Shared Neural Dynamics Between Overt and Inner Speech: An EEG Signal Analysis

1st Kumar Kaustubh
HCLTech, Noida
kumar.kaustubh@hcltech.com

2nd Aastha Bidhen Kachhi
HCLTech, Noida
aasthabidhen.kachhi@hcltech.com

3rd Ashish Issac
HCLTech, Noida
ashishis@hcltech.com

4th Ajay Kumar Sharma
HCLTech, Noida
ajayks@hcltech.com

***Abstract*—Inner speech, the internal representation of language without overt articulation, is a promising paradigm for brain-computer interfaces (BCIs), but its relationship to overt speech at the neural signal level remains unclear. In this work, we analyze EEG signals from the Thinking Out Loud dataset to examine whether neural signatures of overt speech are preserved in inner speech between spectral, spatial, and temporal domains. We find that inner speech preserves the relative spectral organization and exhibits strong spatial similarity with overt speech, particularly in the alpha band (spatial correlation r = 0.99). In contrast, temporal coupling is weak (peak correlation r = 0.15), and frequency-domain analysis reveals low coherence and near-zero phase locking, indicating the absence of precise temporal synchronization. Inter-subject analysis further shows that cross-subject overt-inner similarity is comparable to overt-overt similarity, suggesting a shared spatial structure across individuals. These results indicate that inner speech retains structural neural organization while lacking precise temporal dynamics of overt speech, suggesting that robust BCI systems may benefit more from stable spectral and spatial features than from temporally precise representations.**

***Index Terms*—inner speech, overt speech, EEG, brain-computer interface (BCI), imagined speech.**

I. Introduction

Inner speech, also referred to as verbal thinking, covert self-talk, or internal monologue, denotes the silent experience of language without overt articulation and reflects internal representations of thought [1]. It plays a fundamental role in human cognition, including reasoning, planning, and self-regulation [2]. Although terms such as imagined speech and silent speech are often used interchangeably with inner speech, they differ conceptually. Silent speech involves articulatory movements without acoustic output [3], whereas imagined speech corresponds to the mental simulation of speaking without any articulatory activity [4]. This distinction is important when interpreting neural signals associated with these paradigms. Brain–computer interface (BCI) research has traditionally relied on paradigms such as P300, steady-state visual evoked potentials (SSVEP), and motor imagery for communication and control. While effective, these approaches are limited by user fatigue and dependence on external stimuli or sustained attention [5]. Speech-based paradigms offer a more natural alternative by directly leveraging neural processes underlying language production.

Approaches based on visual articulatory cues, such as lip-reading and speech reconstruction from video, have demonstrated the feasibility of decoding speech from external observations [6], [7], [8], [9]. However, such methods are restricted to articulatory speech and cannot generalize to imagined or inner speech, where no observable movement exists. Additionally, visual ambiguity arising from visemes (e.g., /p/, /b/, /m/) limits the discriminability of such approaches [10]. These limitations motivate the use of neural recording modalities such as electroencephalography (EEG), magnetoencephalography (MEG), and electrocorticography (ECoG), which provide direct access to underlying brain activity.

Prior studies have shown that EEG signals exhibit structured relationships with acoustic speech features such as pitch and intensity during overt speech [11]. However, whether similar structure is preserved in inner speech remains unclear, particularly in the absence of observable behavioral or acoustic references. This challenge is compounded by the difficulty of temporally aligning neural activity with internally generated events [12]. Addressing this gap is critical for developing scalable and data-efficient speech-based BCI systems.

In this work, we analyze the relationship between overt (pronounced) speech and inner speech using EEG signals. Specifically, we investigate whether shared neural structure exists across spectral, spatial, and temporal domains.

The main contributions of this work are as follows:

- A multi-domain analysis to systematically compare overt and inner speech using EEG across spectral, spatial, temporal, and frequency domains.
- A detailed empirical study showing that inner speech preserves spectral and spatial structure of overt speech, while lacking precise temporal and phase alignment.
- Evidence of cross-subject spatial consistency in inner speech, demonstrating that cross-subject overt-inner similarity is comparable to overt-overt similarity, indicating a shared neural representation.

The next section briefly describes related work in this area and provides background.

II. Related Work and Background

Classification of inner speech from EEG has been widely explored. However, ambiguity persists in the distinction between inner, imagined, and silent speech paradigms, often leading to inconsistencies in experimental design and interpretation. Many studies employ task-driven protocols in which subjects are instructed to internally produce specific words or

syllables, aligning more closely with imagined speech than spontaneous inner speech.

Early work demonstrated that EEG signals contain discriminative information for imagined speech under controlled conditions, establishing the feasibility of decoding internally generated linguistic content [13]. Subsequent studies have focused on improving feature representations, incorporating time–frequency analysis and phase-based measures such as inter-trial coherence to capture consistent neural patterns across trials [14].

Both, traditional machine learning and deep learning approaches have been explored for inner speech classification. Recurrent architectures such as LSTM and BiLSTM have shown modest improvements over conventional methods, although overall performance remains limited due to low signal-to-noise ratio and high variability in EEG signals [15]. Addressing these challenges, Tasin et al. [16] emphasize subject-specific modeling and demonstrate that ensemble learning with feature selection can outperform deep learning approaches in low-data settings.

Another line of work focuses on improving signal representation. Wang et al. [17] highlight the importance of frequency-specific cortical representations, showing that spatio-spectral features are essential for capturing the structure of inner speech EEG signals, which are not adequately modeled using purely temporal approaches.

Despite these advances, existing methods primarily focus on improving classification performance through feature engineering or model design, with limited investigation into the underlying relationship between overt and inner speech at the signal level. In particular, it remains unclear whether inner speech preserves structural characteristics of overt speech across spectral, spatial, and temporal domains. To address this, we perform a systematic analysis of EEG signals across overt and inner speech using a controlled experimental dataset, as described in the following section.

III. Dataset and Pre-processing

In this section, we describe the dataset used in this study and the preprocessing steps applied to the EEG signals.

A. Dataset

We utilize the publicly available Thinking Out Loud EEG dataset [18], which provides synchronized recordings of overt (pronounced) and inner speech from 10 healthy subjects using a 128-channel BioSemi EEG system. The experimental paradigm consists of a four-class directional task, where participants either pronounce (overt speech) or internally imagine the semantics (inner speech) of one of the words: *up, down, left, right*. Each trial includes a cue, action, and relaxation interval, with neural activity of interest occurring during the action phase (1.0-3.5 s post cue) [18]. Multiple trials per subject are collected across three sessions for both conditions, enabling paired comparisons. After segmentation, the data can be represented as:

$$X \in \mathbb{R}^{N \times C \times T} \tag{1}$$

where N is the number of trials, $C = 128$ is the number of channels, and T is the number of temporal samples per trial. Across all subjects, the dataset contains approximately 1128 overt and 2236 inner speech trials, providing sufficient samples for subject-level analysis.

B. Pre-processing

We use the preprocessed and epoched EEG signals provided with the dataset, which have undergone standard preprocessing steps including re-referencing, artifact removal, and signal cleaning, as described in [18]. Specifically, the raw EEG signals are re-referenced using ear-lobe electrodes to remove common-mode noise, followed by bandpass filtering (0.5-45 Hz) and notch filtering to suppress power-line interference. Independent Component Analysis (ICA) is further applied to remove artifacts related to eye movements, blinks, and muscle activity.

The continuous recordings are segmented into trials based on the provided task labels. The signals are sampled at 256 Hz, yielding 640 time points per trial for the 2.5-second action phase.

IV. Experiments and Results

We design a series of experiments to investigate whether inner speech preserves the neural characteristics of overt speech across spectral and spatial dimensions. All analyses are based on trial-averaged means, ensuring no systematic bias is introduced.

1) ***Experiment 1: Spectral Profile Analysis****:* The power spectral density (PSD) of EEG signals is estimated using Welch's method. For each trial signal $x_c[n]$ from channel c, the signal is segmented into overlapping 1-second windows, and the PSD is computed as:

$$P_c(f) = \frac{1}{K} \sum_{k=1}^{K} \frac{1}{L} \left| \sum_{n=0}^{L-1} x_c^{(k)}[n]\, w[n]\, e^{-j2\pi f n} \right|^2 \tag{2}$$

where $w[n]$ is a Hann window and K is the number of segments. The final PSD is obtained by averaging across channels, trials, and subjects:

$$P(f) = \frac{1}{SNC} \sum_{s=1}^{S} \sum_{i=1}^{N} \sum_{c=1}^{C} P_{s,i,c}(f) \tag{3}$$

Fig. 1 shows the grand-average PSD for overt and inner speech. Both conditions exhibit the characteristic $1/f$ spectral profile of EEG [19], with dominant low-frequency activity and a clear alpha-band peak. The spectral shapes are highly similar across conditions, indicating that the overall spectral envelope of EEG activity is comparable between overt and inner speech.

Overt speech shows higher power in low-frequency bands compared to inner speech, reflecting differences in magnitude while maintaining a similar spectral structure.

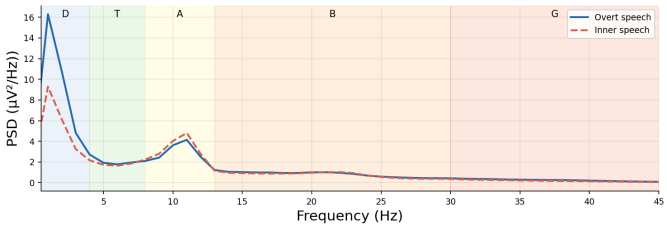

Fig. 1: Grand-average power spectral density (PSD) of overt and inner speech, averaged across subjects, trials, and channels.

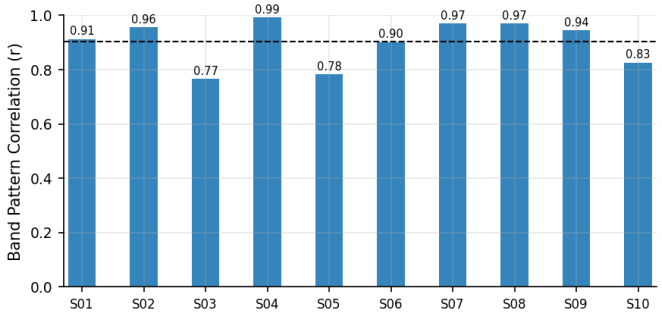

Fig. 2: Band pattern correlation between overt and inner speech for each subject. Bars represent individual subjects S01–S10; the y-axis shows mean band power in $\mu V^2/Hz$. The dashed line indicates the mean correlation across subjects.

*2) **Experiment 2: Band-wise Power Distribution**:* To quantify spectral differences at the band level, mean band power is computed for each subject by integrating the PSD within canonical EEG frequency bands and averaging across trials and channels. This yields two five-dimensional vectors per subject:

$$\mathbf{x} = \left[P_\delta^{\text{ov}}, P_\theta^{\text{ov}}, P_\alpha^{\text{ov}}, P_\beta^{\text{ov}}, P_\gamma^{\text{ov}}\right] \tag{4}$$

$$\mathbf{y} = \left[P_\delta^{\text{in}}, P_\theta^{\text{in}}, P_\alpha^{\text{in}}, P_\beta^{\text{in}}, P_\gamma^{\text{in}}\right] \tag{5}$$

Band-wise differences are evaluated using paired t-tests across subjects, and spectral similarity is quantified using the Pearson correlation between $\mathbf{x}$ and $\mathbf{y}$.

Delta and theta bands show significantly higher power in overt speech ($p < 0.01$), while no significant differences are observed in beta, and gamma bands (Table I). Alpha power is slightly higher in inner speech, although the difference is not statistically significant. The similarity of spectral organization is further supported by strong correlation between band-power profiles across subjects (mean $r = 0.90 \pm 0.08$), as shown in Fig. 2. This indicates that the relative spectral distribution is highly preserved between overt and inner speech despite differences in absolute magnitude.

*3) **Experiment 3: Spatial Topography Analysis**:* To examine whether spatial patterns of neural activation are preserved, we compute band-wise channel power and visualize scalp topographies for both overt and inner speech. For each frequency band, channel-wise power is averaged across trials and subjects to obtain grand-mean spatial maps. Fig. 3 shows the topographic distribution of band-limited EEG power for both conditions across all frequency bands. The spatial patterns exhibit strong similarity between overt and inner speech, with consistent activation distributions across channels. The difference maps further indicate that variations are primarily in magnitude rather than spatial organization.

TABLE I: Band power comparison between overt and inner speech (mean $\pm$ std, $N = 10$)

Band	Overt	Inner	t-statistic	p-value
Delta	31.85 $\pm$ 14.60	18.66 $\pm$ 6.56	$t(9) = 4.531$	0.001
Theta	8.36 $\pm$ 3.53	7.38 $\pm$ 3.08	$t(9) = 3.476$	0.007
Alpha	14.77 $\pm$ 7.77	16.64 $\pm$ 9.87	$t(9) = -2.026$	0.073
Beta	14.10 $\pm$ 7.05	13.36 $\pm$ 7.82	$t(9) = 0.885$	0.399
Gamma	3.84 $\pm$ 2.02	2.85 $\pm$ 1.80	$t(9) = 2.135$	0.062

To quantify spatial similarity, Pearson correlation is computed between overt and inner channel-power vectors for each band. The resulting correlations are high across all bands (delta: $r = 0.684$, theta: $r = 0.859$, alpha: $r = 0.993$, beta: $r = 0.876$, gamma: $r = 0.855$), confirming strong preservation of spatial structure.

To further assess consistency at the individual level, Fig. 4 shows per-subject topographies in the alpha band. Each subject exhibits highly similar spatial patterns between overt and inner speech, with consistently high correlations (mean $r = 0.97 \pm 0.03$).

*4) **Experiment 4: Temporal Coupling Analysis**:* To investigate the temporal relationship between overt and inner speech, we analyze the coupling between their neural dynamics using normalized cross-correlation (XC) of EEG signal envelopes.

For each channel, the analytic signal is computed using the Hilbert transform, and the amplitude envelope is extracted as:

$$e(t) = |\mathcal{H}(x(t))| \tag{6}$$

where $\mathcal{H}(\cdot)$ denotes the Hilbert transform. Trial-averaged envelopes are then computed separately for overt and inner speech.

Temporal similarity is quantified using normalized cross-correlation:

$$CC(\tau) = \frac{\sum_t a(t)\, b(t+\tau)}{\sqrt{\sum_t a(t)^2}\,\sqrt{\sum_t b(t)^2}} \tag{7}$$

where $a(t)$ and $b(t)$ denote zero-mean overt and inner speech envelopes, respectively, and τ represents temporal lag. Positive lag values indicate that overt speech leads inner speech.

For each channel, the peak correlation and corresponding lag are obtained as:

$$\tau^* = \arg\max_\tau |CC(\tau)|, \quad r^* = CC(\tau^*) \tag{8}$$

Channel-level values are averaged within each subject to obtain one peak correlation and one peak lag per subject. Group-level statistics are then computed across subjects using one-sample t-tests against zero.

As shown in Table II, the peak cross-correlation is low but statistically significant, indicating weak temporal coupling

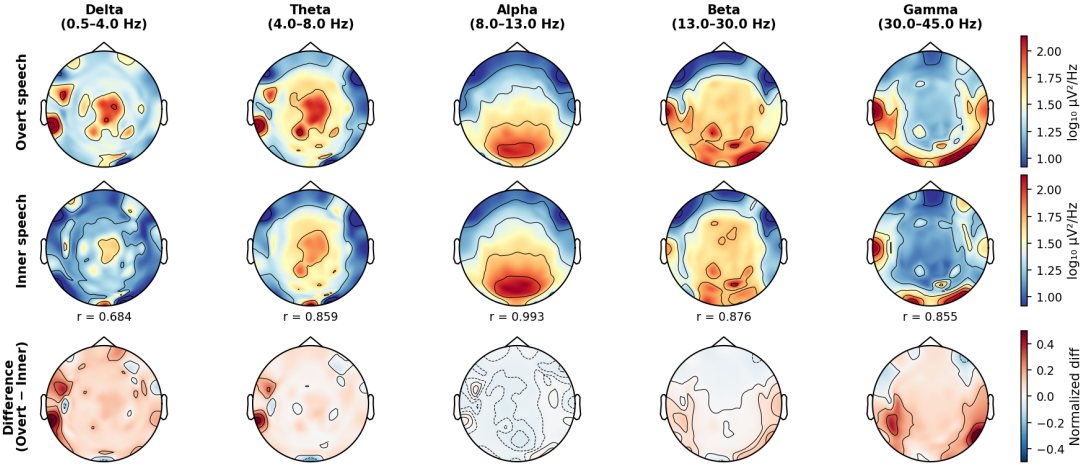

Fig. 3: Grand-mean topographic maps of band-wise EEG power for overt and inner speech across frequency bands, averaged over all trials and subjects. Power is computed for each channel and mapped to its known 3D scalp position using the standard BioSemi-128 montage. Rows correspond to overt speech, inner speech, and their difference.

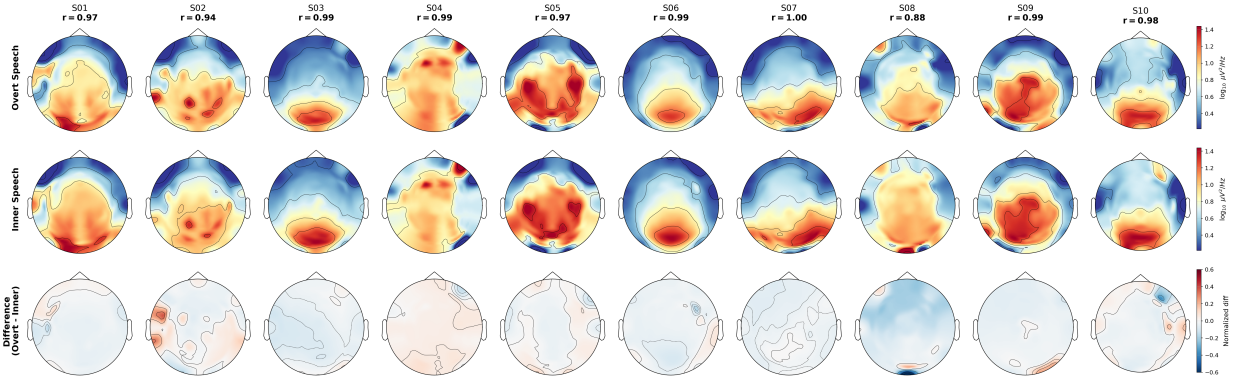

Fig. 4: Per-subject scalp topographies of alpha-band EEG power for overt and inner speech. Each column corresponds to one subject, with rows representing overt speech, inner speech, and their difference.

TABLE II: Temporal coupling between overt and inner speech

Metric	Mean ± Std	t-statistic	p-value
Peak XC	0.145 ± 0.171	$t(9) = 2.55$	0.031
Peak lag (ms)	47.24 ± 43.85	$t(9) = 3.23$	0.010

between overt and inner speech. The positive and significant lag suggests that overt speech tends to precede inner speech. However, the low correlation magnitude and high inter-subject variability indicate that temporal alignment is weak and inconsistent.

These results suggest that inner speech does not preserve the precise temporal dynamics of overt speech, but instead reflects a loosely aligned, higher-level neural process.

*5) **Experiment 5: Spectral Coherence and Phase Synchronisation**:* To examine frequency-specific coupling between overt and inner speech, we compute trial-pair spectral coherence and phase locking value (PLV) [20]. Coherence captures shared amplitude-frequency structure, while PLV measures the consistency of phase alignment across trials. All statistics are computed at the subject level by averaging trial-level measures within each subject.

TABLE III: **Frequency-domain coupling between overt and inner speech**

Measure	Metric	Mean ± Std	t-stat	p-value
Coherence	Low-freq	0.2670 ± 0.0026	$t(9) = -9.15$	< 0.001
	High-freq	0.2565 ± 0.0016		
PLV	Low-freq	0.0288 ± 0.0036	$t(9) = -13.00$	< 0.001
	High-freq	0.0127 ± 0.0021		

As shown in Table III, coherence values are low across all frequency bands, indicating limited shared variance between overt and inner speech at the trial level. Although low-frequency coherence is significantly higher than high-frequency coherence, the absolute difference is small, indicating weak coupling overall.

PLV values are near zero across all bands, confirming the absence of consistent phase synchronisation between overt and inner speech. While low-frequency PLV exceeds high-frequency PLV, the magnitude remains negligible. Together, these results indicate that shared neural activity between overt and inner speech is weak and predominantly confined to low-frequency components. The absence of phase locking further suggests that inner speech does not preserve the precise

temporal structure of overt speech, supporting the view that it reflects a higher-level neural representation rather than a temporally precise motor process.

*6) **Experiment 6: Inter-Subject Spatial Universality**:* To evaluate whether inner speech preserves a shared spatial structure across subjects, we compute pairwise Pearson correlations of band-limited scalp topographies. For each frequency band, three comparisons are considered: overt–overt (between subjects), inner-inner (between subjects), and overt-inner (cross-subject).

TABLE IV: Inter-subject spatial correlation across frequency bands

Band	Overt–Overt	Inner–Inner	Overt–Inner (cross)
Delta	0.024 ± 0.196	0.043 ± 0.142	0.092 ± 0.260
Theta	0.163 ± 0.202	0.203 ± 0.173	0.246 ± 0.265
Alpha	$\mathbf{0.537 \pm 0.189}$	$\mathbf{0.525 \pm 0.191}$	$\mathbf{0.574 \pm 0.225}$
Beta	0.185 ± 0.272	0.141 ± 0.247	0.227 ± 0.326
Gamma	0.157 ± 0.290	0.129 ± 0.259	0.212 ± 0.338

As shown in Table IV, cross-subject overt-inner correlations are comparable to, and in some cases slightly higher than, overt-overt between-subject correlations, indicating that inner speech preserves a shared spatial structure across individuals. This effect is most pronounced in the alpha band, where cross-subject overt-inner similarity ($r = 0.574$) closely matches overt-overt similarity ($r = 0.537$) and exceeds that of other bands. This may reflect greater trial-level variability in overt speech due to motor execution differences, or the higher inner speech trial count yielding more stable averaged topographies. In contrast, delta and high-frequency bands (beta, gamma) exhibit low correlations with high variability, indicating weaker and less stable shared structure.

These findings suggest that spatial consistency is frequency-dependent and strongest in the alpha band, supporting the interpretation that inner speech preserves a common neural architecture in frequency ranges associated with large-scale cognitive processing.

V. Conclusion and Future Work

Across all experiments, we observe a consistent pattern: inner speech exhibits reduced signal magnitude compared to overt speech, while preserving its spectral and spatial structure. Strong band-wise correlations and highly similar scalp topographies, particularly in the alpha band, indicate shared underlying neural organization. However, temporal analysis reveals only weak coupling, and frequency-domain results show low coherence and near-zero phase locking, indicating the absence of precise temporal and phase alignment. These findings suggest that inner speech reflects a higher-level neural representation that retains structural characteristics of overt speech without reproducing its detailed temporal dynamics. These observations have direct implications for BCI design, indicating that inner speech decoding may benefit from relying on stable spectral and spatial features rather than precise temporal or phase-based representations, which appear less reliable across conditions.

Future work will focus on using these stable spectral and spatial features for robust inner speech decoding, as well as exploring subject-independent models to better utilize the observed population-level consistency. Similar analyses can also be extended to imagined and silent speech.

References

[1] B. Alderson-Day and C. Fernyhough, "Inner speech: Development, cognitive functions, phenomenology, and neurobiology." *Psychological bulletin*, vol. 141, no. 5, p. 931, 2015.

[2] C. Gauker, "Inner speech as the internalization of outer speech," *Inner speech: New voices*, vol. 53, 2018.

[3] B. Denby, T. Schultz, K. Honda, T. Hueber, J. M. Gilbert, and J. S. Brumberg, "Silent speech interfaces," *Speech Communication*, vol. 52, no. 4, pp. 270–287, 2010.

[4] T. Schultz, "Bio signal-based spoken communication," in *Proc. IberSPEECH 2018*, 2018.

[5] R. Abiri, S. Borhani, E. W. Sellers, Y. Jiang, and X. Zhao, "A comprehensive review of eeg-based brain–computer interface paradigms," *Journal of neural engineering*, vol. 16, no. 1, p. 011001, 2019.

[6] K. Prajwal, R. Mukhopadhyay, V. P. Namboodiri, and C. Jawahar, "Learning individual speaking styles for accurate lip to speech synthesis," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 13 796–13 805.

[7] A. Ephrat and S. Peleg, "Vid2speech: speech reconstruction from silent video," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017, pp. 5095–5099.

[8] C. Hu, Q. Tian, T. Li, W. Yuping, Y. Wang, and H. Zhao, "Neural dubber: Dubbing for videos according to scripts," *Advances in neural information processing systems*, vol. 34, pp. 16 582–16 595, 2021.

[9] Y. Kumar, R. Jain, K. M. Salik, R. R. Shah, Y. Yin, and R. Zimmermann, "Lipper: Synthesizing thy speech using multi-view lipreading," in *Proceedings of the AAAI Conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 2588–2595.

[10] H. L. Bear and R. Harvey, "Decoding visemes: Improving machine lip-reading," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 2009–2013.

[11] P. A. Abhang and B. W. Gawali, "Correlation of eeg images and speech signals for emotion analysis," *British Journal of Applied Science & Technology*, vol. 10, no. 5, pp. 1–13, 2015.

[12] S. Martin, I. Iturrate, J. d. R. Millán, R. T. Knight, and B. N. Pasley, "Decoding inner speech using electrocorticography: Progress and challenges toward a speech prosthesis," *Frontiers in neuroscience*, vol. 12, p. 422, 2018.

[13] M. D'Zmura, S. Deng, T. Lappas, S. Thorpe, and R. Srinivasan, "Toward eeg sensing of imagined speech," in *International Conference on Human-Computer Interaction*. Springer, 2009, pp. 40–48.

[14] D. Lopez-Bernal, D. Balderas, P. Ponce, and A. Molina, "Exploring inter-trial coherence for inner speech classification in eeg-based brain–computer interface," *Journal of Neural Engineering*, vol. 21, no. 2, p. 026048, 2024.

[15] F. Gasparini, E. Cazzaniga, and A. Saibene, "Inner speech recognition through electroencephalographic signals," *arXiv preprint arXiv:2210.06472*, 2022.

[16] S. M. Tasin, M. E. Chowdhury, S. Pedersen, M. Chabbouh, D. Bushnaq, R. Aljindi, S. Kabir, and A. Hasan, "Ensemble machine learning model for inner speech recognition: A subject-specific investigation," *IEEE Access*, vol. 14, pp. 10 811–10 827, 2026.

[17] Y. Wang, H. Cai, L. Zhou, Z. Yang, L. Wang, J. Zhao, Z. Kuang, H. Wang, and D. Xu, "Decoding inner speech via frequency-specific cortical eeg representations," *Information Processing & Management*, vol. 63, no. 5, p. 104676, 2026.

[18] N. Nieto, V. Peterson, H. L. Rufiner, J. E. Kamienkowski, and R. Spies, "Thinking out loud, an open-access eeg-based bci dataset for inner speech recognition," *Scientific data*, vol. 9, no. 1, p. 52, 2022.

[19] C. Demanuele, C. J. James, and E. J. Sonuga-Barke, "Distinguishing low frequency oscillations within the 1/f spectral behaviour of electromagnetic brain signals," *Behavioral and Brain Functions*, vol. 3, no. 1, p. 62, 2007.

[20] J.-P. Lachaux, E. Rodriguez, J. Martinerie, and F. J. Varela, "Measuring phase synchrony in brain signals," *Human brain mapping*, vol. 8, no. 4, pp. 194–208, 1999.