

## Research on Feature Recognition on mVEP BCI

Teng Ma <sup>1, a</sup>, Hui Li <sup>1, b</sup>, Dezhong Yao <sup>1, 2, c</sup>, Peng Xua <sup>2, d</sup>

<sup>1</sup>Key Laboratory for NeuroInformation of Ministry of Education, School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, 610054, China

<sup>2</sup>Center for Information in BioMedicine, University of Electronic Science and Technology of China, Chengdu, 610054, China

<sup>a</sup>1178016887@qq.com; <sup>b</sup>465123149@qq.com; <sup>c</sup>dyao@uestc.edu.cn; <sup>d</sup>xupeng@uestc.edu.cn.

**Keywords:** locally liner embedding; kernel entropy component analysis; kernel entropy liner embedding.

**Abstract.** The recognition to element N2 is the fundamental basis determining the stability and the practicability of the mVEP-BCI systems. The feature extraction to EEG signals is the key for the accuracy on recognition of element N2. Different from the traditional down sampling feature extraction method, the method in this article is designed by utilizing the over-complete dictionary based compressed sensing method, to conduct dimension reduction processing to EEG signals in the time window by utilizing the row echelon observation matrix of different compression ratios for multiple times, to acquire the ordinary features of mVEP signals. It further conducts sparse noise reduction to ordinary features. It also conducts LDA classifications on down sampling features and the features extracted in this article. According to verification, the classification accuracy rate of the features extracted in this article has significant improvement than that of the traditional method. The feature extraction method in this article improves classification effect by taking the avoidance of overfitting and the retaining of useful information of original signal to the maximum extent into consideration, which improves the classification effect and enhances the stability and practicability of mVEP-BCI system in a more efficient way.

### 1. Introduction

The brain-computer interface (BCI) is the direct link between brains and external equipment <sup>[1]</sup>. The system collects the brain activity signals reflecting the intention of users, and decodes specific brain patterns with computer, and “translates” them to control instructions to control external equipment. It is widely applied for living assistant [2, 3] for the disabled and the rehabilitation of functions such as movement [4-6], because it avoids the direct connection of nerve cells between the brain and the other part. Besides, BCI system is divided into two types as per the detection means for brain activities: the BCI system detecting electro-discharge activities of nerve cells such as Electroencephalography (EEG) [7], Electrocorticography (ECoG) [8] and Local field potentials (LFP); the BCI system detecting the metabolic activities of nerve cells such as Magnetoencephalography (MEG) [9] and functional magnetic resonance imaging (fMRI) [10]. Among the above BCI systems, the EEG based BCI system is the BCI system with the lowest cost and the widest practical applications. This article mainly conducts research on motion-onset visual evoked potential (mVEP) potential based BCI system.

Element N2 is the most important component of the waveform features of mVEP, which often appears at about 130 ms, and the element N2 generated at the temporo-occipital area often appears at 160-200 ms, which mainly reflects the nerve activities involved in movement visual processing <sup>[11]</sup>. However, there are large differences on the ranges of N2 collected on the left and the right hemispheres on the brain, with no relevance to left handedness or right handedness <sup>[12]</sup>. According to numerous research findings, compared with traditional visual evoked potential, mVEP has larger range and smaller differences between different subjects and within the same subject <sup>[11]</sup>. mVEP has been widely applied due to the above feature.

The essence of mVEP based BCI system classification recognition is the recognition to the EEG signals of target stimulation and non-target stimulation, i.e., the problem of dichotomy. The traditional method conducts dimension reduction for the EEG signals in the time window under target stimulation and non-target stimulation in the way of down sampling, to acquire the feature vector. However, although pure down sampling can reduce the dimension, it inevitably leads to the large quantities of useful information; what's more, it results in the reduction of the ultimate classification effect.

By adopting the over-complete dictionary based compressed sensing method, this article conducts dimension reduction processing to the EEG signals in the time window through row echelon observation matrix; it compresses the original signals to the maximum extent by utilizing the sparseness of the signals in the over-complete dictionary, and reserves the features of the original signals to the maximum extent. The article then conducts noise reduction with the method of sparse noise reduction to restrain the random noise during the experiment process. It takes the vector quantity after down sampling as the ordinary features of mVEP signals, and calls the mVEP signals after compressed sensing and sparse decomposition noise reduction as the ordinary features. Among various classification methods, it selects LDA classifier to recognize the ordinary features and the sensing features of mVEP signals.

## 2. Method for Feature Extraction

The determination of the sparseness of signals is the first step for the compressed sensing method. The classification of mVEP signals mainly involves in the signals of two labels: the target stimulation signal and the standard stimulation signal. In which the target stimulation signal is considered to contain the information of N2 and P2 elements, and the standard stimulation signal is considered to contain no information of any useful element of mVEP. Therefore, as for the recognition of mVEP signals, compared with the target stimulation signal, the standard stimulation signal is 0; it can be considered that under the same dictionary, the useful information of the sparseness of target stimulation signal is higher than that of standard stimulation signal. Under the same compression degree, standard stimulation signal has higher reconstruction probability than that of target stimulation signal. Therefore, the sparseness of target stimulation signal can be taken as the common sparseness of the signals of the two labels.

As for a group of mVEP signal data, the target stimulation signals with different trials are taken, to superpose all the trial target stimulation signals. The superposed signals are conducted with sparse decomposition in the over-complete dictionary. The decomposition stops when the reconstruction error of the sparse decomposition achieves fixed threshold value, and the number of atoms acquired at this time is the sparseness of the superposed signals, i.e., the sparseness of the group of mVEP stimulation experiment signals.

In which the selected over-complete dictionary is the Gabor element word stock with preferable performance on voice signal processing [16].

$$g_r = \frac{1}{\sqrt{s}} g\left(\frac{t-u}{s}\right) \cos(vt+w) \quad (1)$$

In the above formula,  $g(\cdot)$  is Gaussian window function, and  $g(t) = e^{-\pi t^2}$ ;  $\gamma = f(s, u, v, w)$  is time-frequency parameter, and  $s, u, v$  and  $w$  is scale factor, shift factor, frequency factor and phase factor, respectively. Discretization is conducted as per the method in (4):

$$\gamma = f(a^j, pa^j \Delta_u ka^{-j} \Delta_v i \Delta_w) \quad (2)$$

And then the MP decomposition algorithm [17] with the widest application is utilized to conduct sparse decomposition on superposed signals in Gabor element word stock.

The reconstruction error of sparse decomposition is noted as:  $Er = \|f - fK\| / \|f\|$ ;  $f$  is the original signal, and  $fK$  is the reconstruction signal, representing for the ratio between the reconstruction error energy and the original signal energy. The threshold value at the stop of the sparse decomposition is set

as 0.01, and the sparse decomposition stops when the reconstruction error  $Er$  is less than 0.01; record the number of current atoms, to determine the sparseness.

Although the compressed sensing method can reduce dimension to the original signal to a large extent, it is especially important to select proper observation matrix, in order to avoid the defect of the down sampling feature that is cannot embody the feature information of the original signal. Traditional observation matrixes generally include Gaussian random matrix, partial Hadamard matrix and partial Fourier matrix. Among voice signal compressed sensing methods, row echelon matrix has an excellent performance on retaining original signal feature information. Therefore, the echelon matrix is selected in this article as the observation matrix, and the row echelon matrix with the compression ratio of 1:4 is shown as follows:

$$\phi = \begin{pmatrix} 111100000000...0 \\ 000011110000...0 \\ ..... \\ 0...000000001111 \end{pmatrix} \quad (3)$$

According to what mentioned in previous paragraphs, it can be known that the projection of signal  $X \in R^{N \times 1}$  by the observation matrix  $\phi \in R^{M \times N}$  ( $M < N$ ) can be utilized to acquire the observation vector  $Y \in R^{M \times 1}$ , which meets  $Y = \phi X$ . Compared with the original signal  $X$ , the dimension of the observation vector  $Y$  is reduced from  $N$  to  $M$ . But the reconstruction of high accuracy is the premise for the compression of the compressed algorithm; therefore, it is needed to reconstruct the signal after compression, in order to inspect the effect of dimension reduction. It is generally regarded that under the premise that matrixes  $\phi$  and  $\psi$  meet RIP, when  $M > k \times \log N$ , it is available to reconstruct the original signal  $X$  through the observation to the vector  $Y$ , in which  $k$  is the sparseness of the original signal. Known as the sensing matrix, Matrix  $\Theta = \phi \times \psi$  can be regarded as a group of sensing atoms. The MP algorithm<sup>[17]</sup> is utilized to decompose the observation vector  $Y$  under this group of sensing atoms, to acquire the coefficient vector  $\theta$ , and the product vector of  $\theta$  and matrix  $\Psi$  can be regarded as the accurate reconstruction to the original signal. Based on the above procedures, this article conducts reconstruction to all the compressed signals, and works out the reconstruction error:  $Er = \|f - fK\| / \|f\|$ , to verify the dimension reduction effect.

It is generally regarded that a group of vectors are inevitably sparse under certain transform domain, and the observation vector after the dimension reduction through observation matrix is inevitably in the same condition. Therefore, the vector after dimension reduction has the condition for dimension reduction again, as long as the vector for dimension reduction again includes the feature information of the original signal. On this basis, this article conducts multiple compressed sensing dimension reductions to the original signal, in order to achieve better dimension reduction effect; at the same time, it checks each dimension reduction through compressed sensing reconstruction error, to ensure that the vector for further dimension reduction includes sufficient original signal feature information. In order to eliminate the interference from the random noise introduced during the experiment, this article adopts the sparse decomposition noise reduction method by selecting the over-complete dictionary and the MP algorithm based sparse decomposition algorithm mentioned before. The determination of coherent ratio threshold value is the key point for sparse decomposition noise reduction, which is defined as:

$$\lambda(R, f) = \sup_{r \in Y} \frac{|R_n f, g_r|}{\|R_n f\|} \quad (4)$$

Generally speaking, the components of noise-free signal have close coherence with the atom dictionary; higher coherence leads to larger coherent ratio. However, the coherence is weak for random noise, because the atom dictionary includes no random atom vector. With the proceeding of sparse decomposition, the coherence between residual error and atom dictionary increasingly reduces, leading to increasingly reduced coherent ratio. The effect of noise elimination can be achieved if the components with large coherence are extracted, and the ones with small coherence are eliminated.

### 3. Results

#### 3.1 Feature Extraction

As for the experimental data acquired from laboratory, the data led by CP1 and Pz including P1, N2 and P2 elements of 131ms to 323ms are utilized for analysis verification. In which the first group of experimental data of each subject is taken as the training set for analysis verification, and the second group of experimental data is taken as the test set. Conduct five times of superposition averaging to 36 target stimulation trials and 36 standard stimulation trials after preprocessing of each subject, and then conduct sparse decomposition on the target stimulation superposition averaging data. Set the reconstruction error threshold value as 0.05, to acquire the results of the sparseness  $k$  of different leading data and the minimum observation matrix order  $M$  when  $N=192$ , as shown in Table 1.

Table 1. Sparseness  $k$  of Different Leading Data of Subjects and the Minimum Observation Matrix Order  $M$

Subjects	1	2	3	4	5	6	7	8
P3-k	5	4	5	5	3	5	4	5
Pz-k	5	5	5	5	4	5	5	5
P3-M	40	36	40	40	30	40	36	40
Pz-M	40	40	40	40	36	40	40	40

It can be seen from Table 1 that the sparseness  $k$  of all the leading data of all the subjects is slightly smaller than 6, and the minimum observation matrix order  $M$  is slightly less than 48. Therefore, the observation matrix order of all the leading data of all the subjects can be set as 48; i.e., the compression ratio is 1:4, and the observation matrix is the row echelon matrix with the compression ratio of 1:4. Then conduct observation dimension reduction to target stimulated superposition averaging data of all the subjects by utilizing observation matrix, to acquire the normal feature division of the first dimension reduction of 48. Then reconstruct the signal after compression, and calculate the correlation coefficient between the reconstruction signal and the original signal, with the results as shown in Table 2. It can be known from Table 2 that the reconstruction signal after compressed sensing dimension reduction has very significant correlation with the original signal, and it can be regarded that the signal after dimension reduction includes almost all the information of the original signal.

Table 2. Correlation Coefficient between Reconstruction Signal and Original Signal.

Subjects	1	2	3	4	5	6	7	8
P3	0.9955	0.9932	0.9911	0.9979	0.9903	0.9946	0.9932	0.9985
Pz	0.9935	0.9915	0.9947	0.9904	0.9908	0.9932	0.9946	0.9975

Then conduct the secondary sparse decomposition to the ordinary features acquired after the primary dimension reduction, and the reconstruction error threshold value is set as 0.05, to acquire the correlation result between the sparseness  $k$  and the minimum observation matrix order  $M$  ( $N=48$ ), as shown in Table 3.

Table 3. Sparseness  $k$  and the Minimum Observation Matrix Order  $M$  of Secondary Sparse Decomposition

Subjects	1	2	3	4	5	6	7	8
P3- k	4	3	3	3	3	3	3	4
Pz-k	3	3	3	3	3	3	3	3
P3-M	24	20	20	20	20	20	20	24
Pz-M	20	20	20	20	20	20	20	20

As shown in the above table, the sparseness of all the leading data of all the subjects is less than or equal to 4, and the minimum observation matrix order  $M$  is less than or equal to 24. Set the observation matrix order as 24, i.e., the compression ratio as 1:2, and the observation matrix is the row echelon matrix with the compression ratio of 1:2. Therefore, the ordinary feature dimension of the secondary dimension reduction is 24. Then reconstruct the signal after secondary compression, and calculate the

correlation coefficient between the reconstruction signal and the original signal, as shown in Table 4. According to Table 4, the reconstruction signal of the signal after secondary dimension reduction has very significant correlation with the original signal, and it can be regarded that the signal after dimension reduction includes almost all the information of the original signal.

Table 4. Correlation Coefficient of the Reconstruction Signal after Secondary Compression and the Original Signal

Subjects	1	2	3	4	5	6	7	8
$P_3$	0.9931	0.9927	0.9978	0.9955	0.9948	0.9927	0.9978	0.9955
$P_z$	0.9979	0.9903	0.9946	0.9932	0.9985	0.9979	0.9903	0.9946

Ultimately, we can draw the conclusion that the three leading data of all the subjects can be conducted with two dimension reduction processing with the compression ratios of 1:4 and 1:2, respectively; i.e., the 192 points of each leading are reduced to 24 points.

### 3.2 Feature Classification

In order to verify the classification effects of down sampling features and the features extracted with the method in this article, the two features are extracted in 72 samples of the training sets and the test sets of 8 subjects; each primary sample includes two leading data, and each leading includes 192 points of data. Each data is conducted with 8-time down sampling processing, to acquire 48 points of down sampling features of 24 points in each leading, the same as the dimension of ordinary features. The same LDA classifier is selected to conduct training set training and test set test to the two features, and the acquired test set accuracy rate of the 8 subjects are as shown in Table 5.

Table 5. Test Set Accuracy Rate of 8 Subjects

Subjects	1	2	3	4	5	6	7	8
Down sampling accuracy rate	90.36	77.86	88.75	80.32	79.29	87.32	90.08	78.29
Accuracy rate of the features	94.29	77.86	91.79	86.05	83.39	87.50	91.96	81.79

According to Table 5, it can be seen that the classification accuracy rate of the features extracted in this article is significantly higher than that of the down sampling features.

## 4. Discussion

Different from traditional mVEP features extraction methods such as the feature extraction with down sampling method, this article adopts the over-complete dictionary based compressed sensing method, to conduct dimension reduction processing in EEG signals in the time window by utilizing the row echelon observation matrix of different compression ratios for multiple times, to acquire the ordinary features of mVEP signals, and further conduct sparse noise reduction to ordinary features. It also conducts LDA classifications on down sampling features and the extracted features in this article. According to verification, the classification accuracy rate of the features extracted in this article has significant improvement than that of the down sampling features. The feature extraction method in this article improves classification effect by taking the avoidance of overfitting and the retaining of useful information of original signal to the maximum extent into consideration, greatly promoting better stable application effect to the mVEP based BCI system.

## References

- [1] J. R. Wolpaw, N. Birbaumer, W. J. Heetderks, et al. Brain-computer interface technology: a review of the first international meeting[J]. Rehabilitation Engineering, IEEE Transactions on, 2000, 8(2): 164-173.



- [2] A. H. Do, P. T. Wang, C. E. King, S. N. Chun, and Z. Nenadic, "Brain-computer interface controlled robotic gait orthosis," *J Neuroeng Rehabil*, vol. 10, no. 1, pp. 111, 2013.
- [3] L. R. Hochberg, D. Bacher, B. Jarosiewicz, N. Y. Masse, J. D. Simeral, J. Vogel, S. Haddadin, J. Liu, S. S. Cash, P. van der Smagt, and J. P. Donoghue, "Reach and grasp by people with tetraplegia using a neurally controlled robotic arm," *Nature*, vol. 485, no. 7398, pp. 372-5, May 17, 2012.
- [4] Mulder T. Motor imagery and action observation: cognitive tools for rehabilitation. *J Neural Transm*, 2007, 114(10): 1265-1278
- [5] Page S, Levine P, Leonard A. Effects of mental practice on affected limb use and function in chronic stroke. *Arch Phys Med Rehabil*, 2005, 86(3): 399-402
- [6] Liu K, Chan C, Lee T, et al. Mental imagery for promoting relearning for people after stroke: a randomized controlled trial. *Arch Phys Med Rehabil*, 2004, 85(9): 1403-1408
- [7] J. R. Wolpaw, and D. J. McFarland, "Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans," *Proc Natl Acad Sci U S A*, vol. 101, no. 51, pp. 17849-54, Dec 21, 2004.
- [8] G. Schalk, K. Miller, N. Anderson, J. Wilson, M. Smyth, J. Ojemann, D. Moran, J. Wolpaw, and E. Leuthardt, "Two-dimensional movement control using electrocorticographic signals in humans," *J Neural Eng*, vol. 5, no. 1, pp. 75-84, 2008.
- [9] J. Mellinger, G. Schalk, C. Braun, H. Preissl, W. Rosenstiel, N. Birbaumer, and A. Kubler, "An MEG-based brain-computer interface (BCI)," *Neuroimage*, vol. 36, no. 3, pp. 581-93, Jul 1, 2007.
- [10] B. Sorger, B. Dahmen, J. Reithler, O. Gosseries, A. Maudoux, S. Laureys, and R. Goebel, "Another kind of 'BOLD Response': answering multiple-choice questions via online decoded single-trial brain signals," *Prog Brain Res*, vol. 177, pp. 275-92, 2009.
- [11] M. Kuba, Z. Kubova, J. Kremlacek, et al. Motion-onset VEPs: Features, methods, and diagnostic use[J]. *Vision Research*, 2007, 47(2): 189-202.
- [12] M. A. Hollants-Gilhuijs, J. C. De Munck, Z. Kubova, et al. The development of hemispheric asymmetry in human motion VEPs[J]. *Vision Research*, 2000, 40(1): 1-11.
- [13] M. Kuba, Z. Kubová Visual evoked potentials specific for motion onset[J]. *Documenta Ophthalmologica*, 1992, 80(1): 83-89.
- [14] Candes E J. The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathematique*, 2008, 346(9-10): 589-592
- [15] Candes E J, Eldar Y C, Needell D, Randall P. Compressed sensing with coherent and redundant dictionaries. *Applied and Computational Harmonic Analysis*, 2011, 31(1): 59-73
- [16] Arthur P L, Philips C L. Voiced /unvoiced speech discrimination in noise using Gabor atomic decomposition[A]. *Proc of IEEE ICASSP[C]*. HongKong:IEEE, 2003, I(4): 820-828.
- [17] Mallat S, Zhang Z. Matching pursuit with time-frequency dictionaries[J]. *IEEE Trans on Signal Processing*, 1993, 41(12): 3397-3415.