An Investigation on the Speech Recovery from EEG Signals Using Transformer

Tomoaki MIZUNO* Takuya KISHIDA† Natsue YOSHIMURA‡ Toru NAKASHIKA*
* The University of Electro-Communications, Tokyo, Japan
† Aichi Shukutoku University, Aichi, Japan
‡ Tokyo Institute of Technology, Kanagawa, Japan

Abstract—**In recent years, brain-machine interfaces (BMI) have been researched to enable people who cannot physically speak or who are placed in situations where they cannot speak to engage in speech communication. However, synthesizing complete speech from ElectroEncephaloGraphy (EEG) signals remains a challenging task. In this paper, reconstructing speech from EEG signals, a Transformer-based model has been developed on the basis of data from listening to the speeches of two people, one male and one female. The objective of this study is to investigate the potential to reconstruct speech from EEG signals, including the characteristics of the corresponding speaker's speech and the potential to reconstruct speeches containing the corresponding linguistic content, by training a single model on both male and female speeches. Our findings reveal that our model can generate two distinctly different speakers' speeches from the EEG signals of the two speakers' speeches. Additionally, the EEG signals appear to contain information about speaker characteristics that can be reconstructed from the EEG signals.**

I. Introduction

Brain-Computer Interface (BCI) refers to computer technology that has the potential to enable people to operate a computer simply by thinking. Research on reading language information from brain activity and synthesizing it as speech is also advancing and could have significant implications for enabling those who are unable to speak due to illness or in environments where speech is not possible, such as underwater, or similar conditions, to communicate vocally.

BCI can be broadly classified into two types (invasive and noninvasive) depending on the method used to measure brain activity. Invasive methods involve the direct implantation of electrodes into the brain, providing high-resolution and precise data. This technique can directly measure the activity of specific neurons in the brain, resulting in minimal signal noise and very short temporal delays. However, it requires surgical procedures, which come with risks such as infections and inflammation. Non-invasive methods, on the other hand, allow for the measurement of brain activity without surgery, thereby posing relatively lower risks. These methods can measure electrical activity and changes in blood flow in the brain, but they offer lower resolution compared to invasive methods. Electrocorticography (ECoG), which is measured by placing electrodes on the surface of the brain in the invasive method, is used as a signal related to brain activity that enables highly accurate speech synthesis [1], [2]. Previous studies have successfully synthesized intelligible speech by combining brain activity and mouth movements, but complete speech has not yet been synthesized with signals without mouth movement information. In addition, the invasive type requires surgery to attach electrodes to the surface of the brain, which is costly.In contrast, electroencephalography (EEG) is a noninvasive method that measures electrical signals by placing electrodes on the scalp, making it a more convenient way to measure brain activity. However, synthesizing complete speech sounds with noninvasive EEG is a more challenging task than with ECoG. Nevertheless, a study synthesized two short vowels from EEG with high intelligibility, using a neural network introduced to infer phonetic features from EEG [3]. The structure of the neural network is relatively straightforward, which suggests that it may be a viable approach to infer speech features from EEG. It also suggests the potential for employing more complex network structures to infer speech features from EEG.

To propose this possibility, we attempted to synthesize speech from a single participant's EEG when listening to a single speaker's speech using a diffusion model first. We trained Diffwave [4], a vocoder using a diffusion model, by modifying the input to an EEG converted from a mel-spectrogram to a two-dimensional signal and assessed the quality of the synthesized speech. The results showed that most segments had only small noises and only one tone was mixed in a few times within the segment. The model possibly may not have handled time-series signals as well as it could have done.

Given that the auditory system is generally hierarchically organized, that low-level acoustic features (frequency, intensity, etc.) are processed faster than higher-order semantic processing [5], [6], and that the electrical signals generated from these features are collected using low spatial resolution methods, it seems that not all phonemes in speech are expressed as EEG while preserving their order. Additionally, although we refer to the EEG of one person when listening to the speech of a single speaker, it is important to note that individual differences in EEG expression do exist, such as the difference in reaction time between the elderly and the young [7]. We believe that a model that can encompass these characteristics would be beneficial. In this study, we focus on the Transformer [8].

Transformer, which is a deep learning model applied to the speech field, has been successfully used to synthesize speech with higher quality than conventional neural networks in speech synthesis and voice conversion [9]. By using a deep learning model with Transformer to infer speech features from EEG, speech more complex than two types of short vowels may be possible to generate. In this study, we attempt to infer speech features from EEG using a deep learning model with Transformer. We use the speech of two speakers and the EEG of a person who listened to the speakers as training data, and then we perform inference to reconstruct the speech from the EEG. Another of our goal is to see if the difference in speaker characteristics between the two speakers can be recovered. We explore the possibility to recover the differences in speaker identity between the two speakers by learning from the two speakers' speeches and the speeches of the people who heard them and by performing inference to reconstruct speeches.

II. Architecture Selection

Voice conversion technology generates speech by converting one person's speech into another person's speech while retaining the linguistic information and speaking style of the speech. Since EEG is also a time-series signal, we thought that a deep learning model based on the voice conversion technique could be used to infer features that represent the speech from EEG. Transformer is an architecture used in deep learning models based onvoice conversion technology and has also been extensively applied beyond voice conversion technology.

Transformer excels at handling sequential data and capturing long-range dependencies. These are particularly effective when input data is transformed into dense vector representations that encapsulate semantic information. Positional encoding is added to maintain the order of data within the sequence. The attention mechanisms in Transformer enable the model to weigh the significance of different elements in the sequence, thereby enhancing contextual understanding.

This paper posits that the strengths of Transformer can be leveraged to infer complex patterns from sequential data. On the basis of these strengths, Transformer-based models used in voice conversion techniques are used to synthesize speech from EEG.

A. Transformer-based Text-To-Speech system

Fig. 1 shows the Transformer-based Text-To-Speech(TTS) system [9] and the Voice Transformer Network, which is a voice conversion system based on the TTS system network [10], which is a voice conversion system based on the TTS system.

The Transformer model consists of an encoder and a decoder, and is used to find the mapping from a source sequence $\boldsymbol{x}_{1:n} = (x_1, ..., x_n)$ to a target sequence $\boldsymbol{y}_{1:m} = (y_1, ..., y_m)$. The lengths of the two sequences do not need to be the same.

The encoder Enc maps the source sequence $\boldsymbol{x}_{1:n} = (x_1, ..., x_n)$ to a sequence of hidden states $\boldsymbol{h}_{1:n} = (h_1, ..., h_n)$.

$$\boldsymbol{h}_{1:n} = \text{Enc}\left(\boldsymbol{x}_{1:n}\right) \tag{1}$$

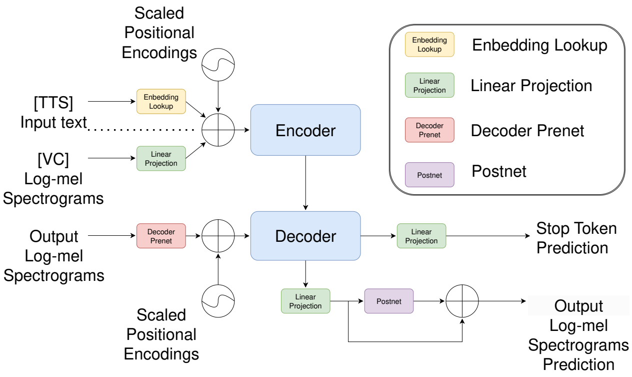

Fig. 1. Architecture of Transformer-based Text-To-Speech system and Voice Transformer Network.

From the hidden states $\boldsymbol{h}_{1:n}$ and the decoder's previous hidden state $\boldsymbol{q}_{t-1}$, we obtain query, key and value.

$$\boldsymbol{Q} = W_Q \boldsymbol{q}_{t-1} \tag{2}$$

$$\boldsymbol{K} = W_K \boldsymbol{h}_{1:n} \tag{3}$$

$$\boldsymbol{V} = W_V \boldsymbol{h}_{1:n} \tag{4}$$

where W_Q, W_K, $W_V \in \mathbb{R}^{d_{model} \times d_{model}}$ are a learnable weight matrix. d_{model} is the input dimension. The weighted sum of these hidden states is the context vector $\boldsymbol{c}_t$.

$$\boldsymbol{c}_t = \sum_{k=1}^{n} a_t^{(n)} \boldsymbol{h}_k \tag{5}$$

Here, the attention weights $\boldsymbol{\alpha}_t$ are calculated using the hidden states $\boldsymbol{h}_{1:n}$ and the decoder's previous hidden state $\boldsymbol{q}_{t-1}$ through an attention mechanism [11], [12]. The attention weights, represented by the vector $\boldsymbol{\alpha}_t = (a_t^{(1)}, \cdots, a_t^{(n)})$, indicate which encoder hidden states the decoder should focus on to determine the output at time t. Each attention probability $a_t^{(k)}$ can be thought of as the importance of the hidden representation $\boldsymbol{h}_k$ at the tth time step.

$$\boldsymbol{a}_t = \text{attention}\left(\boldsymbol{q}_{t-1}, \boldsymbol{h}_{1:n}\right) = \text{softmax}(\frac{\boldsymbol{QK}^T}{\sqrt{d_k}})\boldsymbol{V} \tag{6}$$

d_k is the dimensionality of the key. Scaling by $\sqrt{d_k}$ is used to stabilize the gradient during training.

In practice, multi-head attention is often employed, which involves performing the above computation multiple times in parallel with different learned projections. For h heads:

$$\text{multi-head}\left(\boldsymbol{q}_{t-1}, \boldsymbol{h}_{1:n}\right) = (\text{head}_1, ..., \text{head}_h)W^O \tag{7}$$

$$where\ \text{head}_i = \text{Attention}(W_Q^i \boldsymbol{q}_{t-1}, W_K^i \boldsymbol{h}_{\mathbf{1:n}}, W_V^i \boldsymbol{h}_{\mathbf{1:n}}) \tag{8}$$

and $W^O \in \mathbb{R}^{d_{model} \times d_{model}}$ is a learnable weight matrix.

The decoder Dec takes the context vector $\boldsymbol{c}_t$, the previously generated output sequence $\boldsymbol{y}_{1:t-1} = (y_1, ..., y_{t-1})$, and the previous decoder hidden state as inputs and decodes the decoder hidden state $\boldsymbol{q}_t$ and $\boldsymbol{y}_t$.

$$\boldsymbol{y}_t, \boldsymbol{q}_t = \text{Dec}\left(\boldsymbol{y}_{1:t-1}, \boldsymbol{q}_{t-1}, \boldsymbol{c}_t\right) \tag{9}$$

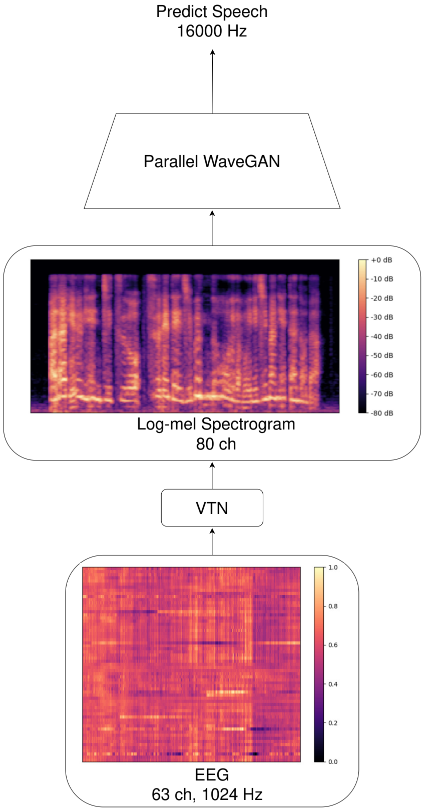

Fig. 2. Overview of the model for inferring speech from EEG. EEG is shown as a two-dimensional signal of channel and time, and the voltage is normalized.

B. Voice Transformer Network: Transformer-based Voice Conversion System

A Voice Transformer Network is proposed as a voice conversion system based on a Transformer-based TTS system [10]. It adapts by passing a linear projection to convert the input from text to acoustic features (in this case, log-scale mel-spectrograms). Otherwise, the system is identical to the Transformer-based TTS system.

III. Method

Since EEG is also a time-series signal, we propose using the Voice Transformer Network to reconstruct speech from EEG. The speech Transformer Network is a model in which the input is replaced by acoustic features. The transformation process is shown in Fig. 2. This model consists of a Voice Transformer Network with 63 channels of EEG input and Parallel WaveGAN [14] as a vocoder, which is trained in an End-2-End fashion. Note that the Parallel WaveGAN has already been trained and the parameters of Parallel WaveGAN are not changed during training.

The following describes the dataset and model for the speech synthesis experiments from EEG using the Voice Transformer Network.

A. EEG measurements and dataset preparation

The EEG during Japanese speech listening was measured and processed to create a dataset of the EEG during speech listening. The speech sounds are the ATR 503 sentences [15] uttered by a male speaker (MMY) and a female speaker (FTK) respectively.

To create the EEG and speech dataset, the EEG was measured while the participants listened to the speech uttered by the two speakers. The Biosemi ActiveTwo system (Biosemi) was used for the measurement. The electrode configuration of the participants was 64 ch, based on the International 10–20 system. The EEG was sampled at 8,192 Hz, bandpass filtered at 1-40 Hz, and downsampled to 1,024 Hz to reduce computational cost. Then, independent component analysis (ICA) was applied to remove the components caused by eye movements. The EEG electrode channel TP8, from which the eye movement-related component could not be removed, was not used in the analysis. Then, all data were normalized. To synchronize the 63 ch EEGs created by the above procedure with the audio, we cut the EEG from the trigger of the start of playback to 1024 samples (1 second) before the trigger of the start of playback of the next audio and created 503 EEG sets corresponding to the 503 sentences of audio. When creating the training sets, the sampling rate of the speech was changed to 16 kHz. The created dataset was divided into training data, validation data, and test data in the ratio of 480:12:11. MMY and FTK datasets were created and then merged, resulting in 960 sentences of training data, 24 sentences of validation data, and 22 sentences of test data in the end.

B. Model details

The parameters of the Voice Transformer Network[1] were used mostly as is. The input dimension was changed to use EEG as input. Also, this implementation did not include Guided Attention [16], but we used guided attention to reduce the risk of using brain wave information from periods when the subject was not listening during the speech reconstruction. Guided Attention means that the attention matrix A is approximately diagonal. The attention matrix A is a list of attention weights $\boldsymbol{a}_t$ at each time step. The loss function is defined as follows

$$L_{att}(A) = \mathbb{E}_{nt}[A_{nt}W_{nt}] \tag{10}$$

$$W_{nt} = 1 - \exp\left(-\frac{(n/N - t/T)^2}{2g^2}\right) \tag{11}$$

Here, n is the position of a character or word in the text, N is the total length of the text, t is the index of the time step, and T is the total length of the audio. A_{nt} refers to the entry of the attention matrix A, representing the correspondence between a specific text position n and a time step t. It indicates the weight of the attention that the text position n should give to the audio at time t, with higher values indicating stronger correlations and lower values indicating weaker correlations. W_{nt} is an element of the weight matrix used to calculate the

[1] https://github.com/unilight/seq2seq-vc/tree/main/egs/arctic/vc1

guided attention loss, as defined in Equation (6). The parameter g controls the variance of the Gaussian function, adjusting the penalty strength on the basis of the distance between the text position n and the time step t. In this experiment, we set $g = 0.4$. For training, we designed and utilized the overall loss function as follows, incorporating the L1-based reconstruction loss L_{L1} used in the existing VTN, the binary cross-entropy loss L_{BCE} to enable the model to learn the timing to stop decoding, and the aforementioned guided attention loss.

$$L = L_{L1} + \alpha L_{BCE} + \beta L_{att} \tag{12}$$

Here, α and β are the weight applied to the binary cross-entropy loss and the guided attention loss. In this experiment, α and β were set to 10 and 1 respectively.

For Parallel WaveGAN [14] used as vocoder, a two-speaker model of MMY and FTK was trained and used for the experiment. The sampling rate was 16 kHz, and the default parameters of Parallel WaveGan [2] were used otherwise.

C. Validation Items

In conducting the experiments, the Voice Transformer Network was trained under the above conditions, and the following items were verified.

- Synthesis quality when inferring speech from EEG training data
- Synthesis quality when inferring speech from EEG test data

To investigate these, we investigated the following indices of speech inferred from EEG training data to examine the linguistic features of the speech.

- Word Error Rate (WER)
- Character Error Rate (CER)
- BertScore [17]

For WER and CER, a smaller value indicates fewer errors. The training data and ground-truth speech were recognized by the Large-v3 model of the speech recognition model Whisper [18] and compared with the text data of the original speech. The BertScore is an index that calculates the similarity of text using BERT, with higher values indicating greater similarity. For the P_{BERT}, R_{BERT}, and $F1_{\text{BERT}}$ values associated with BertScore, Precision (P_{BERT}) is the percentage of predicted correct answers that were actually correct, Recall (R_{BERT}) is the percentage of actual correct answers that were correctly predicted, and F1 score ($F1_{\text{BERT}}$) is the harmonic mean of Precision and Recall.

We also extracted the following data, computed frame by frame, to investigate the representation of talkativeness from the test data and the speech inferred from the ground-truth speech.

- Pitch (F0)
- 1st and 2nd formants (F1, F2)

We use support vector machines (SVM) to verify whether speaker information is reflected in the generated speech. A SVM was trained to calculate mel-frequency cepstral coefficients (MFCC) for each frame of 20 channels from the ground-truth speech to create a speaker identification model (94% accuracy). We use the SVM model to identify whether a male or female speech is used for the generated speech and to verify the accuracy.

TABLE I
WER, CER, and BertScore of the generated speech from the training data.

	WER (↓)	CER (↓)	BertScore(↑) P_{BERT}	R_{BERT}	$F1_{\text{BERT}}$
Train	0.0976	0.0462	0.956	0.954	0.955
Test	2.16	1.13	0.622	0.629	0.625

TABLE II
Mean and variance of F0, F1, and F2 of the generated MMY speech from the test data and the ground truth speech.

MMY	F0[Hz]		F1[Hz]		F2[Hz]	
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
GT	154.6	38.17	773.6	483.2	1898	495.2
Predict	221.2	36.72	793.5	428.0	1940	455.0

TABLE III
Mean and variance of F0, F1, and F2 of the generated FTK speech from the test data and the ground truth speech

FTK	F0[Hz]		F1[Hz]		F2[Hz]	
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
GT	236.2	51.90	759.5	393.9	2010	498.6
Predict	293.1	36.72	730.8	370.2	2019	441.4

IV. Results and Discussion

Fig. 3 compares the mel-spectrograms of the ground-truth speech and the speech generated from the training data, and Fig.4 compares the mel-spectrograms of the ground-truth speech and the speech generated from the test data. The results are shown in Table reftext for the values of WER, CER, and BertScore for the speech synthesized from the EEG that was used for the training data, relative to the ground-truth speech.

The mean values and variances of F0, F1, and F2 calculated for every frame for all speech are shown in Table.II and Table.III. These two tables show the values for ground-truth (GT) speech and generated speech (Predict) for MMY and FTK respectively.

First, comparing the mel-spectrogram in Fig. 3, the speech generated from the training data clearly has a similar overall shape, with many of the wave timings and silence areas similar to the ground-truth speech. However, the speech generated from the test data in the mel-spectrogram in Fig. 4 differs from the truth speech in terms of the timing and characteristics of the silence. From Table I, the WER and CER values for speech inferred from the training data were 0.0976 and 0.0462, and the BertScore value was about 0.955 for both P_{BERT}, R_{BERT} and $F1_{\text{BERT}}$. From these values and the results in Fig. 3 and Fig. 4, we can say that the speech inferred from the training data is synthesized with a high degree of accuracy.

We attempted to perform the same experiment on speech generated from test EEG data, but the WER and CER values

[2] https://github.com/kan-bayashi/ParallelWaveGAN

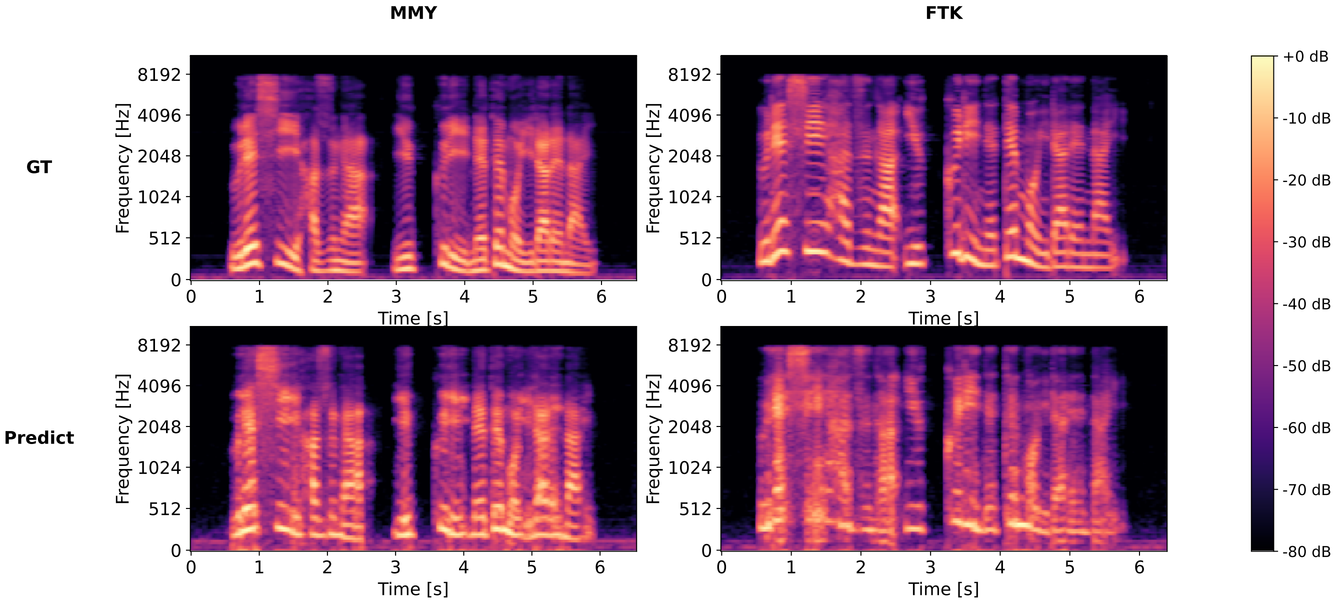

Fig. 3. Comparison of mel-spectrograms of ground-truth speech and speech generated from training data.

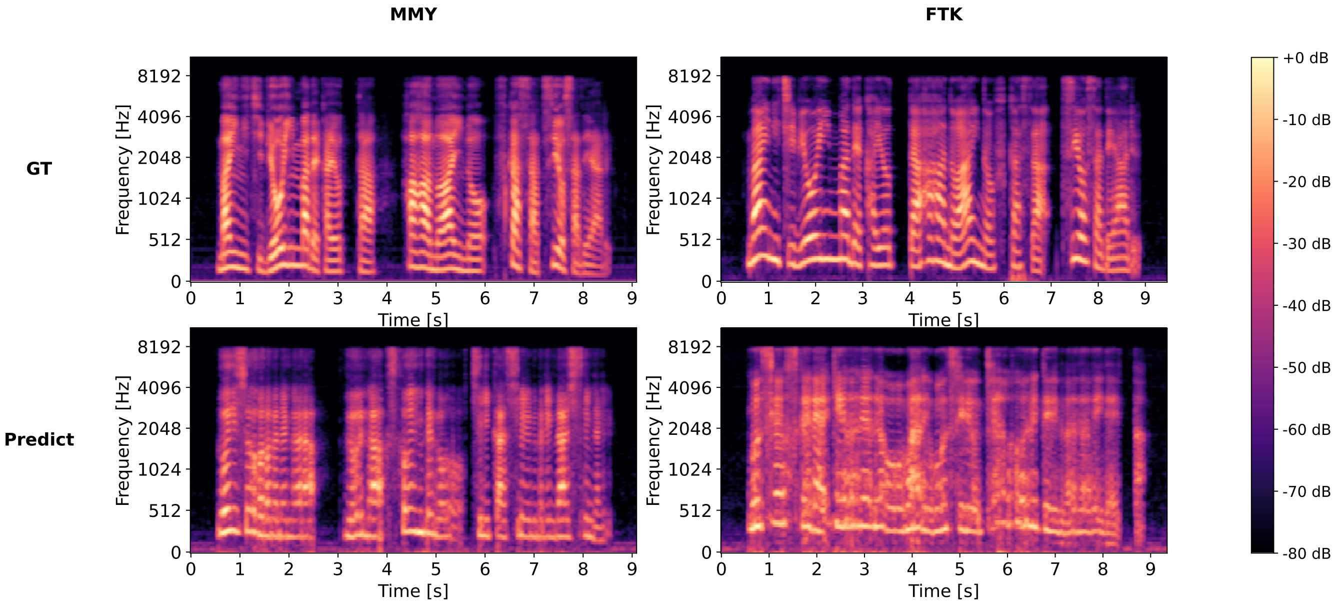

Fig. 4. Comparison of mel-spectrograms of ground-truth speech and speech generated from test data.

exceeded 1, and the BertScore value was about 0.625 for both P_{BERT}, R_{BERT} and $F1_{\text{BERT}}$. This can be described as having significantly low accuracy. In addition to the poor accuracy of generation from EEG signals, the generated speeches do not form coherent sentences, despite Whisper's attempts to infer them as meaningful sentences. This lack of coherence in the generated speech is considered to be a contributing factor. From the above, it can be seen that linguistic information could not be converted into speech.

To conduct the analysis for Tables II and III, we performed Steel-Dwass tests on the F0, F1, and F2 values calculated for each sentence from the original and generated speech for both MMY and FTK. As a result, significant differences were observed in the mean values of F0 and F1 across all four comparisons. For F2, significant differences were also found between the two synthesized speeches, suggesting that the generated voices do not resemble each other closely. In addition, SVM was able to correctly identify the generated speeches as well as the ground-truth speeches with 85.34% accuracy. This indicates that different speeches can be recovered if the speakers of the speeches are different. The average value of MMY of the ground-truth speech and the generated speech increased from 154.6 to 221.2 Hz, The mean value of FTK of the ground-truth speech and FTK of the generated speech increased from 236.2 to 293.1 Hz. These results indicate that the F0 of the generated speech is higher than that of the ground-truth speech about 60 Hz. The values of variance do not change much, indicating that although the distribution of F0 is close to that of the heard sounds, the F0 is not correctly estimated as an average. However, although the generated speech has a different speaker characteristics from the ground-truth speech, MMY and FTK of the generated speech are clearly spoken by different speakers, and this can be seen from the F0 value and the accuracy of speaker identification by SVM. These facts suggest that features representing speaker characteristics can be estimated from EEG. However, we should note that the data in this experiment is only 480 sentences, and the high accuracy of the speech inferred from the training data suggests that the model may be over-trained on the training data. At the same time, the high accuracy of the speech inferred from the training data suggests that the model may be able to synthesize more accurate speech by increasing the training data.

V. Conclusion

As a model for the Transformer, we used the Voice Transfer Network to recover heard speeches from the EEG. This study showed the potential of the Transformer for speech decoding using EEG. Furthermore, using the speeches of two speakers and the EEG signals of the listeners, we were able to distinguish between the two speakers. This indicates that EEG signals contain information related to the vocal characteristics of the heard speeches. It also demonstrated the potential to reconstruct speeches from EEG signals.

As mentioned above, we could not examine the linguistic features of the speeches generated from the test data because the generated speeches were not saying meaningful sentences like the original speeches but rather were making a series of random sounds. Therefore, future work involves investigating in further detail the linguistic features of the speeches and improving the model structure to infer linguistic features from EEG signals.

Another issue that needs to be addressed in future experiments is to increase the amount of data in the experiments. First, the data needs to be increased to avoid over-fitting. We will also need to increase the number of speakers. In this experiment, we used the speeches of two speakers, a man and a woman, and the F0 values of these two speakers are very different. We will investigate the possibility of increasing the number of speakers to generate speeches with different characteristics from a wider variety of speakers.

VI. ACKNOWLEDGMENT

This work was supported by Hirokazu Matsuura, Kaya Chin, and Keito Nakamori of the Tokyo Institute of Technology for measuring and providing EEG data for the execution of this study. This research was supported by JSPS Grant-in-Aid for Scientific Research 23K11161 and 24H00715. This work was supported by JST FOREST Program (Grant Number JPMJFR216W).

References

[1] Anumanchipalli, G. K., et al. (2019). Speech synthesis from neural decoding of spoken sentences. Nature, 568, 493–498.

[2] Metzger, J. C., et al. (2023). A high-performance neuroprosthesis for speech decoding and avatar control. Nature, 620, 1037–1046.

[3] Akashi, W., et al. (2021). Vowel sound synthesis from electroencephalography during listening and recalling. Advanced Intelligent Systems, 2000164, 1-9.

[4] Kong, Z., et al. (2020). DiffWave: A Versatile Diffusion Model for Audio Synthesis. ArXiv, abs/2009.09761.

[5] Keshishian, M., et al. (2023). Joint, distributed and hierarchically organized encoding of linguistic features in the human auditory cortex. Nature human behaviour, 7(5), 740–753. https://doi.org/10.1038/s41562-023-01520-0

[6] Zhang, Y., et al. (2011). Cortical Dynamics of Acoustic and Phonological Processing in Speech Perception. PLOS ONE, 6(7), e20963.

[7] Yang, W., et al. (2022). Auditory attentional load modulates the temporal dynamics of audiovisual integration in older adults: An ERPs study. Frontiers in aging neuroscience, 14, 1007954. https://doi.org/10.3389/fnagi.2022.1007954

[8] Zhang, L., et al. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998–6008.

[9] Li, N., et al. (2019). Neural Speech Synthesis with Transformer Network. In Proceedings of the AAAI Conference on Artificial Intelligence, 33, 6706–6713.

[10] Huang, W.-C., et al. (2020). Voice Transformer Network: Sequence-to-Sequence Voice Conversion Using Transformer with Text-to-Speech Pretraining. In Proc. Interspeech, 4676–4680.

[11] Bahdanau, D., et al. (2014). Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.

[12] Luong, T., et al. (2015). Effective approaches to attention-based neural machine translation. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal, Sep. 2015, 1412–1421.

[13] Shen, J., et al. (2018). Natural TTS Synthesis by Conditioning WaveNet on MEL Spectrogram Predictions. In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 4779–4783.

[14] Yamamoto, R., et al. (2019). Parallel WaveGAN: A fast waveform generation model based on generative adversarial networks with multi-resolution spectrogram. arXiv preprint arXiv:1910.11480.

[15] Akira, K., et al. (1990). ATR Japanese speech database as a tool of speech recognition and synthesis. Speech Communication, 9(4), 357–363.

[16] Tachibana, H., et al. (2018). Efficiently Trainable Text-to-Speech System Based on Deep Convolutional Networks with Guided Attention. In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 4784–4788.

[17] Zhang, T., et al. (2019). Bertscore: Evaluating text generation with bert. arXiv preprint arXiv:1904.09675.

[18] Radford, A., et al. Robust speech recognition via large-scale weak supervision. In International Conference on Machine Learning, 28492-28518.