A Review of Research on Brain-Computer Interface Based on Imagined Speech

Chengyin Wang[1], Wenlong Ding[1], Jianhua Shan[1], and Bin Fang[2](✉)

[1] Anhui University of Technology, Maanshan 243032, Anhui, China
[2] Tsinghua University, Beijing 100085, China
fangbin@tsinghua.edu.cn

Abstract. Brain-computer interface is currently a rapidly developing technology. In recent years, it has received extensive attention and high expectations in the fields of biomedical engineering and rehabilitation medicine engineering. Brain-computer interfaces can enable patients with communication skills or physical disabilities to communicate with machines and equipment, and brain-computer interfaces based on imagined speech can provide patients with normal and effective language communication. At present, its related research has achieved certain results. This article introduces the principles, advantages and disadvantages of several common BCI systems, as well as the two most widely used brain signals EEG and EcoG, and then studies some related feature extraction and data classification algorithms used in current research. Finally, the current problems and future development trends of brain-computer interfaces based on imagined speech are discussed.

Keywords: BCI · Imagined speech · EEG · EcoG

1 Introduction

The brain-computer interface (BCI, hereinafter referred to as BCI) is a direct connection channel created between the human or animal brain and external equipment. The BCI is divided into one-way BCI and two-way BCI; one-way BCI technology means that the computer only accepts information from the brain or transmits information to the brain. The two-way BCI allows two-way information exchange between the brain and external devices. The emergence of BCI has provided great convenience to patients with speech and physical impairments. Nowadays, patients can realize cursor movement through the BCI system, control wheelchairs, letter input and prosthetic movements [1–3].

BCI system includes BCI system based on External Stimulation (Visual P300, SSVEP) and Motor Imagery (SMR, IBK) system. First, the P300 component refers to the positive waveform generated by the EEG signal about 220 to 500 ms after the target stimulus occurs in a stimulation sequence with a small proportion of target stimuli [4]. The P300 paradigm includes auditory P300 and visual P300. At present, the

This work was supported by National Natural Science Foundation of China with Grant No. 91848206 and Natural Science Foundation of university in Anhui Province (No. KJ2019A0086).

F. Sun et al. (Eds.): ICCSIP 2020, CCIS 1397, pp. 370–378, 2021.
https://doi.org/10.1007/978-981-16-2336-3_34

visual P300 paradigm is more widely used [5, 6]. The advantage of P300-BCI is that it is non-invasive, requires less training times, provides communication and control functions, and is a stable and reliable BCI system. So, P300-BCI is the most suitable BCI system for severely disabled patients to independently use in the home environment for a long time. Second, the Steady State Visual Evoked Potential (SSVEP) is another popular visual component used in BCI, SSVEP is also called an optical drive because the generator of this response is located in the visual cortex. The subject must look away and pay attention to the flickering stimulus, not the movement execution or imaginary movement, which requires highly precise eye control, common stimulus sources include flash, light-emitting diodes and a checkerboard pattern of displays. The advantage of SSVEP-BCI is that it has high information transmission rate and many output commands. The subjects only need relatively little training to use it. The disadvantage is that it needs to rely on a stimulus source. However, long-term use of flicker (mainly low frequency) stimulation may cause subject fatigue [7–10]. Then, the SMR paradigm is the most widely used Imagined Motor paradigms. Imagined Motor refers to the imagination of the kinesthetic movement of larger parts of the body such as the hands, feet, and tongue, which can lead to the regulation of brain activity. Specifically, it is the electrophysiological phenomenon of Event-related Desynchronization (ERD/Event-related Synchronization, ERS) to control the device [11]. Finally, Imaginary Body kinematics (IBK) is a motor imagery paradigm derived from invasive BCI techniques [12, 13]. However, non-invasive research pointed out that this mode of information is extracted from low-frequency SMR signals (less than 2 Hz) [14]. Although IBK belongs to SMR, it is classified into a separate category due to its different training and analysis methods from the SMR paradigm. The biggest advantage of MI-BCI is that the BCI control signal generated by the brain action intention is an endogenously induced EEG, so it does not require external stimulation; but it requires multiple training, and the classification accuracy rate is not high, and individual differences cannot be resolved. Imagined speech is similar to motor imagery, and we often use it in our lives, such as silently reading magazines, books, the process of thinking about something in the brain, and recalling conversations with others. The BCI system based on imagined speech extracts the brain signals of the subjects when they imagine pronunciation, and then through a series of data processing, it is finally converted into speech. In order to remove noise, the subjects will be asked not to make a sound when imagining the pronunciation and try not to change their expressions. The BCI system based on imagined speech and the BCI system based on motor imagination are similar in that they both extract the brain signals of the participant during the imagination and convert them into desired actions, such as body movements or voice output, and neither need external stimuli. The BCI system based on P300 and SSVEP requires external stimuli, such as light flicker. The BCI system based on imagined speech also has great prospects in application. It can help patients with language barriers, muscle atrophy, locked-in syndrome and other diseases to communicate and communicate effectively with the outside world. Input letters, cursor selection, etc. are more efficient and more convenient. However, compared with other BCI systems, the technology of the BCI system based on imagined speech is not mature enough, and there are still shortcomings in hardware and brain signal decoding, but the BCI system based on imagined speech has great potential and research significance. It is worth our continued in-depth study.

2 Brain Sensor

The common methods of brain-computer connection can be divided into two types: invasive and non-invasive. Non-invasive methods do not require surgery, mainly including electroencephalography (EEG), magnetoencephalography (MEG), functional magnetic resonance imaging (fMRI), near infrared spectroscopy (NIRS), etc. In addition, it also includes many invasive methods, which may cause certain harm to the human body, including neuron firing signals (spikes), laminar potentials (electrocortex, ECoG), etc. Most brain-computer interfaces have selected EEG signals as the input, which has become the most important part of the brain-computer interface.

2.1 EEG Device

Electroencephalography (EEG) is a widely used non-invasive method for monitoring the brain. It is based on the function of placing conductive electrodes on the scalp, which can measure small electric potentials generated outside the head due to the action of neurons in the brain. The original EEG acquisition device was: a user wore a cap with holes and placed several electrodes next to the scalp. Each electrode had a long wire connected to the recording instrument, the wires are tangled together, which is troublesome to install, and the movement of the unshielded EEG wire will have a great impact on the quality of the collected signal. At present, EEG collection equipment is very advanced, Ahn JW [30] has developed a new wearable device that can measure both electrocardiogram (ECG) and EEG at the same time to realize continuous pressure monitoring in daily life, the developed system is easy to hang on two ears, is light in weight (ie 42.5 g), and has excellent noise performance of 0.12 μVrms. [31] studied a wearable in-ear EEG for emotion monitoring. The device is a low-cost, single-channel, dry contact, in-ear EEG, suitable for non-invasive monitoring, based on the valence and arousal Emotion model, the device can classify basic emotions with 71.07% accuracy, 72.89% accuracy (awakening) and 53.72% (all four emotions). [32] studied the hat-shaped EEG device EEG-Hat with candle-shaped dry microneedle electrodes. The current wearable EEG device has two main problems: 1) it is not adaptable to each participant, 2) in EEG cannot be measured on the hair area. The device can adjust the electrodes according to the size of the subject's head and can be used by multiple people. The device has a louver-like structure to separate hair. After experiments, it was found that the EEG cap successfully measured the EEG of 3 hair parts without manual separation. Currently, the most widely used commercial products are: mBrainTrain Smarting, Brain Products LiveAmp, g.tec g.Nautilus, Cognionics Mobile-128, Emotiv Epoc Flex.

2.2 EcoG Device

Cortical ECoG is used clinically for the detection of epileptic foci, ECoG electrodes are very common and mature clinically, and neurosurgeons need to perform craniotomy or craniotomy to insert them. The electrode disk is inlaid on a silicon rubber sheet, during the operation, the doctor covers the silicon rubber sheet on the patient's cerebral cortex and subdura, which can collect cortical EEG signals. Generally, the clinical detection of epileptic foci ranges from 1–2 weeks, experiments on mice and monkeys have proved that the signals collected by EcoG can remain stable for up to 5 months. [34] designed a novel spiral electric cortex (ECoG) electrode, which consists of three parts: recording electrode, insulator and nut, compared with electroencephalogram (EEG), it has a higher SNR and a wider frequency band, with higher sensitivity, and can capture different responses to various stimuli. [33] et al. studied a flexible EcoG electrode for studying spatiotemporal epilepsy morphological activity and multimodal neural encoding/decoding. The flexible electrode has very little damage to patients and is of great significance for clinical treatment and research. [35] studied a novel flexible and bioabsorbable ECoG device integrated with an intracortical pressure sensor to monitor cortical swelling during operation. The flat and flexible ECoG electrode can minimize the risk of infection and severe inflammation. Its good shape adaptability enables the device to adapt to complex cortical shapes and structures to record brain signals with high spatiotemporal resolution.

3 Research Status

In the experiment of [16], some native speakers were asked to tell a story, and the subjects were asked to listen to the story carefully. After the story was told, the subjects were asked some related questions to ensure that the subjects listened carefully. The classification accuracy of Chinese phoneme clusters according to the pronunciation position and pronunciation mode was tested by using the small Wave sign and support vector machine classifier, with the accuracy of about 60%. In the experiment of [25], RNN and DBN were used to classify and recognize five vowels respectively, and it was found that DBN had a better effect, 8% higher than RNN. In the experiment [19], subjects were asked not to perform any movements or activities, especially lips, tongue and chin, and then brain signals were used to decode whether they were thinking "yes" or "no" with an average accuracy of 69.3%. The subjects will be asked ten questions with answers of "yes" or "no", such as: Are you hungry? The subjects answered "yes" or "no", and the decoding of the brain signal was accurate 92 percent of the time compared to the real thing [22]. In Dash [18]'s experiment, the screen went blank for the first second, then text appeared. The subjects imagined it for one second, then read it aloud for two seconds. They trained on five commonly used phrases and analyzed MEG using CNN with 93% accuracy. Tottrup L [24] use of EMG signal to improve the training effect, every action is secrecy speech or MI (interior), six seconds, then repeat openly talking or ME (external), corresponds to a clock cycle time in front of the subjects had a clock, subjects first imagine some action, such as stopping or walking, bending the left arm, after 6 s, they repeat these movements overtly, the highest with 76% accuracy. At present, many researches are based on monosyllables or monosyllables. It is still difficult to carry out experiments on words or sentences, but some achievements have been made. In the experiments of [26], using the same algorithm, the accuracy rate of the experimental results was 57.4%/57CV, 61.2%/19cons, 88.2%/3VOWels, proving that

Table1. Research status

Signal	Area	Subject	Feature extraction	Classification	Accuracy	Author
EcoG	vSMC	5 epliepsy patients	Hilbert	LSTM	Almost 70%	Anumanchipalli [17]
EcoG	Posterior temporal lobe and inferior parietal cortex	5 epliepsy patients	Wavelet transform	SVM	Mean 60.47%	Song C [16]
MEG	All	8 healthy adults	Wavelet transform	CNN	93%	Dash D [18]
SUA/LFP/ EcoG	vSMC	6 epilepsy patients	Construct feature vector	Sparse Logistic Regression classifier	Mean 59%	Ibayashi [15]
EEG	All	12 healthy adults	Wavelet transform, Autoregressive model	SVM	Mean 69.3%	Sereshkeh [19]
EcoG	Frontal and temporal areas	4 epilepsy patients	FFT	Linear classifier	63% for single phoneme	Mugler [20]
EcoG	Frontal, parietal and temporal regions	8 epliepsy patients	Power Spectral Density	Naive Bayes classifier	Vowel-mean 37.5% Consonant-mean 36.3%	Pei [21]
EEG	Brocas, Geshwind-Wernicke's area	5 healthy adults	PCA	ANN	Best 92.18%	Balaji [22]
EEG	All	6 healthy adults	Wavelet transform	Extreme Learning Machine	Multi-category 49.77%/Two categories 85.57%	Pawar [23]

(*continued*)

Table1. (*continued*)

Signal	Area	Subject	Feature extraction	Classification	Accuracy	Author
EEG	Frontal, central, parietal areas	7 healthy adults	The temporal and spectral features	Random forest classifier	Best 76%	Tottrup [24]
EEG	All	6 healthy subjects	Power Spectral Density	DBN/RNN	DBN80%/RNN72%	Chenggaiyan [25]
EcoG	vSMC	4 epilepsy patients	Deep network	Deep network	Bset: 57.4%/57CV, 61.2%/19 cons, 88.2%/3 vowels	Livezey [26]
EcoG	Left Brain	2 epilepsy patients	Hilbert transform	Linear decoder	81%	Bouchard [27]

the more difficult the task, the lower the accuracy rate. Subjects were asked to read and read the story silently, then to convert the brain signals they collected into speech, and to listen to the final synthetic sentence to complete the test. After hearing 101 sentences, the accuracy rate was about 70% [17]. Anumanchipalli [17] and other experiments also found that reading aloud was more effective than silent reading because sounds were added to aid training. The accuracy of reading aloud was 3% higher in the experiment than in silent reading training. MFCCS features are generally extracted from speech to facilitate training, such as [28] and [17]. Makin JG [28] also used MOCHATIMIT data set to decode and synthesize sentences, and they achieved 97% decoding accuracy by using EcoG signal, and achieved certain results in transfer learning. Pre-training participant A's data improved participant B's performance. For the least effective participant D, there was no improvement, and all individual differences remained difficult to eliminate. See Table 1 for more research status.

4 Conclusion

The brain-computer interface system based on imagined speech has achieved certain results, but there is still a lot to go. At present, it is possible to improve the training effect and improve the test accuracy by extracting the characteristics of the speech signal and fusing the brain signal. Our application target is those who can't speak, so we can only use brain signals for training. Therefore, the brain-computer interface system based on imagined speech has a good development prospect, but further research is needed.

References

1. Kübler, A., Kotchoubey, B., Hinterberger, T., et al.: The thought translation device: a neurophysiological approach to communication in total motor paralysis. Exp. Brain Res. **124**(2), 223–232 (1999)
2. Yahud, S., Abu Osman, N. A.: Prosthetic hand for the brain-computer interface system. In: Ibrahim, F., Osman, N.A.A., Usman, J., Kadri, N.A. (eds.) 3rd Kuala Lumpur International Conference on Biomedical Engineering 2006. IP, vol. 15, pp. 643–646. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-68017-8_162
3. Rebsamen, B., Burdet, E., Guan, C., et al.: Controlling a wheelchair indoors using thought. IEEE Intell. Syst. **22**(2), 18–24 (2007)
4. Abiri, R., Borhani, S., Sellers, E.W., Jiang, Y., Zhao, X.: A comprehensive review of EEG-based brain-computer interface paradigms. J. Neural Eng. **16**(1), 011001 (2019). https://doi.org/10.1088/1741-2552/aaf12e
5. Fabiani, M., Gratton, G., Karis, D., Donchin, E.: Definition, identification, and reliability of measurement of the P300 component of the event-related brain potential. Adv. Psychophysiol. **2**(S 1), 78 (1987).
6. Polich, J.: Updating P300: an integrative theory of P3a and P3b. Clin. Neurophysiol. **118**(10), 2128–2148 (2007)
7. Chang, M.H., Baek, H.J., Lee, S.M., Park, K.S.: An amplitude-modulated visual stimulation for reducing eye fatigue in SSVEP-based brain–computer interfaces. Clin. Neurophysiol. **125**(7), 1380–1391 (2014)

8. Molina, G.G., Mihajlovic, V.: Spatial filters to detect steady-state visual evoked potentials elicited by high frequency stimulation: BCI application. Biomedizinische Technik/Biomed. Eng. **55**(3), 173–182 (2010)
9. Müller, S.M.T., Diez, P.F., Bastos-Filho, T.F., Sarcinelli-Filho, M., Mut, V., Laciar, E.: SSVEP-BCI implementation for 37–40 Hz frequency range. In: Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE, pp. 6352–6355: IEEE (2011)
10. Volosyak, I., Valbuena, D., Luth, T., Malechka, T., Graser, A.: BCI demographics II: how many (and what kinds of) people can use a high-frequency SSVEP BCI? IEEE Trans. Neural Syst. Rehabil. Eng. **19**(3), 232–239 (2011)
11. Morash, V., Bai, O., Furlani, S., Lin, P., Hallett, M.: Classifying EEG signals preceding right hand, left hand, tongue, and right foot movements and motor imageries. Clin. Neurophysiol. **119**(11), 2570–2578 (2008)
12. Hochberg, L.R., et al.: Neuronal ensemble control of prosthetic devices by a human with tetraplegia. Nature **442**(7099), 164–171 (2006)
13. Kim, S.-P., Simeral, J.D., Hochberg, L.R., Donoghue, J.P., Black, M.J.: Neural control of computer cursor velocity by decoding motor cortical spiking activity in humans with tetraplegia. J. Neural Eng. **5**(4), 455 (2008)
14. Yuan, H., He, B.: Brain-computer interfaces using sensorimotor rhythms: current state and future perspectives. IEEE Trans. Biomed. Eng. **61**(5), 1425–1435 (2014)
15. Ibayashi, K., Kunii, N., Matsuo, T., et al.: Decoding speech with integrated hybrid signals recorded from the human ventral motor cortex. Front. Neurosci. **12,** 221 (2018). https://doi.org/10.3389/fnins.2018.00221
16. Song, C., Xu, R., Hong, B.: Decoding of Chinese phoneme clusters using ECoG. In: Conference Proceedings-IEEE Engineering in Medicine and Biology Society 2014, pp. 1278–1281 (2014). https://doi.org/10.1109/EMBC.2014.6943831
17. Anumanchipalli, G.K., Chartier, J., Chang, E.F.: Speech synthesis from neural decoding of spoken sentences. Nature **568**(7753), 493–498 (2019). https://doi.org/10.1038/s41586-019-1119-1
18. Dash, D., Ferrari, P., Wang, J.: Decoding Imagined and spoken phrases from non-invasive neural (MEG) signals. Front. Neurosci. **14**, 290 (2020). https://doi.org/10.3389/fnins.2020.00290
19. Sereshkeh, A.R., Trott, R., Bricout, A., Chau, T.: Online EEG classification of covert speech for brain-computer interfacing. Int. J. Neural Syst. **27**(8), 1750033 (2017). https://doi.org/10.1142/S0129065717500332
20. Mugler, E.M., Patton, J.L., Flint, R.D., et al.: Direct classification of all American English phonemes using signals from functional speech motor cortex. J. Neural Eng. **11**(3), 035015 (2014). https://doi.org/10.1088/1741-2560/11/3/035015
21. Pei, X., Barbour, D.L., Leuthardt, E.C., Schalk, G.: Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans. J. Neural Eng. **8**(4), 046028 (2011). https://doi.org/10.1088/1741-2560/8/4/046028
22. Balaji, A., Haldar, A., Patil, K., et al.: EEG-based classification of bilingual unspoken speech using ANN. In: Conference Proceedings-IEEE Engineering in Medicine and Biology Society 2017, pp. 1022–1025 (2017). https://doi.org/10.1109/EMBC.2017.8037000
23. Pawar, D., Dhage, S.: Multiclass covert speech classification using extreme learning machine. Biomed. Eng. Lett. **10**(2), 217–226 (2020). https://doi.org/10.1007/s13534-020-00152-x
24. Tottrup, L., Leerskov, K., Hadsund, J.T., Kamavuako, E.N., Kaseler, R.L., Jochumsen, M.: Decoding covert speech for intuitive control of brain-computer interfaces based on single-trial EEG: a feasibility study. In: IEEE International Conference on Rehabilitation Robotics 2019, pp. 689–693 (2019). https://doi.org/10.1109/ICORR.2019.8779499

25. Chengaiyan, S., Retnapandian, A., Anandan, K.: Identification of vowels in consonant–vowel–consonant words from speech imagery based EEG signals. Cogn. Neurodyn. **14**(1), 1–19 (2019). https://doi.org/10.1007/s11571-019-09558-5
26. Livezey, J.A., Bouchard, K.E., Chang, E.F.: Deep learning as a tool for neural data analysis: speech classification and cross-frequency coupling in human sensorimotor cortex. PLoS Comput. Biol. **15**(9), e1007091 (2019). https://doi.org/10.1371/journal.pcbi.1007091
27. Bouchard, K.E., Chang, E.F.: Neural decoding of spoken vowels from human sensory-motor cortex with high-density electrocorticography. In: Conference Proceedings-IEEE Engineering in Medicine and Biology Society 2014, pp. 6782–6785 (2014). https://doi.org/10.1109/EMBC.2014.6945185
28. Makin, J.G., Moses, D.A., Chang, E.F.: Machine translation of cortical activity to text with an encoder-decoder framework. Nat. Neurosci. **23**(4), 575–582 (2020). https://doi.org/10.1038/s41593-020-0608-8
29. Akbari, H., Khalighinejad, B., Herrero, J.L., Mehta, A.D., Mesgarani, N.: Towards reconstructing intelligible speech from the human auditory cortex. Sci. Rep. **9**(1), 874 (2019). https://doi.org/10.1038/s41598-018-37359-z
30. Ahn, J.W., Ku, Y., Kim, H.C.: A novel wearable EEG and ECG recording system for stress assessment. Sensors (Basel) **19**(9), 1991 (2019). https://doi.org/10.3390/s19091991
31. Athavipach, C., Pan-Ngum, S., Israsena, P.: A wearable in-ear EEG device for emotion monitoring. Sensors (Basel). **19**(18), 4014 (2019). https://doi.org/10.3390/s19184014
32. Kawana, T., Yoshida, Y., Kudo, Y., Miki, N.: In: EEG-hat with candle-like microneedle electrode. In: Conference Proceedings-IEEE Engineering in Medicine and Biology Society 2019; pp. 1111–1114 (2019). https://doi.org/10.1109/EMBC.2019.8857477
33. Shi, Z., Zheng, F., Zhou, Z., et al.: Silk-enabled conformal multifunctional bioelectronics for investigation of spatiotemporal epileptiform activities and multimodal neural encoding/decoding. Adv. Sci. (Weinh) **6**(9):1801617 (2019). https://doi.org/10.1002/advs.201801617
34. Choi, H., Lee, S., Lee, J., et al.: Long-term evaluation and feasibility study of the insulated screw electrode for ECoG recording. J. Neurosci. Methods. **308**, 261–268 (2018). https://doi.org/10.1016/j.jneumeth.2018.06.027
35. Xu, K., Li, S., Dong, S., et al.: Bioresorbable electrode array for electrophysiological and pressure signal recording in the brain. Adv. Healthc. Mater. **8**(15), e1801649 (2019). https://doi.org/10.1002/adhm.201801649
36. Brumberg, J.S., Pitt, K.M., Burnison, J.D.: A noninvasive brain-computer interface for real-time speech synthesis: the importance of multimodal feedback. IEEE Trans. Neural Syst. Rehabil. Eng. **26**(4), 874–881 (2018). https://doi.org/10.1109/TNSRE.2018.2808425