Internship Report

Neural dynamics of speech production through the lens of independent component analysis

Student:

Mohammad keshtkar

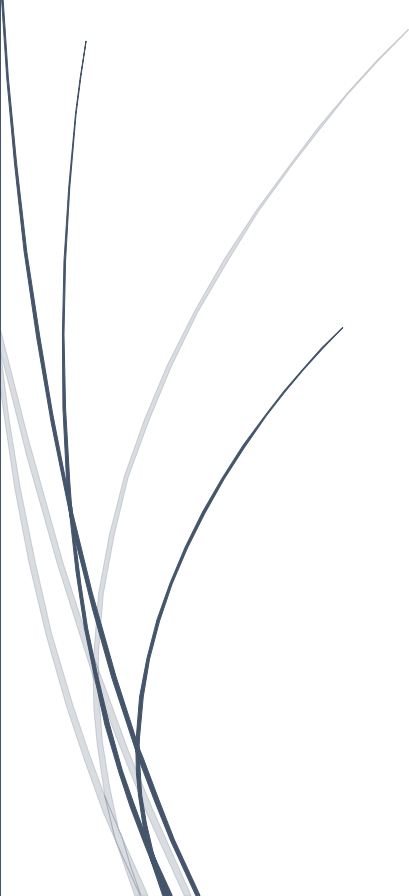M2 modeling of neuro and cognitive system

Supervisor:

Raphaël Fargier

Laboratory:

BCL, Université Côte d'Azur, France

Duration:

Sep 2023 – Feb 2024

First and foremost, I want to express my deepest gratitude to my supervisor, Raphaël. His unwavering support played a crucial role in my journey throughout this internship. He was always available to answer my questions, and he dedicated his time generously to ensure my growth and development.

Thank you for being such an inspiring mentor and for all the opportunities you have afforded me during this time.

Table of Contents

1   Introduction

Speaking represents the most intricate cognitive-motor skill of the human brain (Indefrey & Levelt, 2000). It involves a multifaceted and intricate process wherein different linguistic representations, such as semantic, lexical, phonological, and phonetic knowledge, must be retrieved and integrated (Fairs et al., 2021). The semantic and lexical stages are concerned with the meaning and content of the word, whereas phonological and phonetic processing deal with the form (sound) of the word. Take the word 'cow' as an example; the knowledge that it represents an animal and is a concrete (as opposed to abstract) term pertains to the content of the word, while its initial phoneme /k/ and how it is pronounced relate to the form of the word. Speech production requires the activation of a broad range of brain regions. It is believed that specific brain regions are more closely related to distinct aspects of word processing; for instance, the Wernicke's area, is heavily implicated in semantic processing, while left supramarginal gyrus is more associated with phonological processing (see (Price, 2012) for a comprehensive review).

1.1   Serial versus parallel processing

Regarding the coordination of these regions over time, Speech production models can be broadly categorized into two main models. In sequential models (Indefrey & Levelt, 2004), linguistic knowledge appears to be temporally segregated: initially, the semantic and lexical information (content) of the intended words is retrieved, followed by the activation of the phonological and phonetic code (form), which is subsequently translated into a motor program,

initiating the vocal folds for articulation. While sequential models vary in the number of processing stages required to transition from concept to speech and may allow for cascading and interaction between processing layers instead of strictly serial progression, they consistently maintain that upstream processes (content) commence before downstream processes (form) within the hierarchical architecture. Conversely, parallel models (Strijkers & Costa, 2016) claim that words are represented as integrated cell assemblies, wherein content (semantic-lexical information) and form (phonological-phonetic knowledge) rapidly ignite in parallel. While the temporal dynamics of language processing have been a longstanding question in the field, an ongoing debate persists between proponents of a more sequential framework and those advocating for parallel activation (Fairs et al., 2021).

It has been a longstanding debate as there are empirical evidence supporting both serial (Hulten et al., 2009; Liljeström et al., 2009; Maess et al., 2002; Salmelin et al., 1994; Sörös et al., 2003; Vihla et al., 2006) and parallel (Fairs et al., 2021; Feng et al., 2021; Janssen et al., 2020; Miozzo et al., 2015; Riès et al., 2017; Strijkers et al., 2017; Strijkers et al., 2010) processing. Discrepancies in the methods of assessing the activity of a specific brain source, particularly determining 'when' the source is active, can be the main reason behind these variations in results among previous studies.

1.2 Previous studies and their methodology

1.2.1 Six MEG studies supporting serial processing

A similar analytical framework has been employed in previous investigations using MEG to explore the dynamics of speech production during a picture naming task (Hulten et al., 2009; Liljeström et al., 2009; Maess et al., 2002; Salmelin et al., 1994; Sörös et al., 2003; Vihla et al., 2006). Three main steps were involved in their analysis. As a first step, they averaged the signal across trials and clusters of electrodes. Next, for each participant, the time-varying topographic maps corresponding to the averaged epochs, locked to the stimulus (picture), were scanned for dipolar map profiles. The location of the dipole corresponding to these topographic maps in the brain was then estimated through Equivalent Current Dipole (ECD) fitting procedures. The results identified several neural sources spanning wide brain areas (e.g., around 10 neural sources in (Salmelin et al., 1994)). Finally, the activation time of a specific neural source was determined based on the peak latency. This involves dividing the whole time period into arbitrary time bins

(e.g., three bins: 0-200 ms, 200-400 ms, and 400-800 ms after the stimulus in (Salmelin et al., 1994)), and considering the time bin with the highest peak as the period during which the neural source is active. (Salmelin et al., 1994) observed the presence of sources in occipital areas during a time bin spanning from 0 to 200 ms, in temporo-parietal areas from 200 to 400 ms, and in frontal areas from 400 to 800 ms. Consistent findings have been reported in other MEG studies employing a similar procedure. These outcomes suggest a sequential progression of activation during picture naming, moving from occipital input areas to higher-level temporo-parietal areas, and ultimately to frontal motor output areas (Janssen et al., 2020). These results are in line with the serial models of speech production, stating flow of information from lexical-semantic to phonological-phonetic processing.

Nevertheless, there are at least two noteworthy limitations in the methodology of previous MEG studies. Firstly, there is a limitation in source localization precision (the issue of "where"). A key component of the analysis involved identifying neural sources through the detection of dipolar maps in the epoched MEG data. However, topographic maps derived from the original epoched data often contain a mixture of dipoles, introducing uncertainties in the source localization procedure. Secondly, there is a limitation in the procedure to determine the time bin in which the source is active (the issue of "when"). Their method relied on peak latency as the primary indicator for defining the source activation. This approach may introduce artificial seriality due to the division of epochs into arbitrary time bins, meaning these bins are not determined by any biological factor. Furthermore, evoked potentials often feature multiple significant peaks. Relying on just one of these as the marker of activity may not accurately reflect brain activation (Janssen et al., 2020).

1.2.2 Alternative method of analysis

Considering these challenges, (Janssen et al., 2020) used a new analysis procedure to examine the temporal dynamics of brain regions involved in the picture naming task. To address the challenge of weak source localization, they identified dipolar maps through the application of Independent Component Analysis (ICA), contrary to the conventional approach of finding dipolar maps in the original observed data. Topographic component maps resulting from ICA frequently contain single dipoles which leads to more accurate estimation of ECD fitting procedures. But what is ICA?

ICA or Independent Component Analysis is a blind-source separation technique that separates mixed signals (observed data) into statistically independent components (Comon, 1994). Since the first application of ICA on EEG data (Makeig et al., 1995), it has become a common popular method in neural signal processing. However, it is frequently employed for the detection and removal of stereotyped artifacts such as those arising from eye movements, muscle activity, and line noise (Delorme & Makeig, 2004). The novelty of (Janssen et al., 2020) study lies in the application of ICA for the multi-subject analysis of event-related EEG data, specifically in the context of speech production. This methodology, referred as group ICA (gICA), involves temporally concatenating the data from all trials of multiple subjects and subsequently applying ICA as if it were a single subject with all the trials. The concept of gICA was published in a formal framework initially by (Cong et al., 2013). (Huster et al., 2015) further validated the usefulness of group-level ICA procedures for electroencephalographic data, establishing them as promising approaches for analyzing structure within multi-subject event related datasets.

In order to overcome the second challenge, (Janssen et al., 2020) used another approach to determine "when" the source is active. They determined the onset and offset times of neural activity for each source by calculating the time at full-width half maximum (FWHM) values for each component rather than depending on peak latency. Subsequently, they quantified the percentage of temporal overlap between each component (as we will apply a similar approach, we will explain gICA, FWHM, and overlap computation with more detail in the method section). Using the above methodology, they found fifteen neural sources with around 80 percent temporal overlap, in average, during picture naming (Janssen et al., 2020).

1.3   Referential versus Inferential

As we said, speech production involves the retrieval of a number of different types of knowledge, including semantic, lexical, phonological, and phonetic processing. According to (Marconi et al., 2013), there are two distinct types of semantic processing: referential and inferential. In the referential type, the words are defined through associating them to the external world by visual perception. For instance, consider the words "nose" and "ears". In the case of referential naming, they can be defined and identified by referring to their pictures.. However, in the inferential type, words are identified with words. For example, the words above could be

elicited by their descriptive definition such as "the facial organ responsible for the sense of smell and breathing" and "the sensory organs on the sides of the head that enable the perception of sound".

Hence, word production can be elicited by two distinct tasks, which, from a psycholinguistic theoretical standpoint, likely require different lexical-semantic processing. Referential tasks, such as picture naming, require subjects to produce the word corresponding to a displayed image. On the other hand, inferential tasks, like naming from definition, demand subjects to produce the word based on provided written or oral descriptions. Neuroimaging results (Fargier & Laganaro, 2017) confirm that these tasks elicit different brain activation patterns as well. Nevertheless, our understanding of language production relies on studies that mostly only used picture naming task. In particular, concerning the temporal dynamics of speech production and the debate between serial and parallel processing, there is a lack of studies utilizing inferential naming tasks to date, at least to the best of my knowledge. Therefore, we found it quite important to investigate the temporal dynamics of speech production using the new methodologies, same as (Janssen et al., 2020), but during an inferential naming task.

1.4   The current project

In the current project I studied the degree of parallel processing during auditory naming from definition. In the task, auditory definitions of 108 words were presented to participants, and they overtly produced the target words while their brain activation was recorded with EEG. This phase of the study (data acquisition) was done previously. In my internship, I was responsible for preprocessing and analyzing a proportion of the total recorded EEG data. I used methodology similar to (Janssen et al., 2020), meaning applying gICA and FWHM. The primary goal was to determine if there are similar levels of parallel processing during naming from definition.

2   Methods

2.1   Data acquisition

Participants: The data analyzed in this project pertained to 17 neurotypical senior adults (aged over 70). **Task**: The task involved auditory naming from definition of 108 French words. Each trial commenced with a blank screen presented for 250 ms, followed by an auditory definition.

These definitions consisted of simple sentences, with only the last word disambiguating the defined concept. For instance, for the word "angel," the definition was "creature avec des ailes qui vit au *paradis*" – a creature with wings that lives in *heaven*. The stimulus onset was determined as the moment when the last word of the definition was given. Participants were instructed to overtly produce the word corresponding to the definition as quickly and accurately as possible within a time limit of 3000 ms. There was an inter-trial interval of 1500 ms as well as a self-managed break in the middle of the task (after 54 stimuli). A control order was implemented to prevent consecutive stimuli from same semantic category or having high phonological overlap. Overt word productions were recorded using a dynamic microphone, digitally amplified, and the signal was directed to a computer. Response times, measured in milliseconds (i.e., the time difference between the stimulus onset and the onset of the speech wave), were systematically verified using speech analysis software (Check-Vocal 2.2.6). **EEG recording**: EEG data were acquired using the Active-Two Biosemi EEG system (Biosemi V.O.F. Amsterdam, Netherlands), utilizing a 128-electrode cap. The sampling frequency was configured at 512 Hz, with filters set to operate within the range of DC to 104 Hz, employing a 3 dB/octave slope. The custom online reference for the system comprised the common mode sense (CMS active electrode) and the driven right leg (DRL passive electrode), which worked to minimize the average potentials and maintain them as close as possible to the amplifier zero (Atanasova & Laganaro, 2022). It's important to clarify again that the data acquisition was completed before the start of my internship, and it was not part of my internship responsibilities.

2.2   EEG preprocessing

EEGLAB (Delorme & Makeig, 2004), an open source MATLAB toolbox, was employed for the preprocessing of EEG recordings. Initially, a bandpass filter within the frequency range of 1 Hz to 30 Hz was applied to eliminate low and high-frequency noise, a threshold chosen based on existing literature (Winkler et al., 2015) to optimize signal-to-noise ratio (SNR) and enhance the detectability of Independent Components (ICs) through ICA. Subsequently, I removed the bad channels (average 3.7 (2.9 %) per subject) and noisy segments of recording based on a manual inspection. The ICA was then conducted for each participant using the INFOMAX version of ICA (Makeig et al., 2004), in order to remove eye blinks and movement artifacts. I identified and removed eye-related ICs using the IClabel plugin (Pion-Tonachini et al., 2019) in EEGLAB.

Following this artifact removal, I interpolated the removed channels. After that, the data were epoched, spanning a duration of 200 ms before the stimulus onset and 600 ms after the stimulus onset. In our recording the stimulus event was moment just after the last word of the oral definition. Subsequently, I manually removed noisy trials (average 11.5 (10.6 %) per subject), trials with wrong answers (average 23.5 (21.7 %) per subject), as well as trials exhibiting naming latencies beyond the [400, 2200] ms range (average of 6.5 (6 %) per subject). Finally, I re-referenced the data using an average reference.

2.3 EEG analysis

2.3.1 ICA

ICA is a statistical method for transforming observed multidimensional data (X) with m rows (number of electrodes) and n columns (number of time points) into a source matrix (S) with the same dimensions where their rows (components) are statistically as independent from each other as possible. Mathematically, this transformation is represented as S = WX, where W denotes the coefficient or weight matrix ($m \times m$) that defines the components. We need to minimize mutual information between components in order to find statistically independent components. It means we are searching for a W that results in components with the minimum shared information. This can be achieved through iterative procedures that seek a matrix S and W wherein the rows of S (components) exhibit maximum non-Gaussianarity. Once W and S are determined, the equation can be reformulated as $X = AS$, where matrix A is the inverse of W (w^{-1}). Based on matrix A, we can define the topographic maps of each Independent Components, which from now on will be referred as ICs.

2.3.2 Group ICA (gICA)

Group ICA (gICA) follows a similar procedure, with the key distinction being that the matrix X includes EEG recordings from all participants. In this regard, the EEG datasets of all participants, encompassing their correct trials, were concatenated by columns. This resulted in a matrix of dimensions $m \times k$, where m represents the number of electrodes (128), and k corresponds to the total number of correct epochs across all participants (1129) multiplied by the

number of time points per epoch (409) = 461,761. I computed the ICA of this matrix using the INFOMAX version of ICA that is available in the windows version of EEGLAB.

2.3.3 Dipole fitting

Following the identification of 128 ICs with their respective topographical scalp maps, we proceeded to estimate their source locations through Equivalent Current Dipole (ECD) modeling. ECD modeling involves fitting a dipole model with parameters encompassing location, orientation, and amplitude to the data, employing a least squares algorithm (Delorme et al., 2012). This was made possible by the DIPFIT plug-in (V5.3) within EEGLAB. For forward modeling, we opted for the template boundary element model to represent electrical conductance across 3-D layers (skin, skull, cortex). Each dipole's location, orientation, and amplitude in the MNI space were obtained along with its residual variance (RV), which is a representation of estimation accuracy. Additionally, this version of DIPFIT provided anatomical inferences for each fitted dipole base on the Desikan-Killiany Atlas (Desikan et al., 2006).

2.3.4 ICs classification

Subsequently, we eliminated non-brain-related ICs through a dual approach involving manual inspection—taking into account their topographic map, spectral map, and averaged signal—and an automatic IC labeling, using the IClabel plugin (Pion-Tonachini et al., 2019) within EEGLAB. Additionally, ICs with RV exceeding 12%, were excluded from further analysis.

2.3.5 ICs temporal overlap

The subsequent phase of analyses focused on determining when each IC was active as well as the percentage of temporal overlap between each IC. I used MATLAB scripts (Version: (R2023a)) to execute this last analysis step (the full script is provided in supplementary material A). Firstly, for each IC, the average signal was computed by averaging across all epochs from all participants. Then, the onset and offset times of each IC activation was defined by calculating the time at full-width half maximum (FWHM). This means the onset timepoint of an IC activity was determined as the moment in time (relative to stimulus onset) when it first reached 50% of its absolute maximum amplitude value. Similarly, the offset of IC activation was computed by working backward from the end of the epoch.

The percentage of activity overlap (AO) between one target IC_a relative to another IC_b was calculated using this formula:

$$AO = \frac{I(a,b)}{T_a}$$

Where $I(a,b)$ represents the intersection time between IC_a and IC_b, and T_a is the total number of timepoints (from onset to offset) for target IC_a. This formula computes the sequence of overlapping timepoints between one target IC_a and another IC_b relative to the total number of timepoints in the target IC_a. It is noteworthy that this approach yields asymmetrical overlaps, as the overlap of a given target a with respect to b may differ from the overlap of b with respect to a.

3 Results

3.1 Response time

The total number of trials ultimately included into the analysis amounted to 1129, representing 61.4% of the initial set. The average response time (RT) for these trials was 924.5 ms, with a standard deviation of 376.3 ms. Refer to Figure 1 for a graphical presentation of the RT distribution, presented as a histogram plot.
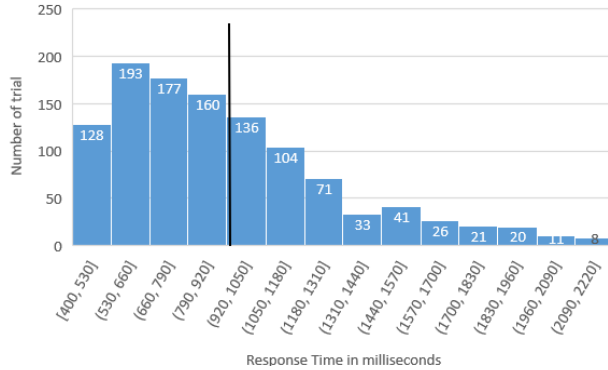
Figure 1. The RT distribution for the trials included in the analysis using histogram. RT time bins (in ms) are depicted on the x-axis, and the total number of trials with relative RT is represented on the y-axis. The black vertical line indicates the average RT (924.5 ms).

3.2 EEG results

Following the inspection of the 128 ICs derived from gICA, 9 ICs were identified as neural sources based on our previously mentioned procedure (see 2.3.4 ICs classification) and,

11

consequently, were included for further analysis. Figure 2 and Figure 3 exhibit the topographic maps and dipoles for each IC. In addition, these figures present the time course of each IC for every trial, ordered by RT, along with the averaged signal across all trials. These figures extracted using the ERP image function of EEGLAB (Delorme et al., 2015). It is important to take into account that the absolute amplitude and polarity of component activations lack inherent meaning, and these activations are not associated with any unit of measure, though they are proportional to microvolts (see (Delorme, 2020) for more detail).

Also note that in order to plot the RTs on these graphs, we normalized their values. This normalization process involved dividing the RTs value by the maximum RT (2190 ms) and then multiplying by the maximum time after the stimulus (600 ms).
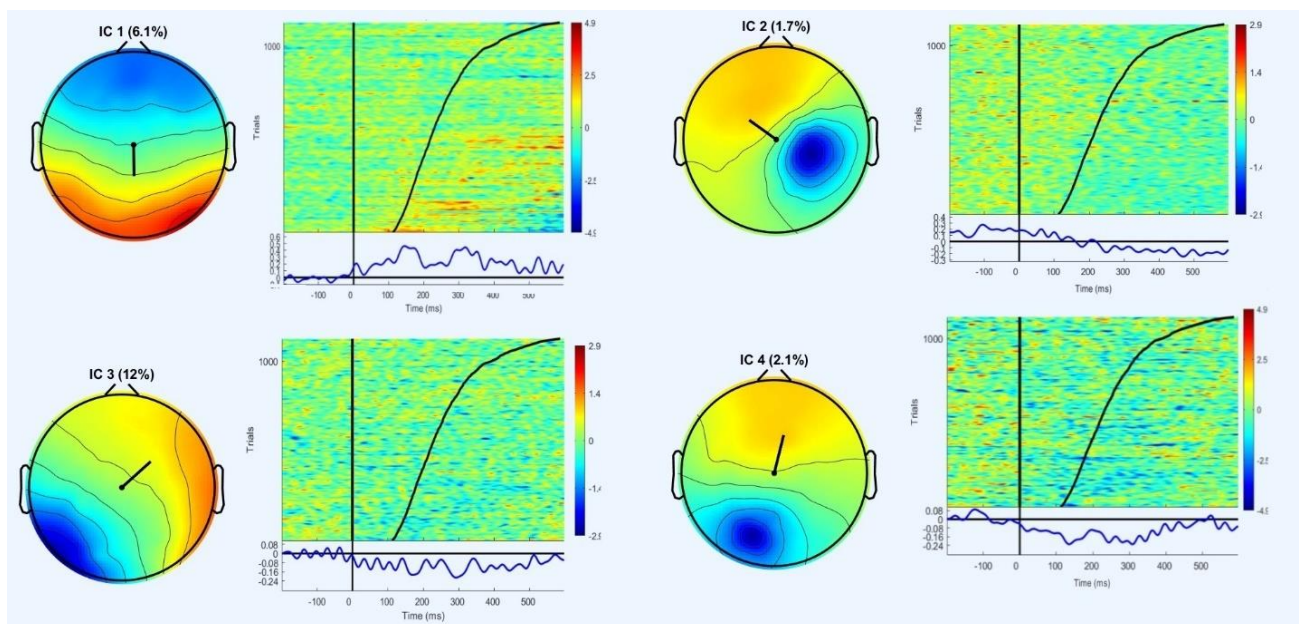

Figure 2. The topographic map and ERP image of ICs 1-4. The topographic maps include the location and moment of each dipole, with the Residual Variance specified in the map's title. The ERP image depicts component time courses for each trial, with high amplitudes represented by warm colors and low amplitudes by cold colors. The order of trials is determined by their normalized RT, illustrated by the curved black line, with shorter naming latencies positioned at the bottom of the graph. Additionally, the average component time course across all trials is displayed below it, indicated by a blue line.

As it can be seen in these graphs, the topographic maps of the ICs are clear (dipolar, not mixed), and they display distinct time course of activity and average signal. The sequence of the ICs corresponds to the extent they accounted for variance in the raw data, indicating that IC 1 explains a more portion of the signal variation compared to IC 9. Refer to Supplementary Material B for comparable figures depicting all ICs that were excluded. Figure 4 presents the result of ECD

fitting procedure along with the anatomical interpretation of ICs dipole based on Desikan-Killiany Atlas as well as their RV. The DIPFIT plugin was unable to assign a brain region for IC 1. After thorough analysis, I believe the thalamus is the nearest brain region to its dipole location. All ICs exhibit RVs below 12%, ranging from a minimum of 1.72% (IC 2) to a maximum of 11.61% (IC 3), with an average of 6.43%. Notably, ICs closer to the scalp (IC 2, 6, and 5) present much lower RVs, indicating more accurate estimations, which is rational as they are closer to EEG electrodes. The ICs are linked to various brain regions: occipital lobe (IC 3, 4, and 8), parietal lobe (IC 2, 5, and 9), temporal lobe (IC 7), limbic lobe (IC 6), and a subcortical area (IC 1). It's crucial to consider that these findings do not necessarily imply that these brain regions were more active, while others were less active or inactive (for such interpretations, other methods such as functional Magnetic Resonance Imaging (fMRI) are required). Instead, the results suggest that gICA was able to identify independent components associated with these brain areas, which exhibited clear, significant, and consistent activity across individuals during naming from definition.
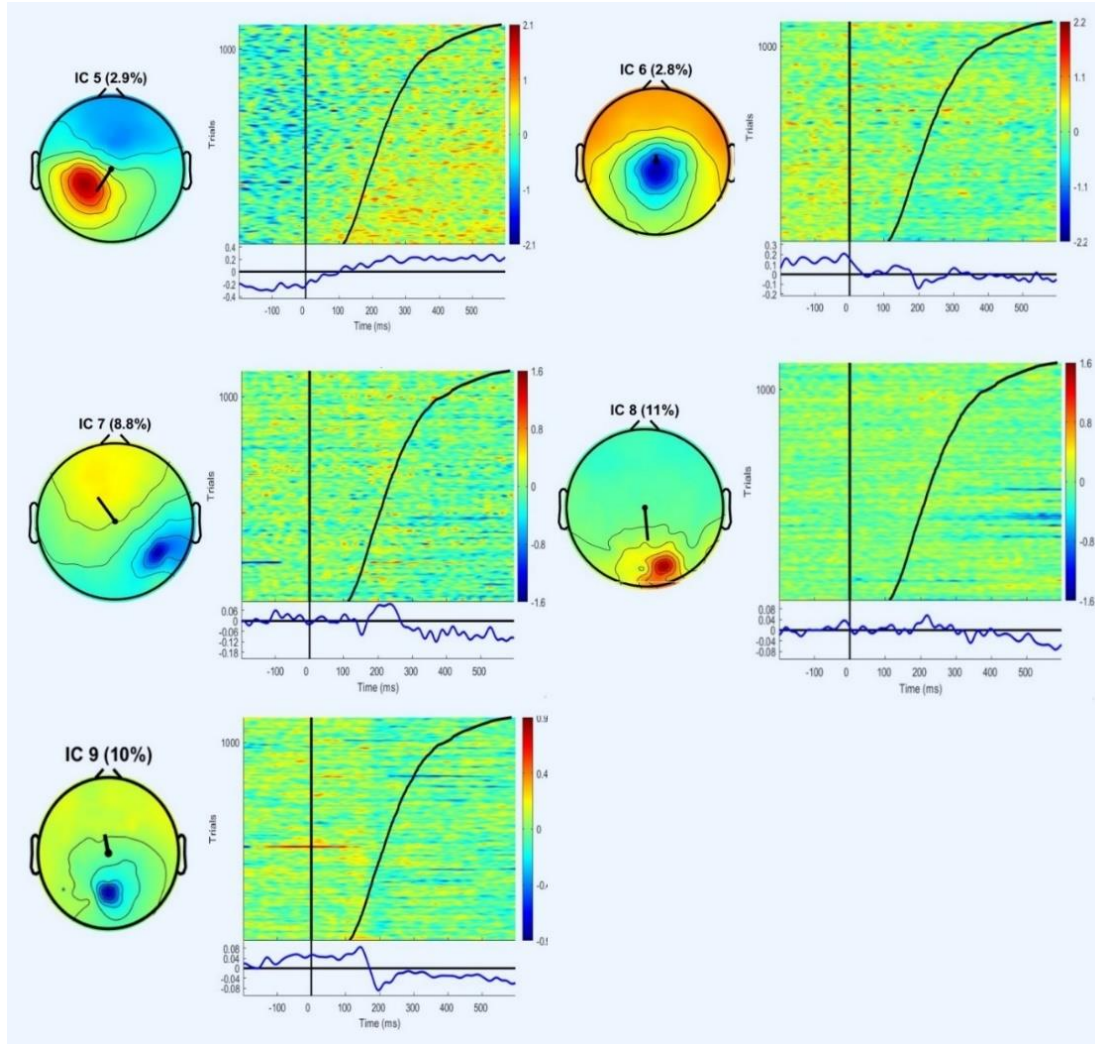
Figure 3. The topographic map and ERP image of ICs 5-9. Refer to the caption of Figure 2 for detail.

Figure 5-A exhibits the averaged signal of each IC along with their active segments determined by FWHM. A graphical representation of the activation of all ICs, ordered according to their peak latency, is presented in Figure 5-B. For each IC, the precise activation onset, offset, peak latency, and their total duration (i.e., the interval between offset and onset) are detailed in Table 1. As can be seen in these figures, a substantial number of ICs are active for a significant portion of the epoch, ranging from a minimum of 213 ms (IC 6) to a maximum of 600 ms (IC 2, 5, and 9), with an average activity span of 457 ms. As you can see, there is significant overlap in the activation of ICs; for instance, observe how nearly all the ICs are active simultaneously between 150 ms and 380 ms. Figure 5-C presents the precise activity overlap (AO) ratio across all ICs through a heatmap for better visualization. Note that I set the diagonal values to zero, since

they correspond to the AO ratio of each IC with itself. On average, ICs exhibited a temporal overlap of 76%, with a minimum overlap of 1% (between IC 6 and IC 8) and a maximum overlap of 100% (mostly related to IC 2, 5, and 9).
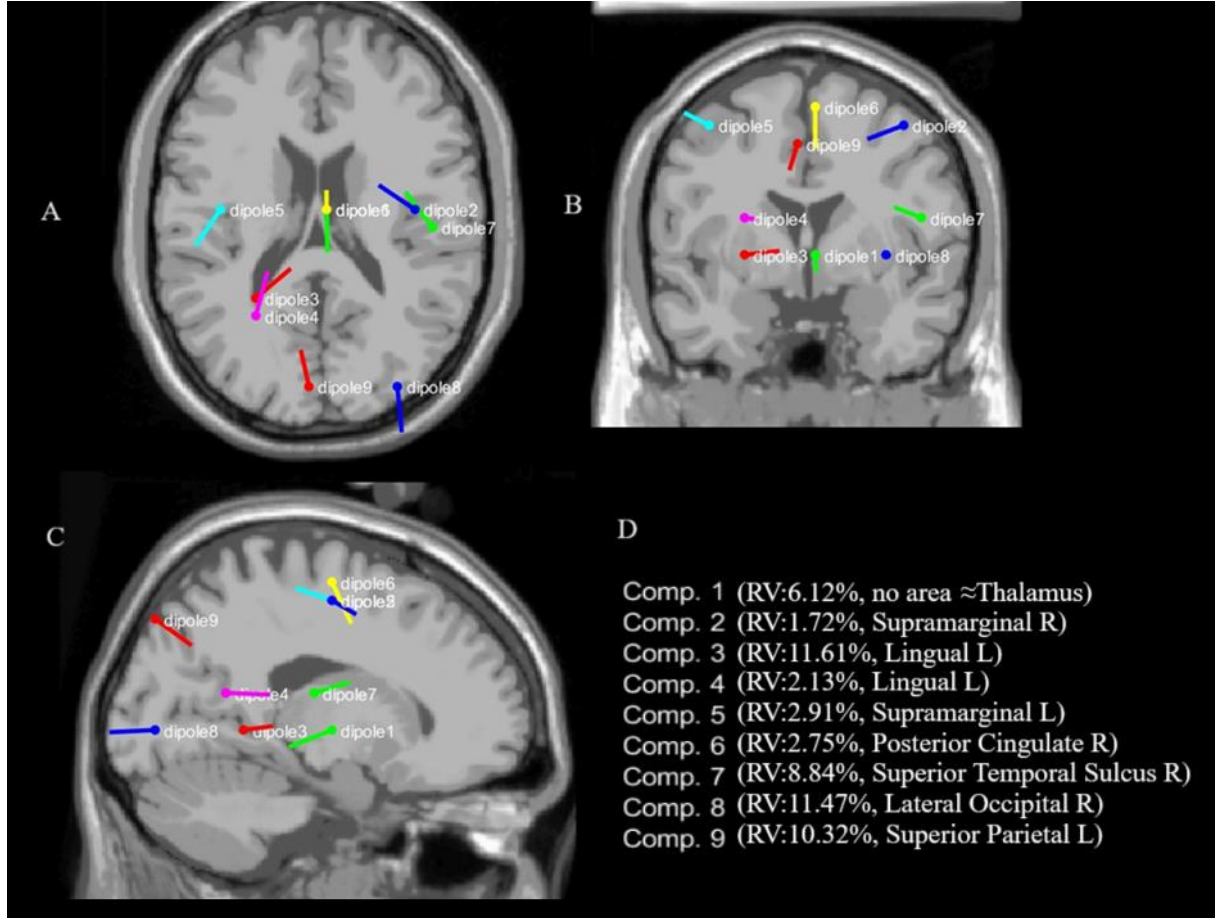

Figure 4. Estimated location of ICs using ECD fitting procedure. The graph in (A) shows top view, (B) coronal view, and (C) sagittal view of the brain. Note that the MRI image serves as the template image within the DIPFIT plug-in and is not derived from our study participants. (D) provides the residual variance (RV) and the anatomical interpretation of ICs dipole based on Desikan-Killiany Atlas.

3.3   Complementary analysis

3.3.1   Latency jitter

One potential concern arises from including trials with a wide range of reaction times (RT), from 400 ms to 2190 ms. This substantial latency jitter could attenuate the average signal, potentially leading to an overestimation of the active segment of the IC. To address this concern, I conducted the same analysis but limited it to a subset of data with similar RTs. This was achieved in various ways; for instance, one method involved dividing the trials into three subsets (early, median, late). Across all groups, the active segment and average activity overlap ratio of the ICs

15

were quite similar, revealing no significant differences. The results are presented in Supplementary Material C.

3.3.2 Longer epoch

The epoch length was set to 600 ms after stimulus, consistent with common methodology in EEG speech production studies, where epoch lengths are kept shorter than reaction times (RTs) to prevent the inclusion of articulation noises. However, this short epoch might be the reason for the lack of identified ICs in the frontal lobe. Regions in frontal lobe (e.g., Broca and motor cortex) are mainly responsible for planning and execution of motor movements for speech which clearly processed as the final step. To address this, I reanalyzed the data for 13 subjects, extending the epoch length to 800 ms after stimulus, aiming to explore potential findings of ICs in the frontal lobe and assess the activity overlap ratio. This procedure led to finding similar ICs (without any new IC in the frontal lobe) but with a less clear dipolar topographic map as the number of individuals was lower. The results can be found in the supplementary material D.

3.3.3 Individual ICs

There was a concern that averaging ICs' activation across all participants might yield different results compared to averaging the ICs' time courses within each subject. In this regard, after performing gICA, I back-reconstructed the ICs' activation for each individual using a MATLAB script (that is provided as Supplementary Material E). Subsequently, employing a similar methodology as before, I computed the active segment of each IC and, and their AO ratio for each individual. It is important to note that I still applied gICA, and I did not conduct individual ICA for each participant. The average AO ratio per individual closely aligned with what we had previously found for all participants combined (76%). The grand average AO ratio across all individuals amounted to 79%, with a minimum average AO ratio of 67% and a maximum of 91%. For detailed results, please refer to Supplementary Material F.

Figure 5. (A) The averaged signal of each IC across all trials with their active segments based on FWHM. The active segment is depicted in green, while the non-active segment is represented in blue. Note that here, only the period following the stimulus is presented. **(B)** Temporal activation of ICs. The x-axis represents time (in ms), and the y-axis lists all ICs ordered by their peak latency, with those exhibiting early peak positioned at the bottom. The red * is corresponding to ICs peak latency. **(C)** The heatmap of the activity overlap ratio between all ICs. The diagonal deliberately chosen to be zero as they show the AO ratio between ICs with themselves.

IC	onset	duration	offset	peak
1	94	475	568	146
2	2	600	602	471
3	18	482	500	295
4	88	299	387	311
5	2	600	602	557
6	2	213	215	0
7	150	451	602	373
8	211	391	602	584
9	2	600	602	199
Average	63	457	520	326

Table 1. The precise activation onset, offset, peak latency, and total activity duration (i.e., the interval between offset and onset) for each IC.

4   Discussion

4.1   Overview of results

In my internship, I delved into the temporal dynamics of speech production, focusing particularly on the extent of parallel processing in naming from definition. To achieve this, I engaged in preprocessing and analyzing EEG datasets that had been previously recorded. Following applying gICA, 9 ICs identified as brain sources through a dual approach involving both manual inspection and using an automatic IC labeling algorithm. Source localization of ICs, using ECD fitting procedure, led to estimated dipole locations in various regions of the brain. Based on FWHM criteria, the ICs active segment and subsequently their activity overlap (AO) ratio were computed. Our findings indicate a notable AO ratio among ICs with an overall average of 76%. Additionally, we obtained consistent results when accounting for latency jitter, employing longer epochs, and analyzing individual ICs. The observed level of parallel processing in naming from definition closely aligns with the findings of (Janssen et al., 2020) in picture naming with an average of 81% AO ratio. Hence, we found similar level of parallel processing in different stimuli while applying a similar procedure.

4.2   Results in detail

In this section, I will explore the potential roles of each component in speech production and analyze their activation timings as indicated by our findings. I will particularly address the possible cognitive stages linked to these brain sources, exploring if their activation timings correspond with their roles, and the possible interpretations that can be drawn. I believe this exploration can help us to examine our results and see if it is in line with previous literature. This discussion will be structured based on ICs peak latency, starting with ICs peaking at the onset of the epoch and progressing to those with later peaks.

Before proceeding this section, I found it quite important to review what we outlined in introduction about the speech production models and cognitive stages in this task. Speech production cognitive stages can be broadly categorized into two main categories. The first pertains to the processing of a word's content and meaning, such as semantic processing and lexical selection, while the second involves the processing of the word's form and sounds, like phonological processing. It is believed that our brain first processes the meaning of the word and then form of the word. Serial models suggest that these processes occur sequentially in the brain,

moving from content (around first 200-300 ms) to form (mostly after 300 ms)**,** without temporal overlap. In contrast, parallel models claim that the process of these stages has overlap with each other.

4.2.1   IC 6 - Right posterior cingulate cortex

As illustrated in Figure 5-B, IC 6, associated with the right posterior cingulate cortex (rPCC), emerges as the earliest active component. Although the rPCC is not predominantly involved in speech production, it is thought to play a role in memory (Lega et al., 2017) and semantic processing (Weber et al., 2016). Hence, we can assume that this component relates to processing the content of the word. IC 6 temporal activation which display an activation within the first 200 ms of the epoch with an early peak (time zero) aligns well with our expectations of its role. In addition, it has been suggested (Wilson et al., 2007) that the rPCC is important in speech comprehension, especially at higher levels where context understanding and inference are necessary, as is the case in our task. This aligns with the observation of a high amplitude signal for IC 6 preceding the stimulus, as depicted in Figure 3.

4.2.2   IC 1- Thalamus

The second component is IC 1, for which the DIPFIT plugin was unable to assign a corresponding brain region based on the Desikan-Killiany Atlas. However, after thorough analysis, I believe the thalamus is the nearest brain region to its dipole location. Although the thalamus and other subcortical structures are not often included in speech and language models, there is growing recognition of their involvement. For instance, a recent study by (Dengyu et al., 2022) provided direct evidence from intracranial thalamic recordings, demonstrating the region's sensitivity to the lexical aspects of word production. Furthermore, the symptoms of Thalamic Aphasia, such as anomia, further emphasize the thalamus link to lexical processing during speech production (Bulleid et al., 2018). In this context, associating IC 1 with lexical processing, our results indicate lexical processing peaking around 145 ms post-stimulus, which aligns closely with serial model assumptions. However, the extension of this activity until the later stages of speech production (with an offset at 568 ms) and consequently its high overlap with other ICs supports parallel models.

4.2.3   IC 9- Left superior parietal lob

IC 9, the subsequent component, is representing the activity of left superior parietal lob. This component exhibited activity throughout the entire epoch, peaking at around 200 ms. The involvement of the superior parietal lobule in working memory (Michael et al., 2009) and reasoning (Javaheripour et al., 2019) might explain its continuous activity observed in our results. Therefore, its high degree of overlap with other components might not, at least directly, indicate parallel processing in speech production.

4.2.4   IC 3,4 - Left lingual gyrus

The next components, IC3 and IC4, are linked to the left lingual gyrus (LLG). While the LLG is known for its involvement in visual processing (Bogousslavsky et al., 1987), it also plays a significant role in processing the semantic aspects of words (Kim et al., 2011)) and is identified as the origin of the recognition potential (RP) (Martin-Loeches et al., 2001). RP is an electrophysiological response of the brain that occurs during lexical processing. It manifests as a component of event-related potentials (ERPs) with a peak around 250 ms post-visual stimulus (like a written word), although this can vary with task difficulty (Martin-Loeches et al., 2001). This finding aligns well with the peak latency of IC3 and IC4, which is around 300 ms in our task, a slightly delayed response that could be attributed to the task's complexity. However, it's noteworthy that our stimuli are auditory rather than visual. Based on these observations, it can be inferred that IC3 and IC4 play a role in lexical selection (part of processing the content of the word), with a peak latency generally consistent with serial models, but with extended activation to both the start and end of the epoch, resulting in overlap with other components.

4.2.5   IC 7- Right superior temporal sulcus

The subsequent component, IC 7, is estimated to represent the right superior temporal sulcus (rSTS). Prior research has highlighted the significant activity of the rSTS in relation to the acoustic characteristics of words, both in speech comprehension (Kriegstein & Giraud, 2004) and speech production (Yamamoto et al., 2019). (Yamamoto et al., 2019) observed notable activity in the right posterior superior temporal sulcus (rpSTS) during speech production, especially with auditory stimuli, and suggested that the rpSTS plays a role in the short-term memory of speech sounds. In this regard, IC 7 is related to processing the form of the word. This is in line with our observation for IC 7 as it is active predominantly in the second half of the epoch with a peak at

373 ms. However, it is important to note that the rSTS is not recognized as the primary area responsible for phonological processing or for handling the acoustic characteristics of words during speech production, particularly when compared to dominant areas in the left hemisphere such as the Left Superior Temporal Gyrus(Peeva et al., 2010) and Left Inferior Frontal Gyrus. Nonetheless, it is indeed a need for further research into the role of the rSTS during speech production, especially in the context of auditory naming from definition.

4.2.6   IC 2,5 - Right and left supramarginal gyrus

The IC2 and IC5 are corresponded with the right and left supramarginal gyrus (rSMG and lSMG) respectively. The lSMG is known to be crucial for phonological processing (Oberhuber et al., 2016)) and plays a role in bridging phonetic and articulatory representations of words. The rSMG, on the other hand, is involved in adjusting speech motor plans in response to changes in somatosensory inputs (Li et al., 2019). Therefore, both ICs are related to processing the form of the word. Based on our finding, both regions display a late peak in activity (occurring at 470 ms for the rSMG and 576 ms for the lSMG) which is quite match with their assumed role in speech production. However, our results claim that these regions are concurrently active with other areas throughout the entire epoch. It's also noteworthy that, given the roles of the lSMG and rSMG, one would expect the first peak of activity to occur in the lSMG, followed by the rSMG. However, our results do not reflect this sequence.

4.2.7   IC 8 - Right lateral occipital

The final brain source, IC 8, is estimated to represent the activity of the right lateral occipital (rLO) region. This particular brain area is primarily associated with visual processing (Jonas & David, 2006) and is not conventionally linked to speech and language functions. However, considering the age of the participants (senior adults) in this analysis, the occipital activity might possibly indicate compensatory recruitment of novel brain regions due to aging. This phenomenon, for instance, has been observed in speech comprehension(Wingfield & Grossman, 2006). Therefore, conducting further analysis involving other age groups might leads to a better understanding.

4.2.8   Summary

Our discussion leads us to classify IC 6, 1, 3, and 4 under the content category, and IC 7, 2, and 5 under the form category. Our findings indicate that while the sequence of peak activation in

these ICs aligns with the serial models, there is notable concurrent activity among them. For instance, consider the activation of ICs 200 ms post-stimulus, where ICs from both categories are active concurrently. This interpretation is True even if we exclude IC 8 and 9 as they might not represents brain activity directly related to speech production.

4.3 Results compare to other studies

As previously mentioned, we observed a similar degree of overlap in naming from definition compared to (Janssen et al., 2020), where stimuli was picture naming. comparing two results, the ICs are mostly associated with inconsistent brain regions possibly due to difference in the task. For example, there were more ICs in the occipital lobe in their study, which aligns with the visual nature of picture naming. Nonetheless, the flow of activity is comparable: ICs related to the meaning of the word showed early peaks, while ICs associated with the form of the word exhibited later peaks, alongside a similar level of overlap.

Our findings present a contrast to the MEG studies discussed in the introduction, primarily due to differing approaches in determining when a brain source is active (peak latency versus FWHM). As previously argued, segmenting the epoch by peak latency introduces an artificial seriality and serves as a poor indicator of neural activity. For example, consider the averaged signal for IC 1, which exhibits two peaks: the first around 150 ms (power ≈ 0.46) and the second around 320 ms (power ≈ 0.43). Using peak latency, this component would be classified as active in the first 200 ms, yet it's evident there's another peak of similar power following 200 ms. Furthermore, our study demonstrates more accurate source localization, as we conducted ECD fitting procedures on data decomposed by ICA, rather than on raw epoch data.

4.4 Limitation and further studies

4.4.1 Regarding the result

The limitation, considering our result, is that we ended up with only 9 ICs representing a neural source with a clear dipole out of initial set of 128 ICs (as we had 128 channels). This outcome was unexpected, especially when considering that (Janssen et al., 2020) identified 15 ICs out of initial set of 32 ICs (as they had 32 channels). Several factors could account for this limitation. Firstly, the quality (signal-to-noise ratios) and quantity (length) of data inputted into the ICA algorithm can significantly influence its results (Delorme et al., 2012). The age of the participants and the complexity of the task could be other contributing factors. Naming from

definition, being a complex task, might lead to each individual employing different cognitive strategies to identify and articulate the word. This likelihood increases with participants aged over 70, given the compensatory approaches in old people during language processing (Wingfield & Grossman, 2006). Such inconsistency can influence the output of data decomposition by ICA. Consequently, further research incorporating more data, a higher signal-to-noise ratio, and participants from other age groups might facilitate the identification of clearer ICs.

4.4.2   Regarding the method

It is crucial to acknowledge that the application of gICA relies on certain assumptions that may not always hold true. When employing gICA, we assume that there exists the exact same number of independent sources with identical topographic maps across all individuals. However, this assumption may not be accurate, as participants may employ different cognitive strategies to achieve a common goal. Additionally, we assume that each electrode in each individual is positioned in precisely the same location, which is also not accurate due to natural variations in brain structures. Nonetheless, such challenges are inherent in group-level analyses.

Another important consideration relates to the spatial resolution of EEG. Although source localization on decomposed data offers improved accuracy compared to other analysis methods, such as fMRI, the spatial resolution of EEG remains relatively low. However, this resolution can be enhanced, for instance, by incorporating MRI images of all participants in conjunction with EEG recording. With the MRI images of each participant's brain, we can enhance our ECD modeling, particularly benefiting older participants, as aging is known to alter the anatomical structure of the brain.

4.5   Conclusion

Speech production is a complex cognitive-motor skill that requires the neural activation of a wide range of brain areas. In this project, we delved into the neural dynamics of brain activation over time during auditory naming from definition by analyzing previously recorded EEG datasets. Our findings indicate that while the sequence of peak activation in these ICs aligns with the serial models, there is notable concurrent activity among them, in line with previous study during picture naming (Janssen et al., 2020).

Our study was constrained by the identification of only nine ICs with clear dipoles and representing neural activity. Future research with larger data, higher signal-to-noise ratios, and

diverse age groups may lead to improved outcomes in data decomposition. Additionally, it is important to take into account the limitations of gICA and EEG spatial resolution. We believe that incorporating MRI images of participants in future studies could enhance source localization.

5 Supplementary material s

All the supplementary materials can be found in our GitHub repository using the following link:

https://github.com/MohammadMMK/neural-dynamics-of-speech-production-through-the-lens-of-ICA

6 References

Atanasova, T., & Laganaro, M. (2022). Word Production Changes through Adolescence: A Behavioral and ERP Investigation of Referential and Inferential Naming. *Developmental Neuropsychology*, *47*(6), 295-313.

Bogousslavsky, J., Miklossy, J., Deruaz, J. P., Assal, G., & Regli, F. (1987). Lingual and fusiform gyri in visual processing: a clinico-pathologic study of superior altitudinal hemianopia. *Journal of Neurology, Neurosurgery &amp; Psychiatry*, *50*(5), 607. https://doi.org/10.1136/jnnp.50.5.607

Bulleid, L., Hughes, T., & Leach, P. (2018). A case of transient thalamic dysphasia—considering the role of the thalamus in language. *Child's Nervous System*, *34*(12), 2345-2346. https://doi.org/10.1007/s00381-018-3967-7

Comon, P. (1994). Independent component analysis, a new concept? *Signal processing*, *36*(3), 287-314.

Cong, F., He, Z., Hämäläinen, J., Leppänen, P. H., Lyytinen, H., Cichocki, A., & Ristaniemi, T. (2013). Validating rationale of group-level component analysis based on estimating number of sources in EEG through model order selection. *Journal of neuroscience methods*, *212*(1), 165-172.

Delorme, A. (2020). *Plotting ICA components*. https://eeglab.org/tutorials/08_Plot_data/Plotting_ICA_components.html#component-erp-contributions

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

Delorme, A., Miyakoshi, M., Jung, T.-P., & Makeig, S. (2015). Grand average ERP-image plotting and statistics: A method for comparing variability in event-related single-trial EEG activities across subjects and conditions. *Journal of neuroscience methods*, *250*, 3-6.

Delorme, A., Palmer, J., Onton, J., Oostenveld, R., & Makeig, S. (2012). Independent EEG sources are dipolar. *PloS one*, *7*(2), e30135.

Dengyu, W., Witold, J. L., Alan, B., Anna, C., Christina, A. D.-H., Michael, D., Julie, A. F., & Richardson, R. M. (2022). Lateralized and Region-Specific Thalamic Processing of Lexical Status during Reading Aloud. *The Journal of Neuroscience*, *42*(15), 3228. https://doi.org/10.1523/JNEUROSCI.1332-21.2022

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., & Hyman, B. T. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*, *31*(3), 968-980.

Fairs, A., Michelas, A., Dufour, S., & Strijkers, K. (2021). The same ultra-rapid parallel brain dynamics underpin the production and perception of speech. *Cerebral Cortex Communications*, *2*(3), tgab040.

Fargier, R., & Laganaro, M. (2017). Spatio-temporal dynamics of referential and inferential naming: Different brain and cognitive operations to lexical selection. *Brain topography*, *30*, 182-197.

Feng, C., Damian, M. F., & Qu, Q. (2021). Parallel Processing of semantics and phonology in spoken production: evidence from blocked cyclic picture naming and EEG. *Journal of cognitive neuroscience*, *33*(4), 725-738.

Hulten, A., Vihla, M., Laine, M., & Salmelin, R. (2009). Accessing newly learned names and meanings in the native language. *Human brain mapping*, *30*(3), 976-989.

Huster, R., Plis, S., & Calhoun, V. (2015). Group-level component analyses of EEG: validation and evaluation [Original Research]. *Frontiers in Neuroscience*, *9*. https://doi.org/10.3389/fnins.2015.00254

Indefrey, P., & Levelt, W. J. (2000). The neural correlates of language production. *The new cognitive neurosciences; 2nd ed.*, 845-865.

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

Janssen, N., Meij, M. v. d., López-Pérez, P. J., & Barber, H. A. (2020). Exploring the temporal dynamics of speech production with EEG and group ICA. *Scientific reports*, *10*(1), 3667.

Javaheripour, N., Shahdipour, N., Noori, K., Zarei, M., Camilleri, J. A., Laird, A. R., Fox, P. T., Eickhoff, S. B., Eickhoff, C. R., Rosenzweig, I., Khazaie, H., & Tahmasian, M. (2019). Functional brain alterations in acute sleep deprivation: An activation likelihood estimation meta-analysis. *Sleep Medicine Reviews*, *46*, 64-73. https://doi.org/https://doi.org/10.1016/j.smrv.2019.03.008

Jonas, L., & David, J. H. (2006). Two Retinotopic Visual Areas in Human Lateral Occipital Cortex. *The Journal of Neuroscience*, *26*(51), 13128. https://doi.org/10.1523/JNEUROSCI.1657-06.2006

Kim, K. K., Karunanayaka, P., Privitera, M. D., Holland, S. K., & Szaflarski, J. P. (2011). Semantic association investigated with functional MRI and independent component analysis. *Epilepsy & Behavior*, *20*(4), 613-622. https://doi.org/10.1016/j.yebeh.2010.11.010

Kriegstein, K. V., & Giraud, A.-L. (2004). Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *Neuroimage*, *22*(2), 948-955. https://doi.org/https://doi.org/10.1016/j.neuroimage.2004.02.020

Lega, B., Germi, J., & Rugg, M. D. (2017). Modulation of Oscillatory Power and Connectivity in the Human Posterior Cingulate Cortex Supports the Encoding and Retrieval of Episodic Memories. *Journal of cognitive neuroscience*, *29*(8), 1415-1432. https://doi.org/10.1162/jocn_a_01133

Li, W., Zhuang, J., Guo, Z., Jones, J. A., Xu, Z., & Liu, H. (2019). Cerebellar contribution to auditory feedback control of speech production: Evidence from patients with spinocerebellar ataxia. *Human brain mapping*, *40*(16), 4748-4758. https://doi.org/https://doi.org/10.1002/hbm.24734

Liljeström, M., Hulten, A., Parkkonen, L., & Salmelin, R. (2009). Comparing MEG and fMRI views to naming actions and objects. *Human brain mapping*, *30*(6), 1845-1856.

Maess, B., Friederici, A. D., Damian, M., Meyer, A. S., & Levelt, W. J. (2002). Semantic category interference in overt picture naming: Sharpening current density localization by PCA. *Journal of cognitive neuroscience*, *14*(3), 455-462.

Makeig, S., Bell, A., Jung, T.-P., & Sejnowski, T. J. (1995). Independent component analysis of electroencephalographic data. *Advances in neural information processing systems*, *8*.

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

Marconi, D., Manenti, R., Catricala, E., Della Rosa, P. A., Siri, S., & Cappa, S. F. (2013). The neural substrates of inferential and referential semantic processing. *Cortex*, *49*(8), 2055-2066.

Martin-Loeches, M., Hinojosa, J. A., Gomez-Jarabo, G., & Rubia, F. J. (2001). An early electrophysiological sign of semantic processing in basal extrastriate areas. *Psychophysiology*, *38*(1), 114-124.

Michael, K., Aron, K. B., Bradley, R. P., & Jordan, G. (2009). Superior Parietal Cortex Is Critical for the Manipulation of Information in Working Memory. *The Journal of Neuroscience*, *29*(47), 14980. https://doi.org/10.1523/JNEUROSCI.3706-09.2009

Miozzo, M., Pulvermüller, F., & Hauk, O. (2015). Early parallel activation of semantics and phonology in picture naming: Evidence from a multiple linear regression MEG study. *Cerebral cortex*, *25*(10), 3343-3355.

Oberhuber, M., Hope, T. M. H., Seghier, M. L., Parker Jones, O., Prejawa, S., Green, D. W., & Price, C. J. (2016). Four Functionally Distinct Regions in the Left Supramarginal Gyrus Support Word Processing. *Cerebral cortex*, *26*(11), 4212-4226. https://doi.org/10.1093/cercor/bhw251

Peeva, M. G., Guenther, F. H., Tourville, J. A., Nieto-Castanon, A., Anton, J.-L., Nazarian, B., & Alario, F. X. (2010). Distinct representations of phonemes, syllables, and supra-syllabic sequences in the speech production network. *Neuroimage*, *50*(2), 626-638. https://doi.org/https://doi.org/10.1016/j.neuroimage.2009.12.065

Pion-Tonachini, L., Kreutz-Delgado, K., & Makeig, S. (2019). ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *Neuroimage*, *198*, 181-197.

Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage*, *62*(2), 816-847.

Riès, S. K., Dhillon, R. K., Clarke, A., King-Stephens, D., Laxer, K. D., Weber, P. B., Kuperman, R. A., Auguste, K. I., Brunner, P., & Schalk, G. (2017). Spatiotemporal dynamics of word retrieval in speech production revealed by cortical high-frequency band activity. *Proceedings of the National Academy of Sciences*, *114*(23), E4530-E4538.

Salmelin, R., Hari, R., Lounasmaa, O., & Sams, M. (1994). Dynamics of brain activation during picture naming. *Nature*, *368*(6470), 463-465.

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

Strijkers, K., & Costa, A. (2016). The cortical dynamics of speaking: Present shortcomings and future avenues. *Language, Cognition and Neuroscience*, *31*(4), 484-503.

Strijkers, K., Costa, A., & Pulvermüller, F. (2017). The cortical dynamics of speaking: Lexical and phonological knowledge simultaneously recruit the frontal and temporal cortex within 200 ms. *Neuroimage*, *163*, 206-219.

Strijkers, K., Costa, A., & Thierry, G. (2010). Tracking lexical access in speech production: electrophysiological correlates of word frequency and cognate effects. *Cerebral cortex*, *20*(4), 912-928.

Vihla, M., Laine, M., & Salmelin, R. (2006). Cortical dynamics of visual/semantic vs. phonological analysis in picture confrontation. *Neuroimage*, *33*(2), 732-738.

Weber, K., Lau, E. F., Stillerman, B., & Kuperberg, G. R. (2016). The Yin and the Yang of Prediction: An fMRI Study of Semantic Predictive Processing. *PloS one*, *11*(3), e0148637. https://doi.org/10.1371/journal.pone.0148637

Wilson, S. M., Molnar-Szakacs, I., & Iacoboni, M. (2007). Beyond Superior Temporal Cortex: Intersubject Correlations in Narrative Speech Comprehension. *Cerebral cortex*, *18*(1), 230-242. https://doi.org/10.1093/cercor/bhm049

Wingfield, A., & Grossman, M. (2006). Language and the Aging Brain: Patterns of Neural Compensation Revealed by Functional Brain Imaging. *Journal of Neurophysiology*, *96*(6), 2830-2839. https://doi.org/10.1152/jn.00628.2006

Winkler, I., Debener, S., Müller, K.-R., & Tangermann, M. (2015). On the influence of high-pass filtering on ICA-based artifact reduction in EEG-ERP. 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC),

Yamamoto, A. K., Parker Jones, O., Hope, T. M. H., Prejawa, S., Oberhuber, M., Ludersdorfer, P., Yousry, T. A., Green, D. W., & Price, C. J. (2019). A special role for the right posterior superior temporal sulcus during speech production. *Neuroimage*, *203*, 116184. https://doi.org/https://doi.org/10.1016/j.neuroimage.2019.116184