BRAIN-TO-SPEECH SYNTHESIS

Author(s) Liu HAO, Ma Zihan

1. INTRODUCTION

Brain-Computer Interface (BCI) is a technology that directly connects the brain to external devices, and can control computers, mechanical devices or other electronic devices by reading the brain's neural activity. The core goal of BCI is to interact with computer systems through EEG signals or other neural signals, thereby achieving a control method that does not rely on traditional body movements. It is widely used in medical, rehabilitation, communication and other fields.

In recent years, with the advancement of neural recording, computer technology and robotics, BMI has developed very rapidly. BMI is generally divided into three types: sensory, motor and bidirectional, which are used for movement, sensation and sensorimotor functions respectively. In addition, cognitive BMI has appeared in the field of advanced brain function. BMI is also divided into non-invasive or invasive according to the degree of interference of BMI with biological tissue. Although non-invasive BMI is safe and easy to implement, its information bandwidth is limited. Invasive BMI is expected to increase bandwidth by utilizing multi-channel recordings from brain neuron groups. BMI has a wide range of clinical goals, as well as the goal of enhancing normal brain function.

The working process of brain-computer interface usually includes the following steps:

1.Signal acquisition: First, various neural signal acquisition devices (such as EEG, functional magnetic resonance imaging fMRI, implanted electrodes, etc.) are used to capture the electrical activity of the brain.

2.Signal processing: The collected signals are usually very complex and contain noise and interference, so signal preprocessing is required, including denoising, filtering, amplification, etc., to extract information useful for controlling external devices.

3.Signal decoding: The processed signal is decoded into specific instructions or control information through an algorithm. This process is the most critical part of the brain-computer interface system, and the decoding algorithm needs to be able to accurately identify the user's intention.

4.Device control: The decoded signal is used to control external devices (such as robotic arms, mice, wheelchairs, prostheses, etc.). This enables users to directly operate these devices through brain electrical activity.

However, brain-computer interfaces have faced huge challenges in recent years. The first is signal quality and decoding accuracy. The signal quality of non-invasive BCI is poor and is greatly affected by noise, resulting in low decoding accuracy. Although invasive BCI has good signal quality, the risk of invasive surgery is high. The second is real-time and stability. BCI systems need to have real-time decoding and control capabilities, which are essential for fast and precise operation of external devices. However, the response speed and stability of most systems are not yet sufficient to meet the needs of high-precision applications. However, with the continuous advancement of technology, brain-computer interfaces have very broad prospects, and will play a big role in both medical treatment and communication. The future development of BCI will rely on close cooperation in the fields of neuroscience, computer science, artificial intelligence, and medicine to achieve more intelligent and precise brain-computer interaction.

2. METHODS OF TRAINING

We use deep neural networks instead of linear models to train, validate, and test a single speaker. We choose artificial neural networks for experiments. Later, we use the conformer model for cross-speaker training and synthesis. The following is a detailed description of the training method:

2.1. Artificial Neural Network (ANN)

Artificial Neural Network (ANN) is an algorithm that simulates the working mode of human brain nervous system. It is composed of a large number of neurons connected to each other, each neuron receives input signals and generates output signals through activation functions. Artificial neural network is based on the simulation of the neuron connection mode of the human brain, and is composed of input layer, hidden layer and output layer.

Principle: Each node is a specific function, which is equivalent to a neuron. The connection between each two nodes represents a weighted value. These weighted values are called weights as the connection between two nodes. Many nodes and weights form a complete network structure, and a visual nonlinear equation can be established through the network structure.

Advantages and disadvantages: The complex structure of

Thanks to XYZ agency for funding.

the artificial neural network itself can almost perfectly display any complex nonlinear relationship, and has a certain degree of adaptive ability, which can handle some unknown and uncertain situations; but if there is not enough data set, the artificial neural network cannot perform the next step of calculation and prediction.

Parameter settings:

```
model = models.Sequential()
model.add(layers.Dense(512, activation='relu', input_shape=(input_shape,)))
model.add(layers.Dropout(0.2))
model.add(layers.Dense(256, activation='relu'))
model.add(layers.Dropout(0.2))
model.add(layers.Dense(128, activation='relu'))
model.add(layers.Dense(64, activation='relu'))
```

FIG 1:

dense is the number of neurons

dropout is to randomly drop

neurons to prevent overfitting activation function is relu function

The optimizer uses adam

2.2. Conformer model

In this experiment, we use the Conformer model for training and evaluation. The Conformer (Convolution-augmented Transformer) model is an architecture that combines convolutional neural networks (CNNs) with Transformers, and is mainly used in sequence-to-sequence tasks such as speech recognition. It enhances the model's local feature capture capabilities by introducing convolutional modules, while retaining the Transformer model's advantages in modeling long-term dependencies. The Conformer encoder model architecture is shown in Figure 1:

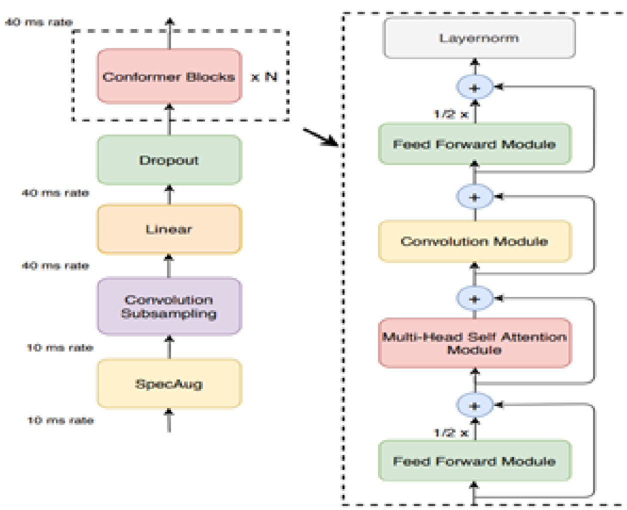

FIG 2

Parameter settings:

Learning Rate: model.optim.lr=5.0

The learning rate (LR) determines the step size or learning speed each time the model parameters are updated. A suitable learning rate can speed up training and help the model converge to a better solution.

Optimizer: AdamW model.optim.betas=[0.9,0.98]

The optimizer is used to calculate and update model parameters to minimize the loss function. AdamW is a variant of the Adam optimizer with weight decay added. It combines the momentum method (through the estimation of the first-order 0.9 and second-order moment 0.999) and the adaptive learning rate to help accelerate training and avoid local optimality.

batch size: 16 Batch Size specifies how many samples are used in each iteration to calculate the gradient and update the model parameters. Here we choose 16.

3. EVALUATE

3.1. ANN Data Visualization

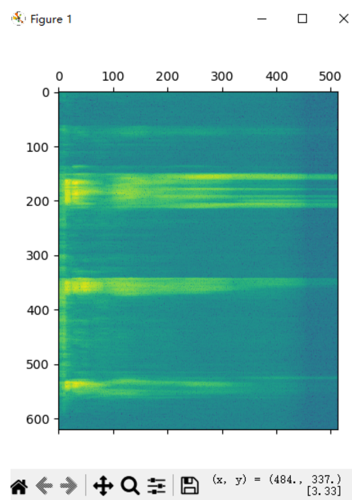

FIG 3:

Shape of complex spectrum is (621,513)

0.4977860450744629

3.2. Data visualization after the Conformer model

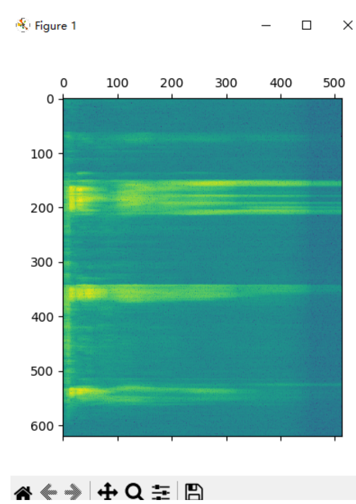

FIG 4:
Shape of complex spectrum is (621,513)
0.5293223857879639

4. CONCLUSION

Brain-computer interface (BCI) technology is a promising and transformative approach that enables direct communication between the brain and external devices through neural signals. It has a wide range of applications in medical rehabilitation, communication, and cognitive enhancement. In this experiment, we focused on improving the accuracy of the model by replacing the linear model with deep neural networks and train/valid/test set from single speaker. We then further implemented cross-speaker train and synthesis. Despite significant progress in BCI technology, it still faces challenges such as signal quality, decoding accuracy, real-time performance, and system stability. The future prospects of BCI technology are promising, and the continued progress in neuroscience, artificial intelligence, and signal processing algorithms will promote the widespread application of BCI systems in clinical and non-clinical fields.

5. REFERENCES

[1] Lebedev, Mikhail A., and Miguel AL Nicolelis. ”Brain–machine interfaces: past, present and future.” TRENDS in Neurosciences 29.9 (2006)

[2] Lebedev, M. A., Tate, A. J., Hanson, T. L., Li, Z., O’Doherty, J. E., Winans, J. A., ... Nicolelis, M. A. Future developments in brain-machine interface research. Clinics.(2011)

[3] Lebedev, Mikhail. ”Brain-machine interfaces: an overview.” Translational Neuroscience 5 (2014)

[4] Andersen, Richard A., Sam Musallam, and Bijan Pesaran. ”Selecting the signals for a brain–machine interface.” Current opinion in neurobiology 14.6 (2004)

[5] Musk, Elon. ”An integrated brain-machine interface platform with thousands of channels.” Journal of medical Internet research 21.10 (2019)

[6] Patil, Parag G., and Dennis A. Turner. ”The development of brain-machine interface neuroprosthetic devices.” Neurotherapeutics 5 (2008)

[7] Lebedev, Mikhail A., and Miguel AL Nicolelis. ”Brain-machine interfaces: from basic science to neuroprostheses and neurorehabilitation.” Physiological reviews 97.2 (2017)

[8] Ramos-Murguialday, Ander, et al. ”Brain–machine interface in chronic stroke rehabilitation: a controlled study.” Annals of neurology 74.1 (2013)

[9] Shanechi, Maryam M. ”Brain–machine interfaces from motor to mood.” Nature neuroscience 22.10 (2019)