

scientific data



OPEN
DATA DESCRIPTOR

An open-access EEG dataset for speech decoding: Exploring the role of articulation and coarticulation

João Pedro Carvalho Moreira^{1,7}, Vinícius Rezende Carvalho¹,
 Eduardo Mazoni Andrade Marçal Mendes¹, Aria Fallah², Terrence J. Sejnowski^{3,4,5},
 Claudia Lainscsek^{3,4} & Lindy Comstock^{6,7}

Electroencephalography (EEG) holds promise for brain-computer interface (BCI) devices as a non-invasive measure of neural activity. With increased attention to EEG-based BCI systems, publicly available datasets incorporating the complex stimuli found in naturalistic speech are necessary to establish a common standard of performance within the BCI community. Effective solutions must overcome noise in the EEG signal and remain reliable across sessions and stimuli that reflect types of real-world linguistic complexity without overfitting to a dataset or task. We present two validated datasets ($N=8$ and $N=16$) for classification at the phoneme and word level and by the articulatory properties of phonemes. EEG signals were recorded from 64 channels while subjects listened to and repeated six consonants and five vowels. Individual phonemes were combined in different phonetic environments to produce coarticulated variation in 40 consonant-vowel pairs, 20 real words, and 20 pseudowords. Phoneme pairs and words were presented during a control condition and during transcranial magnetic stimulation (TMS) to assess whether stimulation would augment the EEG signal associated with specific articulatory processes.

Background & Summary

Brain-computer interface (BCI) devices aim to restore communicative abilities to individuals who have lost motor function. BCI speech decoding devices must balance the need for a user-friendly interface with task performance¹. Neural signals from electrocorticography (ECoG) provide a measure of task-relevant brain activity with a minimal processing latency and a superior signal-to-noise ratio^{2,3}. However, ECoG studies must be conducted on surgery patients who differ in many respects from the end user of BCI devices. Alternatively, electroencephalography (EEG) allows for BCI devices to be tested directly among the target population. But because neural signals become distorted as they pass through the skull and scalp to reach EEG sensors⁴, BCI applications are often limited to the detection of stimulus-evoked potentials^{5–8}, resulting in non-naturalistic paradigms that remain comparatively slow and inflexible^{9,10}. Although a practical solution to speech decoding will ultimately require the quick and accurate classification of a large inventory of context-dependent speech sounds in rapid succession¹¹, this goal remains out of reach for state-of-the-art EEG decoding methods¹².

To utilize EEG effectively for these purposes, researchers would benefit from datasets that incrementally increase the degree of “naturalness” of stimulus items, such that existing models may be tested with similar yet successively more complex datasets and adjusted to compensate for this increase in complexity¹³. Such data would comprise stimuli that incorporate systematic linguistic regularities and rules regarding phonological, phonotactic, or semantic content that could affect the decoding of naturalistic speech. This would entail stimuli in which the same sounds are presented in different linguistic environments, and word items that comprise real or pseudowords, which differentially engage semantic and motor neural networks. While much of the decoding

¹Postgraduate Program in Electrical Engineering, Federal University of Minas Gerais, Belo Horizonte, MG, 31270-901, Brazil. ²Department of Neurosurgery, University of California, Los Angeles, Los Angeles, CA, 90095, USA.

³Computational Neurobiology Laboratory, The Salk Institute for Biological Studies, La Jolla, CA, 92037, USA.

⁴Institute for Neural Computation University of California San Diego, La Jolla, CA, 92093, USA. ⁵Division of Biological Sciences, University of California San Diego, La Jolla, CA, 92093, USA. ⁶Department of Psychiatry & Biobehavioral Studies, University of California, Los Angeles, Los Angeles, CA, 90095, USA. ⁷These authors contributed equally: João Pedro Carvalho Moreira, Lindy Comstock. e-mail: lcomstock@ucla.edu

literature is based on what we assume is neural motor activation, decoding from semantic activation remains a goal within the field¹⁴, and to transition to the type of predictive networks that have been proposed for real-time decoding of propositional or conversational content, decoding that incorporates semantic neural networks may be necessary.

Currently, methodological paradigms for speech decoding generally include covert speech and auditory comprehension tasks^{15–27}, with innovations consisting largely as modifications of these paradigms. The resulting analyses are heterogeneous in nature and few authors make their data publicly available, complicating efforts to reproduce and compare different decoding methods: the publication of EEG datasets allows researchers to benchmark model performance^{28,29}. The available datasets typically focus on a task of high difficulty (e.g., imagined or inner speech) yet remain minimal in scope, with a small number of trials, stimulus types, and subjects¹². Moreover, there is little coherence between different stimulus types that would allow researchers to test their models against progressively more naturalistic data. BCI devices tend to rely upon the neural signals associated with muscle movements³⁰, and although the utility of classification schemes based on speech articulation has been noted by numerous researchers^{31–35}, this factor has yet to be included as an organizing principle in published datasets. Finally, the overfitting of highly complex machine learning models is frequently discussed in the literature³⁶ but rarely tested overtly. Independently collected datasets of the same data types are needed to ensure that models can generalize across subjects and sessions. We present two datasets for EEG speech decoding that partially address these limitations:

- Naturalistic speech is not comprised of isolated speech sounds³⁷. The phonetic environment surrounding phonemes affects their quality^{38,39}, complicating accurate category designation^{40,41}. Existing datasets lack a wide range of co-articulated phonemes. We provide single, double, and phoneme triplets for six consonants (/b/, /p/, /d/, /t/, /s/, /z/) and five vowels (/i/, /ɛ/, /a/, /u/, /oʊ/) to assess classification accuracy in progressively more complex phonetic environments.
- The decoding literature devotes considerable attention to the articulatory properties of phonemes^{31–35}. The most successful decoding models integrate this knowledge^{22,42,43}. Available datasets do not select phonemes that fall within overlapping articulatory categories (place, manner, voicing). We provide consonants that represent unique combinations of features (bilabial/alveolar; stop/fricative; voiced/unvoiced) to assess articulatory features as a classification parameter.
- The integration of probabilistic language models into BCI devices has predisposed researchers to anticipate better results when training on real words^{44–47}. EEG datasets tend to comprise word stimuli that are real¹². Yet functional magnetic resonance imaging (fMRI) studies reveal a greater hemodynamic response to pseudowords^{48,49}. Effortful processing may strengthen neural signals and facilitate decoding⁵⁰. We provide real and pseudowords to test this hypothesis.
- Speech decoding papers typically publish one or more analyses conducted on a single dataset¹², raising concerns about overfitting and how well the model might perform on additional data without significant modification⁵¹. We provide two datasets ($N=8$ and $N=16$) collected at different time points for the same stimulus types and/or participants. The second serves as an external validation set to allow model performance to be assessed on independently derived data⁵².
- Modification of the EEG signal may prove useful for the decoding of neural signals as an initial training aid or as an augmentation technique. We provide data from a successful study which found that transcranial magnetic stimulation (TMS) may have the potential to increase successful decoding from EEG signals⁵³. We provide data from a control condition and two TMS conditions by phoneme type and stimulation target⁵⁴ for further experimentation and testing.
- Robust speech decoding models must compensate for a high degree of noise from various sources. In addition to the limitations of EEG data (noise from within the signal), sounds are often heard inaccurately in the presence of environmental noise or when the participant is inattentive to stimuli^{55–57}. Noise may also occur as a feature of accompanying techniques, such as TMS.

Methods

Our aim was to create a systematically structured, multifaceted dataset that would represent a novel contribution to the publicly available data for EEG speech decoding. This involved the selection of stimuli that allow a scaffolded progression in the type of classification analyses that can be conducted in terms of the choice of parameters and the complexity of the task (single phonemes, phoneme pairs, phoneme triplets/words). The dataset includes comprehension and production tasks collected either during neuromodulation or a control condition. There is evidence that TMS may improve subject performance in phoneme discrimination by administering two closely spaced TMS pulses prior to phoneme presentation while targeting motor cortex regions that control the muscles involved in the articulation of a specific phoneme category⁵⁸. Likewise, errors may be induced when stimulation occurs in an unrelated region of the motor cortex, prompting the perception of a different phoneme category⁵⁸. The majority of this data is published here for the first time. A subset of the data (CV pairs) has previously been published in conjunction with a pilot study that investigated whether the perceptual effects would translate into a similar bias in decoding accuracy during the neural decoding of EEG signals recorded during the perception task⁵³. The pilot study illustrated proof-of-concept; by collecting data independently at two separate time points (2019 and 2021) and now offering the full dataset, we invite researchers (i) to develop a model on the larger dataset (2021) and assess the robustness of the model in a second, smaller and noisier validation dataset (2019), and (ii) to test their models on progressively larger units of speech (2021).

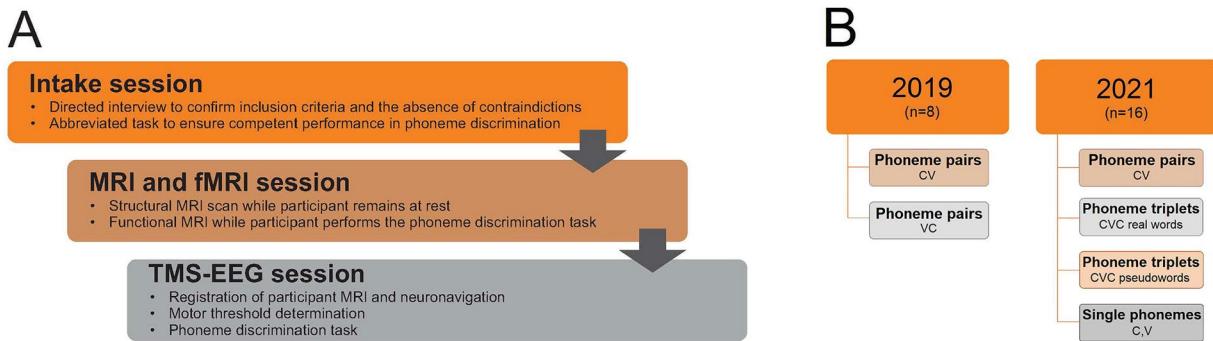


Fig. 1 Session organization. **(A)** The experiment was conducted in three sessions that were held on separate days. Participant eligibility was confirmed in the first session. MRI scanning and TMS-EEG data collection were conducted independently. **(B)** Additional participants and trial types were included in 2021.

Participants. Participants aged between 20 and 40 were recruited from the UCLA campus using flyers. Ten participants (6 female) were recruited for the first round of data collection in 2019 and twenty participants (10 female) were recruited for a second round of data collection in 2021. Inclusion criteria were defined as no diagnosis of any neurological, psychiatric, or developmental disorders, self-reported normal hearing, and no contraindications for TMS or MRI procedures (implanted medical devices, implanted metal, pregnancy, personal or family history of seizures, and exclusionary medications). During an initial screening session, participants completed an abbreviated version of the experimental task to ensure that participants understood the task directions and could perform the phoneme discrimination task. Left-hemisphere lateralization of the language processing regions in all participants was established during an fMRI scan in which participants performed a similar phoneme discrimination task, slightly modified for the MRI scanner. Two individuals (1 female) were excluded from the 2019 dataset due to necessary modifications that were made to the stimulus audio files after their participation. In the 2021 dataset, one participant (male) was excluded when he exhibited a biased response strategy upon stimulation (i.e., failure to select from the full set of phonemes), and three participants (male) were excluded due to complications with the TMS equipment that may have led to imprecise targeting. Participants provided informed consent and were paid for two sessions. The experimental protocol was approved by the UCLA Institutional Review Board (IRB#21-000333). The same participant recruitment and data collection procedures described below are presented in an abbreviated form in our pilot study publication⁵³.

Experimental design. The study was conducted in three sessions (Fig. 1). In the first session, participants underwent a directed interview to ensure that they met the study inclusion criteria and possessed no contraindications. Participants who were able to perform an abbreviated phoneme perception task with at least 75% accuracy were enrolled. In the second session, participants underwent an MRI scan to aid in neuronavigation for the TMS procedure. They performed a modified phoneme discrimination task in the MRI scanner to lateralize their primary language processing areas for consonant and word stimuli. The final session involved the recording of EEG signals while participants performed the phoneme perception task. TMS was targeted to areas of the motor cortex associated with the production of specific phonemes. In the second round of data collection that was conducted in 2021, participants also listened to and repeated single phonemes and performed the perception task for phoneme triplets.

Data collection. *MRI scanning.* Scanning was conducted in the UCLA Center for Cognitive Neuroscience with a Siemens Prisma-FIT 3T Scanner. Participants were provided with ear protectors and headphones for a 45 to 60 dB reduction of the noise associated with scanning to ensure that participants could hear the stimuli clearly and that the noise level was not uncomfortably loud. Participants were asked to lie with their head motionless during all scanning procedures. High-resolution anatomical images were acquired, followed by a functional scan in which participants were directed to either relax passively while looking at a fixation cross or to perform the button-press phoneme discrimination task. Stimuli were grouped by consonant or word type in a block design to increase the statistical power of the fMRI analysis. Functional data were acquired in the block design with a BOLD-weighted echoplanar imaging sequence aligned in parallel to the bicommissural plane, thus yielding 36 slices covering the whole brain, each 3 mm thick with a 1 mm gap between slices. Each slice was acquired as a 64 × 64 matrix yielding an in-plane resolution of 1.5 × 1.5 mm. The total duration of the scanning session was 40 minutes.

TMS-EEG. The TMS-EEG procedure was conducted in the Neuromodulation Division of the Semel Institute for Neuroscience and Human Behavior at UCLA. The TMS equipment utilized for the procedure included a Magstim Super Rapid Plus1 stimulator and a figure-of-eight 40 mm coil. The EEG system included an eego™ sports WaveGuard 64-channel EEG cap and eego mylab system compatible with electromagnetic stimulation. Targeting was completed using the Visor 2 neuronavigation system. The electrode positions were digitized and registered to individual participant MRIs using the ANT Neuro Xensor. EEG signals were bandpass-filtered 0.1–350 Hz, sampled at 2000 Hz, and referenced to the CPz electrode. All electrode impedances were kept <5 kΩ. PsychoPy⁵⁹ stimulus presentation software initiated the audio routines and recorded reaction time data.

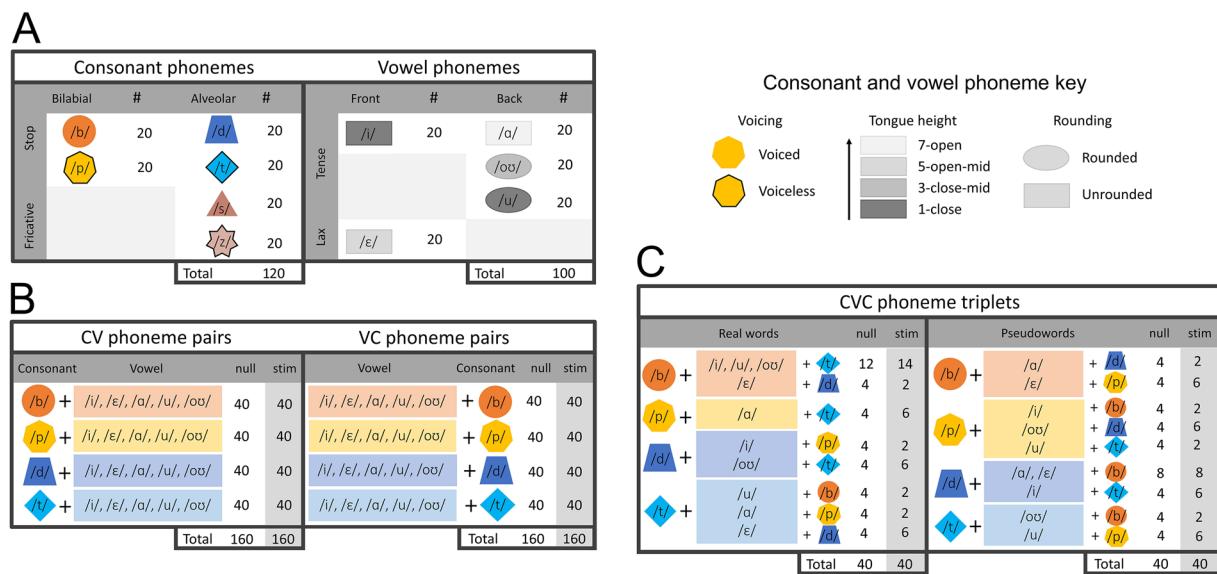


Fig. 2 Stimulus types. (A) Consonant and vowel phonemes possess unique articulatory features. Consonants can be described by three parameters: place of articulation (bilabial, alveolar), manner of articulation (stop, fricative), and voicing (voiced, unvoiced). Vowels can be described by four parameters: tongue height (from close to open), tongue position (front, back), tongue tension (tense, lax), and lip position (rounded, unrounded). (B) Phoneme pairs included eight instances of each combination of stop consonants and vowels. (C) Phoneme triplets were limited by the stipulation to create real and pseudowords. Stimuli included eight instances of each vowel in the real and pseudoword conditions.

The appropriate stimulation intensity for TMS studies is determined on an individual basis⁶⁰. Prior to the experimental session, the motor threshold (rMT) of each participant was determined by eliciting motor-evoked potentials (MEPs) in the first dorsal interosseus (FDI) muscle of the dominant hand at the minimum amount of stimulation needed to evoke an MEP in a hand muscle after a single pulse over M1. Single TMS pulses were delivered to locations in the motor cortex contralateral to the dominant hand. The intensity of the stimulation was gradually lowered until reaching a level of stimulator output at which 5 out of 10 MEPs in the hand muscle had an amplitude of at least 50 microvolts. Potentials evoked during TMS represent the net sum of excitatory and inhibitory stimulation effects^{61–63}. The literature has found that excitation increases at intensities of 110–120% rMT. In accordance with our reference study, stimulation was administered at 110% of the FDR rMT⁵⁸. A physician observed the motor thresholding procedure to ensure that no negative effects were incurred by participants.

TMS targeted areas of the motor cortex involved in (i) lip and tongue movements (which produce bilabial and alveolar consonants, respectively) or (ii) processing of real and pseudowords. Stimulation targets were defined as the MNI coordinates of peak motor cortex activation in LipM1 and TongueM1 during lip and tongue articulatory movements (lips: -56, -8, 46; tongue: -60, -10, 25), taken from the literature⁶⁴ and the reference study⁵⁸. However, cortical functional localization is known to show individual variation⁶⁵. Therefore, the coordinates were overlaid over the activation map of the task results for each participant to ensure an overlap between the targets and individual task localization. The target was taken as the nearest peak to the MNI coordinate. Broca's area (BA 44: -51, 7, 23) was the target for real words and a region implicated in verbal memory (BA 6: -46, 1, 41) was the target for pseudowords⁶⁶.

Behavioral task. The phoneme discrimination task consisted of listening to speech sounds and identifying stimuli with a button-press response. Auditory stimuli (Fig. 2) were presented via laptop speakers: (i) single phonemes, (ii) paired consonant-vowel phonemes (CV, VC), and (iii) real or pseudowords constructed of phoneme triplets (CVC). Consonant stimuli included four phonemes in the pair and triplet conditions (/b/, /p/, /d/, /t/), with the addition of two additional phonemes (/s/, /z/) in the single condition. Vowel stimuli included five phonemes in all conditions (/i/, /ɛ/, /a/, /u/, /oo/). These sets yielded 11 individual phonemes (6 consonants and 5 vowels), 40 phoneme pairs (20 CV/20 VC), and 40 phoneme triplets (20 real/20 pseudowords). Participants were asked to listen to and repeat the phoneme in the single condition, to identify the consonant phoneme in paired conditions, and to identify phoneme triplets as real or pseudowords. Multiple classification analyses may be conducted on each stimulus type (Fig. 3).

TMS elicits a period of excitatory activation with an onset latency of 50–80 ms after stimulation⁶⁷. We reproduced the design of our reference study⁵⁸ to ensure an excitatory neural response that would translate into task facilitation. Each trial delivered paired TMS pulses at one of the stimulation targets, separated by a short interpulse interval (50 ms). Excitation of the cortical region not involved in stimulus production (i.e., TMS at LipM1 during alveolar phoneme presentation) results in neural noise that interferes with the perception task. The audio stimulus followed 50 ms after the second TMS pulse. One target was stimulated per run (counterbalanced across participants). Details of the experimental protocol are illustrated in Fig. 4.

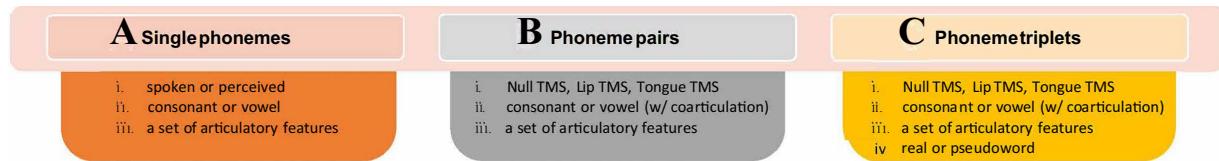


Fig. 3 Analysis. (A) Single phonemes can be classified by modality, category, and articulation. (B) Phoneme pairs – by target, category, and articulation. (C) Phoneme triplets – by target, category, articulation, and word type.

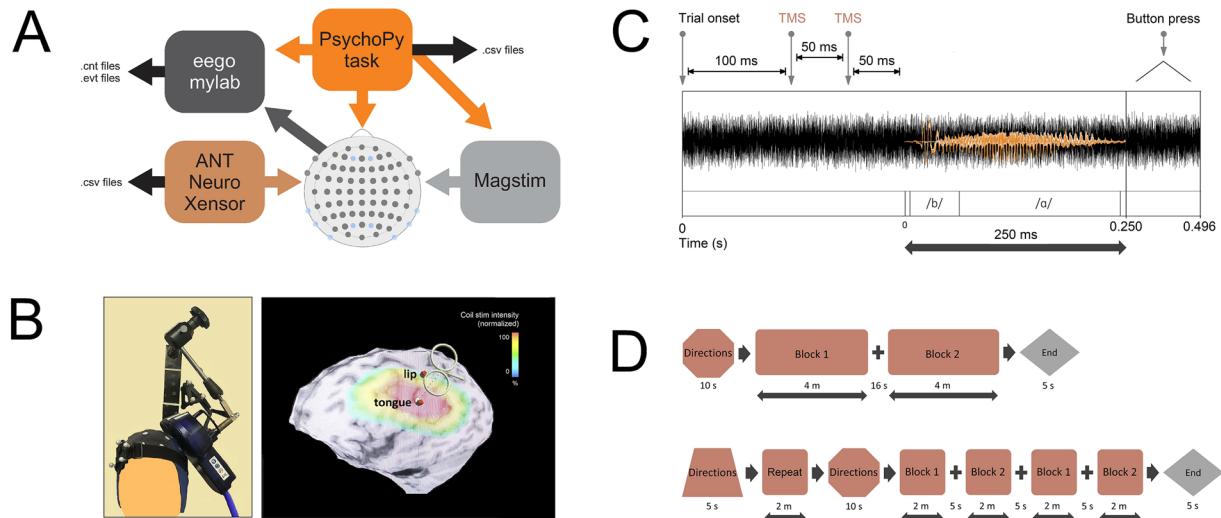


Fig. 4 Experimental protocol. (A) Software controlled the experimental task and sent triggers to initiate TMS and to create timestamps for each pulse. (B) The TMS coil was positioned at the necessary stimulation site. Despite some overlap in the induced magnetic field, only the targeted region received maximum stimulation intensity. (C) The audio stimuli were immersed in white noise. Two TMS pulses were administered 50 ms prior to stimulus onset. (D) The run design in 2019 was reproduced in 2021 with the addition of more frequent breaks between blocks. A block of single phonemes was introduced in 2021.

Participants listened to audio clips immersed in 500 ms of white noise. The white noise created a mild background distraction for participants to ensure that they did not perform the phoneme discrimination task at ceiling. Participants were instructed to respond as fast as possible with a button press after they had identified the phoneme. In the case of multiple button presses, correct trials were determined from the initial button press. Participants who exhibited a non-random response strategy (i.e., failure to select from the full set of phonemes) were excluded. In 2021, participants were instructed to listen to single phonemes without TMS and to repeat the sound they heard immediately after stimulus presentation (300 ms from trial onset).

Two lists of stimulus items were used with one list assigned to each block. In 2019, the runs were split into two blocks. The first block presented CV pairs, followed by a block of VC pairs. In 2021, four blocks were administered per run. The first two blocks presented CV pairs, followed by two blocks of CVC stimuli (real and pseudowords). A five-minute break was provided between runs. Participants completed 120 trials in each run: 80 with TMS and 40 random catch trials. In 2021, each run of the task was preceded by the presentation of 220 trials of single phonemes (20 trials each). Stimuli in all conditions were presented in a pseudo-randomized order. The total run time of the experiment lasted 49 minutes in 2019 and 58 minutes in 2021. As noted in the preceding sections, minimal modifications to the procedure were required for the intake and scanning sessions. For the initial assessment, half of the task was administered. During fMRI scanning, the full-length task was administered with stimuli presented in blocks of the same type (bilabial, alveolar, real words, pseudowords).

Data characterization. *Classification of acquired data.* The procedure required sustained attention during a lengthy TMS procedure. The mean reaction time and standard deviation were calculated to confirm that participants were attentive to the task throughout the procedure. These metrics are documented in .csv files uploaded to the data repository⁶⁸. In the 2019 dataset, some variation in trial numbers is observed due to missed trials and rotation in the list of stimuli administered to each participant. No subjects performed less than 90% of the total list, with the exception of P04 in the VC condition with LipTMS. Here, excluded trials resulted from missed trials. In the 2021 dataset, all trials were uploaded irrespective of a button-press response. Two subjects performed an abbreviated list of phoneme triplets, and one also performed an abbreviated list of single phonemes. The number of tagged trials is shown in Tables 1 and 2.

Target	Task	S01	S02	S03	S04	S05	S06	S07	S08	Total
Lip	CV pairs	79	81	80	76	77	80	81	77	631
TMS	VC pairs	77	79	79	57	79	79	78	78	606
Lip	CV pairs	39	37	38	38	36	37	38	38	301
NULL	VC pairs	41	43	42	33	41	41	42	42	325
Tongue	CV pairs	80	81	81	78	77	79	79	76	631
TMS	VC pairs	80	78	78	78	77	79	78	76	624
Tongue	CV pairs	39	38	38	38	36	38	38	36	301
NULL	VC pairs	41	42	41	41	38	41	40	38	322
	Total	476	479	477	439	461	474	474	461	

Table 1. Trial Numbers: 2019 Dataset.

Target	Task	S01	S02	S03	S04	S05	S06	S07	S08	S09	S10	S11	S12	S13	S14	S15	S16	Total
NA	C/V spoken	110	220	220	220	220	220	220	220	220	220	220	165	220	220	220	220	3355
NA	C/V perceived	110	220	220	220	220	220	220	220	220	220	220	165	220	220	220	220	3355
Lip	CV pairs	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80	1280
NULL	CV pairs	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	640
Tongue	CV pairs	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80	80	1280
NULL	CV pairs	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	40	640
BA 44	real words	0	30	40	40	40	40	40	40	40	40	40	40	0	40	40	40	550
	pseudowords	0	31	40	40	40	40	40	40	40	40	40	40	0	40	40	40	551
NULL	real words	0	16	20	20	20	20	20	20	20	20	20	20	0	20	20	20	276
	pseudowords	0	16	20	20	20	20	20	20	20	20	20	20	0	20	20	20	276
BA 6	real words	40	40	40	40	40	40	40	40	40	40	40	40	80	40	40	40	680
	pseudowords	40	40	40	40	40	40	40	40	40	40	40	40	80	40	40	40	680
NULL	real words	20	20	20	20	20	20	20	20	20	20	20	20	40	20	20	20	340
	pseudowords	20	20	20	20	20	20	20	20	20	20	20	20	40	20	20	20	340
	Total	690	893	920	920	920	920	920	920	920	920	920	920	810	920	920	920	

Table 2. Trail Numbers: 2021 Dataset.

Data processing. The continuous raw data from the EEG recordings have been uploaded to two data repositories^{68,69} in BIDS format and as .cnt files so that researchers may apply their preferred pre-processing and processing pipeline, as necessary for alternative speech decoding models. However, in order to exemplify the principle of transparency, we have also made available three stages of data cleaning or signal processing that were utilized for our data validation section (i) data normalized within each data window to zero mean and unit variance (for DDA); (ii) data resampled at 256Hz, filtered (notch filter in the 59Hz to 61Hz bands and band-pass filter from 0.1Hz to 100Hz), and separated by trial and with the removal of bad channels and TMS artifacts (for ERPs); and (iii) data that additionally underwent 1-2 rounds of cleaning of unwanted ICA components (for ERPs).

A Matlab routine was designed based on the EEGLAB library⁷⁰ for data analysis and pre-processing. The code^{68,71} has been made available in several separate sections, each responsible for a part of the data processing. This makes it possible to organize, streamline, and automate the analysis, in a process that eliminates the extensive use of EEGLAB interface by the user in cases where it is not strictly necessary. The routine is structured as shown in Fig. 5 and consists of seven main sections that include the removal of unwanted channels, event setup based on the information tables, resampling and filtering, separation of trials based on events, visual inspection for cleaning up bad trials, ICA decomposition, and manual removal of unwanted ICA components. In addition, the code offers optional sections for interpolation of TMS signals, generation of signal state images, and other tools that provide information on changes made during processing.

After each data processing stage, the data were stored in specific folders that consist of .set files containing the pre-processing steps, a summary of the manual modifications made, and .mat files that document the event-related potential and spectral power density obtained for the analysis conducted. As a result, the outputs are organized and clearly annotated to ensure reproducibility and access to all the stages of the pipeline. For DDA, all data were utilized without filtering or downsampling.

Load data, remove bad channels, and set events. The pipeline was built to run one participant at a time. Initially, all the variables used to store the location of the necessary files were defined, then the folders for organizing the outputs were created, and the .cnt file was imported and saved as a .set file. Only the EEG recordings are of interest for this analysis, so the EOG and BIPs channels were removed. In addition, M1 and M2 were discarded. In total, 62 channels were processed, with CPz used as the reference and AFz as the ground electrode. Next, the database for each subject was updated with seven events for each phoneme pair task-related trial, namely: baseline, trial onset, first TMS pulse, second TMS pulse, first auditory stimulus, second auditory stimulus, and trial end.

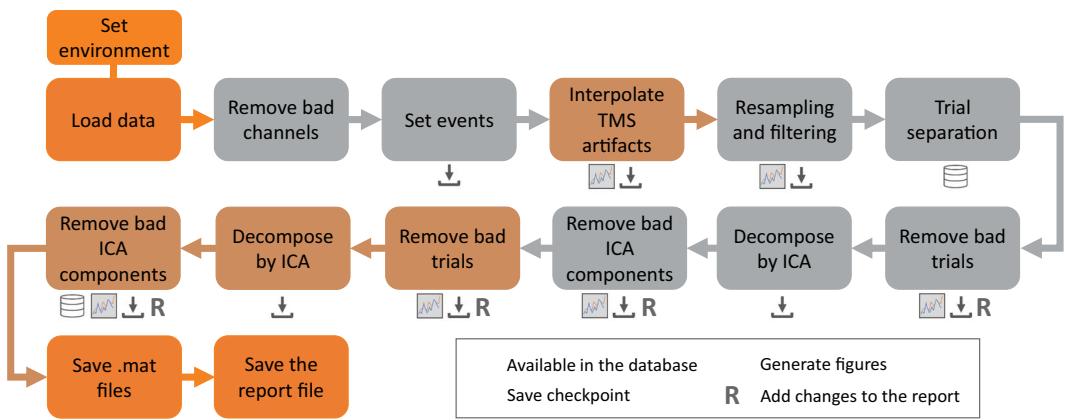


Fig. 5 Code structure for processing the data. The seven fundamental sections are shown in gray, the optional sections in brown, and the extra functions for organizing the data in orange. The icons below each section indicate operations applied after the data processing step is complete.

Resampling and filtering. The data were resampled to 256Hz and two filters were applied: a notch filter with cutoff frequencies at 59Hz and 61Hz and a band pass filter with cutoff frequencies at 0.1Hz and 100Hz. The output generated by this last step is used for the subsequent analyses, differentiated only by the type of trial studied: (i) control, (ii) TMS applied to the lip target region, or (iii) with TMS applied to the tongue target region.

Trials separation. This section selects the analysis type and defines the first sound stimulus as the base event for ERP construction. This event sets an epoch separation that avoids undesired effects due to the high-amplitude TMS spikes.

Trials inspection and data cleaning. Two rounds of visual inspection were conducted on the data in search of trials with contaminated signals and channels with high-amplitude artifacts. To do so, the spectral power density (PSD) plots generated from the data were analyzed, in addition to the EEG recordings themselves. Using the EEGLAB interface, the unwanted segments were selected and removed from the analysis. After each round, ICA decomposition was applied using the library-adapted infomax ICA algorithm⁷². During ICA, 35 components were inspected and those that clearly exhibited an artifact signal were removed.

Data Records

The entire dataset⁶⁹ can be found at OpenNeuro (<https://doi.org/10.18112/openneuro.ds006104.v1.0.0>) in BIDS format. In addition, the complete data records can be found in the Open Science Framework repository⁶⁸. Studies are labeled chronologically. Each primary folder contains subfolders for the raw data in .cnt format, processed data, and trial characteristics. The raw and processed data are grouped individually, with one subject per folder, and labeled as per Tables 1 and 2.

The routines for the analyses in the technical validation section and the results for both signal processing techniques in the data processing section are located in the Study/EEG_Data_Processing/Code folder. ERPs are obtained using only ICA, and signal cleaning was performed using the pipeline described in Fig. 1, based on the EEGLab library versions 2022.0 and 2022.1 native to MATLAB. The EEG data records can be found in the OpenNeuro repository⁶⁹ in BIDS format. The dataset follows BIDS convention with the following structure: /sub-[subject]/ses-[session]/eeg/. Subject labels are P01-P08 for Study 1 and S01-S16 for Study 2 to avoid confusion about the origin of the data files. Session is 01 for Study 1 and 02 for Study 2.

Raw and pre-processed EEG data. Raw EEG files were stored in the .cnt format. This format contains continuous EEG recordings saved over the EEG-TMS sessions. 66 channels were recorded, with electrode placement according to Fig. 4. Pre-processed EEG data has also been made available in .set and .mat files, according to steps described in Fig. 5.

Event timestamps and behavioral data. For each trial, event timestamps are provided in .csv format, with one file for each recording session (Fig. 6). The events include (i) the second (final) TMS pulse of the pair, (ii) the sound stimulus onset, and (iii) the subsequent phoneme onsets. In addition to timestamps, the files provide labels for presented (true) and identified sound stimuli (phoneme or real/nonce word).

Technical Validation

Two sets of analyses were performed to support the technical quality of the datasets. Firstly, we extracted the grand mean event-related potentials (ERPs) by means of independent component analysis (ICA)⁷² to illustrate evidence of a stimulus-locked response across participants in each condition. We selected this method primarily due to its widespread use in the investigation of human cognitive information processing and therefore its familiarity among the electrophysiology research community. However, ICA can be subjective in its implementation

A	Stimulus	TMS	Correct_Key	TMStarget	Task	P1_TSidx	TMS_TSidx	TrialN	Phoneme1	Phoneme2	Order	P2_TSidx	Place	Manner	Voicing	Category
iT_angry2	0	4	control_lip	phonemes	14408	14308	4	i	t	vC	14652	alveolar	stop	no	alveolar	
Bo_angry2	1	1	lip	phonemes	22374	22274	5	b	o	Cv	22428	bilabial	stop	yes	bilabial	
iD_angry2	1	3	lip	phonemes	30372	30272	6	i	d	vC	30714	alveolar	stop	no	alveolar	
eB_angry1	1	1	lip	phonemes	38372	38272	7	e	b	vC	38692	alveolar	stop	no	alveolar	
aD_happy1	1	3	lip	phonemes	46371	46271	8	a	d	vC	46628	alveolar	stop	no	alveolar	
Pa_angry1	1	2	lip	phonemes	54370	54270	9	p	a	Cv	54504	bilabial	stop	no	bilabial	
eT_happy2	0	4	control_lip	phonemes	62399	62299	10	e	t	vC	62662	alveolar	stop	no	alveolar	

B	Stimulus	TMS	Correct_key	TMStarget	Task	P1_TSidx	TMS_TSidx	TrialN	Phoneme1	Phoneme2	Phoneme3	P2_TSidx	P3_TSidx	Place	Manner	Voicing	Category
Ah1	0	NaN	control_tongue	single-phoneme	235776	235530	1	a				NaN	NaN	vowels			vowels
Ee1	0	NaN	control_tongue	single-phoneme	239719	239493	2	i				NaN	NaN	vowels			vowels
Eh1	0	NaN	control_tongue	single-phoneme	243655	243488	3	e				NaN	NaN	vowels			vowels
Uh1	0	NaN	control_tongue	single-phoneme	247652	247452	4	u				NaN	NaN	vowels			vowels
Oo1	0	NaN	control_tongue	single-phoneme	251669	251481	5	o				NaN	NaN	vowels			vowels
S1	0	NaN	control_tongue	single-phoneme	255737	255478	6	s				NaN	NaN	alveolar	fricative	no	alveolar
Z1	0	NaN	control_tongue	single-phoneme	259651	259475	7	z				NaN	NaN	alveolar	fricative	yes	alveolar
P1	0	NaN	control_tongue	single-phoneme	263616	263470	8	p				NaN	NaN	bilabial	stop	no	bilabial
B1	0	NaN	control_tongue	single-phoneme	267618	267468	9	b				NaN	NaN	bilabial	stop	yes	bilabial
T1	0	NaN	control_tongue	single-phoneme	271986	271497	10	t				NaN	NaN	alveolar	stop	no	alveolar
D1	0	NaN	control_tongue	single-phoneme	275640	275493	11	d				NaN	NaN	alveolar	stop	yes	alveolar

C	Stimulus	TMS	Correct_key	TMStarget	Task	P1_TSidx	TMS_TSidx	TrialN	Phoneme1	Phoneme2	Phoneme3	P2_TSidx	P3_TSidx	Place	Manner	Voicing	Category
Ti_angry1	0	5	control_lip	phonemes	2461673	2461573	335	t	i			2461847	NaN	alveolar	stop	no	alveolar
Pu_angry1	0	8	control_lip	phonemes	2467267	2467167	336	p	u			2467350	NaN	bilabial	stop	no	bilabial
Do_happy2	0	4	control_lip	phonemes	2472863	2472763	337	d	o			2472915	NaN	alveolar	stop	yes	alveolar
Ta_happy2	1	5	lip	phonemes	2478494	2478394	338	t	a			2478564	NaN	alveolar	stop	no	alveolar
Ba_angry1	0	7	control_lip	phonemes	2484086	2483966	339	b	a			2484139	NaN	bilabial	stop	yes	bilabial

D	Stimulus	TMS	Correct_key	TMStarget	Task	P1_TSidx	TMS_TSidx	TrialN	Phoneme1	Phoneme2	Phoneme3	P2_TSidx	P3_TSidx	Place	Manner	Voicing	Category
Deb_angry1	1	2	BA06	Words	3274821	3274721	421	d	e	b		3274871	3275252				nonce
Pod_happy1	0	2	control_BA06	Words	3280413	3280313	422	p	o	d		3280521	3280859				nonce
Bit_happy1	1	1	BA06	Words	3286011	3285911	423	b	i	t		3286060	3286350				real
Deb_angry2	0	2	control_BA06	Words	3291602	3291502	424	d	e	b		3291652	3291965				nonce

Fig. 6 Files characterizing the EEG data. (A) The 2019 dataset provides labels for CV and VC phoneme pairs. The TMS condition in which control trials were collected is noted, as well as the articulatory features and category of each consonant phoneme per trial. A timestamp is provided for the onset of each phoneme and the final TMS pulse prior to stimulus presentation. Stimulation was conducted at only one cortical site per run. The control items are labeled according to the run in which they were collected. (B) The 2021 dataset provides the same information, when relevant, for single phonemes. (C) CV phoneme pairs are described in full, similar to the 2019 dataset. (D) Each component phoneme in the word trials is indicated and marked with a time stamp. Stimuli are categorized by word type.

by individual researchers, and the method may not be ideal for the analysis of specific types of data^{73,74}. In particular, substantial attention has been paid to the need to remove the TMS artifact from TMS-EEG data^{75–77}. Therefore, we performed a second analysis with delay differential analysis (DDA)^{78–80}, a non-linear signal processing technique that requires minimal pre-processing and is noise insensitive^{81,82}. The two analyses provide complementary evidence for the presence of a condition-dependent response in the EEG data. In particular, the DDA analysis illustrates excitatory activity during the time window of interest for our cognitive task, which differs by TMS condition.

Event-related potentials. The processing steps described in the Data Processing section were applied to the raw data to define the format of the auditory event-related potential (ERP) in the control condition and each TMS condition. The pipeline shown in Fig. 5 was executed for all participants to ensure homogeneity in the analysis. The mean and standard deviation of the potentials in the 1-second window after the first sound stimulus and the standard deviation for each of the 61 channels separated by the participant is represented in Fig. 9, while 8 represents the mean ERP from channel CP5. The reference pictures provided in Fig. 7 are meant to provide general guidance in interpreting the waveform; please refer to the cited papers for their original findings. We observe that the ERPs approximate the expected auditory-evoked potential (AEP) induced by phoneme pairs composed of stop consonants and vowels (see 7A,B). Deviations from the anticipated AEP may occur due to noise (our stimuli were immersed in white noise) and the exact combination of consonants and vowels in each stimulus item^{83,84}. The shape of the TMS-evoked potential (TEP) will depend on the number of pulses delivered, the interpulse interval, and whether stimulation is subthreshold or suprathreshold. A wide variety of TMS paradigms have been tested with conflicting results, such that it may be better to observe the TEP in order to identify whether the paradigm was excitatory or inhibitory, or to consider the effect by means of an additional measure, such as a behavioral task (see 7C)⁸⁵. The TMS paradigm used for collection of the dataset produced a facilitatory effect on performance in a phoneme perception task^{54,58}.

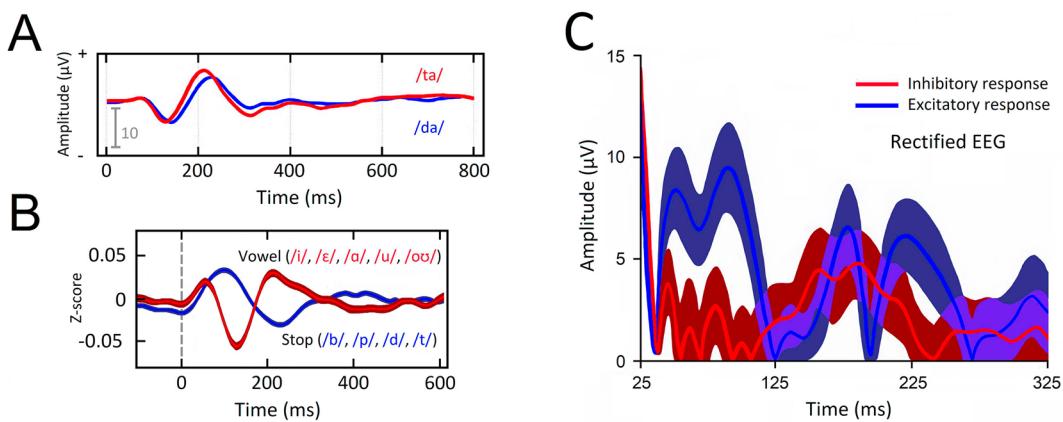


Fig. 7 Expected waveforms. All plots were modified and reproduced with permission from the publishers. (A) The auditory-evoked potential (AEP) for stop (/d/, /t/) consonant-vowel pairs exhibits a small N100 potential followed by a larger P200 potential. Variation in timing will occur depending on the stimulus type and in the presence of noise⁸³. (B) Vowels and consonants each produce a unique waveform⁸⁴, such that the overall shape is dependent on the contribution of each to the waveform. (C) The shape of the TMS-evoked potential (TEP) will differ according to the cortical region targeted. Whether TMS creates an excitatory or inhibitory response can be observed in the shape of the resulting TEP. In the motor cortex, greater activity between 25–125 ms accompanies excitatory paradigms⁸⁵. This figure illustrates the characteristic shape of the waveform for each type of neural response in the line and its standard deviation in the darker envelope.

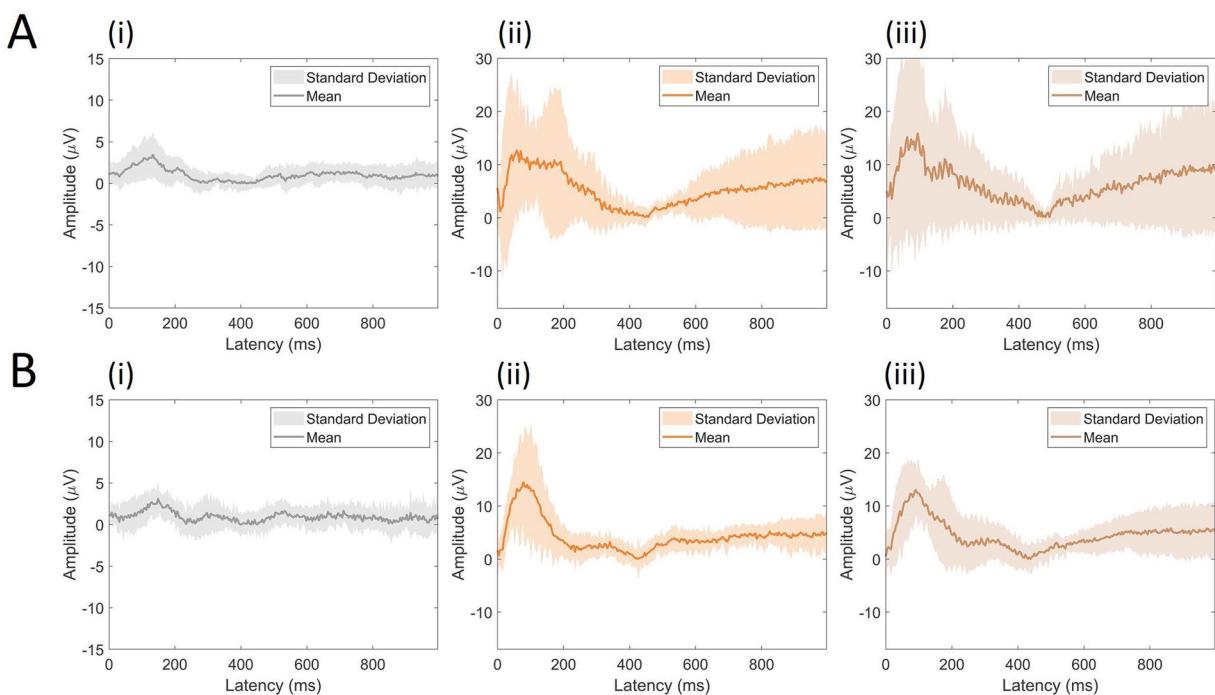


Fig. 8 Event-related potentials (ERPs) from trial onset for the channel CP5. The grand mean average ERPs to (i) control stimuli, (ii) stimuli with LipM1 stimulation, and (iii) stimuli with TongueM1 stimulation are displayed for (A) the 2019 and (B) 2021 datasets. The TMS conditions show rectified EEG activity to allow for comparison with the reference study.

Fig. 9 provides an overview of the analyzed data in which each recorded channel is represented by its mean and standard deviation. Note that the value of the dispersion of the control trial group is smaller than the other values, as expected, and that some channels of certain subjects have high signal variation, especially those related to the frontal and medial portions of the brain, where stimulation occurred. This is unsurprising, given that TMS affects a unique subset of cortical neurons in each individual based on the position and orientation of neurons relative to the stimulation coil^{86–91}.

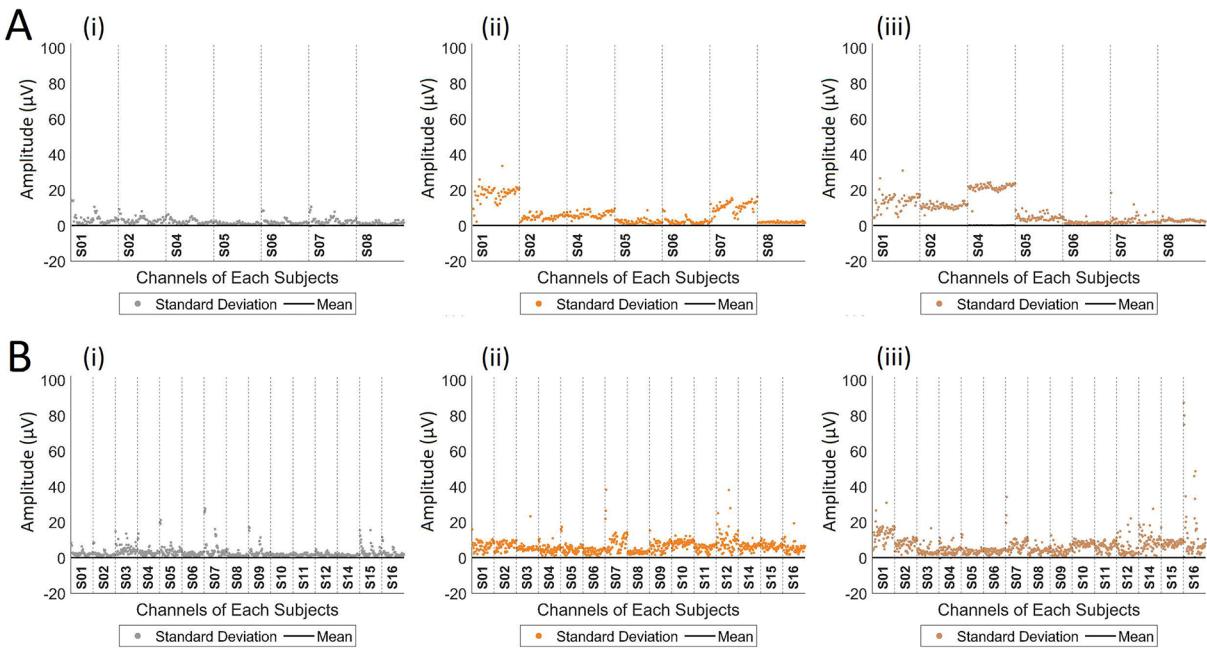


Fig. 9 Characterization of the data by the mean and standard deviation of each channel per participant to (i) control stimuli, (ii) stimuli with LipM1 stimulation, and (iii) stimuli with TongueM1 stimulation are displayed for (A) the 2019 and (B) the 2021 datasets. The channels are arranged for each experimental subject from most anterior on the left to most posterior on the right.

Delay differential analysis. Delay differential analysis (DDA) is a signal processing technique that combines differential embeddings with linear and nonlinear nonuniform functional delay embeddings. The integration of nonlinear dynamics allows information from the data to be detected which may not be observable in traditional linear methods. DDA requires minimal pre-processing, which eliminates a highly subjective step in the data analysis. Sparse DDA models have several advantages over the high dimensional feature spaces of other signal processing techniques: (i) the risk of overfitting is greatly reduced; (ii) the sparse model concentrates on the overall dynamics of the system and cannot additionally model noise; (iii) DDA is computationally fast; (iv) there is no need of pre-processing except normalization to zero mean and unit variance for each data window in order to ignore amplitude information and concentrate on system dynamics. The DDA model consists of two sets of parameters: (i) the delays and model form are the fixed parameters that are kept constant throughout the analysis; (ii) the coefficients (a_1, a_2, a_3) and the fitting error of the model are the free parameters. The coefficients are used as features to distinguish different dynamics in the data.

The DDA model used in this analysis is

$$\dot{x} = a_1 x_1 + a_2 x_2 + a_3 x_1^2, \quad (1)$$

where $x_i = x(t - \tau_i)$. In this analysis, the fixed parameters are the same as in Ref. ⁹². We found that one of the free parameters, namely a_3 , can be used to describe neural activity in a manner similar to ERPs. However, an ERP and a_3 are not strictly the same phenomenon. For details, see Ref. ⁹². Note that in most cases, there is no direct relation between frequencies and any of the model parameters, as explained in Ref. ⁷⁸. In the current analysis, the delays are $\tau_1 = 6\delta t$ and $\tau_2 = 16\delta t$, with $\delta t = \frac{1}{f_s}$, where the sampling rate is $f_s = 2000$ Hz. These are double to the delays in Ref. ⁹² because the sampling rate is double. The window length is 30 ms and the window shift is 1 ms. In Fig. 10, we observe waveforms that display the same dynamics as the reference studies in Fig. 7. We observe neural activity 200 ms and 400 ms after stimulus-onset, the same time window where activity is observed in Fig. 7A. We also observe a sharp spike in activity 25–125 ms after the final TMS pulse, which corresponds to the results illustrated in Fig. 7B. This finding suggests excitatory activity.

Limitations and final remarks. This dataset is intended to provide researchers with a means to systematically test the classification accuracy of speech decoding models against naturalistic speech stimuli of increasing complexity, within and across datasets that manipulate the cortical state of participants. To our knowledge, this is the first EEG dataset for neural speech decoding that (i) augments neural activity by means of neuromodulation and (ii) provides stimulus categories constructed in accordance with principles of phoneme articulation and coarticulation. Nonetheless, several limitations of the dataset can be noted.

First of all, the experimental task involves aspects of comprehension, production, and motor activity (in the form of a button-press response), which may be subject to some overlap in the neural signal. In particular, in single phoneme trials, speech may have been produced while potentials relevant to comprehension of the speech sound were still ongoing. However, it is well known that the neural networks underlying motor and language

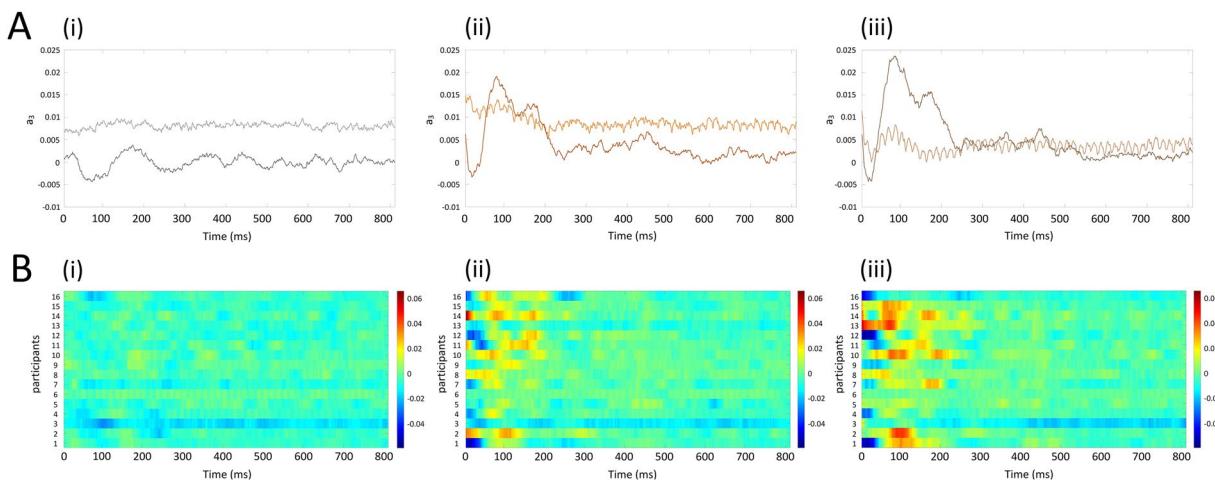


Fig. 10 DDA coefficient a_3 from Eq. (1) from trial onset. **(A)** The grand mean average DDA coefficient to (i) control stimuli, (ii) stimuli with LipM1 stimulation, and (iii) stimuli with TongueM1 stimulation are displayed. In each plot, the lighter line represents the 2019 dataset, and the darker line represents the 2021 dataset. **(B)** The heatmaps for all individual participants are shown for the 2021 dataset.

functions are not strictly dissociable. They are frequently coactivated, even in covert speech or comprehension paradigms^{48,49}. Therefore, we believe this phenomenon underlies most if not all speech decoding paradigms, to a greater or lesser degree, and likely represents neural processing in a naturalistic context.

Secondly, inner speech is widely adopted in the speech decoding literature, where it is often considered to be the most intuitive way of controlling a BCI device. However, inner speech decoding paradigms may not accurately mark the onset of individual stimuli or component phonemes in the recorded data. We believe that prior to transitioning to an inner speech paradigm, researchers would benefit from developing models that can target specific features of the speech stream, such as articulatory features and coarticulation. This kind of systematic study of the speech input may lead to more robust models overall and a better understanding of how this process occurs, rather “black box” models that must be trained on huge amounts of data or rely more heavily on predictive language models than on actual decoding accuracy. Both of these trends in the literature address issues that are orthogonal to the improvement of the actual decoding model.

Questions may also arise as to how neuromodulation may be integrated into a BCI device. We believe that greater attention is needed to study the possibilities for applying neuromodulation during speech decoding. At this time, there is no viable BCI device for evoked paradigms that can be used by the target user population, even with an inner speech paradigm, and therefore any talk of a full-functional device that is ready for end users would be premature. As the study of this procedure continues, we may find a means to utilize neuromodulation for model or participant training, and new possibilities for fast and accurate neuromodulation techniques that could be integrated into a headset continue to be developed.

Finally, we have noted that the two datasets differ in size and quality. We recommend that the larger dataset be used for model development and that the smaller dataset be used to validate the model in more stringent conditions. Our own tests have shown that, by means of DDA, we can successfully perform a classification analysis with both data sets⁵⁴.

Usage Notes

The raw .cnt EEG files can be read in MATLAB with the FieldTrip Toolbox⁹³ and in the Brainstorm⁹⁴ eepv4_read.m function, or in Python with the libeep library. The pre-processed files can be read in MATLAB with EEGLab⁷⁰, FieldTrip⁹³, or in Python with MNE⁹⁵. The already published data corresponding to the experimental pilot study may also be found at OSF (<https://doi.org/10.17605/OSF.IO/E82P9>). This repository⁹⁶ duplicates just one subset of the larger dataset: all data, folders, and code pertaining to CV and VC pair stimuli collected in 2019 and 2021.

Code availability

The data and codes used in this work are publicly available to allow for the reproducibility and sharing of information under the CC-BY 4.0 license. The entire dataset⁶⁹ can be found at OpenNeuro (<https://doi.org/10.18112/openneuro.ds006104.v1.0.0>) in BIDS format. The entire dataset⁶⁸ with the routines and code can be found at OSF (<https://doi.org/10.17605/OSF.IO/3ZFKU>) in .cnt file format. The code⁷¹ is also available on GitHub (<https://doi.org/10.5281/zenodo.15281494>) to allow for version control and discussion of the implementation and analysis carried out in this work.

Received: 22 November 2022; Accepted: 12 May 2025;

Published online: 17 June 2025

References

- Mudgal, S. K., Sharma, S. K., Chaturvedi, J. & Sharma, A. Brain computer interface advancement in neurosciences: Applications and issues. *Interdisciplinary Neurosurgery* **20**, 100694, <https://doi.org/10.1016/j.inat.2020.100694> (2020).
- Ramsey, N. *et al.* Decoding spoken phonemes from sensorimotor cortex with high-density ECoG grids. *NeuroImage* **180**, 301–311, <https://doi.org/10.1016/j.neuroimage.2017.10.011> (2018). New advances in encoding and decoding of brain signals.
- Pasley, B. N. *et al.* Reconstructing Speech from Human Auditory Cortex. *PLoS Biology* **10**, e1001251, <https://doi.org/10.1371/journal.pbio.1001251> (2012).
- Buzsáki, G., Anastassiou, C. A. & Koch, C. The origin of extracellular fields and currents — EEG, ECoG, LFP and spikes. *Nature Reviews Neuroscience* **13**, 407–420, <https://doi.org/10.1038/nrn3241> (2012).
- Gao, P., Huang, Y., He, F. & Qi, H. Improve P300-speller performance by online tuning stimulus onset asynchrony (soa). *Journal of Neural Engineering* **18**, 056067, <https://doi.org/10.1088/1741-2552/ac2f04> (2021).
- Gao, X., Wang, Y., Chen, X. & Gao, S. Interface, interaction, and intelligence in generalized brain–computer interfaces. *Trends in cognitive sciences* **25**, 671–684, <https://doi.org/10.1016/j.tics.2021.04.003> (2021).
- Xiao, X. *et al.* Enhancement for P300-speller classification using multi-window discriminative canonical pattern matching. *Journal of neural engineering* **18**, 046079, <https://doi.org/10.1088/1741-2552/ac028b> (2021).
- Xu, L., Xu, M., Jung, T.-P. & Ming, D. Review of brain encoding and decoding mechanisms for EEG-based brain–computer interface. *Cognitive Neurodynamics* **15**, 569–584, <https://doi.org/10.1007/s11571-021-09676-z> (2021).
- Huggins, J. E., Wren, P. A. & Gruis, K. L. What would brain–computer interface users want? opinions and priorities of potential users with amyotrophic lateral sclerosis. *Amyotrophic Lateral Sclerosis* **12**, 318–324, <https://doi.org/10.3109/17482968.2011.572978> (2011).
- Townsend, G. & Platsko, V. Pushing the P300-based brain–computer interface beyond 100 bpm: Extending performance guided constraints into the temporal domain. *Journal of neural engineering* **13**, 026024, <https://doi.org/10.1088/1741-2560/13/2/026024> (2016).
- Scagliola, C. Language models and search algorithms for real-time speech recognition. *International Journal of Man-Machine Studies* **22**, 523–547, <https://doi.org/10.1073/pnas.92.22.9956> (1985).
- Carvalho, V., Mendes, E., Fallah, A., Sejnowski, T., Comstock, L. & Lainscsek, C. Decoding imagined speech with delay differential analysis. *Frontiers In Human Neuroscience* **18**, 1398065 (2024).
- Stanley, K. & Miikkulainen, R. Evolving neural networks through augmenting topologies. *Evolutionary Computation* **10**, 99–127 (2002).
- Rybár, M. & Daly, I. Neural decoding of semantic concepts: A systematic literature review. *Journal Of Neural Engineering* **19**, 021002 (2022).
- Denby, B. *et al.* Silent speech interfaces. *Speech Communication* **52**, 270–287, <https://doi.org/10.1016/j.specom.2009.08.002> (2010). Silent Speech Interfaces.
- Gonzalez-Lopez, J. A., Gomez-Alanis, A., Martín Doñas, J. M., Pérez-Córdoba, J. L. & Gomez, A. M. Silent speech interfaces for speech restoration: A review. *IEEE Access* **8**, 177995–178021, <https://doi.org/10.1109/ACCESS.2020.3026579> (2020).
- Panachakel, J. T. & Ramakrishnan, A. G. Decoding covert speech from eeg—a comprehensive review. *Frontiers in Neuroscience* **392**, <https://doi.org/10.3389/fnins.2021.642251> (2021).
- Lopez-Bernal, D., Balderas, D., Ponce, P. & Molina, A. A state-of-the-art review of eeg-based imagined speech decoding. *Frontiers in Human Neuroscience* **16**, <https://doi.org/10.3389/fnhum.2022.867281> (2022).
- Zhao, S. & Rudzicz, F. Classifying phonological categories in imagined and articulated speech. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 992–996, <https://doi.org/10.1109/ICASSP.2015.7178118> (IEEE, 2015).
- Pressel Coretto, G. A., Gareis, I. E. & Rufiner, H. L. Open access database of EEG signals recorded during imagined speech. *12th International Symposium on Medical Information Processing and Analysis* **10160**, 1016002, <https://doi.org/10.1117/12.2255697> (2017).
- DaSalla, C. S., Kambara, H., Sato, M. & Koike, Y. Single-trial classification of vowel speech imagery using common spatial patterns. *Neural Networks* **22**, 1334–1339, <https://doi.org/10.1016/j.neunet.2009.05.008> (2009).
- Saha, P., Fels, S., & Abdul-Mageed, M. Deep learning the EEG manifold for phonological categorization from active thoughts. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2762–2766, <https://doi.org/10.1109/ICASSP.2019.8682330> (IEEE, 2019).
- Nguyen, C. H., Karavas, G. K. & Artemiadis, P. Inferring imagined speech using EEG signals: A new approach using Riemannian manifold features. *Journal of Neural Engineering* **15**, <https://doi.org/10.1088/1741-2552/aa8235> (2018).
- Nieto, N., Peterson, V., Rufiner, H. L., Kamienkowski, J. E. & Spies, R. Thinking out loud, an open-access EEG-based BCI dataset for inner speech recognition. *Scientific Data* **9**, 1–17, <https://doi.org/10.1038/s41597-022-01147-2> (2022).
- Suppes, P., Lu, Z.-L. & Han, B. Brain wave recognition of words. *Proceedings of the National Academy of Sciences* **94**, 14965–14969, <https://doi.org/10.1073/pnas.94.26.14965> (1997).
- Suppes, P., Han, B. & Lu, Z.-L. Brain-wave recognition of sentences. *Proceedings of the National Academy of Sciences* **95**, 15861–15866, <https://doi.org/10.1073/pnas.95.26.15861> (1998).
- Wellington, S. & Clayton, J. Fourteen-channel eeg with imagined speech (feis) dataset. *University of Edinburgh* <https://doi.org/10.5281/zenodo.3554128> (2019).
- Glaser, J. I. *et al.* Machine learning for neural decoding. *Eneuro* **7**, <https://doi.org/10.1523/ENEURO.0506-19.2020> (2020).
- Prechelt, L. *et al.* Proben1: A set of neural network benchmark problems and benchmarking rules (1994).
- Vallabhaneni, A., Wang, T. & He, B. Brain—computer interface. In *Neural engineering*, 85–121 (Springer, 2005).
- Chartier, J., Anumanchipalli, G. K., Johnson, K. & Chang, E. F. Encoding of articulatory kinematic trajectories in human speech sensorimotor cortex. *Neuron* **98**, 1042–1054, <https://doi.org/10.1016/j.neuron.2018.04.031> (2018).
- Correia, J. M., Jansma, B. M. & Bonte, M. Decoding articulatory features from fmri responses in dorsal speech regions. *Journal of Neuroscience* **35**, 15015–15025, <https://doi.org/10.1523/JNEUROSCI.0977-15.2015> (2015).
- Comstock, L. *et al.* Developing a real-time translator from neural signals to text: An articulatory phonetics approach. *Proceedings of the Society for Computation in Linguistics* **2**, 322–325, <https://doi.org/10.7275/z2k5-r779> (2019).
- Lakretz, Y., Ossmy, O., Friedmann, N., Mukamel, R. & Fried, I. Single-cell activity in human stg during perception of phonemes is organized according to manner of articulation. *NeuroImage* **226**, 117499, <https://doi.org/10.1016/j.neuroimage.2020.117499> (2021).
- Zhang, C., Liu, Y. & Lee, C.-H. Detection-based accented speech recognition using articulatory features. In *2011 IEEE Workshop on Automatic Speech Recognition & Understanding*, 500–505, <https://doi.org/10.1109/ASRU.2011.6163982> (IEEE, 2011).
- Ying, X. An overview of overfitting and its solutions. In *Journal of physics: Conference series*, vol. 1168, 022022, <https://doi.org/10.1088/1742-6596/1168/2/022022> (IOP Publishing, 2019).
- Davis, M. H. & Johnsrude, I. S. Hearing speech sounds: top-down influences on the interface between audition and speech perception. *Hearing research* **229**, 132–147, <https://doi.org/10.1016/j.heares.2007.01.014> (2007).
- Ladd, D. R. “distinctive phones” in surface representation. *Laboratory Phonology* **8**, 3–26 (2006).
- Ostry, D. J., Gribble, P. L. & Gracco, V. L. Coarticulation of jaw movements in speech production: is context sensitivity in speech kinematics centrally planned? *Journal of Neuroscience* **16**, 1570–1579, <https://doi.org/10.1523/JNEUROSCI.16-04-01570.1996> (1996).
- Bouchard, K. E. & Chang, E. F. Control of spoken vowel acoustics and the influence of phonetic context in human speech sensorimotor cortex. *Journal of Neuroscience* **34**, 12662–12677, <https://doi.org/10.1523/JNEUROSCI.1219-14.2014> (2014).

41. Liberman, A. M., Harris, K. S., Hoffman, H. S. & Griffith, B. C. The discrimination of speech sounds within and across phoneme boundaries. *Journal of experimental psychology* **54**, 358, <https://doi.org/10.1037/h0044417> (1957).
42. Anumanchipalli, G. K., Chartier, J. & Chang, E. F. Speech synthesis from neural decoding of spoken sentences. *Nature* **568**, 493–498, <https://doi.org/10.1038/s41586-019-1119-1> (2019).
43. Moses, D. A. *et al.* Neuroprosthesis for decoding speech in a paralyzed person with anarthria. *New England Journal of Medicine* **385**, 217–227, <https://doi.org/10.1056/NEJMoa2027540> (2021).
44. Armeni, K., Willems, R. M. & Frank, S. L. Probabilistic language models in cognitive neuroscience: Promises and pitfalls. *Neuroscience & Biobehavioral Reviews* **83**, 579–588, <https://doi.org/10.1016/j.neubiorev.2017.09.001> (2017).
45. Koskinen, M., Kurimo, M., Gross, J., Hyvärinen, A. & Hari, R. Brain activity reflects the predictability of word sequences in listened continuous speech. *NeuroImage* **219**, 116936, <https://doi.org/10.1016/j.neuroimage.2020.116936> (2020).
46. Moses, D. A., Mesgarani, N., Leonard, M. K. & Chang, E. F. Neural speech recognition: continuous phoneme decoding using spatiotemporal representations of human cortical activity. *Journal of neural engineering* **13**, 056004, <https://doi.org/10.1088/1741-2560/13/5/056004> (2016).
47. Orhan, U. *et al.* Improved accuracy using recursive bayesian estimation based language model fusion in erp-based bci typing systems. In *2012 annual international conference of the ieee engineering in medicine and biology society*, 2497–2500, <https://doi.org/10.1109/EMBC.2012.6346471> (IEEE, 2012).
48. Wilson, S. M., Saygin, A. P., Sereno, M. I. & Iacoboni, M. Listening to speech activates motor areas involved in speech production. *Nature neuroscience* **7**, 701–702, <https://doi.org/10.1016/j.neuroimage.2006.05.032> (2004).
49. Wilson, S. M. & Iacoboni, M. Neural responses to non-native phonemes varying in producibility: evidence for the sensorimotor nature of speech perception. *NeuroImage* **33**, 316–325, <https://doi.org/10.1016/j.neuroimage.2006.05.032> (2006).
50. Roy, A. C., Craighero, L., Fabbri-Destro, M. & Fadiga, L. Phonological and lexical motor facilitation during speech listening: a transcranial magnetic stimulation study. *Journal of Physiology-Paris* **102**, 101–105, <https://doi.org/10.1016/j.jphysparis.2008.03.006> (2008).
51. Hawkins, D. M. The problem of overfitting. *Journal of chemical information and computer sciences* **44**, 1–12, <https://doi.org/10.1021/ci0342472> (2004).
52. Ho, S. Y., Phua, K., Wong, L. & Goh, W. W. B. Extensions of the external validation for checking learned model interpretability and generalizability. *Patterns* **1**, 100129, <https://doi.org/10.1016/j.patter.2020.100129> (2020).
53. Comstock, L., Carvalho, V., Lainscsek, C., Fallah, A. & Sejnowski, T. Transcranial Magnetic Stimulation Facilitates Neural Speech Decoding. *Brain Sciences* **14**, 895 (2024).
54. Comstock, L. *et al.* Exploratory methods for high-performance eeg speech decoding. *bioRxiv* <https://doi.org/10.1101/2021.11.16.468876> (2021).
55. Eyherabide, H. G. & Samengo, I. When and why noise correlations are important in neural decoding. *Journal of Neuroscience* **33**, 17921–17936, <https://doi.org/10.1523/JNEUROSCI.0357-13.2013> (2013).
56. MacKay, D. G. Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia* **8**, 323–350, [https://doi.org/10.1016/0028-3932\(70\)90078-3](https://doi.org/10.1016/0028-3932(70)90078-3) (1970).
57. O'Sullivan, J. *et al.* Neural decoding of attentional selection in multi-speaker environments without access to clean sources. *Journal of neural engineering* **14**, 056001, <https://doi.org/10.1088/1741-2552/aa7ab4> (2017).
58. D'Ausilio, A. *et al.* The motor somatotopy of speech perception. *Current Biology* **19**, 381–385, <https://doi.org/10.1016/j.cub.2009.01.017> (2009).
59. Peirce, J. *et al.* PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods* **51**, 195–203, <https://doi.org/10.3758/s13428-018-01193-y> (2019).
60. Herbsman, T. *et al.* Motor threshold in transcranial magnetic stimulation: The impact of white matter fiber orientation and skull-to-cortex distance. *Human brain mapping* **30**, 2044–2055, <https://doi.org/10.1002/hbm.20649> (2009).
61. Day, B. *et al.* Motor cortex stimulation in intact man: 2. multiple descending volleys. *Brain* **110**, 1191–1209, <https://doi.org/10.1093/brain/110.5.1191> (1987).
62. Day, B. *et al.* Electric and magnetic stimulation of human motor cortex: surface emg and single motor unit responses. *The Journal of physiology* **412**, 449–473, <https://doi.org/10.1113/jphysiol.1989.sp017626> (1989).
63. Di Lazzaro, V. *et al.* The physiological basis of transcranial motor cortex stimulation in conscious humans. *Clinical neurophysiology* **115**, 255–266, <https://doi.org/10.1016/j.clinph.2003.10.009> (2004).
64. Pulvermüller, F. *et al.* Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences* **103**, 7865–7870, <https://doi.org/10.1073/pnas.0509989103> (2006).
65. Ojemann, G. A. Individual variability in cortical localization of language. *Journal of neurosurgery* **50**, 164–169, <https://doi.org/10.3171/jns.1979.50.2.0164> (1979).
66. Pisoni, A. *et al.* Cognitive enhancement induced by anodal tdcS drives circuit-specific cortical plasticity. *Cerebral Cortex* **28**, 1132–1140, <https://doi.org/10.1093/cercor/bhx021> (2018).
67. Wassermann, E. M. *et al.* Topography of the inhibitory and excitatory responses to transcranial magnetic stimulation in a hand muscle. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section* **89**, 424–433, [https://doi.org/10.1016/0168-5597\(93\)90116-7](https://doi.org/10.1016/0168-5597(93)90116-7) (1993).
68. Comstock, L., Lainscsek, C., Moreira, J. & Carvalho, V. Speech decoding dataset. *Open Science Framework* <https://doi.org/10.17605/OSFIO/3ZFKU> (2025).
69. Moreira, J., Carvalho, V., Mendes, E., Fallah, A., Sejnowski, T., Lainscsek, C. & Comstock, L. EEG dataset for speech decoding. *OpenNeuro* <https://doi.org/10.18112/openneuro.ds006104.v1.0.0> (2025).
70. Delorme, A. & Makeig, S. Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of Neuroscience Methods* **134**, 9–21, <https://doi.org/10.1016/j.jneumeth.2003.10.009> (2004).
71. Moreira, J. Speech decoding. *Github* <https://doi.org/10.5281/zenodo.15281494> (2025).
72. Bell, A. J. & Sejnowski, T. J. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation* **7**, 1004–1034, <https://doi.org/10.1162/neco.1995.7.6.1129> (1995).
73. Albera, L. *et al.* Ica-based eeg denoising: a comparative analysis of fifteen methods. *Bulletin of the Polish Academy of Sciences: Technical Sciences* **60**, 407–418, <https://doi.org/10.2478/v10175-012-0052-3> (2012).
74. Davies, M. Identifiability issues in noisy ica. *IEEE Signal processing letters* **11**, 470–473, <https://doi.org/10.1109/LSP.2004.826508> (2004).
75. Bertazzoli, G. *et al.* The impact of artifact removal approaches on tms-eeg signal. *NeuroImage* **239**, 118272, <https://doi.org/10.1016/j.neuroimage.2021.118272> (2021).
76. Hernandez-Pavon, J. C., Kugiumtzis, D., Zrenner, C., Kimiskidis, V. K. & Metsomaa, J. Removing artifacts from tms-evoked eeg: A methods review and a unifying theoretical framework. *Journal of Neuroscience Methods* **109591**, <https://doi.org/10.1016/j.jneumeth.2022.109591> (2022).
77. Rogasch, N. C., Biabani, M. & Mutanen, T. P. Designing and comparing cleaning pipelines for tms-eeg data: a theoretical overview and practical example. *Journal of Neuroscience Methods* **109494**, <https://doi.org/10.1016/j.jneumeth.2022.109494> (2022).
78. Lainscsek, C. & Sejnowski, T. J. Delay differential analysis of time series. *Neural computation* **27**, 594–614, https://doi.org/10.1162/NECO_a_00706 (2015).

79. Lainscsek, C., Hernandez, M. E., Poizner, H. & Sejnowski, T. J. Delay differential analysis of electroencephalographic data. *Neural computation* **27**, 615–627, https://doi.org/10.1162/NECO_a_00656 (2015).
80. Gonzalez, C. E., Lainscsek, C., Sejnowski, T. J. & Letellier, C. Assessing observability of chaotic systems using delay differential analysis. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **30**, 103113, <https://doi.org/10.1063/5.0015533> (2020).
81. Lainscsek, C., Weyhenmeyer, J., Hernandez, M. E., Poizner, H. & Sejnowski, T. J. Non-linear dynamical classification of short time series of the rössler system in high noise regimes. *Frontiers in Neurology* **4**, 182, <https://doi.org/10.3389/fneur.2013.00182> (2013).
82. Lainscsek, C., Hernandez, M. E., Weyhenmeyer, J., Sejnowski, T. J. & Poizner, H. Non-linear dynamical analysis of EEG time series distinguishes patients with Parkinson's disease from healthy individuals. *Frontiers in Neurology* **4**, <https://doi.org/10.3389/fneur.2013.00200> (2013).
83. Digeser, F. M., Wohlberedt, T. & Hoppe, U. Contribution of spectrotemporal features on auditory event-related potentials elicited by consonant-vowel syllables. *Ear and Hearing* **30**, 704–712, <https://doi.org/10.1097/AUD.0b013e3181b1d42d> (2009).
84. Khalighinejad, B., da Silva, G. C. & Mesgarani, N. Dynamic encoding of acoustic features in neural responses to continuous speech. *Journal of Neuroscience* **37**, 2176–2185, <https://doi.org/10.1523/JNEUROSCI.2383-16.2017> (2017).
85. Rogasch, N. C. & Fitzgerald, P. B. Assessing cortical network properties using tms-eeg. *Human brain mapping* **34**, 1652–1669, <https://doi.org/10.1002/hbm.22016> (2013).
86. Macabbee, P. et al. Influence of pulse sequence, polarity and amplitude on magnetic stimulation of human and porcine peripheral nerve. *The journal of physiology* **513**, 571–585, <https://doi.org/10.1111/j.1469-7793.1998.571bb.x> (1998).
87. Nagarajan, S. S., Durand, D. M. & Warman, E. N. Effects of induced electric fields on finite neuronal structures: a simulation study. *IEEE Transactions on Biomedical Engineering* **40**, 1175–1188, <https://doi.org/10.1109/10.245636> (1993).
88. Nagarajan, S. S. & Durand, D. M. A generalized cable equation for magnetic stimulation of axons. *IEEE Transactions on Biomedical Engineering* **43**, 304–312, <https://doi.org/10.1109/10.486288> (1996).
89. Nagarajan, S. S., Durand, D. M. & Hsuing-Hsu, K. Mapping location of excitation during magnetic stimulation: Effects of coil position. *Annals of biomedical engineering* **25**, 112–125 (1997).
90. Pell, G. S., Roth, Y. & Zangen, A. Modulation of cortical excitability induced by repetitive transcranial magnetic stimulation: influence of timing and geometrical parameters and underlying mechanisms. *Progress in neurobiology* **93**, 59–98, <https://doi.org/10.1016/j.pneurobio.2010.10.003> (2011).
91. Silva, S., Bassar, P. & Miranda, P. Elucidating the mechanisms and loci of neuronal excitation by transcranial magnetic stimulation using a finite element model of a cortical sulcus. *Clinical neurophysiology* **119**, 2405–2413, <https://doi.org/10.1016/j.clinph.2008.07.248> (2008).
92. Lainscsek, C. et al. Nonlinear dynamics underlying sensory processing dysfunction in schizophrenia. *Proceedings of the National Academy of Sciences* **116**, 3847–3852, <https://doi.org/10.1073/pnas.1810572116> (2019).
93. Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J.-M. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational intelligence and neuroscience* **2011**, 156869, <https://doi.org/10.1155/2011/156869> (2011).
94. Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D. & Leahy, R. M. Brainstorm: A user-friendly application for MEG/EEG analysis. *Computational Intelligence and Neuroscience* **2011**, 1–13, <https://doi.org/10.1155/2011/879716> (2011).
95. Gramfort, A. et al. MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience* **7**, 1–13, <https://doi.org/10.3389/fnins.2013.00267> (2013).
96. Comstock, L., Carvalho, V. R., Lainscsek, C., Mendes, E. and Moreira, J. P. C. TMS and speech decoding. *Open Science Framework* <https://doi.org/10.17605/OSFIO/E82P9> (2025).

Acknowledgements

This research was funded by the U.S. Russia Foundation through award No. 20-AUG-19-UCLA. The article is an output of a research project implemented as part of the Basic Research Program at the National Research University Higher School of Economics (HSE University). We would like to thank Tyler Wishard, Bela Syed, Sophia Mourad, and Panagiota Loizidou for their assistance in collecting the data. Figure 3a is adapted with permission from Wolters Kluwer Health, Inc.: Digeser FM, Wohlberedt T, Hoppe U. Contribution of spectrotemporal features on auditory event-related potentials elicited by consonant-vowel syllables. *Ear and Hearing*. 2009 Dec 1;30(6):704-12. <https://doi.org/10.1097/AUD.0b013e3181b1d42d>. The Creative Commons license does not apply to this content. Use of the material in any format is prohibited without written permission from the publisher, Wolters Kluwer Health, Inc. Please contact permissions@lww.com for further information. Figure 3b is adapted from Khalighinejad B, da Silva GC, Mesgarani N. Dynamic encoding of acoustic features in neural responses to continuous speech. *Journal of Neuroscience*. 2017 Feb 22;37(8):2176-85. <https://doi.org/10.1523/JNEUROSCI.2383-16.2017>. Figure 3c is adapted with permission from John Wiley & Sons, Inc.: Rogasch NC, Fitzgerald PB. Assessing cortical network properties using TMS-EEG. *Human brain mapping*. 2013 Jul;34(7):1652-69. <https://doi.org/10.1002/hbm.22016>.

Author contributions

L.C. conceived the experiments, ran the experiments, created figures, and wrote the manuscript. J.P.C.M. developed codes, analyzed the results (ICA), created figures, and wrote the manuscript. V.R.C developed codes, analyzed the results, and wrote the manuscript. C.L. analyzed the results (DDA), created figures, provided technical feedback, and reviewed the manuscript. E.M.A.M.M., A.F., and T.J.S provided technical feedback and reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to L.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025