Bimodal pilot study on inner speech decoding reveals the potential of combining EEG and fMRI

Foteini Simistira Liwicki[1*], Vibha Gupta[1], Rajkumar Saini[1†], Kanjar De[1†], Nosheen Abid[1†], Sumit Rakesh[1†], Scott Wellington[2†], Holly Wilson[2†], Marcus Liwicki[1] and Johan Eriksson[3]

[1*]Department of Computer Science, Luleå University of Technology, Sweden.
[2]Department of Computer Science, University of Bath, United Kingdom.
[3]Department of Integrative Medical Biology (IMB), Umeå University, Sweden.

*Corresponding author(s). E-mail(s): foteini.liwicki@ltu.se; Contributing authors: vibha.gupta@ltu.se; rajkumar.saini@ltu.se; kanjar.de@ltu.se; nosheen.abid@ltu.se; sumit.rakesh@ltu.se; sdlw20@bath.ac.uk; H.L.Wilson@bath.ac.uk; marcus.liwicki@ltu.se; johan.eriksson@umu.se;
†These authors contributed equally to this work.

Abstract

This paper presents the first publicly available bimodal electroencephalography (EEG) / functional magnetic resonance imaging (fMRI) dataset and an open-source benchmark for inner speech decoding. Decoding inner speech – or thought (expressed through a voice without actual speaking) – is a challenge with typical results close to chance level. The dataset comprises of 1'280 trials (4 subjects, 8 stimuli = 2 categories * 4 words, and 40 trials per stimuli) in each modality. The pilot study reports for the binary classification, a mean accuracy of 71.72% when combining the two modalities (EEG and fMRI), compared to 62.81% and 56.17% when using EEG, resp. fMRI alone, following a subject dependent

approach. The same improvement in performance for word classification (8 classes) can be observed (30.29% with combination, 22.19% and 17.50% without). As such, this paper demonstrates that combining EEG with fMRI is a promising direction for inner speech decoding.

Keywords: brain–computer interface (BCI); inner speech; electroencephalography (EEG); functional magnetic resonance imaging (fMRI); bimodal dataset

1 Introduction

Thought is strongly related to inner speech, through a "voice inside the head" without actual speaking [1, 2]. Inner speech decoding from brain activity is a burgeoning research area, with applications in brain-computer interface (BCI) paradigms such as speech prosthesis, clinical contexts - for example informing models of psychiatric disorders in which inner-speech is disturbed (e.g schizophrenia), and neuroscience, by deepening our understanding of the spatiotemporal neural dynamics of inner speech.

There are several technologies available for measuring brain activity, using invasive (e.g., electrocorticography (ECoG)) and non-invasive methods (e.g., electroencephalography (EEG), functional magnetic resonance imaging (fMRI), functional near-infrared spectroscopy (fNIRS), magnetoencephalography (MEG)).

EEG and fMRI, are non-invasive methods, although with drawbacks when used alone, in terms of signal quality and resolution. EEG has a high temporal resolution, while fMRI has a high spatial resolution that allows for a deeper look into the subcortical structures of the brain. A combination of EEG and fMRI is beneficial for brain analysis, as it combines both the high temporal resolution of EEG and the high spatial resolution of fMRI.

Recent studies have shown the potential of combining different modalities for brain analysis. Cooney et al. [3] proposed a bimodal deep learning architecture for imagined speech recognition using EEG (64-channel) and fNIRS (8-channel). The simultaneous data from 19 subjects (age: 26.63±2.13) was collected for overt and inner speech. The EEG and fNIRS signals were preprocessed and used to train CNN-GRU based deep learning architecture to predict the overt and inner speech. The combined modality performed better than the individual unimodal architectures. The bimodal performance as 49.61% (46.04%/32.46%), and 31.78% (30.25%/30.62%) were recorded for overt and inner speech, respectively (chance: 25%)). Berezutskaya et al. [4] developed a publicly available multimodal dataset consisting of ECoG and fMRI naturalistic simulation with a short audiovisual film. The dataset contains ECoG data of 51 subjects (between 5-55 years) and fMRI data of 30 subjects (between 7-47 years) for the same task. Scrivener [5] provided a set of questions to understand whether simultaneous EEG-fMRI are necessary or not. Mano et al. [6] suggested how a Hybrid Neurofeedback Platform combining EEG and fMRI can be built. Dong et al. [7] developed a toolbox to study EEG-fMRI multimodal fusion. The tool offers fMRI analysis, EEG analysis, and their fusion. The tool is developed in matlab and is an open source. Lioi et al. [8] proposed a bimodal motor imagery neurofeedback based dataset aquired with simultaneous EEG-fMRI recordings. The data were recorded from 30 subjects performed kinaesthetic motor imagery of the right hand to bring the ball to the target. Zotev et al. [9] analyzes the human brain activity (happy, rest) with EEG and fMRI using high-beta band. EEG-neurofeedback was designed to reduce the effect of artifacts in EEG-fMRI. The neurofeedback study done

by Perronnet et al. [10] suggests that the hemodynamic and electrophsyiological activities are higher when combining EEG-fMRI compared to EEG and fMRI alone.

This pilot study shows that a combination of EEG and fMRI is beneficial for inner-speech decoding, as it combines both high temporal and high spatial resolution. To the best of our knowledge, this is the first study that combines EEG and fMRI for inner speech detection. We describe the acquisition of a bimodal dataset of EEG and fMRI and provide baseline classification results using machine-learning methods.

2 Methods

The study was approved by the Swedish Ethical Review Authority (Etikprövning myndigheden, ID:2021-06710-01)[1] in accordance with the Swedish Ethical Review Act (SFS 2003:460). All subjects provided a written consent form for their participation in the study.

2.1 Dataset

The chosen dataset was selected to decode inner speech using EEG and fMRI recordings. Two categories with 4 words each were selected following the study in [11]. Category *social* contains the words *child, daughter, father, wife* and category *numeric* the words *four, three, ten, six*. There are in total 2´200 fMRI scans per subject, divided into 2 sessions, that generates a volume size of 100x100x68 per scan. The EEG recordings per subject gave in total 320 x64 x 1024 data samples.

2.2 Participants

In total five healthy right-handed subjects within the age group of 33-52, participated in this study. Three of them were female and the other two male. None of them is native English speaker. All subjects were asked to sign a written consent form and followed the same experimental protocol. In this study, we refer to the data of the different subjects with the following name convention: *Subj01, Subj02, Subj03, Subj04, Subj05*. During the EEG acquisition the data of Subj04 were excluded from the study due to high fluctuations during the recording.

2.3 fMRI acquisition

The data were collected using a Siemens Magnetom Prisma with a 20 channel head coil. Anatomical images were acquired using a sagittal T1 3D MPRAGE sequence with the following parameters: TR = 2,3 ms; TE = 2.98 ms; TI = 900 ms; flip angle = 9○; 208 slices; matrix size 256×256; voxel size 1×1×1 mm. After the anatomical scan, two field maps were obtained (A, B) with the following parameters: TR=6,62ms, TE: A=4,92ms, B=7,38ms; voxel size 3x3x2 mm.

After the field maps, functional maps were obtained using double-echo gradient echo imaging BOLD sequences parallel to the bi-commissural plane with the following parameters: TR = 2,16 ms; TE = 30 ms; 68 slices; matrix size 100×100; voxel size 2×2×2 mm.

The visual stimuli were presented from the stimuli computer to the participant via an Ultra HD LCD display from NordicNeuroLab [2]. The screen has a width/height of 88/48 cm (3840/2160 pixels when on full resolution) and

[1] https://etikprovningsmyndigheten.se/
[2] https://www.nordicneurolab.com/

a built-in camera (640 x 480, 30 fps) that can be accessed from the stimuli computer.

2.4 EEG acquisition

The EEG data were acquired using the BioSemi Active2[3] measuring system with a 16 bits resolution and a sampling rate of 512 Hz. A BioSemi EEG head cap with 64 electrodes and 6 external sensors was used (see Figure 1). The experiment started by choosing the appropriate cap size for each participant. Measurement tape is used for measuring the participant's head circumference from nasion to inion in order to determine the appropriate cap size. We also made sure that the cap was properly centered; the Cz (Vertex) was in the center of the head: halfway between the nasion and inion and halfway between the ears. SIGNAGEL[4] was used for each electrode to reinsure the electrode connectivity with the subject's head. All six EXG electrodes were placed using stickers. The location of the six electrodes are as follows:

- EXG1: On left mastoid behind right ear
- EXG2: On right mastoid behind left ear
- EXG3: 1cm left of the left eye (aligned to center of eye)
- EXG4: 2cm above the left eye (aligned to center of eye)
- EXG5: 2cm below the right eye (aligned to center of eye)
- EXG6: 1cm right of the right eye (aligned to center of eye)

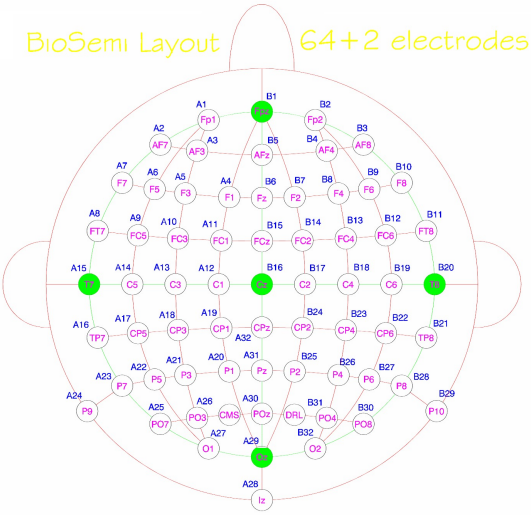

Fig. 1 Location of the electrodes in the BioSemi EEG head cap.

[3] https://www.biosemi.com/products.htm
[4] https://www.parkerlabs.com/products/signagel-electrode-gel/

The recording software was ActiView, which was also developed by BioSemi. ActiView allows to verify the electrode impedance as well as the overall quality of the incoming data. Before starting the recording session, the impedance of each electrode was carefully examined to ensure that it was between -20 µV and 20 µV. The electrodes were adjusted in the case that they were not within these range, by adding/removing some gel, moving hair below the electrode or wiggling the electrode. ActiView screen should appear like the one in Figure 2 when all electrodes are connected correctly. We also dimmed the lights and make sure that the room is not too dark, as a high contrast between the room and the display can cause eye flickering.

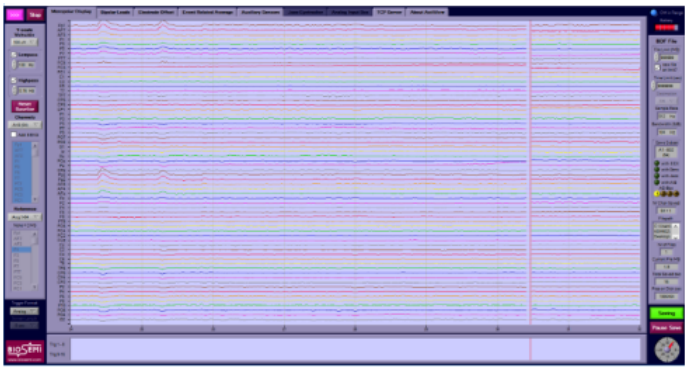

Fig. 2 Stable data while starting recording.

We start the experiment after filling out the date under the Local Recording Information (behavioral or other data from humans), and when the experiment is finished, we stop the button in the upper left corner. The data are saved (.bdf format) under the selected folder. The continuous recording of the 64 EEG channels, the 6 external channels, and the labeled events is preserved in the saved file.

2.5 Experimental protocol

The experimental protocol for both modalities (fMRI and EEG, see Figure 3) was designed using E-Prime 3.0[5] and was executed in a computer (not manual). The fMRI recordings consist of 2 sessions with 20 trials each. In the beginning of each session a textual information regarding the experiment was presented on the screen followed by a fixation period of 2 (seconds) where the participants were instructed to fixate their eyes to the center of the screen. The stimulus were 8 different words belonging to 2 categories (social/numeric) randomly presented to maintain participant attention. Each trial consists of a task and a rest period. The first screen presents the textual representation of the stimulus with a duration of 4 s, while the second the rest state with a duration of

[5] https://pstnet.com/products/e-prime/

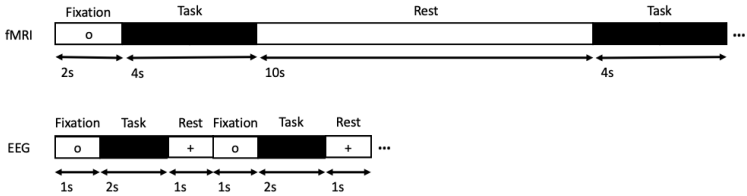

Fig. 3 fMRI and EEG experimental protocol.

10 s. During the task period, the participants were asked to repeat the given stimulus as many times as possible (approximately 4 times) using their inner speech, without any articulation or muscle movement. During the rest period, the participants could relax and prepare for the next trial. The total duration of the recording for the 320 repetitions was 74.6 min per participant.

The EEG recordings consist of 1 session with 40 trials, using the same stimuli as for the fMRI protocol. In the beginning of the session, a textual information regarding the experiment was presented on the screen until the space bar was pressed to indicate the starting point of the experiment. Each trial includes a fixation, task, and rest period. The fixation period with a duration of 1 s allows the participant to fixate their eyes in the middle of the screen where a small circle is located. The task period presents the textual representation of the stimulus with a duration of 2 s, followed by a rest period of 1 s. During the task period, the participants were asked to repeat the given stimulus using their inner speech, without any articulation or muscle movement. During the rest period, the participants could relax and prepare for the next trial. The total duration of the recording for the 320 repetitions was 21.33 min per participant. Figure 3 represents the experimental protocol for each modality. Note that the rest period for the fMRI is higher than for the EEG. This is due to the fact that the neuronal activity of the brain during fMRI typically rises to a peak at about 5s after the stimulus and it takes approximately 14s to recover to the baseline. For EEG it is not required, but due to the short periods, an extra fixation period is added before each task.

2.5.1 fMRI Preprocessing

The fMRI data were preprocessed using SPM12 [6]. First, spatial displacement maps were calculated for each session. These were used for motion correction of the functional data. As they were acquired in an inter-leaved order, we next performed slice time correction. This was followed by co-registration to the T1 structural scan with a normalised mutual information cost function. These images were then used in the classification pipeline. For group-level fixed effect analyses, the origin was set to the anterior commisure, followed by normalisation to the Montreal Neurological Instistute (MNI) space. smoothing using a 8mm FWHM (full width at half maximum) Gaussian kernel was applied. We

[6] https://www.fil.ion.ucl.ac.uk/spm/software/spm12/

estimated a General Linear Model (GLM) convolved with a canonical haemodynamic response function. This was alongside nuisance regressors including movement parameters calculated in the previous alignment step.

2.6 EEG Preprocessing

Preprocessing of the EEG signal, which is virtually a set of signal processing steps preceding main EEG data analyses, is essential to obtain only brain activity from the noisy EEG recordings. Preprocessing of EEG largely includes a number of processes, such as line noise removal, adjustment of referencing, elimination of bad EEG channels, and artifact removal.

EEGLAB [12] is a Matlab toolbox for processing continuous and event-related EEG, MEG, and other electrophysiological data, with features such as Independent Component Analysis (ICA), time/frequency analysis, artifact rejection, event-related statistics, and several useful visualization modes for averaged and single-trial data. In the current work, we have utilized EEGLAB for preprocessing the EEG data. Raw Biosemi EEG data (.bdf format) was imported EEGLAB using reference channel 48 (Cz). Baseline correction was applied using a zero-phase FIR high-pass filter at 0.1 Hz. A lowpass filter at 50 Hz was applied, relieving the requirement for a notch filter. Re-referencing was performed using the average reference to Cz, excluding channels 65-70 (the ocular and mastoid electrodes). Channels were manually inspected: bad channels were rejected, not interpolated. Time-locked epochs were extracted using start-stop limits fixed within the interval [-0.1, 0.1]. Independent component analysis (ICA) was used to zero-out signal components related to EOG artifacts, via manual inspection.

To summarize, the following steps are included in the EEG preprocessing step:

- Importing the raw data: Import the Biosemi raw data file (.bdf). Use refrence channel 48 (Cz) upon import.
- Baseline Correction: A multitude of internal and environmental causes can generate temporal drifts, which change over time and across the electrodes. To reduce the impact of such variances, it is usual practice to perform a so-called baseline correction.
- Bandpass filter: Low-frequency and high-frequency signal, which is commonly caused by environmental/muscle noise in scalp EEG and is not usually the focus of analysis, must be filtered out. Noise below and above a given frequency can be kept in the data using low-pass and high-pass filtering. Here, we apply zero-phase FIR bandpass filtering with 0.1 for lower edge of the frequency band pass and 50 for higher edge of the frequency of pass band.
- Re-reference data: Re-referencing aids data cleaning by providing an estimate of physiological noise baseline activity. We have used $ComputeAverageReference(CRR)$ for re-refrencing. While doing re-referencing, we excluded channel 65-70 (ocular and mastoid electrodes), and add current reference channel Cz (48)back to the data.

- Epoch the data: We extract the epochs choosing $Time - locingeventtypes$. Also put the input the epoch limit [-0.1,0.1].
- Run independent component analysis and reject noisy components: After cleaning Epochs, we run it via Independent Component Analysis. This enables us to discard data segments that appear to be strongly influenced by motor-related artifacts such as blinking, jaw, neck, arm, or upper back movement. Figure 4 and 5 illustrate the two component (out of 64) which are decomposed by ICA algorithm. Here, in this study, we examine the topography along with spectrogram and frequency variation to keep the component.

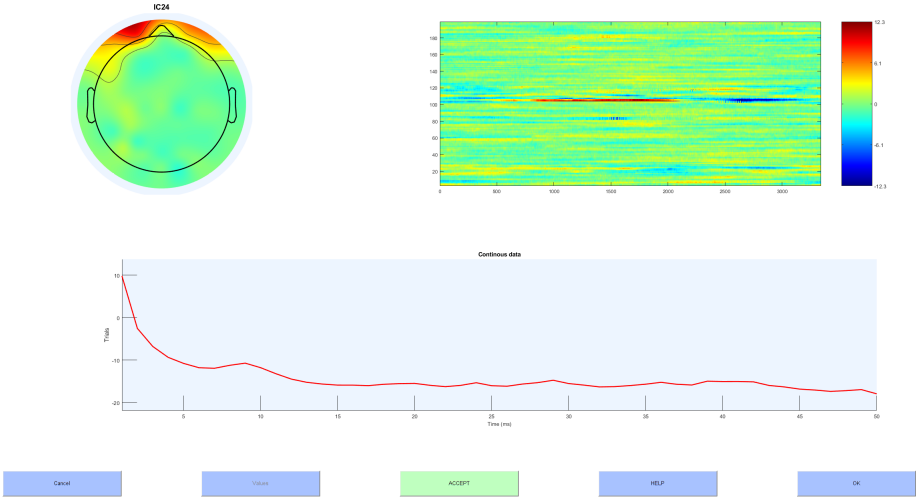

Fig. 4 Decomposed component by ICA, shows eye blinks.

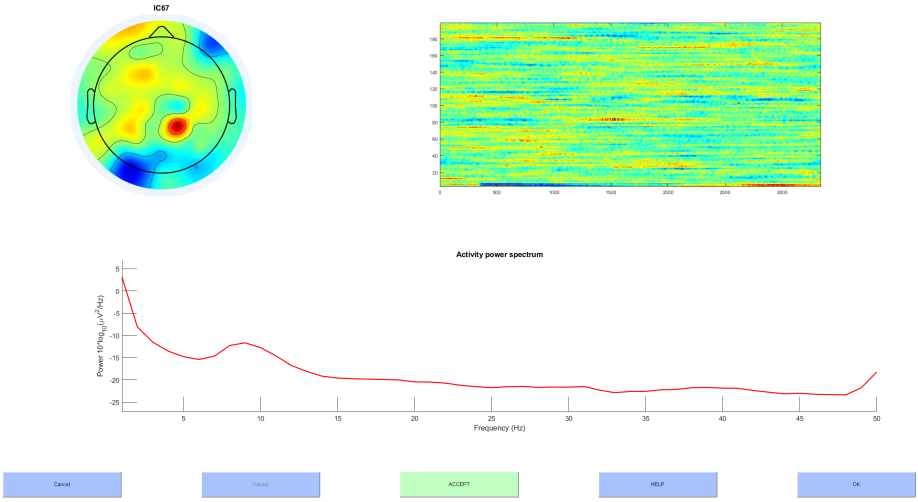

Fig. 5 Decomposed component by ICA, shows brain activity.

3 Results

The current section discuss the results following a subject-dependent approach when using only EEG, only fMRI and fusing EEG and fMRI. The subject-dependent approach is a baseline approach which is commonly followed in almost all EEG and fMRI related paradigm. Knowing that, brain signals change heavily when along different subjects, the current study mainly focuses on evaluating the subject dependent performance. Here, individual models are trained for each subject. In other words, subject-dependent models are trained; data from the same subject are used for the training, validation, and testing set. This method effectively examines how much the data of a single subject changes (or fluctuates) over time. In this study, we also use subject-dependent approach to evaluate the proposed framework. After validation of several standard machine learning methods, the following benchmark was created: A Random Forest (RF) classifier for the EEG and a k-nearest neighbor (Knn) classifier for the fMRI data. To evaluate the performance of the proposed bimodal approach, two classification tasks were employed: (i) a binary classification task for classifying the 2 categories *social* and *numeric* and (ii) a multiclass classification task for classifying the 8 words, see section 2.1. A 5-fold cross validation approach followed for all classification results. The mean accuracy and standard deviation are used to report the classification performance for all tasks.

3.1 EEG

A RF classifier for the EEG data were used (classifier selection based on validation set performance). The features used for the RF classifier were: mean, standard deviation, and root mean square values. The dataset is split into 2 sets, one for training (80%) and one for testing (20%). Results on the RF classifier, are reported for the word classification (8 classes) and for the category classification tasks (2 classes) in Figures 6, 7, respectively.

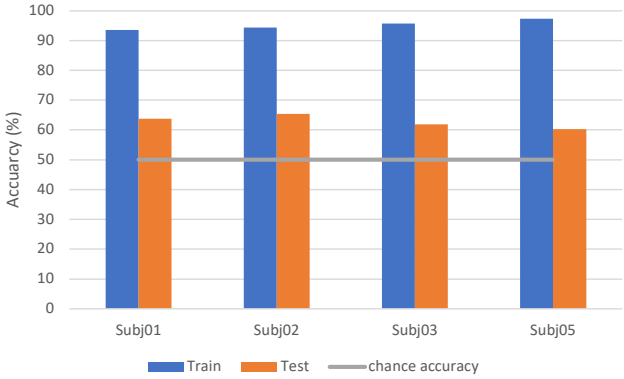

Fig. 6 Binary (category) test and validation classification results using only EEG. Chance accuracy 50%.

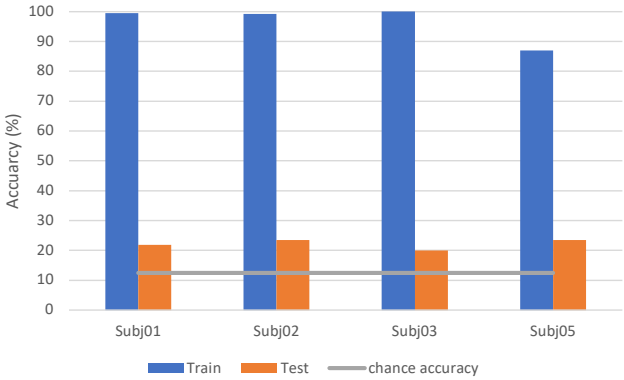

Fig. 7 Multiclass (8 words) test and validation classification results using only EEG. Chance accuracy 12.5%.

Following observations can be made from figure 6 and 7: (i) For all classification tasks; word and category, the individual accuracies (corresponds to each subject) are above the chance level accuracy. This shows that the chosen features along with RF are good enough to capture discriminability among individual subjects. (ii) The low variation across mean accuracies and standard deviations ensures the stability of the proposed framework.

3.2 fMRI

An Knn classifier for the fMRI data were used (classifier selection based on validation set performance). Figures 8 and 9 reports the results for the word classification (8 classes) and for the category classification tasks (2 classes).

Based on figure 8 and 9, following observations can be made: (i) Lower performance than EEG can be noticed for fMRI. However, expect one, all three subjects show the above chance level accuracy for binary classification. For multi-class classification, all subjects perform above the chance level. (ii) Higher variation (as compared to EEG) among the subjects´ performance can be seen here. The reason for such low performance both in terms of accuracy and variation, is likely due to low temporal resolution of fMRI. As observed in the related literature temporal features may play important role while making inner-speech decoding.

Section 3.3 shows that the fusion of both modalities sufficiently improves the performance which is the main aim of the study. This high performance improvement confirms that error made by one modality is captured by the other modality which further ensures the discriminability, acquired by different modality.

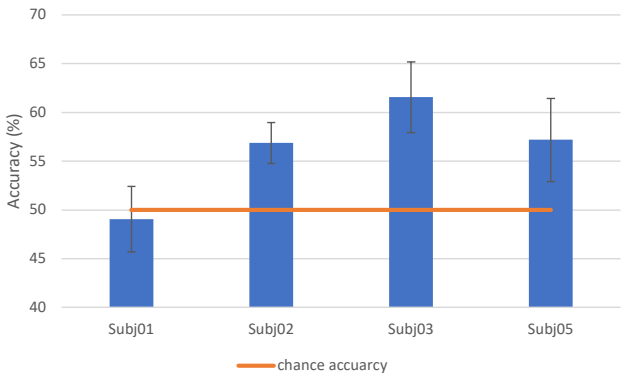

Fig. 8 Binary (category) test classification results using only fMRI. Chance accuracy 50%.

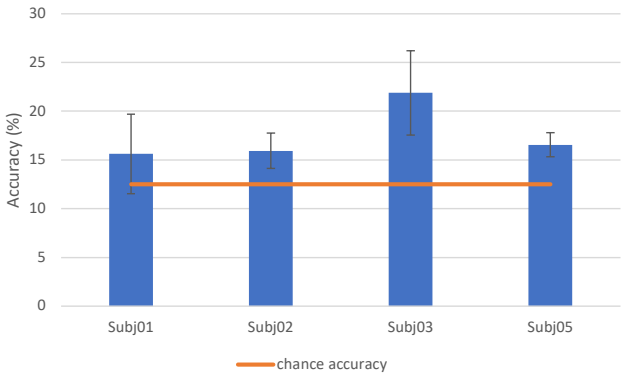

Fig. 9 Multiclass (8 words) test classification results using only fMRI. Chance accuracy 12.5%.

3.3 EEG and fMRI fusion

In this section, we construct and combine a set of hypotheses to obtain better predictive performance than could be obtained from any of the modalities alone. Knowing that EEG captures a good temporal resolution while fMRI good spatial resolution, combining theses two modalities will lead to get a better analysis of the brain activity.

In the fusion process, input specific models; for EEG and fMRI, are learned first, and a bimodal representation vector is created by concatenating the input specific decision probability vectors. In other words, probability vectors which are received in form of decision from modality-specific models are concatenated. Hence, now for each data sample, we have bimodal representation vector. The length of this bimodal representation vector will be four (two dimensional vector from each modality, hence total concatenated dimension will be four) for binary classification (social vs. numbers), and sixteen (eight dimensional vector from each modality, hence total concatenated dimension will be sixteen) for multi-class classification (total 8 words). After splitting the data samples into training-validation-test (as done for individual modality), an additional classifier (here, RF) is employed to predict inner speech using the bimodal representation vector. Results on the fusion RF classifier are reported for the word classification (8 classes) and for the category classification tasks (2 classes) in Figure 10 and Figure 12, respectively.

The results obtained by combining the two modalities are improved significantly, compared to the results when using only one of the modalities ($p < 0.0005$). An oracle classifier (which knows the true label) was also utilised to show the maximum possible classification performance of all combinations (see Figure 11 and 13). The reported accuracies are 84.78%±3.91% and 35.08%±5.81% for the 2 categories and 8 words classification problem, respectively.

Our results demonstrate that the performance of the fused model by combining the two modalities are improving significantly, compared to the results when using only one of the modalities ($p < 0.0005$), and show the potential of applying this methodology for inner-speech decoding.

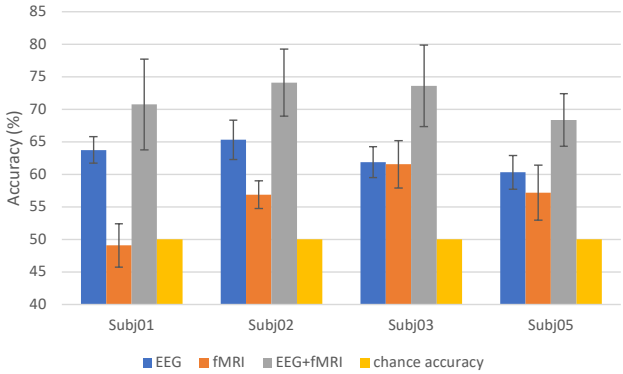

Fig. 10 Binary (category) test classification results using only EEG, only fMRI and fusing EEG and fMRI. Chance accuracy 50%.

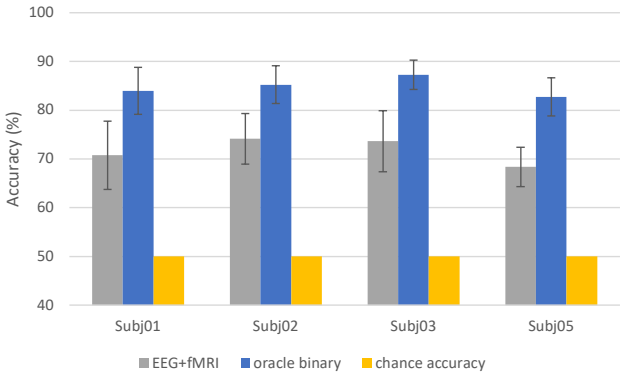

Fig. 11 Binary (category) test classification results fusing EEG and fMRI. The oracle shows the upper bound performance for the combination of the two modalities. Chance accuracy 50%.

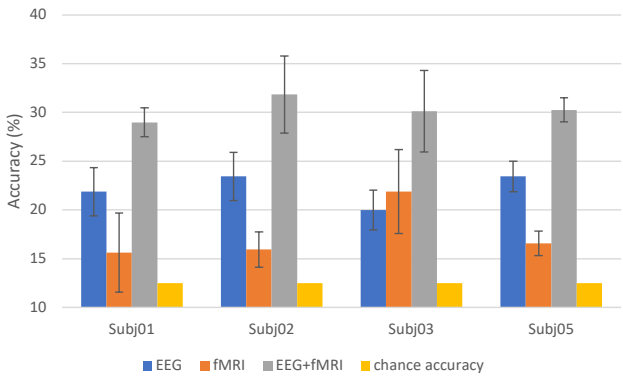

Fig. 12 Multiclass (8 words) test classification results using only EEG, only fMRI and fusing EEG and fMRI. Chance accuracy 12.5%.

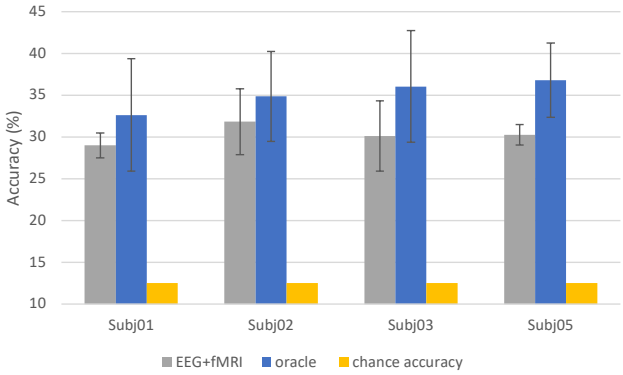

Fig. 13 Multiclass (8 words) test classification results fusing EEG and fMRI. The oracle shows the upper bound performance for the combination of the two modalities. Chance accuracy 12.5%.

4 Discussion

The novel bimodal dataset we contribute, demonstrates the promise of combining EEG and fMRI for inner-speech decoding. The results from our fusion approach indicate that fMRI combined with EEG, outperforms uni-modal classification in both a binary and multi-class task context.

In both binary and multiclass (8) unimodal classification, performance was consistently lower for fMRI data than for EEG. This is initially surprising as Ryba et al. [13] observed that fMRI decoders show better performance on semantic categories with higher conceptual similarity, such as 'child', 'wife', 'father', than neuroimaging techniques with lower spatial resolution like EEG. We suggest that in this current study superior multiclass performance with EEG is facilitated by the temporal features of inner-speech, and EEG's high temporal resolution. This would indicate that the type of input impacts fMRI vs. EEG superiority. For inner speech EEG is a preferable uni-modal neuroimaging technique for inner-speech decoding than fMRI, due to its temporal features, however fMRI may be superior for visual imagery decoding due to the importance of spatial features in visual representation.

Still, fMRI whilst limited in capturing the temporal dynamics, complements EEG by boosting classification performance as seen in the fusion results. This is likely through its high spatial resolution. The two semantic categories we selected words from: numbers vs. social, are represented in distinguishable neural regions as shown by Huth et al. [11]. By mapping semantically grouped words onto the cortex based on voxel semantic selectivity. Huth et al. [11] demonstrated that whilst the central regions of lateral and medial parietal cortex, including the angular gyrus were selective for social words, number words were selected for in the supramarginal gyrus.

We anticipate our work to be a starting point for further studies in the use of multimodal technologies in inner-speech decoding: (i) this dataset can be useful for deeper investigation or the application of advanced machine-learning algorithms; (ii) the insights motivate collection of larger datasets with more subjects, to also perform subject-independent approaches in the future.

5 Conclusions

This work mainly focuses on collecting the brain data from different modalities, such as EEG and fMRI, and their fusion to better integrate spatial and temporal information. In recent years, significant progress has been achieved on various fronts in combined EEG and fMRI studies of human brain function, but not in the research area inner speech decoding. Based on the obtained fusion results, it has been noticed that cross-modality provides the sufficient information which is necessary to process the spatio-temporal aspects in the human brain. As EEG and fMRI represent the complementary imaging techniques, using information from them to analyze the spatial and temporal dynamics of brain processes is particularly beneficial.

Appendix A Results

Table A1 Binary (category) test classification results using only EEG, only fMRI and fusing EEG and fMRI. Chance accuracy 50%.

Subjects	EEG	fMRI	EEG+fMRI
Subj01	63.75%±2.04%	49.06%±3.36%	70.75%±6.99%
Subj02	65.31%±3.01%	56.87%±2.11%	74.14%±5.16%
Subj03	61.88%±2.37%	61.56%±3.64%	73.63%±6.25%
Subj05	60.31%±2.61%	57.19%±4.26%	68.38%±4.03%
Overall	62.81%±2.51%	56.17%±3.34%	71.72%±5.61%

Table A2 Multiclass (8 words) test classification results using only EEG, only fMRI and fusing EEG and fMRI. Chance accuracy 12.5%.

Subjects	EEG	fMRI	EEG+fMRI
Subj01	21.88%±2.47%	15.63%±4.07%	29.98%±1.49%
Subj02	23.44%±2.47%	15.94%±1.82%	31.84%±3.94%
Subj03	20.00%±2.04%	21.88%±4.31%	30.11%±4.18%
Subj05	23.44%±1.56%	16.56%±1.25%	30.25%±1.24%
Overall	22.19%±2.14%	17.50%±2.86%	30.29%±2.71%

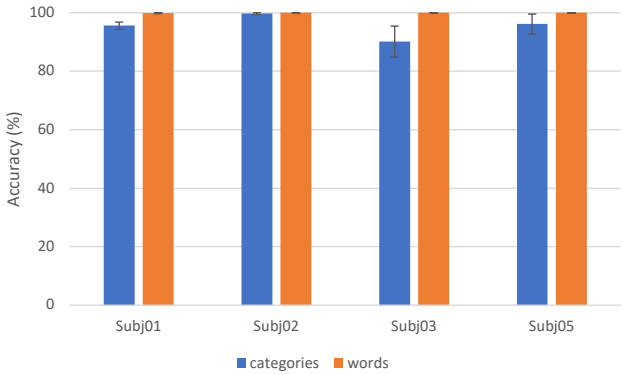

Fig. A1 Train classification results fusing EEG and fMRI.

References

[1] Alderson-Day, B., Fernyhough, C.: Inner speech: development, cognitive functions, phenomenology, and neurobiology. Psychological bulletin **141**(5), 931 (2015)

[2] Whitford, T.J., *et al.*: Neurophysiological evidence of efference copies to inner speech. Elife **6**, 28197 (2017)

[3] Cooney, C., Folli, R., Coyle, D.: A bimodal deep learning architecture for eeg-fnirs decoding of overt and imagined speech. IEEE Transactions on Biomedical Engineering (2021)

[4] Berezutskaya, J., Vansteensel, M.J., Aarnoutse, E.J., Freudenburg, Z.V., Piantoni, G., Branco, M.P., Ramsey, N.F.: Open multimodal ieeg-fmri dataset from naturalistic stimulation with a short audiovisual film. Scientific Data **9**(1), 1–13 (2022)

[5] Scrivener, C.L.: When is simultaneous recording necessary? a guide for researchers considering combined eeg-fmri. Frontiers in Neuroscience **15**, 774 (2021)

[6] Mano, M., Lécuyer, A., Bannier, E., Perronnet, L., Noorzadeh, S., Barillot, C.: How to build a hybrid neurofeedback platform combining eeg and fmri. Frontiers in neuroscience **11**, 140 (2017)

[7] Dong, L., Luo, C., Liu, X., Jiang, S., Li, F., Feng, H., Li, J., Gong, D., Yao, D.: Neuroscience information toolbox: An open source toolbox for eeg–fmri multimodal fusion analysis. Frontiers in Neuroinformatics **12**, 56 (2018)

[8] Lioi, G., Cury, C., Perronnet, L., Mano, M., Bannier, E., Lécuyer, A., Barillot, C.: Simultaneous eeg-fmri during a neurofeedback task, a brain imaging dataset for multimodal data integration. Scientific data **7**(1), 1–15 (2020)

[9] Zotev, V., Phillips, R., Yuan, H., Misaki, M., Bodurka, J.: Self-regulation of human brain activity using simultaneous real-time fmri and eeg neurofeedback. NeuroImage **85**, 985–995 (2014)

[10] Perronnet, L., Lécuyer, A., Mano, M., Bannier, E., Lotte, F., Clerc, M., Barillot, C.: Unimodal versus bimodal eeg-fmri neurofeedback of a motor imagery task. Frontiers in Human Neuroscience **11**, 193 (2017)

[11] Huth, A.G., *et al.*: Natural speech reveals the semantic maps that tile human cerebral cortex. Nature **532**(7600), 453–458 (2016)

[12] 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**(1), 9–21 (2004)

[13] Rybář, M., Daly, I.: Neural decoding of semantic concepts: A systematic literature review. Journal of Neural Engineering (2022)