. J Memory Lang 47(1):145–171

Kohler KJ (1999) "German", Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet. Cambridge University Press, Cambridge, pp. 86–89. doi:10.1017/S0025100300004874

Lins OG, Picton TW, Berg P, Scherg M (1993) Ocular artifacts in EEG and event-related potentials I: scalp topography. Brain Topogr 6(1):51–63. doi:10.1007/BF01234127

Maess B, Koelsch S, Gunter TC, Friederici AD (2001) Musical syntax is processed in Broca's area: an MEG study. Nat Neurosci 4(5):540–545

Murray MM, Brunet D, Michel CM (2008) Topographic ERP analyses: a step-by-step tutorial review. Brain Topogr 20(4):249–264. doi:10.1007/s10548-008-0054-5

Muthukumaraswamy SD (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

Nunez PL, Srinivasan R, Westdorp AF, Wijesinghe RS, Tucker DM, Silberstein RB, Cadusch PJ (1997) EEG coherency. I: statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales. Electroencephalogr Clin Neurophysiol 103(5):499–515. doi:10.1016/S0013-4694(97)00066-7

O'Donnell RD, Berkhout J, Adey WR (1974) Contamination of scalp EEG spectrum during contraction of craniofacial muscles. Electroencephalogr Clin Neurophysiol 37(2):145–151. doi:10.1016/0013-4694(74)90005-4

Ouyang G, Herzmann G, Zhou C, Sommer W (2011) Residue iteration decomposition (RIDE): a new method to separate ERP components on the basis of latency variability in single trials. Psychophysiology 48:1631–1647. doi:10.1111/j.1469-8986.2011.01269.x

Ouyang G, Schacht A, Zhou C, Sommer W (2013) Overcoming limitations of the ERP method with residue iteration decomposition (RIDE): a demonstration in go/no-go experiments. Psychophysiology 50(3):253–265. doi:10.1111/psyp.12004

Ouyang G, Sommer W, Zhou C (2015a) A toolbox for residue iteration decomposition (RIDE)—a method for the decomposition, reconstruction, and single trial analysis of event related potentials. J Neurosci Methods 250:7–21. doi:10.1016/j.jneumeth.2014.10.009

Ouyang G, Sommer W, Zhou C (2015b) Updating and validating a new framework for restoring and analyzing latency-variable ERP components from single trials with residue iteration decomposition (RIDE). Psychophysiology 52(6):839–856. doi:10.1111/psyp.12411

Piai V, Roelofs A, van der Meij R (2012) Event-related potentials and oscillatory brain responses associated with semantic and Stroop-like interference effects in overt naming. Brain Res 1450:87–101. doi:10.1016/j.brainres.2012.02.050

Piai V, Roelofs A, Jensen O, Schoffelen JM, Bonnefond M (2014) Distinct patterns of brain activity characterise lexical activation and competition in spoken word production. PLoS One. doi:10.1371/journal.pone.0088674

Porcaro C, Medaglia MT, Krott A (2015) Removing speech artifacts from electroencephalographic recordings during overt picture naming. Neuroimage 105:171–180. doi:10.1016/j.neuroimage.2014.10.049

Protopapas A (2007) CheckVocal: a program to facilitate checking the accuracy and response time of vocal responses from DMDX. Behav Res Methods 39(4):859–862. doi:10.3758/Bf03192979

Schmitt BM, Schiltz K, Zaake W, Kutas M, Munte TF (2001) An electrophysiological analysis of the time course of conceptual and syntactic encoding during tacit picture naming. J Cogn Neurosci 13(4):510–522. doi:10.1162/08989290152001925

Sommer W, Leuthold H, Ulrich R (1994) The lateralized readiness potential preceding brief isometric force pulses of different peak force and rate of force production. Psychophysiology 31:503–512. doi:10.1111/j.1469-8986.1994.tb01054.x

Strijkers K, Costa A, Thierry G (2010) Tracking lexical access in speech production: electrophysiological correlates of word frequency and cognate effects. Cereb Cortex 20(4):912–928. doi:10.1093/cercor/bhp153

Stürmer B, Ouyang G, Zhou C, Boldt A, Sommer W (2013) Separating stimulus-driven and response-related LRP components with residue iteration decomposition (RIDE). Psychophysiology 50(1):70–73. doi:10.1111/j.1469-8986.2012.01479.x

van Turennout M, Hagoort P, Brown CM (1998) Brain activity during speaking: from syntax to phonology in 40 ms. Science 280:572–574

Vanhatalo S, Voipio J, Dewaraja A, Holmes MD, Miller JW (2003) Topography and elimination of slow EEG responses related to tongue movements. Neuroimage 20(2):1419–1423. doi:10.1016/S1053-8119(03)00392-6

Verleger R, Metzner MF, Ouyang G, Smigasiewicz K, Zhou CS (2014) Testing the stimulus-to-response bridging function of the oddball-P3 by delayed response signals and residue iteration decomposition (RIDE). Neuroimage 100:271–289. doi:10.1016/j.neuroimage.2014.06.036

Wang F, Ouyang G, Zhou CS, Wang SP (2015) Re-examination of Chinese semantic processing and syntactic processing: evidence from conventional ERPS and reconstructed ERPS by residue iteration decomposition (RIDE). PLoS One. doi:10.1371/journal.pone.0117324

Wirth M, Abdel Rahman R, Kuenecke J, Koenig T, Horn H, Sommer W, Dierks T (2011) Effects of transcranial direct current stimulation (tDCS) on behaviour and electrophysiology of language production. Neuropsychologia 49(14):3989–3998. doi:10.1016/j.neuropsychologia.2011.10.015

Wohlert AB (1993) Event-related brain potentials preceding speech and nonspeech oral movements of varying complexity. J Speech Lang Hear Res 36(5):897–905