

# Spatial and Temporal Analysis of the fNIRS-based Decoding Models for Speech Perception

Santiago Posso-Murillo

*Dept. of Electrical Engineering  
University of Kentucky  
Lexington, USA  
spo230@uky.edu*

Luis G. Sanchez-Giraldo

*Dept. of Electrical Engineering  
University of Kentucky  
Lexington, USA  
luis.sanchez@uky.edu*

Jihye Bae

*Dept. of Electrical Engineering  
University of Kentucky  
Lexington, USA  
jihye.Bae@uky.edu*

**Abstract**—Speech decoding extracts speech information from neural activity. Previous studies have demonstrated that functional near-infrared spectroscopy (fNIRS) contains information suitable for language decoding. However, the interpretability of neural decoders for speech decoding remains unexplored. This study aims to identify the relevant spatial regions and time points that enable neural decoders to distinguish speech perception from silence. We trained a linear support vector classifier (SVC) and a linear discriminant analysis (LDA), and conducted post-hoc analysis using the Shapley additive explanations (SHAP) technique to identify the spatiotemporal patterns contributing to the neural decoder’s performance. We utilized a publicly available fNIRS dataset, which comprises recordings from eight adults during the auditory perception of speech and silence. Our results indicate that features from oxyhemoglobin (HbO) and deoxyhemoglobin (HbR) are relevant and that the inferior frontal gyrus (IFG) and Wernicke’s area (auditory-B) are key for differentiating speech perception from silence, which aligns with established neurophysiological processes. Temporally, we observed a subtle increase in feature relevance at the stimulus onset ( $t=0$  s) and around 6–10 s across subjects, which may be related to the initial dip and the peak of the hemodynamic response as measured by fNIRS. These results suggest that studies could use key spatiotemporal fNIRS features to improve speech decoding performance.

**Index Terms**—fNIRS, neural decoding, post-hoc analysis, SHAP, perceived speech.

## I. INTRODUCTION

Brain-machine interfaces (BMIs) have emerged to restore communication capabilities of people with neurological disorders and traumatic injuries [1]. Noninvasive techniques such as functional magnetic resonance imaging (fMRI) [2], magnetoencephalography (MEG) [3], electroencephalography (EEG) [3], and functional near-infrared spectroscopy (fNIRS) [4] have shown their potential to capture linguistic information, making them suitable candidates for language decoding. fNIRS, in particular, measures brain activity indirectly by detecting changes in oxyhemoglobin (HbO) and deoxyhemoglobin (HbR) concentration. fNIRS offers higher spatial resolution than EEG and better temporal resolution than fMRI [5]. Furthermore, fNIRS’s high portability and motion-tolerance make it a practical choice for BMI applications.

Previous studies have shown that fNIRS contains information suitable for differentiating perceived concepts (words) using machine learning (ML) models [6]–[8], and have provided

compelling evidence that complex speech, including sentences, can also be decoded via fNIRS [4], [9]. Despite such advances, it remains unclear which brain areas contribute the most to speech decoding and whether the learned feature spaces align with known neurophysiological processes underlying speech perception, as measured by fNIRS. In addition, it has not been explored whether ML models identify informative features temporally, and how the relevance of these features evolves throughout the stimulus.

Speech processing in the brain is a highly complex function involving multiple parallel and hierarchical pathways across cortical and subcortical regions. Key cortical areas include the superior temporal gyrus (STG), which is essential for phonological processing [10]. Within the STG, Heschl’s gyrus, located in its rostral aspect, is responsible for transforming basic acoustic features into more complex variations [11], while Wernicke’s area, located in the caudal aspect, plays a crucial role in the comprehension of meaningful speech [12]. In addition, the inferior frontal gyrus (IFG), which houses Broca’s area, is involved in phonological analysis and supports working memory processes during language tasks [13]. Furthermore, subcortical structures such as the cingulate gyrus, striatum, and thalamus contribute to auditory memory. These insights have been primarily derived from studies that combine transcranial magnetic stimulation with neuroimaging techniques, such as EEG [14] and fMRI [10].

Although fNIRS is limited by shallow light penetration and lower spatial resolution compared to fMRI [5], it has been effectively used to assess cortical responses to speech. For instance, cortical activation in both normal-hearing adults [15] and children [16] were examined while they listened to clear and noise sentences with varying levels of intelligibility. Their findings revealed a linear relationship between speech intelligibility and bilateral superior temporal cortex (STG) activation in adults. Notably, both adult and children groups showed a peaked hemodynamic response in the left IFG in response to heavily degraded speech. Similarly, [17] found that the left IFG showed greater activation to degraded speech compared to clear speech, with the hemodynamic response peaking significantly later in the left IFG than in the superior temporal cortex. Moreover, [18] reported strong activation near Heschl’s gyrus during auditory speech perception and [19]

investigated single-word perception and observed significant activation during passive listening in bilateral middle and superior temporal gyri (MTG and STG), as well as in left-lateralized regions, including the inferior temporal gyrus (ITG) and the temporopolar cortex.

It is important to note that the aforementioned fNIRS studies examined auditory brain responses to speech using a generalized linear model (GLM). GLM incorporates prior knowledge about the experimental design and expected signal morphology to model brain activity. Unlike decoding approaches, such as the one presented in this study, GLM does not require training a model to recover the perceived stimuli; instead, it functions as an encoder, training a model to estimate how brain activity responds to known stimuli. This distinction is important because our study investigates whether specific information related to speech processing is localized in particular brain regions using a decoder framework. We then compare these results with those obtained from traditional encoding models to assess their neurophysiological consistency.

While spatial information is fundamental to decoding tasks, temporal dynamics also play a critical role. fNIRS studies have shown that hemodynamic response delays vary significantly depending on the type of task. For instance, motor tasks, such as finger tapping, exhibit a peak hemodynamic response around 8 seconds post-stimulus, whereas foot tapping peaks earlier, at approximately 5 seconds [20]. Similarly, task-related variability has been observed in speech tasks. It has been reported that the HbO signal peaked at 6 seconds during a naming task and at 10 seconds during a verbal fluency task, while the corresponding HbR signals peaked much later at 17 and 24 seconds, respectively, highlighting that HbR responses typically exhibit a more delayed peak compared to HbO [21]. In addition, subject-dependent variability was reported in imagined speech responses, with two primary peak time windows emerging: one between 4–6 seconds and another between 6–8 seconds [4]. Moreover, in the auditory domain, it was found that during passive listening in normal-hearing adults, peak hemodynamic responses appear between 5 and 7 seconds after the onset of speech and noise stimuli [22].

Given the complexity of speech processing, our study investigates which spatiotemporal features are most relevant for speech decoding using Shapley additive explanations (SHAP) technique. SHAP is a game-theoretic approach that assigns impact values to learned features based on their contribution to model predictions [23]. SHAP has been widely adopted in fNIRS research [24], [25]. For instance, [26] used gradient-boosted decision trees and a multilayer perceptron with SHAP to identify biomarkers associated with major depressive disorder. Their findings highlighted key channels in the frontopolar area, dorsolateral prefrontal cortex, and Broca's area as significant for the diagnosis of major depressive disorder. Furthermore, by applying SHAP to an support vector classifier (SVC) for early diagnosis of Parkinson's disease, the importance of the frontopolar cortex was emphasized [25]. Inspired by these studies, our SHAP-based analysis aims to identify the most informative spatial features for the decoding task and

validate them against established neurophysiological patterns observed in fNIRS research. Additionally, we examine whether the decoder leverages temporal information to distinguish between speech and silence conditions. This work represents a step toward the development of fNIRS-based BMIs designed to enhance communication access and quality of life for individuals with communication impairments. All codes to generate the results will be available upon publication on author's GitHub repository.

## II. METHODS

*1) fNIRS dataset:* Experiments were conducted using the public dataset in [18], which contains fNIRS signals from eight healthy adults responding to auditory speech segments from a children's story. Our analysis focused on 18 trials involving auditory-only stimuli, each containing a different segment of the story, as well as 10 trials of silence control. The fNIRS signal is comprised of 44 channels with a sampling rate of 3.9 Hz, covering left IFG, bilateral auditory (Auditory-A, including Heschl's gyrus, and Auditory-B, including caudal temporal lobe, encompassing the planum temporale, also known as Wernicke's area), and visual regions (Visual-A, including Cuneus and superior occipital gyrus, and Visual-B, including middle occipital gyrus). We followed the same preprocessing pipeline in [18]. First, we convert the raw intensity signals to optical density. Then, we calculate the scalp coupling index value to identify optodes that were not well connected to the scalp. Motion artifacts were then removed using temporal derivative distribution repair. Then, the short-separation channel signal was subtracted from the long-separation channel signal. The signal was subsequently transformed to HbO and HbR using the modified Beer-Lambert Law. Next, 0.02–0.4 Hz bandpass filtering was applied to remove slow drifts in the signal. The signal was then enhanced using a correlation-based signal improvement method to reduce head motion noise and improve signal quality [27]. Finally, an epoch rejection criterion was applied to exclude any epoch with a peak-to-peak value exceeding  $100\mu M$ . The selected time window for our analysis ranged from stimulus onset to 18 seconds to encompass the 12.5 seconds during which participants perceived the audio stimulus and the additional  $\sim 6$  seconds required for the hemodynamic response to peak after stimulus onset [5].

*2) Neural Decoding:* We use linear SVC and linear discriminant analysis (LDA) models. SVC is well-suited for high-dimensional data [25], which is particularly useful for neuroimaging datasets that often include thousands of features with limited samples. LDA offers a low computational cost and relatively good classification performance, making it suitable for online BMI systems [28].

To address class imbalance (18 trials for the auditory-only condition and 10 for the resting condition), we applied the Adaptive Synthetic (ADASYN) sampling method [29] to oversample the minority class during training. Model performance was then evaluated using the area under the curve (AUC) metric instead of accuracy, since the AUC is more sensitive to false positives and negatives. To train the decoders, we used

the implementation in Scikit-learn [30]. Following the ML methodology outlined in [31] for correctly evaluating models applied to fNIRS data, we tested regularization parameter values for the SVC: 0.001, 0.01, 0.1, and 1, with a maximum of 250,000 iterations to ensure convergence.

Additionally, we employed a nested cross-validation to optimize and evaluate the models without bias. The outer cross-validation consisted of 5-fold cross-validation to define the test set. Meanwhile, the inner cross-validation, designed for hyperparameter tuning, used a 3-fold cross-validation to separate the training and validation sets. Following the statistical recommendations in [31], we evaluated the performance of each model using a one-tailed *t*-test to determine whether its accuracy is above chance level at a 0.05 significance level. Next, we conducted a statistical analysis to compare the performance of the models against each other. One-way analysis of variance (ANOVA) test was performed if the assumptions of normality (assessed using the Shapiro test) and homoscedasticity (assessed using a Bartlett test) were met. If these assumptions were violated, a non-parametric Kruskal-Wallis test was employed instead. If a significant effect of the model was identified, one-tailed paired *t*-tests with Bonferroni correction were used to perform pairwise comparisons between the models.

3) *Post-hoc Analysis by SHAP*: We employ SHAP, which deconstructs a prediction into a sum of contributions from each of the model input variables, to interpret the trained neural decoders. The fNIRS data notation is structured to facilitate the understanding of the input data in each experiment. The data is arranged into a tensor  $X \in \mathbb{R}^{trial \times C \times S}$  initially, where *trial* is the number of trials for auditory and silence conditions, *C* is twice the number of fNIRS channels (accounting for HbO and HbR signals), and *S* =  $f_s \times T$  corresponds to the number of sampling points, with  $f_s = 3.9$  Hz being the sampling frequency, and  $T = 18$  seconds is the sampling duration;  $S = 70.2$ , so we round it up to 71. To implement the SHAP-based analysis, we reshape  $X$  into  $X' \in \mathbb{R}^{trial \times C \cdot S}$ , to allow the simultaneous analysis of the temporal and spatial features learned by the decoders.

After the SVC and LDA models were trained to distinguish between the silent condition and speech perception, we performed the SHAP-based analysis. The SHAP explainer was applied to all training samples, and SHAP values were computed for all test samples in each fold. Since this is a binary classification task, SHAP decomposes each prediction score into contributions toward the positive and negative classes. For each sample, the SHAP values indicate how individual features contribute to the prediction; positive SHAP values for samples from positive class (silence condition) and negative SHAP values for samples of the negative class (auditory condition) help predictions. To assess the global importance of each feature, we multiply each instance SHAP values by its corresponding label  $-1, 1$ ; in this way, we ensure that positive values per instance mean the feature contributes to the prediction. Next, we average them across all samples from the test set in each folder. The resulting shape after averaging

is ( $\text{number of test samples} \times \text{number of features}$ ). This array is then reshaped back to the shape of  $X$  and its values are normalized to the  $[0, 1]$  range. Subsequently, we sum the coefficients separately along the spatial (channel) and temporal (time point) dimensions for each subject across all folds from the outer cross-validation. Since the number of channels varies across subjects due to preprocessing (optodes with poor scalp connectivity were discarded), the spatial analysis is conducted at the brain region level rather than at the individual channel level. Features derived from HbO and HbR signals are analyzed separately to provide a more detailed evaluation of their contributions to the decoding task.

### III. RESULTS

1) *Neural Decoder's Performance*: We computed the overall AUC for each subject by averaging performance across the 5 outer folds, including the standard deviation (Table I). All models achieved AUCs significantly above the chance level (50%), with  $p$ -values  $\leq 0.05$ . In addition, when we statistically compared the performance between LDA and linear SVC models using ANOVA, no significant difference in performance between the two models was found. Notably, the models for subjects 1 and 4 achieve the highest AUC scores ( $\geq 0.9$  with SVC). Thus, we emphasize results from these two subjects, as the correct interpretation of the model behavior depends on the model performances [32].

TABLE I: Decoding performances.

	sub.0	sub.1	sub.2	sub.3	sub.4	sub.5	sub.6	sub.7
SVC	0.76 ± 0.17	0.90 ± 0.12	0.63 ± 0.16	0.63 ± 0.16	0.98 ± 0.05	0.72 ± 0.07	0.85 ± 0.20	0.63 ± 0.14
LDA	0.79 ± 0.17	0.84 ± 0.16	0.63 ± 0.11	0.58 ± 0.07	0.94 ± 0.07	0.64 ± 0.1	0.73 ± 0.16	0.67 ± 0.16

2) *Post-hoc analysis*: The Fig. 1 displays the spatiotemporal distribution of relevant features learned by SVC and LDA, respectively. The subject-wise relevant spatiotemporal features indicate that both HbO and HbR signals contribute almost equally to the decoding task for all subjects. Subject 4, who achieved the highest classification performance, exhibited a distinct pattern characterized by broad relevance across the left IFG, bilateral auditory-A regions (Au\_A\_L and Au\_A\_R), and bilateral auditory-B regions (Au\_B\_L and Au\_B\_R), with a slightly stronger emphasis on the left IFG and Au\_B\_L, for both classifiers. This pattern contrasts with that of other subjects, whose relevant features tended to be more localized, primarily within either the IFG or the auditory-B region. Temporally, subject 4 showed a peak in feature relevance within the 7–10 second window post-stimulus onset. Similarly, subject 1 exhibited a comparable distribution of relevant features but with a more pronounced focus on the left IFG and Au\_B\_R areas and stimulus onset, followed by a second peak approximately 5 seconds later for both classifiers.

Overall, the most informative features identified by the LDA model generally aligned with those highlighted by the SVC model. Table II summarizes the two most relevant spatiotemporal features for each subject based on Fig. 1. Spatially, the

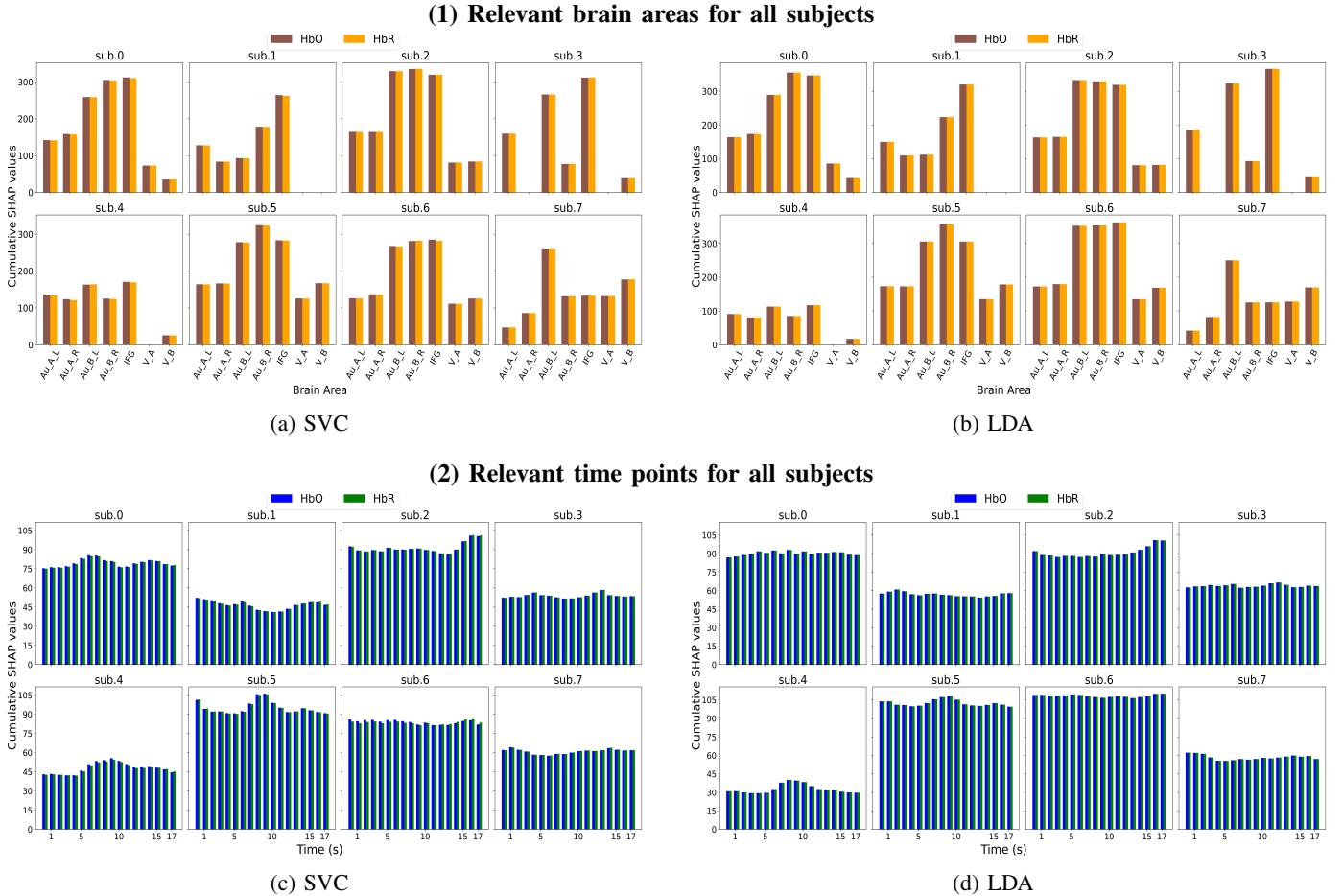


Fig. 1: Subject-wise relevant spatiotemporal features identified by SHAP for Linear SVC and LDA.

SHAP results from both models were largely consistent, except for subject 0, where the SVC model emphasized the left IFG as the most relevant region, whereas the LDA model highlighted (Au\_B\_R). Temporally, both models selected closely aligned time points for most subjects, except for subject 3. Although the exact time points differed in this case, both models showed similar temporal patterns as depicted in Fig. 1.

3) *Group-level post-hoc analysis:* Based on the SHAP analyses for both SVC and LDA models (Fig. 1), we averaged the SHAP values of all spatiotemporal features, as shown in Fig. 2, which also includes the corresponding standard deviations. In terms of spatial relevance, both models consistently highlight three key brain regions as the most informative: the bilateral auditory-B regions (Au\_B\_L and Au\_B\_R) and the left IFG, while assigning minimal importance to the visual region. Regarding temporal relevance, both models indicate that relevant features are distributed throughout the entire event duration, with a subtle peak observed during the onset of the event (0–1 second) and at the mid-event period (6–10 seconds).

#### IV. DISCUSSION

Our results show partial alignment with the findings reported in [18], whose dataset is used in this study. For Subject

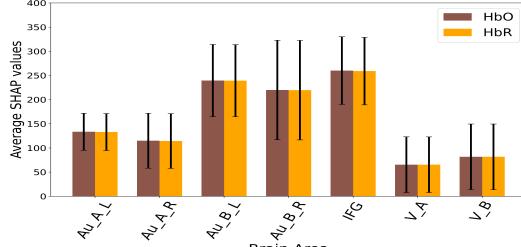
TABLE II: Top 2 most relevant brain areas and time points.

Method	Spatial		Temporal (seconds)	
	SVM	LDA	SVM	LDA
sub.0	IFG *	Au_B_R *	6 HbO 7 HbO	8 *
sub.1	IFG *	IFG *	0 *	2 *
sub.2	Au_B_R *	Au_B_L *	16 HbR 17 HbR	16 *
sub.3	IFG *	IFG *	13*	8 *
sub.4	IFG *	IFG *	9 *	8*
sub.5	Au_B_R *	Au_B_R *	9 *	9 *
sub.6	IFG HbO Au_B_R HbR	IFG *	0 HbO 16 HbR	16,17 *
sub.7	Au_B_L *	Au_B_L *	1 *	0,1*

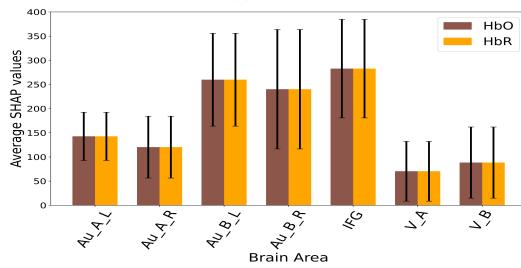
Au\_A\_R and Au\_A\_L refers to right and left auditory A. Au\_B\_R and AU\_B\_L refers to right and left auditory B, and V\_A refers to visual A region. \* indicates both HbR and HbO

4, who achieved the highest classification performance, both models indicate high relevance of the left auditory-A region, and features from the visual areas (V\_A and V\_B) were not relevant, aligning with findings in [18]. However, in contrast to [18], our results showed that the left IFG and auditory-B

### (1) Relevant brain areas across all subjects

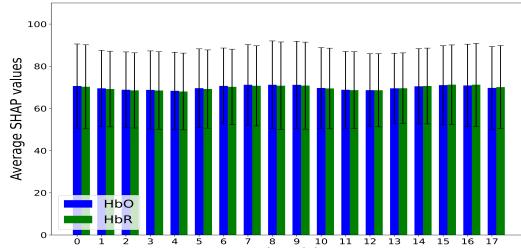


(a) SVC

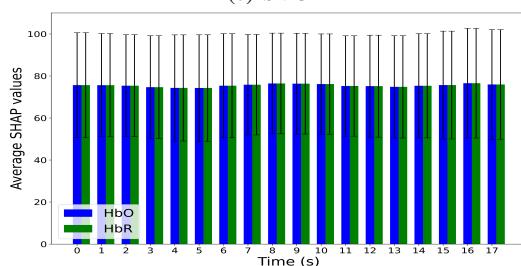


(b) LDA

### (2) Relevant time points across all subjects



(c) SVC



(d) LDA

Fig. 2: Averaged SHAP-based spatiotemporal feature relevance across all subjects for linear SVC and LDA .

regions exhibited the highest relevance. A similar pattern was observed for Subject 1, where the most relevant features were located in the left IFG and Au\_B\_R, along with contributions from the auditory-A region.

Regarding temporal dynamics, Subject 4 showed a peak in feature relevance within the 7–10 second window following stimulus onset, which aligns with the peak hemodynamic response reported by [22] for speech perception. In contrast, Subject 1 exhibited an early peak at stimulus onset, potentially reflecting the initial dip described by [33], followed by the second peak approximately 5 seconds later, close to the expected

peak of the hemodynamic response. Similarly, Subjects 6 and 7 also showed elevated feature relevance at stimulus onset (Table II), which may likewise be associated with the initial dip of the fNIRS signal.

At the group level, the most consistently relevant regions were the left IFG and auditory-B, with minimal contribution from auditory-A. This is consistent with other fNIRS studies that report significant activity in the left IFG in response to auditory speech stimuli [15]–[17]. It is well-known that IFG, which includes Broca’s area, plays a key role in phonological processing and verbal working memory during language tasks [13]. Regarding the auditory-B region, which corresponds to the caudal aspect of the superior temporal gyrus (known as Wernicke’s area), previous fNIRS studies have reported significant activation in response to auditory stimuli in the superior temporal lobe, a broader region that contains Wernicke’s area [15], [16], [19]. Notably, Wernicke’s area is known to be critical for speech comprehension [12]. Moreover, spatial relevance patterns derived from both models suggest limited cross-modal activation in the visual cortex across subjects, which matches the findings in [18], [34].

Interestingly, our results showed that neither the individual nor the group-level analyses revealed a distinct temporal pattern for HbR relative to HbO, despite previous reports suggesting a delayed onset of the HbR response compared to HbO. Moreover, temporal results show a subtle peak in feature relevance at the onset of the event (0–1 second) and another during the mid-event period (6–10 seconds), which may correspond to the initial dip and the delayed peak of the hemodynamic response, respectively. These findings underscore the slow nature of the hemodynamic response and the delay between neural activity and the fNIRS signal, which limits its suitability for real-time BMIs. To address this limitation, recent studies have explored the utility of the initial dip, an early change in oxygenation that occurs before the subsequent rise in cerebral blood flow as response of any stimuli. For instance, [35] leveraged the initial dip to enhance the decoding of finger-tapping tasks, showing its potential for improving the responsiveness of fNIRS-based BMIs.

## V. CONCLUSIONS

In this study, we trained linear SVC and LDA models to decode perceived speech from silence using fNIRS signals. We evaluated the spatial and temporal relevance of learned features using SHAP. Our findings consistently highlight left IFG and Wernicke’s area (Auditory B region) as a key region for speech decoding, particularly in high-performing subjects such as subject 4 and subject 1 and across all subjects. These findings align with established neurophysiological processes for which left IFG, which includes Broca’s area, plays a key role in phonological processing and verbal working memory during language tasks [13] and Wernicke’s area, which is known to be critical for speech comprehension [36]. For temporal relevance, in general, relevant features are distributed throughout the entire event duration for all subjects, with a subtle peak observed during the onset of the event (0–1 second)

and at the mid-event period (6–10 seconds) that may be related with the initial dip and the delay of the hemodynamic response measured by fNIRS.

Further research should investigate whether prioritizing activity from critical brain regions, particularly the auditory-B region and the left IFG, and incorporating attention mechanisms to emphasize features extracted during the onset and mid-event periods, when the peak of the hemodynamic response happen, can enhance decoding accuracy. Additionally, it is important to validate the analyses proposed in this study using a larger dataset, as the current evaluation is based on only 18 auditory trials and 10 silence control trials.

## REFERENCES

- [1] Y. Wang, Y. Tang, Q. Wang, M. Ge, J. Wang, X. Cui, N. Wang, Z. Bao, S. Chen, J. Wang *et al.*, “Advances in brain computer interface for amyotrophic lateral sclerosis communication,” *Brain-X*, vol. 3, no. 1, p. e70023, 2025.
- [2] Z. Ye, Q. Ai, Y. Liu, M. de Rijke, M. Zhang, C. Lioma, and T. Ruotsalo, “Generative language reconstruction from brain recordings,” 2024.
- [3] A. Défossez, C. Caucheteux, J. Rapin, O. Kabeli, and J.-R. King, “Decoding speech perception from non-invasive brain recordings,” *Nature Machine Intelligence*, vol. 5, no. 10, pp. 1097–1107, 2023.
- [4] S. Zhang, E. Alam, J. Baber, F. Bianco, E. Turner, M. Chamanzar, and H. Dehghani, “Mindgpt: Advancing human-ai interaction with non-invasive fnirs-based imagined speech decoding,” *arXiv preprint arXiv:2408.05361*, 2024.
- [5] P. Pinti, I. Tachtsidis, A. Hamilton, J. Hirsch, C. Aichelburg, S. Gilbert, and P. W. Burgess, “The present and future use of functional near-infrared spectroscopy (fnirs) for cognitive neuroscience,” *Annals of the New York Academy of Sciences*, vol. 1464, no. 1, pp. 5–29, 2020.
- [6] M. Rybář, R. Poli, and I. Daly, “Decoding of semantic categories of imagined concepts of animals and tools in fnirs,” *Journal of Neural Engineering*, vol. 18, no. 4, 2021.
- [7] L. Cao, D. Huang, Y. Zhang, X. Jiang, and Y. Chen, “Brain decoding using fnirs,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 14, 2021, pp. 12 602–12 611.
- [8] S. Posso-Murillo, L. G. Sanchez-Giraldo, and J. Bae, “Semantic reconstruction from fnirs using recurrent neural networks,” in *2025 IEEE 22nd International Symposium on Biomedical Imaging*. IEEE, 2025.
- [9] S. Zhang, E. Alam, J. Baber, F. Bianco, E. Turner, M. Chamanzar, and H. Dehghani, “Mindspeech: Continuous imagined speech decoding using high-density fnirs and prompt tuning for advanced human-ai interaction,” *arXiv preprint arXiv:2408.05362*, 2024.
- [10] J. Obleser, R. J. Wise, M. A. Dresner, and S. K. Scott, “Functional integration across brain regions improves speech perception under adverse listening conditions,” *Journal of Neuroscience*, vol. 27, no. 9, pp. 2283–2289, 2007.
- [11] B. Khalighinejad, P. Patel, J. L. Herrero, S. Bickel, A. D. Mehta, and N. Mesgarani, “Functional characterization of human heschl’s gyrus in response to natural speech,” *Neuroimage*, vol. 235, 2021.
- [12] J. R. Binder, “The wernicke area: Modern evidence and a reinterpretation,” *Neurology*, vol. 85, no. 24, pp. 2170–2175, 2015.
- [13] V. J. Hodgson, M. A. L. Ralph, and R. L. Jackson, “Multiple dimensions underlying the functional organization of the language network,” *NeuroImage*, vol. 241, p. 118444, 2021.
- [14] J. A. Schroën, T. C. Gunter, O. Numssen, L. O. Kroczeck, G. Hartwigsen, and A. D. Friederici, “Causal evidence for a coordinated temporal interplay within the language network,” *Proceedings of the National Academy of Sciences*, vol. 120, no. 47, p. e2306279120, 2023.
- [15] R. J. Lawrence, I. M. Wiggins, C. A. Anderson, J. Davies-Thompson, and D. E. Hartley, “Cortical correlates of speech intelligibility measured using functional near-infrared spectroscopy (fnirs),” *Hearing research*, vol. 370, pp. 53–64, 2018.
- [16] R. J. Lawrence, I. M. Wiggins, J. C. Hodgson, and D. E. Hartley, “Evaluating cortical responses to speech in children: A functional near-infrared spectroscopy (fnirs) study,” *Hearing research*, vol. 401, p. 108155, 2021.
- [17] P. Wijayasiri, D. E. Hartley, and I. M. Wiggins, “Brain activity underlying the recovery of meaning from degraded speech: A functional near-infrared spectroscopy (fnirs) study,” *Hearing research*, vol. 351, pp. 55–67, 2017.
- [18] M. J. Shader, R. Luke, N. Gouailhardou, and C. M. McKay, “The use of broad vs restricted regions of interest in functional near-infrared spectroscopy for measuring cortical activation to auditory-only and visual-only speech,” *Hearing research*, vol. 406, p. 108256, 2021.
- [19] I. M. Butera, E. D. Larson, A. J. DeFreese, A. K. Lee, R. H. Gifford, and M. T. Wallace, “Functional localization of audiovisual speech using near infrared spectroscopy,” *Brain Topography*, vol. 35, no. 4, pp. 416–430, 2022.
- [20] Z. Wang, J. Fang, and J. Zhang, “Rethinking delayed hemodynamic responses for fnirs classification,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2023.
- [21] M. Uga, I. Dan, T. Sano, H. Dan, and E. Watanabe, “Optimizing the general linear model for functional near-infrared spectroscopy: an adaptive hemodynamic response function approach,” *Neurophotonics*, vol. 1, no. 1, pp. 015 004–015 004, 2014.
- [22] R. Luke, E. Larson, M. J. Shader, H. Innes-Brown, L. Van Yper, A. K. Lee, P. F. Sowman, and D. McAlpine, “Analysis methods for measuring passive auditory fnirs responses generated by a block-design paradigm,” *Neurophotonics*, vol. 8, no. 2, pp. 025 008–025 008, 2021.
- [23] S. Lundberg, “A unified approach to interpreting model predictions,” *arXiv preprint arXiv:1705.07874*, 2017.
- [24] C. J. Shibu, S. Sreedharan, K. Arun, C. Kesavadas, and R. Sitaram, “Explainable artificial intelligence model to predict brain states from fnirs signals,” *Frontiers in Human Neuroscience*, vol. 16, 2023.
- [25] P. Hui, Y. Jiang, J. Wang, C. Wang, Y. Li, B. Fang, H. Wang, Y. Wang, and S. Qie, “Exploring the application and challenges of fnirs technology in early detection of parkinson’s disease,” *Frontiers in Aging Neuroscience*, vol. 16, p. 1354147, 2024.
- [26] L. Mao, X. Hong, and M. Hu, “Identifying neuroimaging biomarkers in major depressive disorder using machine learning algorithms and functional near-infrared spectroscopy (fnirs) during verbal fluency task,” *Journal of Affective Disorders*, vol. 365, pp. 9–20, 2024.
- [27] X. Cui, S. Bray, and A. L. Reiss, “Functional near infrared spectroscopy (fnirs) signal improvement based on negative correlation between oxygenated and deoxygenated hemoglobin dynamics,” *Neuroimage*, vol. 49, no. 4, pp. 3039–3046, 2010.
- [28] N. Naseer and K.-S. Hong, “fnirs-based brain-computer interfaces: a review,” *Frontiers in human neuroscience*, vol. 9, 2015.
- [29] H. He, Y. Bai, E. Garcia, and S. A. Li, “Adaptive synthetic sampling approach for imbalanced learning. ieee international joint conference on neural networks,” in *IEEE World Congress On Computational Intelligence*, vol. 2008. IEEE Hong Kong, 2008.
- [30] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, “Scikit-learn: Machine learning in python,” *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.
- [31] J. Benerradi, J. Clos, A. Landowska, M. F. Valstar, and M. L. Wilson, “Benchmarking framework for machine learning classification from fnirs data,” *Frontiers in Neuroergonomics*, vol. 4, 2023.
- [32] N. Kriegeskorte and P. K. Douglas, “Interpreting encoding and decoding models,” *Current opinion in neurobiology*, vol. 55, pp. 167–179, 2019.
- [33] A. Zafar and K.-S. Hong, “Detection and classification of three-class initial dips from prefrontal cortex,” *Biomedical optics express*, vol. 8, no. 1, pp. 367–383, 2016.
- [34] L.-C. Chen, P. Sandmann, J. D. Thorne, C. S. Herrmann, and S. Debener, “Association of concurrent fnirs and eeg signatures in response to auditory and visual stimuli,” *Brain topography*, vol. 28, pp. 710–725, 2015.
- [35] M. U. Ali, A. Zafar, K. D. Kallu, M. A. Yaqub, H. Masood, K.-S. Hong, and M. R. Bhutta, “An isolated cnn architecture for classification of finger-tapping tasks using initial dip images: A functional near-infrared spectroscopy study,” *Bioengineering*, vol. 10, no. 7, 2023.
- [36] M. Roll, “Heschl’s gyrus and the temporal pole: The cortical lateralization of language,” *NeuroImage*, vol. 303, 2024.