Smart Health 36 (2025) 100568

Contents lists available at ScienceDirect

Smart Health

journal homepage: www.elsevier.com/locate/smhl

A low-channel EEG-to-speech conversion approach for assisting people with communication disorders

Kunning Shen [a], Huining Li [b,*]

[a] Department of Electrical and Computer Engineering, Pratt School of Engineering, Duke University, Durham, 27708, NC, USA
[b] Department of Computer Science, North Carolina State University, Raleigh, 27606, NC, USA

ARTICLE INFO

Keywords:
Brain–Computer Interface
EEG-to-speech

ABSTRACT

Brain–Computer Interface (BCI) technology has emerged as a promising solution for individuals with communication disorders. However, current electroencephalography (EEG) to speech systems typically require high-channel EEG equipment (64+ channels), limiting their accessibility in resource-constrained environments. This paper implements a novel low-channel EEG-to-speech framework that effectively operates with only 6 EEG channels. By leveraging a generator-discriminator architecture for speech reconstruction, our system achieves a Character Error Rate (CER) of 64.24%, outperforming baseline systems that utilize 64 channels (68.26% CER). We further integrate Undercomplete Independent Component Analysis (UICA) for channel reduction, maintaining comparable accuracy (64.99% CER) while reducing computational complexity from 6 channels to 4 channels. This breakthrough demonstrates the feasibility of efficient speech reconstruction from minimal EEG inputs, potentially enabling more widespread deployment of BCI technology in resource-limited healthcare settings.

1. Introduction

Speech communication disorders significantly impact the quality of life for millions of individuals worldwide. Recent global burden studies revealed that among 2.6 billion children and adolescents, over 291.2 million (11.2%) are affected by various disabilities including sensory impairments that affect communication (Olusanya et al., 2020). Notably, 95% of these individuals live in low- and middle-income countries, where healthcare resources are often limited. The prevalence of these disabilities shows an alarming increasing trend with age, rising from 6.1% in infants to 13.9% in adolescents (Olusanya et al., 2020). For patients who have lost their ability to speak due to stroke, Amyotrophic Lateral Sclerosis (ALS), or other neurological conditions, the inability to engage in basic verbal communication not only affects their daily lives but also leads to significant psychological burden and social isolation (Palmer et al., 2019).

Traditional assistive communication solutions primarily rely on physical movements, such as eye-tracking devices (Vessoyan, Smart, Steckle, & McKillop, 2023), gesture recognition systems (Ascari, Silva, & Pereira, 2019), or specialized keyboard devices (Light & McNaughton, 2014). However, these methods require patients to retain some degree of motor control, which is often impossible for severely affected individuals (Pinheiro et al., 2011). Learning to use these systems is a complex process that demands significant time and effort (Light & McNaughton, 2014).

In recent years, Brain–Computer Interface technology, particularly EEG-based speech reconstruction systems, has shown tremendous potential (Lopez-Bernal, Balderas, Ponce, & Molina, 2022a; Luo, Rabbani, & Crone, 2022). By directly decoding brain signals

* Corresponding author.
E-mail addresses: kunning.shen@duke.edu (K. Shen), hli83@ncsu.edu (H. Li).

https://doi.org/10.1016/j.smhl.2025.100568
Received 7 March 2025; Accepted 15 March 2025
Available online 26 March 2025

to reconstruct speech, BCI systems can bypass the reliance on physical movement capabilities. These systems are particularly attractive due to their non-invasive nature and good temporal resolution, although they face challenges such as limited signal-to-noise ratio and spatial resolution (Lopez-Bernal, Balderas, Ponce, & Molina, 2022b). Current mainstream EEG-to-Speech systems employ end-to-end deep learning architectures requiring high-channel EEG equipment (64+ channels) to capture detailed spatial information (Panachakel & Ramakrishnan, 2021).

While these systems have shown promising results in laboratory settings, they face significant challenges in practical applications, particularly in resource-constrained environments where healthcare infrastructure is limited and the cost of high-channel EEG systems is prohibitive (Armand Larsen, Klok, Lehn-Schiøler, Gatej, & Beniczky, 2024). Moreover, the limited number of channels in low-channel EEG systems restricts the choice of feature extraction methods, as mainstream approaches like Common Spatial Patterns (CSP) become ineffective with reduced spatial information. These limitations create major barriers to technology deployment in resource-constrained environments, such as primary healthcare facilities.

To address these challenges, we implement an innovative framework that enables effective speech reconstruction from low-channel EEG signals. Our approach includes several key implementations: first, we apply optimized signal processing and feature extraction techniques to enable the efficient use of low-channel EEG signals for speech reconstruction; second, the undercomplete independent component analysis (UICA) technology is employed to further reduce signal channels while preserving critical neural activity features, which significantly reduces subsequent processing complexity; finally, we develop a generator-discriminator-based modular architecture that effectively extracts features from these low-channel EEG signals, accurately reconstructing voice characteristics with high computational efficiency (Mishra, Sharma, Jha, & Bhavsar, 2023). The effectiveness of our framework has been validated on a Spanish dataset. Using 6-channel EEG signals, our framework achieves a CER of 64.24%, outperforming the baseline (68.26% with 64-channel EEG). After applying UICA processing to reduce to 4 channels, the performance (64.99% CER) remains comparable despite using fewer resources. This demonstrates our approach can effectively reconstruct voice from EEG signals while significantly reducing hardware requirements.

The main contribution of our work is the development of an accurate continuous speech reconstruction framework using low-channel (6-channel) EEG signals. It achieves superior performance compared to the baseline which only reconstructs isolated words and phrases using 64-channel EEG. Moreover, through the integration of UICA-based dimension reduction technique with our architecture, we enable further channel reduction to 4-channel EEG while maintaining comparable performance. Our system not only enables continuous speech reconstruction that better reflects real-world usage scenarios, but also significantly reduces hardware and computational requirements, making this technology more accessible in resource-constrained environments while maintaining robust performance.

2. Methodology

2.1. Overview

As shown in Fig. 1, we implement an architecture for reconstructing speech from EEG signals. The system processes both imagined and spoken EEG inputs through multiple processing stages: beginning with 6-channel EEG signal acquisition, followed by UICA for channel reduction from 6 to 4 channels while preserving essential neural patterns. The reduced signals are then fed into a generator network to reconstruct mel-spectrograms, which are optimized by a discriminator network during adversarial training. The discriminator compares the generated mel-spectrograms against ground truth speech mel-spectrograms to improve reconstruction quality. For final speech synthesis, the generated mel-spectrograms are processed through a pre-trained HiFi-GAN vocoder to produce audio waveforms, which are then transcribed using a pre-trained Automatic Speech Recognition (ASR) model. During real-world application (shown by dashed arrows in Fig. 1), only imagined EEG signals are required as input (Angrick et al., 2021), making the system suitable for users who have difficulty in producing vocal speech.

2.2. Channel reduction for neural signal processing

We first employ undercomplete independent component analysis (UICA) to reduce the channel of the EEG signals while preserving essential neural patterns (Naik & Kumar, 2011). UICA is particularly suitable for our application as it performs channel reduction while maintaining the statistical independence of the extracted components, given by:

$$X_{UICA} = f_{UICA}(X_{EEG}, n = 4) \tag{1}$$

where X_{EEG} is input EEG data with shape $(n_{trials}, n_{samples}, n_{channels})$, X_{UICA} is reduced data with shape $(n_{trials}, n_{samples}, n'_{channels})$, $n = 4$ is the number of independent components. The UICA transformation reduces the channel while preserving the most informative independent neural components.

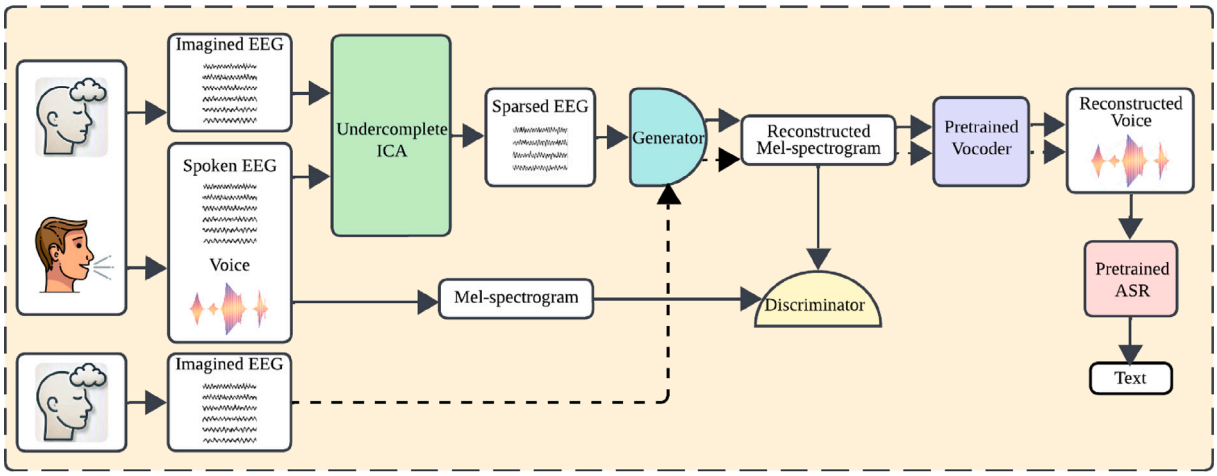

Fig. 1. Overview of the low-channel EEG-to-speech framework. The solid arrows represent the model optimization process. The dashed arrows indicate the inference pathway for real-world speech reconstruction.

2.3. Neural-to-speech feature generation

The generator network employs a multi-stage architecture designed to transform 4-channel ICA-processed EEG signals into mel-spectrograms (Xu et al., 2024). The network begins with a pre-convolution layer that processes the input through a weight-normalized 1D convolution, expanding the input channels. This is followed by a bidirectional GRU layer that captures temporal dependencies in the signal, with the output concatenated with the pre-convolution features to preserve both local and global temporal information. The core of the generator consists of multiple upsampling stages implemented through weight-normalized transposed convolutions. Each upsampling stage is followed by a Multi-Receptive Field (MRF) module containing parallel residual blocks with varying kernel sizes and dilation rates. These residual blocks employ dual-path convolutions with LeakyReLU activation and skip connections. The network concludes with a post-processing convolution layer that produces the mel-spectrogram output.

The generator is optimized using three loss components:

(1) **Reconstruction Loss** measures the mean squared error between generated and target mel-spectrograms, formulated as:

$$L_{rec}(G) = \mathbb{E}_s[(G(s) - x)^2], \tag{2}$$

where $G(s)$ is the generated mel-spectrogram from input EEG signal s, and x is the target mel-spectrogram.

(2) **Adversarial Loss** optimizes the generator to fool the discriminator, given by:

$$L_{adv}(G; D) = \mathbb{E}_s[\log(1 - D(G(s)))], \tag{3}$$

where $D(G(s))$ represents the discriminator's prediction on generated samples.

(3) **CTC Loss** optimizes character-level reconstruction through the ASR model, given by:

$$L_{ctc}(G) = -\sum_{t=1}^{T} \log \sum_{\pi \in \mathcal{B}^{-1}(y)} \prod_{t=1}^{T} p(\pi_t | G(s)), \tag{4}$$

where $\mathcal{B}$ is the CTC many-to-one mapping, y is the target transcript, and $p(\pi_t|G(s))$ is the probability of path π at time t (Krishna, Han, Tran, Carnahan, & Tewfik, 2019).

The total generator loss is formulated as a weighted combination:

$$L_{total}(G) = \lambda_{g1} L_{rec}(G) + \lambda_{g2} L_{adv}(G; D) + \lambda_{g3} L_{ctc}(G), \tag{5}$$

where λ_{g1}, λ_{g2}, and λ_{g3} are the weighting factors.

2.4. Speech quality enhancement through adversarial learning

The discriminator follows a complementary structure to analyze both real and generated mel-spectrograms. It processes inputs through an initial weight-normalized 1D convolution layer followed by multiple downsampling stages, which progressively increase the channel dimension while reducing the temporal resolution. Similar to the generator, each downsampling stage incorporates MRF modules with varying kernel sizes and matching dilation rates. The processed features are then passed through a bidirectional GRU layer to capture sequence-level information, with the output concatenated with the convolutional features. The discriminator has two parallel output branches: one for real/fake classification using sigmoid activation, and another for speech content classification

using softmax activation. This dual-task design enables the discriminator to evaluate both the quality and content accuracy of the generated samples.

The discriminator is trained using two loss components:

(1) **Adversarial Loss** evaluates the authenticity prediction, given by:

$$L_{adv}(D) = \frac{1}{2}(L_{real} + L_{fake}), \tag{6}$$

where $L_{real} = BCE(D_{valid}(x), 1)$ and $L_{fake} = BCE(D_{valid}(G(s)), 0)$, with BCE denoting binary cross-entropy loss.

(2) **Classification Loss** measures content classification accuracy, given by:

$$L_{cl}(D) = CE(D_{cl}(x_{real}), c), \tag{7}$$

where x_{real} is the target mel-spectrogram and c is the target class label, with CE denoting cross-entropy loss.

The total discriminator loss is written as:

$$L_{total}(D) = \lambda_{d1} L_{adv}(D) + \lambda_{d2} L_{cl}(D), \tag{8}$$

where λ_{d1} and λ_{d2} are the weighting factors.

2.5. Speech synthesis and recognition

For speech synthesis and recognition, our framework employs two specialized pre-trained models. The vocoder component uses HiFi-GAN (Kong, Kim, & Bae, 2020), a high-fidelity generative adversarial network designed for efficient and high-quality audio synthesis. Operating at a sampling rate of 22050 Hz, the vocoder transforms mel-spectrograms into waveforms through a series of transposed convolutions.

The ASR component utilizes an automatic speech recognition model based on cross-lingual speech representations (XLSR) (Conneau, Baevski, Collobert, Mohamed, & Auli, 2020). The model first learns cross-lingual speech representations through pretraining on raw waveforms from 53 languages. This pretrained XLSR-53 model is then fine-tuned specifically on Spanish speech data to enable speech recognition in Spanish (Grosman, 2021).

3. Evaluation

3.1. Dataset

The experiments utilized a publicly available EEG dataset collected by Coretto, Gareis, and Rufiner (2017), consisting of EEG recordings from 15 healthy subjects (8 males, 7 females, mean age 25 years) during imagined and spoken speech tasks. While the original dataset contains both vowels and command words, this study focused exclusively on the command word portion, which includes six Spanish directional commands: “arriba” (up), “abajo” (down), “derecha” (right), “izquierda” (left), “adelante” (forward) and “atrás” (backward).

The EEG signals were recorded using six Ag/AgCl electrodes positioned according to the international 10–20 system at locations F3, F4, C3, C4, P3, and P4, with references placed on the left and right mastoids. These electrode positions were specifically selected to capture language-related cortical activities while minimizing muscle artifacts. The EEG signals were sampled at 1024 Hz and filtered between 0.3 Hz and 35 Hz using analog band-pass filters.

During the recording sessions, subjects were seated approximately one meter from a display screen where visual cues were presented. Each trial followed a structured protocol consisting of steps: (1) A 2-second ready interval, (2) A 2-second stimulus presentation interval, (3) A 4-second imagine/pronounce interval, and (4) A 4-second rest interval.

In total, the dataset contains 4,025 imagined speech EEG samples and 1,088 spoken speech EEG samples with corresponding ground truth audio. During each trial, EEG data was recorded along with audio signals when subjects actually pronounced the words. The dataset was split with 70% for training, 15% for validation and 15% for testing.

3.2. Baseline

We adopt the NeuroTalk model as our baseline (Lee, Lee, Kim, & Lee, 2023), which is a state-of-the art EEG-to-Speech conversion system that uses similar generator-discriminator architecture.

This framework achieved significant advances in EEG-based speech reconstruction and was evaluated on an English language dataset consisting of thirteen phrases collected from six participants, including twelve words/phrases (ambulance, clock, hello, help me, light, pain, stop, thank you, toilet, TV, water, and yes) and a silent phase. Using 64-channel EEG signals, their system achieves an RMSE of 0.175 in mel-spectrogram reconstruction and demonstrates robust performance with a Character Error Rate (CER) of 68.26% for imagined speech reconstruction.

Table 1
Model architecture and training parameters.

Parameter	Value
Generator Architecture	
Input channels	4
Output channels	80
Initial channels	1024
Upsampling rates	[3,2,2,2]
Upsampling kernels	[6,4,4,4]
MRF kernel sizes	[3,7,11,15]
MRF dilation rates	[1,3,5]
Discriminator Architecture	
Input channels	80
Initial channels	64
Downsampling rates	[3,3,3]
Downsampling kernels	[6,6,6]
MRF kernel sizes	[11,7,3]
MRF dilation rates	[1,3,5]
Number of classes	6
Training Parameters	
Epochs	100
Batch size	4
Learning rate	1×10^{-4}
Weight decay	0.01
Beta1, Beta2	0.8, 0.99
LR decay factor	0.999
Generator loss weights (λ_{g1}, λ_{g2}, λ_{g3})	1, 0.01, 0.01
Discriminator loss weights (λ_{d1}, λ_{d2})	1, 1

3.3. Implementation

The network training was conducted using the AdamW optimizer with an initial learning rate of 10^{-4} and betas $(0.8, 0.99)$ (Loshchilov, Hutter, et al., 2017). A learning rate decay factor of 0.999 was applied per epoch. Networks were trained for 100 epochs on a single NVIDIA 4090 GPU.

The generator network architecture details are shown in Table 1. The input 4-channel EEG signals are first expanded to 512 channels through the pre-convolution layer. Four upsampling stages and kernel sizes progressively increase the temporal resolution. Each MRF module contains four parallel residual blocks.

The discriminator follows a mirrored structure with three downsampling stages. Its MRF modules use three kernel sizes with matching dilation rates. The initial convolution expands the 80 mel-spectrogram channels to 64 feature channels.

3.4. Evaluation metrics

This study adopts two standard metrics for performance assessment: Root Mean Square Error (RMSE) for mel-spectrogram quality and Character Error Rate (CER) for speech recognition accuracy (Settle, Le Roux, Hori, Watanabe, & Hershey, 2018). The RMSE between ground truth and generated mel-spectrograms is calculated as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}, \tag{9}$$

where y_i and $\hat{y}_i$ represent the ground truth and generated mel-spectrogram values respectively, and N is the total number of elements. The CER quantifies the minimum edit distance between predicted and reference text, formulated as:

$$CER = \frac{S + D + I}{N} \times 100\%, \tag{10}$$

where S, D, and I represent the number of character substitutions, deletions, and insertions required, and N denotes the reference text length. These metrics evaluate both spectrogram reconstruction fidelity and ultimate speech recognition performance.

3.5. Results

Fig. 2 provides a comprehensive visualization of our EEG-to-speech conversion pipeline, illustrating the transformation from neural signals to intelligible speech output. The generated mel-spectrograms and waveforms demonstrate our framework's capability in reconstructing speech from imagined EEG signals. These examples present reconstruction results with varying CER levels across different classes. Notably, the reconstructed samples show diverse temporal patterns with varying onset times, ending times, and intervals, which effectively capture the inherence difference in imagined speech from different subjects, e.g., the duration of thought

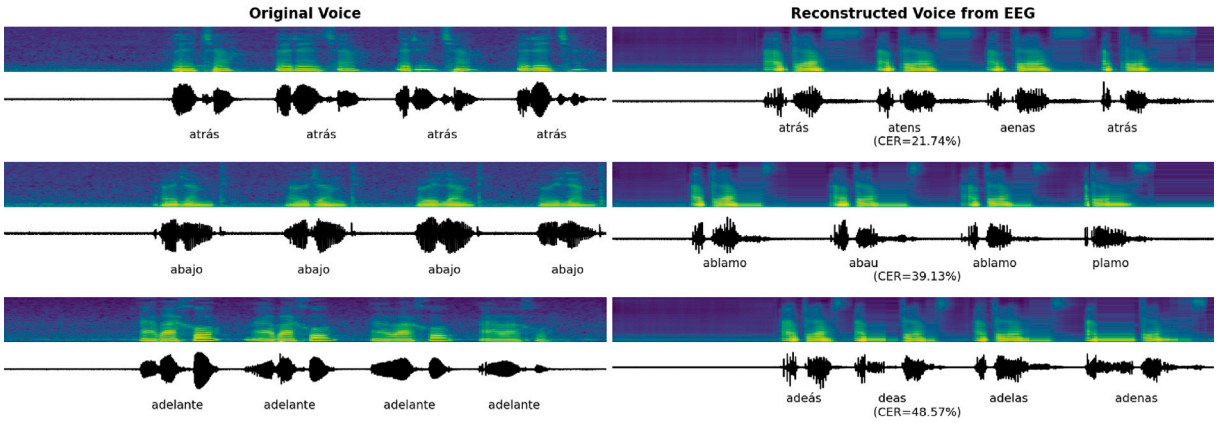

Fig. 2. Visualization of speech reconstruction from EEG signals. Each row shows paired comparisons of original and reconstructed speech samples, with mel-spectrograms (upper), waveforms (middle), and ASR transcriptions (lower) displayed for both. The left panels present the original voice recordings while the right panels show the corresponding reconstructions from imagined EEG signals. Character Error Rate values are annotated below each reconstructed transcription.

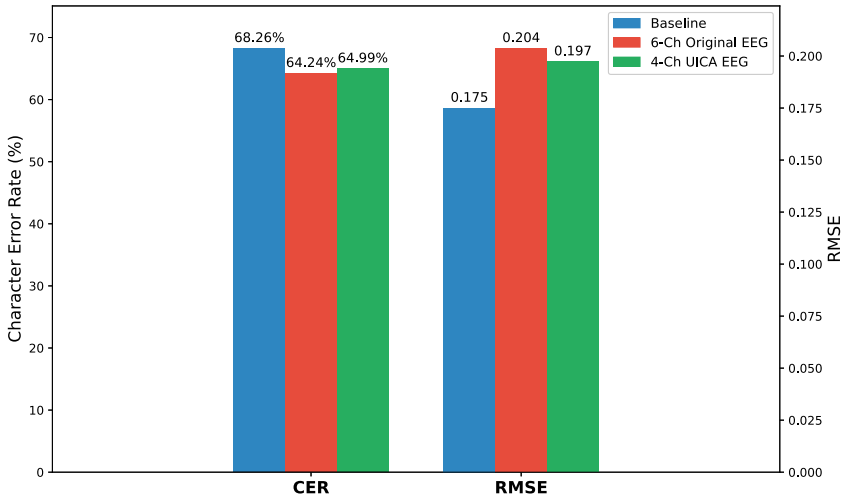

Fig. 3. The comparison of CER and RMSE among Baseline, 6-channel EEG signals and UICA-processed 4-channel EEG signals.

before speaking varies. In general, all reconstructions maintain high intelligibility, demonstrating our model's robust ability to decode imagined speech into recognizable acoustic patterns. We observe that the CER value increases from "atrás" (2 syllables) to "abajo" (3 syllables) to "adelante" (4 syllables), which reveals the growing challenge in accurate reconstruction as syllable count increases. Higher CER values are typically associated with multi-syllabic words, where more syllables need to be articulated or imagined within similar time frames, requiring more precise decoding of the EEG signals.

As shown in Fig. 3, our system achieves a CER of 64.24% using 6-channel EEG input, while the UICA-processed 4-channel EEG signals maintain comparable performance with a CER of 64.99%. This outperforms the baseline's CER of 68.26% despite using significantly fewer channels. In terms of mel-spectrogram reconstruction quality, the UICA-processed 4-channel signals achieve an RMSE of 0.197, slightly better than the 6-channel input's RMSE of 0.204, though marginally higher than the baseline's RMSE of 0.175 achieved with 64 channels.

Our framework not only demonstrates remarkable efficiency but also addresses more challenging scenarios compared to the baseline in the following two aspects: (1) While the baseline works with high-channel 64 EEG channels in their dataset, our dataset only has 6 channels, and we even further reduce it to 4 channels through UICA processing, handling much less spatial information with inherently limited signal quality and spatial resolution; (2) Unlike the baseline's focus on isolated single words or short phrases, we tackle the more challenging task of processing continuous sequences of four repeated utterances, requiring sophisticated signal segmentation and temporal alignment. Despite these substantially more challenging conditions, our framework still achieves superior performance, demonstrating the robustness and effectiveness of our approach.

4. Conclusion & future works

This paper presents a significant advancement in making EEG-based speech reconstruction technology more accessible through a low-channel neural computing framework. By implementing an efficient generator-discriminator architecture operating on only 6 EEG channels, we achieve superior performance (64.24% CER) compared to existing 64-channel systems (68.26% CER). The subsequent integration of UICA successfully maintains this performance level (64.99% CER) while further reducing computational requirements through channel reduction from 6 to 4. These results, along with our system's demonstrated capability in processing continuous speech sequences, represent a crucial step toward making assistive communication technology more accessible in resource-constrained environments. It opens new possibilities for future research in optimizing low-channel neural signal decoding techniques across different languages and clinical conditions.

In our future work, while our research has demonstrated promising results in channel reduction, we will focus on improving the system's accuracy. We will investigate the crucial relationship between user satisfaction and system performance through extensive user studies with individuals who have communication disorders, aiming to enhance the overall effectiveness of our approach. Moreover, we recognize the inherent challenges in training stability and convergence with our current generator-discriminator architecture and will explore alternative neural network architectures that offer more stable training dynamics and better performance. Through our continued research in both model optimization and channel configuration, we aim to develop more reliable and practical solutions for real-world applications.

CRediT authorship contribution statement

Kunning Shen: Investigation, Methodology, Validation, Writing – original draft. **Huining Li:** Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank all anonymous reviewers for their insightful comments on this paper.

Data availability

The authors do not have permission to share data.

References

Angrick, M., Ottenhoff, M. C., Diener, L., Ivucic, D., Ivucic, G., Goulis, S., et al. (2021). Real-time synthesis of imagined speech processes from minimally invasive recordings of neural activity. *Communications Biology*, *4*(1), 1055.

Armand Larsen, S., Klok, L., Lehn-Schiøler, W., Gatej, R., & Beniczky, S. (2024). Low-cost portable EEG device for bridging the diagnostic gap in resource-limited areas. *Epileptic Disorders*, *26*(5), 694–700.

Ascari, R. E. S., Silva, L., & Pereira, R. (2019). Personalized interactive gesture recognition assistive technology. In *Proceedings of the 18th Brazilian symposium on human factors in computing systems* (pp. 1–12).

Conneau, A., Baevski, A., Collobert, R., Mohamed, A., & Auli, M. (2020). Unsupervised cross-lingual representation learning for speech recognition. arXiv preprint arXiv:2006.13979.

Coretto, G. A. P., Gareis, I. E., & Rufiner, H. L. (2017). Open access database of EEG signals recorded during imagined speech. In *12th international symposium on medical information processing and analysis: vol. 10160*, SPIE, Article 1016002.

Grosman, J. (2021). XLSR-53 large model for speech recognition in spanish. *HuggingFace Models*, https://huggingface.co/jonatasgrosman/wav2vec2-large-xlsr-53-spanish.

Kong, J., Kim, J., & Bae, J. (2020). HiFi-GAN: Generative adversarial networks for efficient and high fidelity speech synthesis. *Advances in Neural Information Processing Systems*, *33*, 17022–17033.

Krishna, G., Han, Y., Tran, C., Carnahan, M., & Tewfik, A. H. (2019). State-of-the-art speech recognition using eeg and towards decoding of speech spectrum from eeg. arXiv preprint arXiv:1908.05743.

Lee, Y. E., Lee, S. H., Kim, S. H., & Lee, S. W. (2023). Towards voice reconstruction from EEG during imagined speech. In *Proceedings of the AAAI conference on artificial intelligence: vol. 37*, (no. 5), (pp. 6030–6038).

Light, J., & McNaughton, D. (2014). Communicative competence for individuals who require augmentative and alternative communication: A new definition for a new era of communication? *Augmentative and Alternative Communication*, *30*(1), 1–18.

Lopez-Bernal, D., Balderas, D., Ponce, P., & Molina, A. (2022a). A state-of-the-art review of EEG-based imagined speech decoding. *Frontiers in Human Neuroscience*, *16*, Article 867281.

Lopez-Bernal, D., Balderas, D., Ponce, P., & Molina, A. (2022b). A state-of-the-art review of EEG-based imagined speech decoding. *Frontiers in Human Neuroscience*, *16*, Article 867281.

Loshchilov, I., Hutter, F., et al. (2017). Fixing weight decay regularization in adam. arXiv preprint arXiv:1711.05101 5.

Luo, S., Rabbani, Q., & Crone, N. E. (2022). Brain-computer interface: Applications to speech decoding and synthesis to augment communication. *Neurotherapeutics*, *19*(1), 263–273.

Mishra, R., Sharma, K., Jha, R. R., & Bhavsar, A. (2023). NeuroGAN: Image reconstruction from EEG signals via an attention-based GAN. *Neural Computing and Applications*, *35*(12), 9181–9192.

Naik, G. R., & Kumar, D. K. (2011). An overview of independent component analysis and its applications. *Informatica*, *35*(1).

Olusanya, B. O., Wright, S. M., Nair, M., Boo, N.-Y., Halpern, R., Kuper, H., et al. (2020). Global burden of childhood epilepsy, intellectual disability, and sensory impairments. *Pediatrics*, *146*(1).

Palmer, A. D., Carder, P. C., White, D. L., Saunders, G., Woo, H., Graville, D. J., et al. (2019). The impact of communication impairments on the social relationships of older adults: Pathways to psychological well-being. *Journal of Speech, Language, and Hearing Research*, *62*(1), 1–21.

Panachakel, J. T., & Ramakrishnan, A. G. (2021). Decoding covert speech from EEG-a comprehensive review. *Frontiers in Neuroscience*, *15*, Article 642251.

Pinheiro, C. G., Naves, E. L., Pino, P., Losson, E., Andrade, A. O., & Bourhis, G. (2011). Alternative communication systems for people with severe motor disabilities: A survey. *Biomedical Engineering Online*, *10*, 1–28.

Settle, S., Le Roux, J., Hori, T., Watanabe, S., & Hershey, J. R. (2018). End-to-end multi-speaker speech recognition. In *2018 IEEE international conference on acoustics, speech and signal processing* (pp. 4819–4823). IEEE.

Vessoyan, K., Smart, E., Steckle, G., & McKillop, M. (2023). A scoping review of eye tracking technology for communication: Current progress and next steps. *Current Developmental Disorders Reports*, *10*(1), 20–39.

Xu, X., Wang, B., Yan, Y., Zhu, H., Zhang, Z., Wu, X., et al. (2024). ConvConcatNet: A deep convolutional neural network to reconstruct mel spectrogram from the EEG. In *2024 IEEE international conference on acoustics, speech, and signal processing workshops* (pp. 113–114). IEEE.