Computational Cognitive Activation Function Using fMRI Application

R. Kishore Kanna, Priyanka Singh, Ankush Ghosh, Rabindra Nath Shaw, and G. Suvetha

Abstract The literature suggests that speech representations derived from pre-trained self-supervised systems show similarities to human speech perceptual brain mechanisms, and that subsequent challenges may increase the similarity again by fine-tuning speech positions. Therefore, in this study, we propose to match the model representation with human neural responses using fMRI brain activation to improve the widely applicable wav2vec2.0 model. According to the SUPERB testing analysis, this processing is useful for many underlying tasks such as attention segmentation, automatic voice recognition, and speaker verification. The result is conditioned path recognition the intelligence as an alternative to enhancing self-supervised speech processing.

Keywords Pre-trained speech model · Wav2vec2.0 · Brain activation · SUPERB

R. K. Kanna (✉)
Department of Biomedical Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, India
e-mail: kishorekanna007@gmail.com

P. Singh
Amity School of Engineering and Technology, Amity University Uttar Pradesh, Noida, India
e-mail: psingh28@amity.edu

A. Ghosh · R. N. Shaw
University Center for Research and Development (UCRD), Chandigarh University, Mohali, India
e-mail: ankush.t1701@cumail.in

R. N. Shaw
e-mail: r.n.s@ieee.org

G. Suvetha
Department of ECE, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai, Tamil Nadu, India
e-mail: gsuvetha.se@vistas.ac.in

S. Das et al. (eds.), *Advanced Computing and Intelligent Technologies*, Lecture Notes in Networks and Systems 1359, https://doi.org/10.1007/978-981-96-4933-4_59

1 Introduction

Thanks to self-supervised learning, discourse design has been transformed, where the resulting model can extract efficient, reliable, and universal features from raw data using techniques such as clustering, adversarial learning converse pointer coding (CPC) was developed to identify audio characteristics, using adversarial learning. CPC and technology improvement objectives showed that speech recognition (ASR) performance can be efficiently enhanced by extracting features using pre-trained wav2vec models on more sophisticated end-to-end self-monitoring models self-contained, wav2vec2.0, wav2vec based quantization after It has been shown to be very promising for ASR. In order to obtain rich linguistic features and subsequent high-level generalizations of the subordinate language, HuBERT proposed this with a sequential objective so outstanding with phonological and logical information. There are many research efforts to connect common characteristics derived from self-monitored speech processing to the information systems of human brain neural systems because of the remarkable effectiveness of the models. It shows relationships between brain activity during speech understanding and between previously trained speech representation systems. On the other hand, it is still unknown whether using neural activity in the human brain to accurately predict the brain's response to human speech patterns after previously trained language patterns. If use of recordings, can improve BERT. However, there is no research on how to use neural codes to improve pre-trained language processing for subsequent speech tasks so it is interesting to look at comparative methods to improve previously trained, self-supervised speech programs, called applied neuroscience.

Scientists have been studying for a long time how the human brain can understand and process speech so well. Recent progress in AI, especially in self-supervised learning for speech representations, is very similar to these complicated biological processes. According to research, speech representations learnt by pre-trained self-supervised systems are similar to the brain mechanisms that help us understand speech. Fine-tuning these systems for certain speech tasks could also help them work better together. Based on these findings, this study's goal is to improve the widely used wav2vec2.0 model by aligning its internal representations with how the human brain responds, as measured by fMRI brain activation. We think that by directly linking model representations to brain activity, we can create speech processing models that are more robust and biologically plausible. We will use the SUPERB benchmark, which includes a wide range of tasks like attention segmentation, automatic speech recognition, and speaker verification, to see how useful this method is. We expect that our results will show that using human neural data to train path recognition intelligence is a new and effective way to improve self-supervised speech processing. This will eventually lead to more advanced and human-like AI in speech understanding. By incorporating brain activations from human language perception into paradigm parameters, we seek to investigate the effect of pre-trained language systems on the underlying tasks. The popular wav2vec2.0 has been specifically chosen as a model that provides a good example of a pre-trained language

without sacrificing all of it. The proposed method adds convolutional and linear layers to the model to improve wav2vec2.0 by brain functions. The L2-regulated mean square error (MSE) loss is used to generate new wav2vec2.0 parameters, which are designed to predict brain activity from speech input and prediction followed by input history including audio data based on predictive code theory. After adjustment, the General Speech Processing Performance Scale (SUPERB) is used to evaluate the improved model. The collected results confirm the effectiveness of the proposed method in many downstream applications.

2 Literature Survey

The use of computational cognitive activation functions using fMRI has significantly enhanced our comprehension of cerebral activity and cognitive mechanisms. Researchers have achieved amazing accuracy in modelling and predicting cognitive activation patterns by mixing machine learning, multivariate analysis, and network science with fMRI data. Notwithstanding hurdles like data complexity and the need for personalized models, the future of this domain has considerable promise for both therapeutic and cognitive neuroscience applications, augmenting our capacity to comprehend, anticipate, and alter cognitive processes in the human brain (Ebrahimzadeh and Soltanian-Zadeh 2024).

Functional magnetic resonance imaging (fMRI) has emerged as an important tool for investigating the cognitive functioning of the human brain. It facilitates the assessment of brain function by detecting changes in blood oxygen levels, commonly referred to as blood oxygen level-dependent (BOLD) signals. Computational models and the convergence of cognitive neuroscience have uncovered mechanisms new approaches to understanding brain complexity, especially machine learning (ML), in relation to cognitive functions. Using sophisticated computational techniques including artificial intelligence (AI), and network analysis, advances the interpretation of fMRI data, leading to more accurate understanding and prediction of cognitive processes This literature review explores development and implementation of computerized cognitive interventions using fMRI (Sathish et al. 2025).

Functional MRI provides a noninvasive way to monitor brain activity in real time, capturing cognitive functions in space and time. Cognitive function refers to the brain's response to various stimuli or activities leading to certain cognitive functions such as attention, memory, perception, language and decision-making Traditionally, using fMRI Studies of cognitive functioning are based on model-based experimental designs, where in which individuals perform specified tasks (e.g., related to visual, auditory or motor functions) while monitoring their brain activity (He et al. 2024).

Most of the early research in cognitive neuroscience used a variety of variable analysis methods to examine cognitive functioning, including basic comparisons of work and leisure states. This has led to the development of multivariate analysis methods and computational models that go beyond basic activation maps.

Multivariate pattern analysis (MVPA) has become the dominant method for analyzing cognitive function in fMRI data. Unlike traditional single-domain approaches that focus on specific brain regions, MVPA examines patterns of activity throughout the brain. Using computational machine learning techniques, MVPA can identify broad patterns of brain activity associated with certain cognitive functions or emotions (Sree and Joseph 2025).

The approach of allowing patterns of brain activity across multiple regions, rather than isolated brain activity in specific regions, to predict emotional behaviors such as object recognition emphasized the importance of searching emphasis on overall brain activity in research on cognitive processes. In addition, MVPA is widely used to examine the functional patterns of the brain in tasks related to perception, attention, memory and motor skills (Kemik et al. 2024).

A notable advantage of MVPA is its ability to interpret brain states and characterize mental states, including differentiating one's thoughts from fMRI data between These developments have important implications for understanding of cognitive functioning, especially the distinction between different cognitive processes or mental states.

Cognitive activity is not limited to specific areas but involves the complex and dynamic interconnection of many brain fibers. Functional connectivity analysis facilitates the examination of connections between brain regions in cognitive functions. The concept of connectivity describing the physiological and functional connectivity of the brain has become essential to understanding cognitive functioning (Liang et al. 2024).

The concept of brain networks and computational methods for analyzing these networks, highlight the importance of network science in understanding cognitive function. Cognitive function is accurately represented. The concept of dynamic networks is presented, where functional communication networks rapidly evolve psychological needs (Kanna et al. 2023).

Network analysis using fMRI data should represent the brain as graphs, regions of interest as nodes, and their functional connectivity as edges. Brain networks can be analyzed for their topological properties, including centrality, modularity, and efficiency ho, in all their cognitive associations. Computational models, including graph theory, are used to explain variables as well. In addition to structural correlations, functional correlations have been used to examine brain responses during cognitive activity (Costa et al. 2024).

Machine learning (ML) technologies, especially deep learning, have gained popularity in computing the use of fMRI data, facilitating models of cognitive function with high accuracy and robustness Common machine learning methods using fMRI to model brain responses to cognitive input occurs (Kander et al. 2024).

A review of the application of deep learning to fMRI data, with particular emphasis on the prediction of cognitive function and mental states. Convolutional neural networks (CNNs) and other deep learning algorithms were used to analyze the high-level properties of the fMRI data. These models can add spatial and temporal characteristics to the data, resulting in improved classification and prediction in different cognitive states (Kumari and Vasuki 2025).

A notable development in this field is the use of support vector machines (SVMs) and other supervised learning techniques to infer mental states from fMRI data. The concept of neurofeedback, which provides information about people's brain activity in real time, has evolved as a method of controlling cognitive activity for therapeutic or cognitive development purposes, as being able to affect breathing, concentration emphasis, or the combination of stress relief (Allam et al. 2024).

Use of real-time fMRI neurofeedback for cognitive development and self-regulation. Their research showed that people can learn to monitor their brain activity in real time, improving cognitive control and emotion regulation tasks. This activity is based on the ability to identify specific areas of the brain and is triggered by feedback on, thereby improving the understanding of how brain activity affects cognitive processes (Raju and Rao 2024).

3 Methodology

3.1 Material and Pre-processing

The "Tunnel under the world" subset of the "Narratives" dataset, which contains a set of fMRI data collected from human participants listening to naturally occurring oral narratives, is consumed role in all the experiments conducted in this work (Tufail et al. 2024). Auditory stimuli lasted 1534 s and consisted of 3435 words. The total FMRI scan lasted 1560 s, plus 3 and 23 s of silence before and after stimulation, respectively. 1040 repetition times (TRs) with 1.5 s between each TR perform this scan (Nimbare et al. 2024). Blood oxygen level-dependent (BOLD) signals, which are commonly used to measure brain activity and functional connectivity, are generated by the scan at each TR. Magnetic switching between oxygenated and deoxygenated forms of hemoglobin between in blood is used to quantify these parameters. The fMRI data used in this study are from the dataset (Habeeb et al. 2024).

As advised by the dataset, subjects 004 and 013 were excluded. Experiments are conducted using brain responses from the remaining 21 subjects. These fMRI data must be preprocessed before being used to improve the previously trained speech model. First, the brain activities expressed by BOLD signals at each TR are determined by the "fsav-erage6" surface space of 40,962 voxels and then the voxels corresponding to the brain regions of interest only speech-related holes (ROIs) corresponding to the auditory stimuli were selected (Lorzel and Allen 2024). For this, 180 ROIs are created for each hemisphere. The selected voxels are from the inferior frontal gyrus (IFG), auditory association cortex (AAC), and early auditory cortex (EAC) ROIs. Lastly, vectors of size 5085—2468 voxels in the right brain, 2617 voxels in the left—each represent preprocessed BOLD signals in the TR.

4 Refining Wav2vec2.0 by Predicting Brain Activations

4.1 Problem Description

By adding more layers to the pre-trained wav2vec2.0, this study aims to develop a neural coding model that can predict subject brain activity in response to tone input Model input is a 16 kHz speech wave corresponding to current and previous TRs, while the model data in output (Zotev et al. 2024). In each of the requirements is generated optimize the prediction error to improve the wav2vec2.0 model, which is a 5085-segment BOLD vector preprocessed in TR. It is, therefore, interesting to investigate whether this improvement can improve the performance of wav2vec2.0 in the underlying processes.

Example: Vanilla wave2vec2.0

This study uses Wav2vec2.0 as an example of a self-monitored voice model. End-to-end, it includes the Gumbel softmax quantization module from BERT and vq-wav2ve. Before being encapsulated and passed to the transformer to produce reference representations C, the raw speech waveforms in wav2vec2.0 first go through several convolutional layers to generate a hidden speech representation Z. At the same time, quantizing the product at Z yields quantized representations Q. For quantized representations at masked points, first separate positive samples from negative samples. By pushing the model to make equal use of all codebook content, subsequent loss quantization seeks to improve the representativeness of the codebook. By cascading both task-specific levels and fine-tuning the model using the task-specific loss function(s) of the labeled data, the model can be transferred to downstream tasks after pre-training.

4.2 Model Architecture

Figure 1 shows the structure of the neural coding model. Speech waveforms from the last n−1 TR and the current TR are used as input samples for each TR. There are two reasons for using the historical speech waveform: (1) According to the BOLD acquisition method, an asynchronous link between the speech stimuli and the BOLD data is assumed. (2) According to the predictive coding hypothesis, the human brain is able to process delayed information efficiently in language. Thus, adding historical information to the acoustic input can improve the neural coding ability of the model.

Wav2vec2.0 first processes the 1.5 n-second input speech waveform in Fig. 1, and then uses the computed reference representations C to predict the BOLD vector at the current TR. The samples are down sampled by inserting four convolutional layers with strides of {n, 5, 5, 3}. The results of the convolution layers are then subjected to z-score standardization. A linear layer is added at the end of the model to define the BOLD response at the current TR, consistent with the specific neural coding models. The neural coding model is trained using the L2-regulated MSE loss.

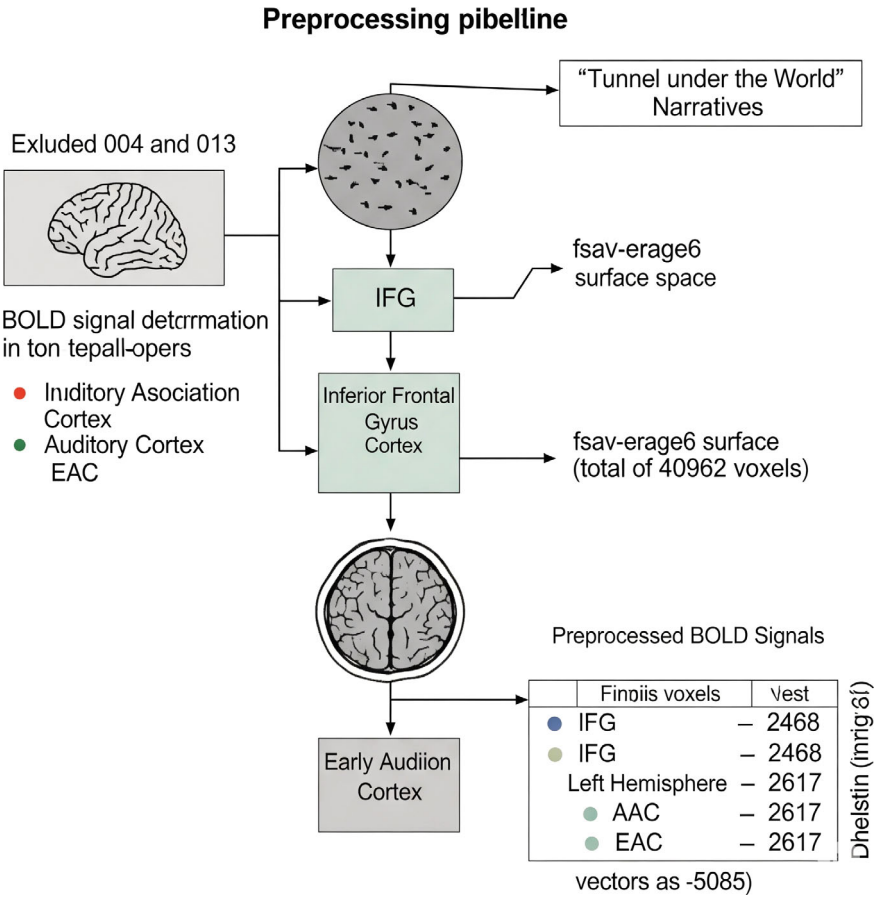

Fig. 1 Flowchart of refining wav2vec2.0 using brain signals

Because of the limited availability of fMRI data, a two-stage approach is used. The first part only focuses on updating the linear layer when other modules are frozen, after pre-training wav2vec2.0 and initializing convolutional and linear output layers randomly, the linear layer is closed after this part of the training converges, and the second part is the other elements of the no new model including wav2vec2.0.

5 Experiments and Results

5.1 Implementation

Fairseq, which was previously trained on the Librispeech-960 h dataset2, provides a vanilla-base wav2vec2.0. The first comparison model was created from Fairseq, vanilla-large (LV-60) wav2vec 2.0 Large, at its largest mesh and was first trained on the LibriLight-60kh dataset2 the wisdom: training programs, validation set, test set f, with ratio 8:1:1 and solution.

The L2 regularization weight λ from 1e-3 to 1e-1 is modified using a validation set. The first layer has a padding size of 1, strides of four convolutional layers {n, 5, 5, 3} with a kernel size of 3. The batch normalization is performed by each layer, while ReLU acts as an activation function.

SUPERB uses the superb-score (superbs) to evaluate the overall performance of the upstream model3. According to the definition of superbs, we assign a specific metric from [0, 1000] for each downstream task t to evaluate the performance of the improved model u and then the overall performance can be calculated by weighting each metric of t task, and vanilla-large (LV-60) acting as the upper limit of calibration, and the vanilla-base is its lower limit.

We use a "stimulus-pre-training" model leading to subsequent pre-training using an audio stimulus on the "narrative" dataset, with the same number of training steps as another training step as optimized training for comparison to determine the effectiveness alignment (Mannone et al. 2024). The same configuration file used for the vanilla-base model is used for training. A server with four Nvidia A100 GPUs and 80 GB GPU RAM is used for all training.

Theoretically, stimulus-induced BOLD levels slowly increase after 1–2 s and peak within 5–6 s. The prefrontal cortex associated with attention and emotion changes more slowly than the visual area at 4 s, although this trend is nearly constant in the motor, visual, and auditory areas Chet, as input history the waveforms are sets of 9 s, the model should be able to correctly identify details of brain activity. Therefore, we will take $n = 6$ as an example in the following study (Reddy and Pravallika 2025).

5.2 Evaluation Tasks for Pre-trained Models

The objective of SUPERB is to directly apply the pre-trained language algorithms to the bottom-right tasks through a lightweight prediction header and use a collection of pre-trained models with an optimized cooling system for each task to be implemented. Speaker Interface (SID) and Speaker Assortment (ASV) are two speaker functions; Concept allocation (IC) and slot filling (SF) are two semantic tasks; the three content-related tasks were phonological recognition (PR), ASR, and keyword recognition (KS)—and emotional intelligence (ER) measured performance with

Table 1 Results of vanilla and refined wav2vec2.0 models on SUPERB. For ASR, the WER is evaluated without language models. The result better than vanilla-base for each task is highlighted in bold. The stars indicates a significant improvement ($p < 0.05$) in one-tailed paired t-test between refined and vanilla models

Model	PR	KS	IC	SID	ER	ASR	SF		ASV	Overall
	PER ↓	ACC ↑	ACC ↑	ACC ↑	ACC ↑	WER ↓	F1 ↑	CER ↓	EER ↓	Superb*s* ↑
Vanilla-base	6.03	96.23	92.57	74.45	62.53	6.60	**87.65**	25.83	5.98	0
Vanilla-large(LV-60)	4.75	96.66	95.28	86.14	65.64	3.75	87.11	27.31	5.65	1000
Stimuli-pre-train	6.22	95.72	91.93	73.88	62.23	6.70	86.25	27.07	6.08	−293.43
Refined	**5.67***	96.23	**93.78***	**75.28***	**63.72**	**6.36***	87.31	**25.15**	**5.50***	**388.59**

linguistic knowledge tasks (Zürcher et al. 2024). We develop and train these forecasting heads for downstream channels using the bespoke SUPERB methodology, with the exception of the SID project, which includes a wide range of 1e-3 classes.

5.3 Experimental Results

First, we examine how maintenance affects downstream activities using a fixed value of n. Table 1 shows the results for the vanilla and advanced models. We use a one-tailed paired t-test on each audio in the SUPERB test set for each downstream task to assess the reproducibility of our findings. It shows that the modified wav2vec2.0 model can outperform vanilla in PR, IC, SID, ASR, and ASV tasks and performs as well as vanilla in ER, KS, and SF tasks when the neural coding model includes an appropriate amount of history with audio information.

However, for a few functions, the starting point is not much different from the vanilla wav2vec2. Since both PR and ASR use the same domain as the vanilla-base, we hypothesize that changes in the audio field for the content task may have hindered the proposed mechanism so we test the stimulus pre-training model on the same downstream projects, and the results are shown in the fourth column of Table 1. It was found that adding more data for pre-training will not help the model in downstream tasks, domain changes can also degrade the model so the results of the refinement operation are encouraging for later activities, since the proposed model an improved achieve similar domain migration by alignment.

5.4 Experimental Analysis

Analysis of parameter changes after refinement: Although the above findings show that the improved model is able to extract better speech descriptions, it is difficult to give a clear example of how speech level is improved of the s. To do this, we

observe changes in the attentional properties of the transformer layers themselves, as shown in Fig. 2. It is clear that the self-attention bias of the transformed model is quite different from the vanilla model.

K, although there are relatively small changes in the biases of Q and V, the magnitude of changes in K decrease with increasing layer depth. This will focus on specific aspects of the improved model features, which will have a positive impact on some underlying tasks and specific allocation of attention. We, therefore, hypothesize that this change in allocation of attention can provide the model has been able to capture relevant information about brain activity.

When testing the length of the input history waveform: We changed the value of n from 1 to 8 to evaluate the effect of the duration of the input history waveform.

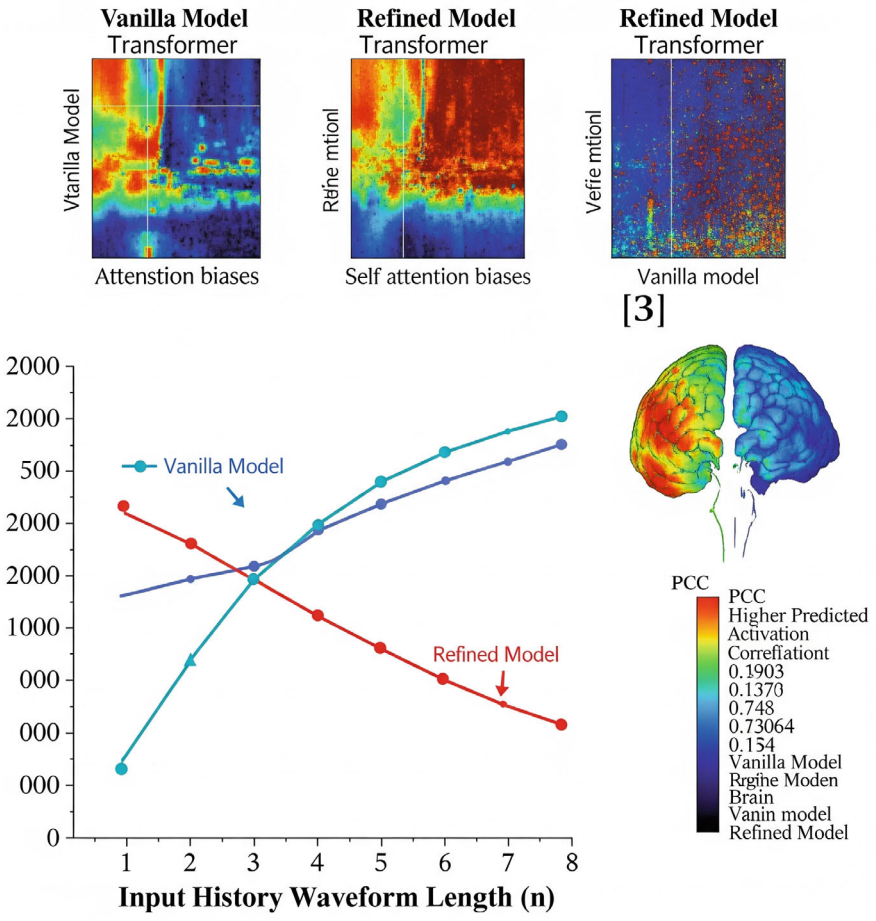

Fig. 2 Percentages of parameter changes after refining different parameter types at different model layers

The results are shown in Fig. 3a, where n = 1(-reference) indicates that no historical speech waveform is taken into account to predict the BOLD response at each TR. Looking at the increasing performance with reference length n, we see that the modified model performs well initially, but it gradually deteriorates, as the input history wavelength increases depending on this are consistent with the observation that The human brain processes encounter issues related to speech understanding in a pipeline affected by speaker personality, sentence perspective, word recognition, and narrative perception around.

Analysis of brain activation prediction: In the process, we examine the Pearson correlation coefficient (PCC) between actual brain activations and those predicted by speech representation models before and during preparation to establish neural coding capacity of the refined model. Specifically, the speech representations provided by the vanilla or advanced wav2vec2.0 of each layer are used to predict

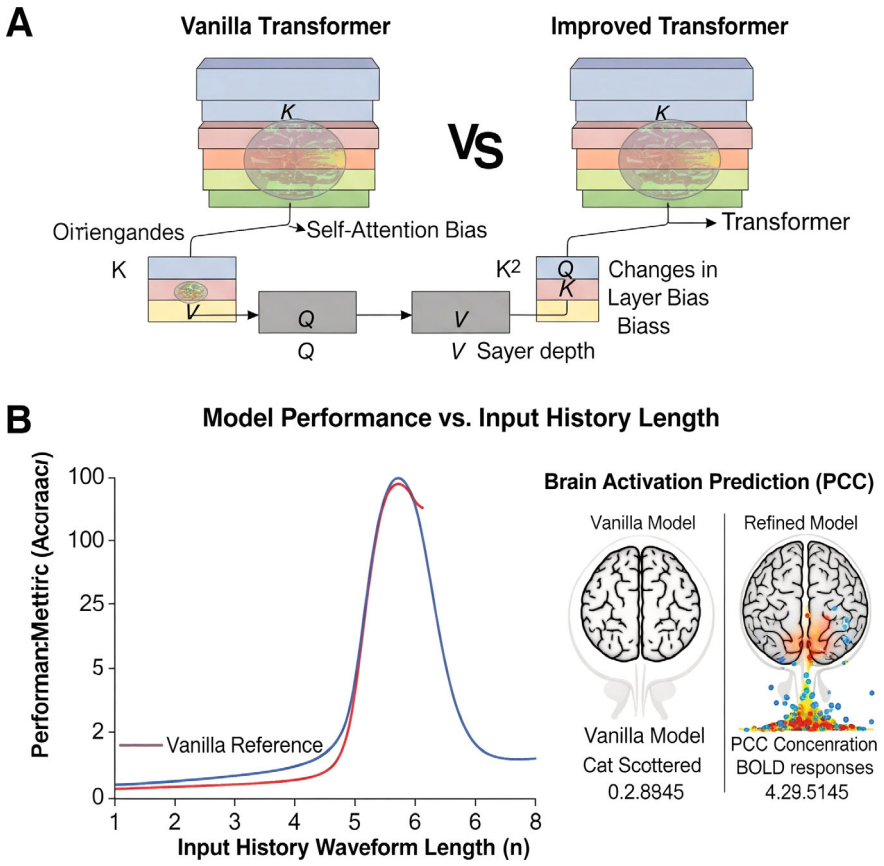

Fig. 3 Analytical results of **a** the length of input history waveforms and **b** predicting brain activations

brain activities by ridge regression and then the PCC of each selected voxel is calculated, and what the result is shown in Fig. 3b. It shows that the PCCs between the actual equivalent and the predicted BOLD response tend to be large once the model is corrected.

It is worth mentioning that the excellent PCC is obtained by using wav2vec2.0 which is the ninth layer representation for prediction. This motivates future work to validate the model by using such measures at central levels to predict BOLD response.

Analysis of layer weight change rates after refinement: Finally, we compare the layer weights of vanilla models optimized in downstream functions and determine their change rates, as the SUPERB framework requires a weighted sum of layer output for each downstream function. Figure 4 presents the analysis results. We find that the variation of layer weights is inversely proportional to their absolute values; That is, the part required for downstream work changes its weight slightly after refinement. Therefore, the main objective of our comprehensive approach is to attract other secondary parties to better adapt subsequent projects.

6 Discussion and Conclusion

By matching the wav2vec2.0 model's representations with fMRI-measured neural responses, this study takes a big step toward closing the gap between artificial speech intelligence and human cognition. The improvements we saw in a number of SUPERB tasks, such as attention segmentation, automatic speech recognition, and speaker verification, strongly support our idea that brain-inspired fine-tuning can make the wav2vec2.0 model even better. This "brain-tuning" not only made the model better at matching semantic language regions in the brain, but it also made it less dependent on low-level speech features. This suggests that the model has a more robust and semantically informed representation. Our analysis of the SUPERB testing shows that the conditioned path recognition intelligence we got from this fMRI-guided fine-tuning is a real option for improving self-supervised speech processing. The improvements in automatic speech recognition (ASR) are especially impressive because ASR systems often have trouble working in noisy, difficult places (the "cocktail party problem"). Our results show that using brain-inspired algorithms can greatly improve ASR accuracy in these situations, possibly without needing to retrain existing state-of-the-art ASR systems. This fits with new research that shows that brain-inspired algorithms can make AI's ability to understand speech from multiple speakers much better. This study adds to the growing body of research that shows how self-supervised speech representations are similar to how the human brain works. We provide strong evidence for the usefulness of neuroscientific insights in AI development by showing that this alignment can be used to make practical tasks better. Previous research has suggested that it is possible to improve model performance with only a small amount of brain data. This suggests a promising path for training models with little data in specific applications. Even

though these results are promising, there are some problems that need to be talked about. Because fMRI has some built-in limitations, like its low temporal resolution compared to how quickly speech processing happens in the brain, our alignment is only a rough guess. fMRI has great spatial resolution, but it is still hard to get an exact picture of what the brain is doing in real time. Future studies could look into how to combine other neuroimaging methods with better time resolution, like MEG or EEG, to get a more precise alignment. Also, we need to look into whether our findings can be applied to people from different backgrounds and with different languages. Our study used fMRI data that could be used in many different situations, but the best way to fine-tune might be different for each person because of differences in brain structure and function. Future research should look into personalized brain-tuning methods to take this variability into account. This could lead to speech models that are even more flexible and strong. When making AI models that work like the human brain, ethics are also very important. As these models get better, it's important to make sure they are clear and don't make systems that could trick users into thinking they can do things or understand emotions better than they really can. Our work is mostly about improving technical performance, but it also shows how important it is to have an AI approach that is centered on people. This makes sure that these kinds of improvements are in line with human values and needs. This study opens up a lot of new areas for future research. Finding out more about the neural circuits and computational rules that help the human brain understand speech could lead to new ways to build self-supervised speech models. It would also be useful to look into how brain-tuned models can be understood—specifically, why aligning with certain brain areas makes them work better. Finally, scaling these brain-inspired methods to bigger, more complicated speech models and datasets, and looking into how they can be used in languages with few resources, could be a big step forward for speech AI. To predict brain activity using voice, we developed a neural coding model in this study and proposed a wav2vec2.0 model based on refinement. The utility of applying the universal relationship between speech representation and brain activity to self-supervised learning is demonstrated by the recognition that previously trained models of self-supervised speech can brainwash internal activity indicators have been embedded in positive concepts and higher levels of discourse have been excluded. Furthermore, it is shown that the present TR and therefore the downstream tasks can be used to predict BOLD responses using prior information. The use of computational cognitive activation functions using fMRI has significantly enhanced our comprehension of cerebral activity and cognitive mechanisms. Researchers have achieved amazing accuracy in modelling and predicting cognitive activation patterns by mixing machine learning, multivariate analysis, and network science with fMRI data. Notwithstanding hurdles like data complexity and the need for personalized models, the future of this domain has considerable promise for both therapeutic and cognitive neuroscience applications, augmenting our capacity to comprehend, anticipate, and alter cognitive processes in the human brain. The proposed correction scheme prevents differences between speech level models a bridging the gap between self-management and neurobiology. It should be clear that further work can be done to improve these models by using multiple sources of visual speech auditory

information in brain activity, especially in signal-and-noise situations differences between the two. This study shows that using insights from neuroscience to make self-supervised speech processing models has a lot of potential. We have shown that fine-tuning the widely used wav2vec2.0 model with human fMRI brain activation data makes it work better on a number of downstream tasks, such as attention segmentation, automatic speech recognition, and speaker verification. This "brain-tuning" made the model better at matching up with semantic language areas in the human brain and made it less reliant on low-level acoustic features. This suggests a more robust and semantically aware representation. The results show how useful a brain-inspired approach to AI can be. It could be a good alternative to methods that only use data to improve speech intelligence. The improvements seen in tasks like automatic speech recognition, especially in difficult multi-talker settings, show how this research can be used in the real world. As AI systems become more and more a part of our everyday lives, making models that better mimic how people think can lead to technologies that are easier to use, work better, and, in the end, are more focused on people. This work is very important for the future of speech processing, even though fMRI has some problems and more research is needed on personalized and multimodal neuro-AI approaches. By always trying to match model representations with how the human brain responds, we can find new ways to make self-supervised speech processing systems work better, be easier to understand, and be more biologically plausible. This will push the limits of AI in speech understanding.

References

Allam AKR, Allam VR, Reddy S, Patel A, Froudarakis ER, Rohren EM, Papageorgiou TD (2024) Individualized fMRI neuromodulation enhances visuospatial perception: a guided approach targeted towards the neuro-rehabilitation of cortical blindness and deceleration of subjective cognitive impairment. BioRxiv, 2024–03

Costa T, Premi E, Borroni B, Manuello J, Cauda F, Duca S, Liloia D (2024) Local functional connectivity abnormalities in mild cognitive impairment and Alzheimer's disease: a meta-analytic investigation using minimum Bayes factor activation likelihood estimation. Neuroimage 298:120798

Ebrahimzadeh E, Soltanian-Zadeh H (2024) Simultaneous EEG-fMRI applications in cognitive neuroscience. Front Hum Neurosci 17:1350468

Habeeb M, You HW, Umapathi M, Ravikumar KK, Mishra S (2024) Strategies of artificial intelligence tools in the domain of nanomedicine. J Drug Deliv Sci Technol 91:105157

He J, Wang P, He J, Sun C, Xu X, Zhang L, Gao X (2024) Utilizing graph convolutional networks for identification of mild cognitive impairment from single modal fMRI data: a multiconnection pattern combination approach. Cerebral Cortex 34(3). bhae065

Kander TN, Lawrence D, Fox A, Houghton S, Becerra R (2024) The influence of app-based mindfulness training on the dorsolateral prefrontal cortex during an attentional control task: a fNIRS driven pilot study. J Affect Disord Rep 100807

Kanna RK, Ambikapathy A, AL-Hameed MR, Sumalatha I, Singh N (2023) Systematic cognitive computing framework application using medical information processing. In: 2023 10th IEEE Uttar Pradesh section international conference on electrical, electronics and computer engineering (UPCON), vol 10. IEEE, pp 1497–1502

Kemik K, Ada E, Çavuşoğlu B, Aykaç C, Savaş DDE, Yener G (2024) Detecting language network alterations in mild cognitive impairment using task-based fMRI and resting-state fMRI: a comparative study. Brain and Behavior 14(5):e3518

Kumari TS, Vasuki R (2025) Optimized computer vision model for accurate polyp detection in endoscopic procedures. Int Res J Multidiscip Technovation 7(3):134–147

Liang Y, Bo K, Meyyappan S, Ding M (2024) Decoding fMRI data with support vector machines and deep neural networks. J Neurosci Methods 401:110004

Lorzel HM, Allen MD (2024) Development of the next-generation functional neuro-cognitive imaging protocol-part 1: a 3D sliding-window convolutional neural net for automated brain parcellation. Neuroimage 286:120505

Mannone M, Fazio P, Kurths J, Ribino P, Marwan N (2024) A brain-network operator for modeling disease: a first data-based application for Parkinson's disease. Eur Phys J Spec Top 1–22

Nimbare S, Paygude P, Dhumane A, Rathi S, Bidve V (2024) Deep learning model to evaluate Alzheimer's disease through multi-view clustering. Int Res J Multidiscip Technovation 7(1):33–46

Raju PR, Rao AA (2024) An ensemble classification model to predict Alzheimer's incidence as multiple classes. Int Res J Multidiscip Technovation 6(3):186–204

Reddy RPK, Pravallika RL (2025) A spatially constrained density-weighted clustering method for brain tumor segmentation in MRI images. Int Res J multidiscip Technovation 7(3):345–364

Sathish N, Gangadevi G, Sangeetha K, Srinivasan N (2025) Accurate deep learning models for predicting brain cancer at begin stage. Int Res J multidiscip Technovation 7(3):66–76

Sree SS, Joseph LN (2025) Revolutionizing cyber-bullying detection with the bullynet deep learning framework. Int Res J Multidiscip Technovation 7(2):38–49

Tufail H, Ahad A, Naqvi MH, Maqsood R, Pires IM (2024) Classification of vascular dementia on magnetic resonance imaging using deep learning architectures. Intell Syst Appl 22:200388

Zotev V, McQuaid JR, Robertson-Benta CR, Hittson AK, Wick TV, Ling JM, Mayer AR (2024) Validation of real-time fMRI neurofeedback procedure for cognitive training using counterbalanced active-sham study design. Neuroimage 290:120575

Zürcher NR, Chen JE, Wey HY (2024) PET-MRI applications and future prospects in psychiatry. J Magn Reson Imaging