Article:
Chen, K, Liu, Q, Ai, Q et al. (3 more authors) (2016) A MUSIC-based method for SSVEP signal processing. Australasian Physical and Engineering Sciences in Medicine, 39 (1). pp. 71-84. ISSN 0158-9938

https://doi.org/10.1007/s13246-015-0398-6

A MUSIC-based Method for SSVEP Signal Processing

Kun Chen[1,3], Quan Liu[2,3,*], Qingsong Ai[2,3], Zude Zhou[1,3], Sheng Quan Xie[4], Wei Meng[2,3]
[1]School of Mechanical and Electronic Engineering, Wuhan University of Technology, Wuhan 430070, China
[2]School of Information Engineering, Wuhan University of Technology, Wuhan 430070, China
[3]Key Laboratory of Fiber Optic Sensing Technology and Information Processing, Ministry of Education, Wuhan 430070, China
[4]Department of Mechanical Engineering, The University of Auckland, Auckland 1010, New Zealand
[*]Corresponding author: email: quanliu@whut.edu.cn; Tel: +862787655520; Fax: +862787655520

Abstract

The research on brain computer interfaces (BCIs) has become a hotspot in recent years because it offers benefit to disabled people to communicate with the outside world. Steady state visual evoked potential (SSVEP)-based BCIs are more widely used because of higher signal to noise ratio (SNR) and greater information transfer rate (ITR) compared with other BCI techniques. In this paper, a multiple signal classification (MUSIC)-based method was proposed for multi-dimensional SSVEP feature extraction. 2-second data epochs from four electrodes achieved excellent accuracy rates including idle state detection. In some asynchronous mode experiments, the recognition accuracy reached up to 100%. The experimental results showed that the proposed method attained good frequency resolution. In most situations, the recognition accuracy was higher than canonical correlation analysis (CCA), which is a typical method for multi-channel SSVEP signal processing. Also, a virtual keyboard was successfully controlled by different subjects in an unshielded environment, which proved the feasibility of the proposed method for multi-dimensional SSVEP signal processing in practical applications.

Keywords

brain computer interface (BCI), steady state visual evoked potential (SSVEP), multiple signal classification (MUSIC), feature extraction.

1 INTRODUCTION

A brain computer interface (BCI) is a communication system that does not depend on the brain's normal output pathways of peripheral nerves and muscles[1] . The ultimate goal of a BCI is to create a specialized interface that allows an individual with severe motor disabilities to have effective control of devices such as computers, speech synthesizers, assistive appliances and prostheses[2].

Electroencephalography (EEG) is most commonly used for BCIs because it has advantages of portability and ease of use. A steady state visual evoked potential (SSVEP) is a periodic response to a visual stimulus modulated at a frequency higher than 6 Hz[3] (or 4Hz[4]). It can be recorded from the scalp as a nearly sinusoidal oscillatory waveform with the same fundamental frequency as the stimulus, and often includes some higher harmonics. The amplitude and phase characteristics of SSVEPs depend upon the stimulus intensity and frequency. SSVEP-based BCIs are becoming a research hotspot because it has many advantages over other BCI systems including a higher signal to noise ratio (SNR), and faster information transfer rate (ITR). It also does not require intensive training [5].

A variety of methods have been developed and used for feature extraction for SSVEP-based BCIs[6]. Fourier-based transform methods are mostly used for power spectrum density analysis (PSDA). Most research used them to compute the accumulative power at the stimulus frequencies and their harmonics for frequency-coded SSVEPs. Or the average power centered on the stimulus frequency was calculated[7] . An important advantage of Fourier-based transforms is their simplicity and small computation time. However, the time window length of SSVEP signals needs to be long enough to enhance the frequency resolution of FFT when the sampling frequency is confirmed. This might limit practical applications because it has a lower information transfer rate (ITR) [8]. Additionally, a larger window length could lead to classification errors during the changing of stimuli. The studies [8-10] all used wavelet analysis to estimate power at relevant frequency points. A key problem with applying wavelet analysis is how to choose an appropriate mother wavelet to attain good performance. Although the wavelet analysis is better for non-stationary signal processing compared with Fourier transform, it is developed based on Fourier transform, which is fit for processing linear signals. New methods suitable for non-linear and non-stationary signal processing are needed. In this case, Hilbert Huang transform (HHT) is adopted in previous studies [11,12]. HHT has more stability than FFT, which means the recognition accuracy will not change greatly when the data length varies. Although HHT can be well used for non-linear and non-stationary SSVEPs, its computation time is higher compared with Fourier transform. All the methods discussed above were commonly employed to process single channel SSVEPs. Signals from multi-channel EEG are less affected by noise than signals from a unipolar or bipolar system. The combination of signals collected from different channels (electrodes) is also referred to as spatial filtering. Typical methods like minimum energy combination (MEC) and maximum contrast combination (MCC) are the most commonly utilized [13-16]. Another method named canonical correlation analysis (CCA) can be employed to extract features from a different viewpoint. It computes the correlation of two multi-variable datasets [17]. Paper [18] used CCA to recognize SSVEP for the first time. A further comparison between the CCA and PSDA method was done by Hakvoort et al. [19], which showed that CCA had better performance. Spatial filtering methods have the advantage of combining signal pre-processing and feature extraction together.

In this paper, a new method based on multiple signal classification (MUSIC) was proposed for feature extraction of multi-dimensional SSVEPs. One of the typical applications of MUSIC is to solve the problem of harmonic retrieval for one-dimensional signals. The principle of using MUSIC for target frequency recognition of multi-dimensional SSVEPs was explained in detail. Also, a criterion used to determine the number of eigenvectors constructing the signal subspace for power spectrum estimation was proposed. The method was verified with both simulated and real SSVEP data. Experiments in synchronous and asynchronous modes were conducted. The results show that MUSIC achieved a good frequency resolution. Meanwhile, it is capable of suppressing noise because it decomposes the original data into signal subspace and noise subspace. Compared with CCA, MUSIC is more flexible, as the number of eigenvectors constructing the signal subspace is adjustable. The recognition accuracy was better than CCA in most situations. Finally, the proposed method was successfully utilized for users to control a virtual keyboard in an unshielded environment.

This paper is organized as follows: methods including the principle of MUSIC, the procedure of SSVEP signal processing and the setup of experiments are explained in Section 2. The experimental results are illustrated in Section 3. Discussion and conclusions are presented in Section 4 and Section 5.

2 Materials and Methods

2.1 Multiple Signal Classification for SSVEP Signal Processing

MUSIC was first proposed by R. O. Schmidt in 1979[20]. It is often used to solve the problem of harmonic retrieval and direction of arrival. The data model [21] can be expressed as follows.

$$x(n) = \sum_{k=1}^{p} a_k \exp(j2\pi n f_k + j\phi_k) + w(n) \quad (1)$$

Here, $w(n)$ is additive noise. a_k, f_k and ϕ_k represent the amplitude, frequency and phase respectively. p denotes the number of harmonics.

The SSVEP signal is a periodic response to the stimulus frequency and its harmonics, which can be modeled using the above formula. Actually, many other methods like MEC, MCC or CCA employ similar data models to represent SSVEPs. The model includes sum of stimulus frequencies and noise. They are typically used for multi-channel signals. In this situation, the data model should be defined as:

$$XX(n) = [x_1(n), x_2(n), ..., x_{num}(n)] \quad (2)$$

Here, num is the number of channels. It is assumed that the data includes both signals and noises, which are independent of each other[22] . Thus, the covariance matrix can be decomposed into signal and noise components.

As mentioned before, the typical use of MUSIC method is to solve the problem of harmonic retrieval for one-dimensional signals. The principle of using MUSIC for multi-dimensional SSVEP signal processing is illustrated as follows.

A simulated signal according to Formula (1) is generated as:

$$\begin{aligned} x(n) = {} & 0.5*\sqrt{20}*\sin(2\pi*5.384*n/f_S) + \\ & \sqrt{10}*\sin(2\pi*5.871*n/f_S) + 15.5*\mathrm{randn}(1,512) \end{aligned} \quad (3)$$

Where, the value of f_S is 256, and n ranges from 1 to 512. It can be seen that there are two harmonic components with frequencies of 5.384 Hz and 5.871 Hz for the simulated signal. The spatial smoothing method which reconstructs one-dimensional signals to multi-dimensional signals, is usually employed to analyze the signal as Formula (3). The reconstructed signal is an $N \times M$ dimension matrix. As for the MUSIC method, N means the number of snapshots, and M means the number of arrays. The reconstructed multi-dimensional signal can be represented by a matrix X.

$$X = \begin{bmatrix} x(1) & x(2) & ... & x(M) \\ x(2) & x(3) & ... & x(M+1) \\ ... & ... & ... & ... \\ x(N) & x(N+1) & ... & x(M+N-1) \end{bmatrix} \quad (4)$$

The first row of the new matrix includes $x(1), x(2), ..., x(M-1), x(M)$; the second row of the new matrix includes $x(2), x(3), ..., x(M), x(M+1)$, and so on. The last row includes $x(N), x(N+1), ..., x(M+N-2), x(M+N-1)$. Because the total number of sampling points for the original signal $x(n)$ is 512, the maximum sum of M and N is 513.

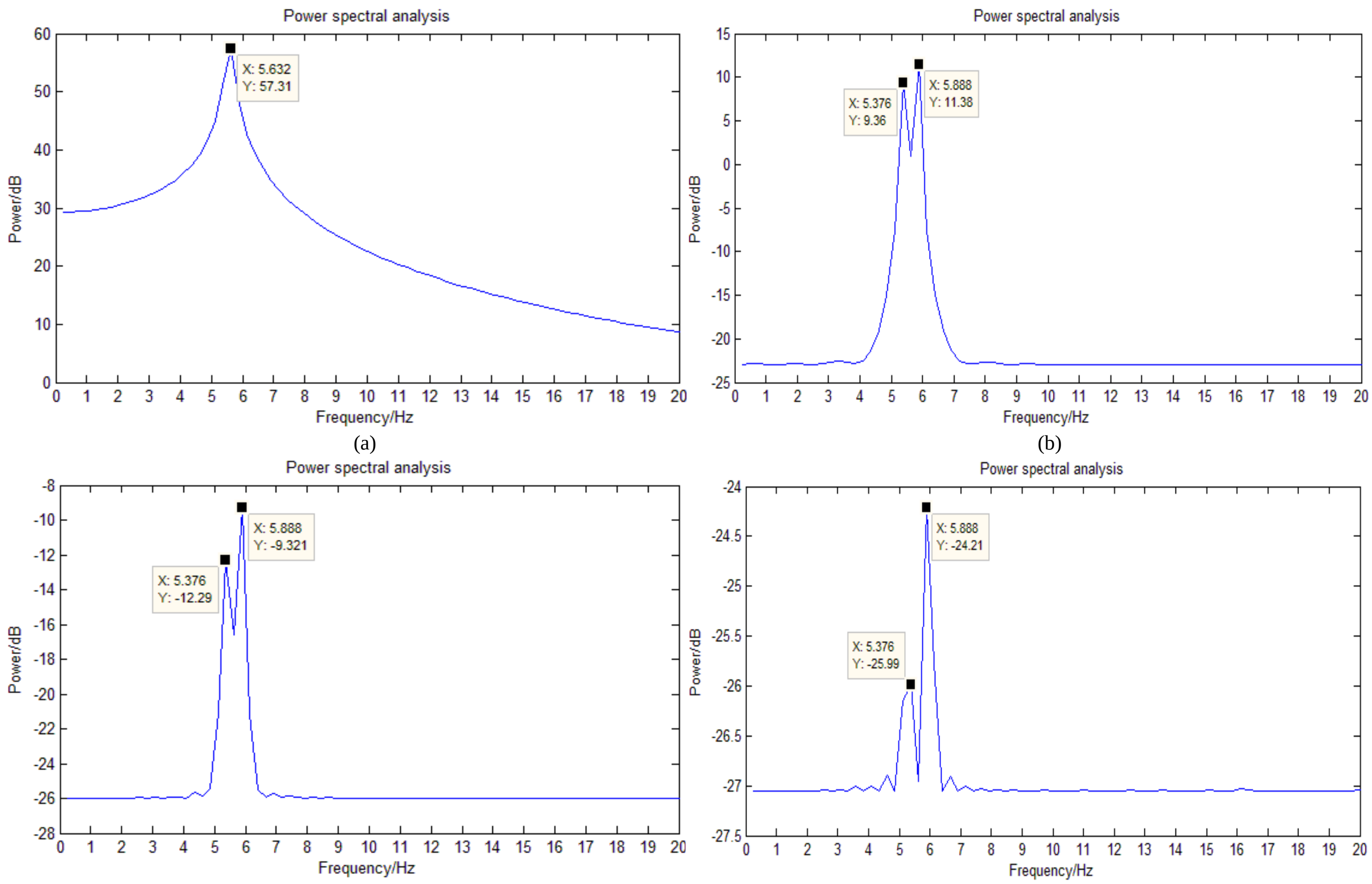
The values of M and N were adjusted to reconstruct the signal. The result of power spectral analysis for the new multi-dimensional signal with various values of M and N is illustrated as Fig. 1.

It is seen in Fig. 1(a) only one harmonic component of 5.632 Hz is recognized. While in Fig. 1(b), Fig. 1(c) and Fig. 1(d) , two harmonic components of 5.376 Hz and 5.888 Hz can be recognized, which are quite close to the harmonic components of the original signal represented by Formula (3). For real SSVEP signals, the recognition accuracy of target frequencies doesn’t normally need to be 0.001 Hz. The harmonic recognition effect became better with the value of M increasing in experiments. However, the result would not improve any more if M increased to some extent.

For the matrix X, each row can be seen as a time series. Compared the next row with the last low, the next one can be regarded as a unit-time delay time series of the last one. The real SSVEP signal as in Formula (2) can be also represented by a matrix XX_{new}.

$$XX_{new} = \begin{bmatrix} x_1(1) & x_1(2) & ... & x_1(sam) \\ x_2(1) & x_2(2) & ... & x_2(sam) \\ ... & ... & ... & ... \\ x_{num}(1) & x_{num}(2) & ... & x_{num}(sam) \end{bmatrix} \quad (5)$$

sam means the sampling points, and num denotes the number of electrodes (channels). Each row of XX_{new} is also a time series. For electrodes are laid at different positions, the time series according to each electrode might be regarded as signals with different transmitting distances from the same sources. It means for XX_{new}, the next row can be regarded as a time-delay series of the last one. From this perspective, XX_{new} as in Formula (5) and X in Formula (4) are quite similar. XX_{new} can be seen as the reconstructed multi-dimensional signal, which is to be analyzed by the MUSIC method for harmonic components recognition. The feasibility of using MUSIC for multi-dimensional SSVEP signal processing were verified by both simulated data and real SSVEP data, and it can be seen in the section Results.



(a) (b)

(c) (d)

Fig. 1 Power spectral analysis for reconstructed signal with various values of M and N. (a)M=5, N=507; (b)M=200, N=312; (c)M=500, N=12; (d) M=508, N=4.

The detailed computation procedure is described as follows:

(1)Assume that the original SSVEP signal is a $sam \times num$ dimension matrix. sam and num represent the number of sampling points and the number of channels.

(2)As explained above, sam can be seen as M in Formula (4), and num can be seen as N .The covariance matrix R_{XX} of the new matrix XX_{new} is a $sam \times sam$ dimensional matrix. Compute the eigenvalues d_1, d_2, ..., d_{sam} and eigenvectors V_1, V_2,..., V_{sam} of the covariance matrix.

(3)The eigenvectors corresponding to the k largest eigenvalues constitute the signal subspace, while the remaining eigenvectors constitute the noise subspace. How to confirm the value of k is also discussed in this paper.

(4)Use the signal subspace or the noise subspace for power spectrum estimation as follows.

$$P(f) = 1/(a_F (I - S_i S_i^T) a_F^T) \tag{6}$$

$$P(f) = 1/(a_F^T (N_o N_o^T) a_F) \tag{7}$$

I is a $sam \times sam$ dimensional unit matrix. S_i and N_o represent the signal and the noise subspace with dimensions of $sam \times k$ and $sam \times (sam - k)$. a_F is a $1 \times sam$ dimensional matrix. The value of each element is:

$$a_F(1, i) = e^{-j*(i-1)*\varpi_{order}}, i \in (1,2,\ldots,sam) \tag{8}$$

$$order = round(f / \Delta f) + 1 \tag{9}$$

$$\varpi_{order} = 2 * \pi * \Delta f * order / f_s \tag{10}$$

Here, f_s is the sampling frequency, Δf is the frequency interval, and f is the target frequency..

The value of k can be defined as a constant which is obtained by conducting experiments. The idea is like setting a threshold. It needs time to adjust the value to maximize the recognition results. Inspired by the principle of minimum energy combination [14], which is another important technique for spatial filtering, the number of eigenvalues is chosen corresponding to the following equation.

$$d_i / d_1 > 0.01, i = 1,2,3,..., k \quad (11)$$

Here, d_1 is the largest eigenvalue of the covariance matrix and d_1, d_2 till d_k are sorted in descending order. Only these eigenvectors whose corresponding eigenvalues are no less than 1% of the largest eigenvalue are classified as the signal subspace for spectrum estimation.

The SSVEP signal processing usually includes data pre-processing, feature extraction and feature classification. In our research, a low-pass filter with a cut-off frequency of 30 Hz and a 50 Hz notch filter were used for pre-processing. The original data were segmented into epochs for feature extraction via the MUSIC method. The basic idea is to compute the power values at stimulus frequencies. As for feature classification, the frequency corresponding to the maximum power value was recognized as the target one. If there are n stimulus frequencies which are f_1, f_2 till f_n. The power value $P(f)$ corresponding to each frequency is computed as shown in Formula (6). And then the target frequency is confirmed as the following formulas.

$$f_{target} = \arg\max(P(f)) \quad (12)$$

$$target \in \{1,2,..., n\} \quad (13)$$

From the above two formulas, it can be seen that we should focus on the magnitude relation of power values corresponding to different stimulus frequencies, not the absolute magnitude of a specific power value. Therefore, the units of power values are not labeled in all figures in this paper.

A dwell time, which means the state of one target kept for a period of time, was employed to reduce the false positives. The detailed explanations about this are given in the Results and Discussion sections.

The recognition results were compared with the CCA method in experiments. Here, the principle of CCA[23] is briefly introduced. If there are two variables X and Y with dimensions of p and q respectively. X and Y are represented as $X = (X_1, X_2,..., X_p)$ and $Y = (Y_1, Y_2,..., Y_q)$. To find out the relation between X and Y, a linear combination is applied to both X and Y. Two new variables are generated which are $U = a_1X_1 + a_2X_2 + ... + a_pX_p$ and $V = b_1Y_1 + b_2Y_2 + ... + b_qY_q$. Then the correlation coefficient $\rho = corr(U,V)$ between U and V is computed. $a = (a_1, a_2,..., a_p)$ and $b = (b_1, b_2,..., b_q)$ are named canonical variables when p has the largest value, and ρ is the canonical correlation coefficient between X and Y.

If there are four stimulus frequencies used in practical applications which are f_1, f_2, f_3 and f_4, four reference signals should be constructed as follows:

$$X_{21} = \begin{bmatrix} \sin(2\pi * f_1 * t) \\ \cos(2\pi * f_1 * t) \end{bmatrix} \quad (14)$$

$$X_{22} = \begin{bmatrix} \sin(2\pi * f_2 * t) \\ \cos(2\pi * f_2 * t) \end{bmatrix} \quad (15)$$

$$X_{23} = \begin{bmatrix} \sin(2\pi * f_3 * t) \\ \cos(2\pi * f_3 * t) \end{bmatrix} \quad (16)$$

$$X_{24} = \begin{bmatrix} \sin(2\pi * f_4 * t) \\ \cos(2\pi * f_4 * t) \end{bmatrix} \quad (17)$$

If the original SSVEP data are denoted by X_1, the aforementioned CCA method is used to compute the correlation coefficients ρ_1, ρ_2, ρ_3, ρ_4 between X_1 and X_{21}, X_{22}, X_{23}, X_{24}. Then the target frequency is dertermined according to the following formula.

$$f_{target} = \arg\max(\rho(f)), target \in \{1,2,3,4\} \quad (18)$$

It is noted that the reference signals only use the fundamental harmonic components corresponding to stimulus frequencies. The second or other harmonics can be also used to calculate the correlation coefficients if necessary.

2.2 Experimental Setup

As the signal and noise components of simulated data are deterministic, the simulated data are suitable for method verification. But the real SSVEP data are more persuasive to prove that the proposed method can be used in practical applications. So both simulated and real SSVEP data were used in our experiments.

1) Simulated data

In experiments, four-channel data were generated. Each channel includes two sine waves to simulate the harmonic components of SSVEP and random noise. The frequencies of two sine components are 5.384 and 5.871 Hz, respectively, just like the harmonic components used in Formula (19). The data are expressed as follows.

$$S = [S_1; S_2; S_3; S_4] = \begin{bmatrix} 0.5 * \sqrt{20} * \sin(2\pi * 5.384 * n / f_s) + \\ \sqrt{10} * \sin(2\pi * 5.871 * n / f_s) + 15.5 * randn(1, N); \\ 0.1 * \sqrt{20} * \sin(2\pi * 5.384 * n / f_s) - \\ 0.8 * \sqrt{10} * \sin(2\pi * 5.871 * n / f_s) + 15.5 * randn(1, N); \\ 0.2 * \sqrt{20} * \sin(2\pi * 5.384 * n / f_s) + \\ \sqrt{10} * \sin(2\pi * 5.871 * n / f_s) + 15.8 * randn(1, N); \\ 0.1 * \sqrt{20} * \sin(2\pi * 5.384 * n / f_s) + \\ \sqrt{10} * \sin(2\pi * 5.871 * n / f_s) + 16.5 * randn(1, N) \end{bmatrix} \quad (19)$$

Here, f_s is the sampling frequency equal to 256 Hz. N is the total number of sampling points with a value of 512. n represents each single sampling point ranging from 1 to 512. Thus, S is a 4×512 dimensional matrix. It is noticed that the frequency interval of two components is 0.487 Hz. It doesn't strictly meet the frequency resolution requirement for traditional FFT method to estimate the spectrum, because the resolution is computed as 0.5 Hz ($f_s / N = 256 / 512$). Also, both of the frequency points 5.384 and 5.871 are not located on the FFT spectral lines. However, MUSIC can better estimate the power values at these two points. It is shown in the Results section.

2) Real SSVEP data

The stimulus panel contains four LED blocks each measured as 2*2 cm. The distance between LED blocks and the flickering frequencies are adjustable. The effects of these parameters on experimental results are discussed in the following section.

A Symtop UE-16B EEG amplifier was used in our experiments as shown in Fig. 2. It has 16 channels and a USB interface. The maximum sampling frequency can be adjusted to 1000 Hz. A low-pass filter and a notch filter are developed in the amplifier.

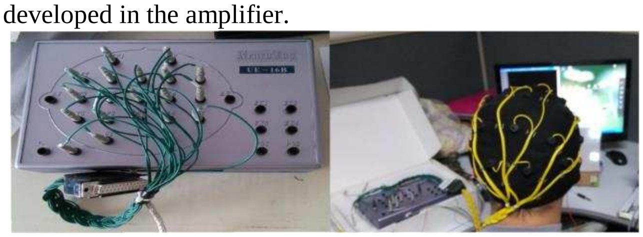


Fig. 2 UE-16B EEG amplifier and the experimental setup

In reality, the sampling frequency was set to 200 Hz, because the frequency of EEG signals is usually lower than 100 Hz. Also, the stimuli frequencies used in experiments are lower than 30 Hz. The choice of 200 Hz is reasonable. Besides, an appropriate sampling frequency rather than a higher frequency saves the computation time. The cut off frequency of the low-pass filter was set to 30 Hz and the 50 Hz power line interference was removed by a notch filter. All 16 channels (Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T3, T4, T5, and T6) were utilized to acquire the original EEG data in early stages. The number of channels was reduced to 4 (P3, P4, O1, O2) in later experiments. Five subjects (one female, four male) participated in the experiments for performance analysis of related methods. All of them have good eye vision. The average age is 25.4 years old. More subjects were asked to conduct the experiment for real time control of a virtual keyboard. All experiments were done in an unshielded environment.

Two different kinds of experiments were designed for the five subjects, namely synchronous paradigm and asynchronous paradigm. Regarding the synchronous one, three sessions were conducted: (i) Only one of the four stimuli was flickering. Subjects looked at the flickering one for 40 seconds from the 20th second; (ii) Only one of the four stimuli was flickering. Subjects looked at each area corresponding to stimuli not flickering for 20 seconds consecutively from the 20th second; (iii) Four stimuli were flickering at the same time. Subjects looked at each one for 20 seconds consecutively from the 20th second. Labels were used to indicate different experiment sessions as illustrated in Fig. 3. As for the asynchronous paradigm, two following sessions were carried out: (i) Subjects looked at different stimuli according to a predefined number sequence. Once a target was recognized, subjects turned to look at the next stimulus. In our experiments, three number sequences were defined as '1-2-3-4', '1-3-2-4' and '1-4-2-3'; (ii) Subjects looked at stimuli randomly, and no sequence was defined before the experiment.

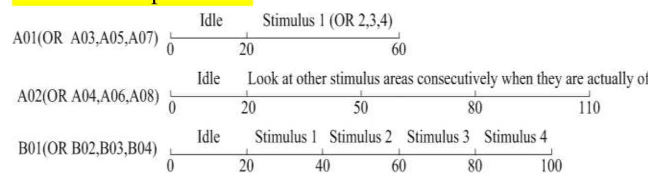


Fig. 3 Explanation of SSVEP dataset labels in the synchronous mode

3) Virtual keyboard layout

In order to verify the aforementioned method, a virtual keyboard was designed for subjects to input words. In this application, four stimuli were used to move the cursor along four different directions: up, down, left and right. The fifth stimulus was used to represent the "confirm" command for selecting a letter or a symbol of the keyboard. It can be seen from Fig. 4 that the keyboard contains 26 letters and other symbols. The layout of the keyboard is based on the frequency of how often these symbols are used [14]. The more frequent they are used, the closer they are placed near the center of the keyboard. When the cursor moved on a letter or a symbol, the background color turns red. When the letter or the symbol is selected, the text box at the bottom of the screen displays the selected symbol with previous symbols.

3 Results

3.1 Simulated Data Analysis

Fig. 5(a) illustrates the power spectrum obtained by MUSIC method. The spectrum area around 6 Hz was zoomed in as shown in Fig. 5(b).

It is obvious in Fig. 5(b) that two peaks appear at around 5.3 and 5.8 Hz, which are almost equal to the harmonic components of the simulated data. Also, the spectral curve is smooth in Fig. 5(b), which means a high frequency resolution can be acquired.

3.2 Real SSVEP Data Analysis

For real SSVEP signals, MUSIC was utilized to calculate the power at those stimulus frequency points. The stimulus frequency corresponding to the maximum power value was recognized as the target one. Taking the processing results of dataset A03 as an example, as shown in Fig. 6, four curves with different colors represent the power values of the four stimulus frequencies (6, 7, 8, 9 Hz).The values corresponding to the second stimulus are higher than the other three after the 22nd seconds. The 2-second delay is mainly caused by the response time when subjects changed from the idle state to the work state. In addition, a dwell time,

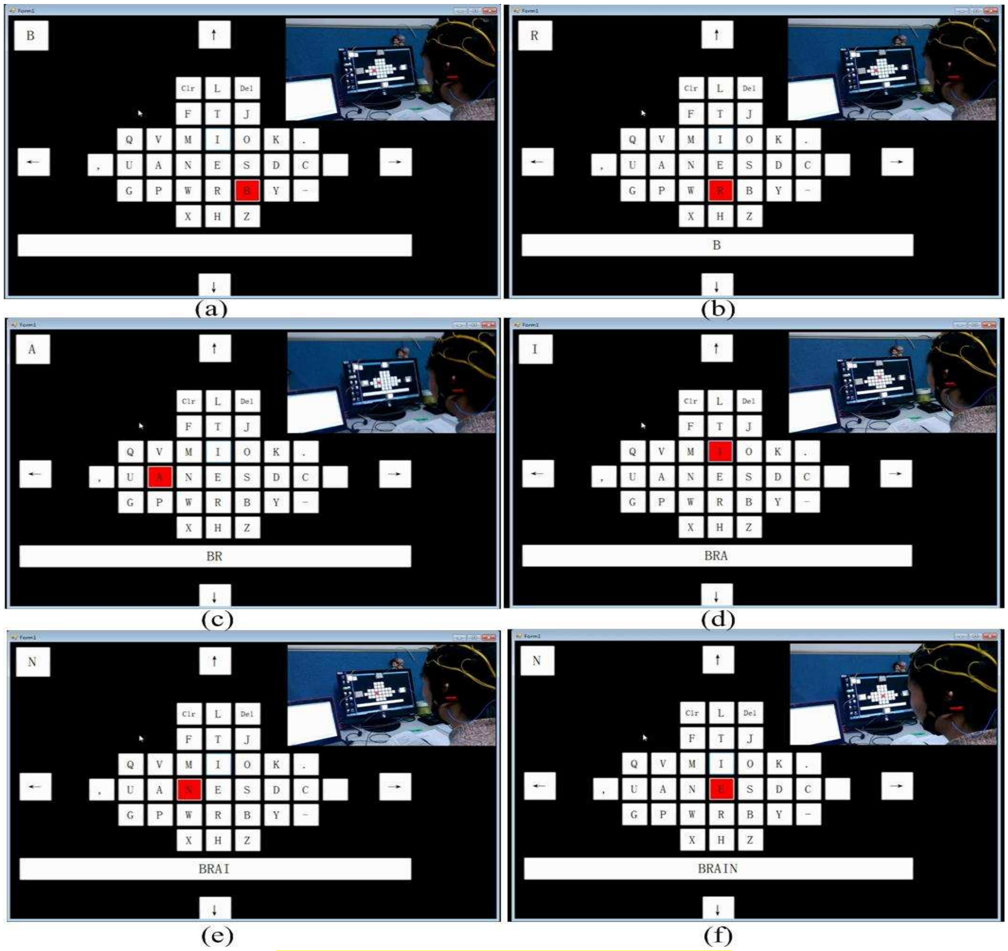


Fig. 4 Control of a virtual keyboard based on SSVEP

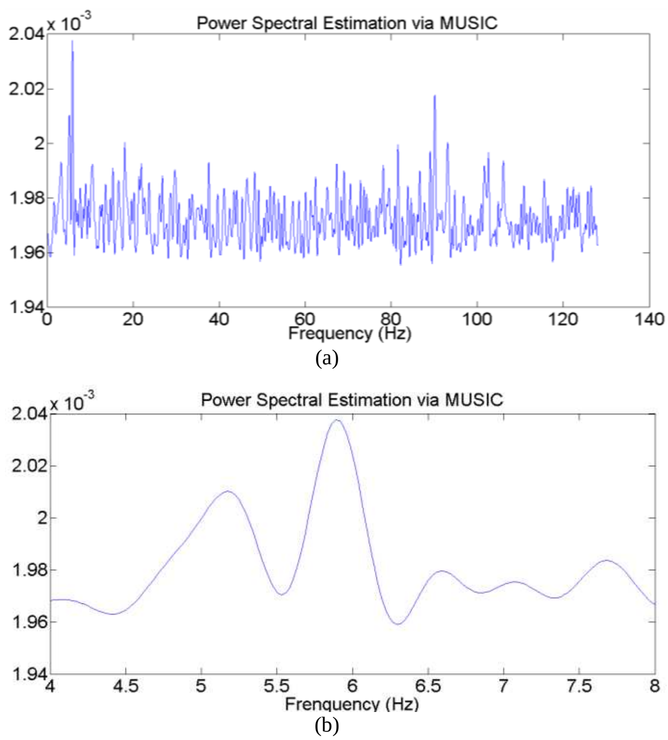


Fig. 5 Spectrum of the simulated data computed by MUSIC

which means SSVEP signals responding to one stimulus have to keep for a period of time, was employed in real applications in order to reduce the false positives. The dwell time was adjusted in experiments. The CCA method was used to compare the processing results. The original SSVEP data were segmented into epochs of 400 sampling points, which means the time duration was 2 seconds (sampling frequency is 200 Hz). Data overlapping was used to make the time window move smoothly. Results of no overlapping and 50% overlapping were compared.

Table 1 shows one subject's recognition rates under different conditions during the first session in the synchronous mode. The frequencies corresponding to A01, A03, A05 and A07 were 6, 7, 8 and 9 Hz. Original data were segmented into epochs of 400 points.

The recognition rates of one subject during the second session are shown in Table 2. In this session, only one stimulus was flickering. The subjects looked at other areas of stimuli not flickering in order, which means the subjects were in idle states all the time. So, the aforementioned dwell time was utilized. Several consecutive decisions made one final target choice. The parameter of consecutive numbers was tested in experiments. Overlapping was also employed to recognize the target smoothly.

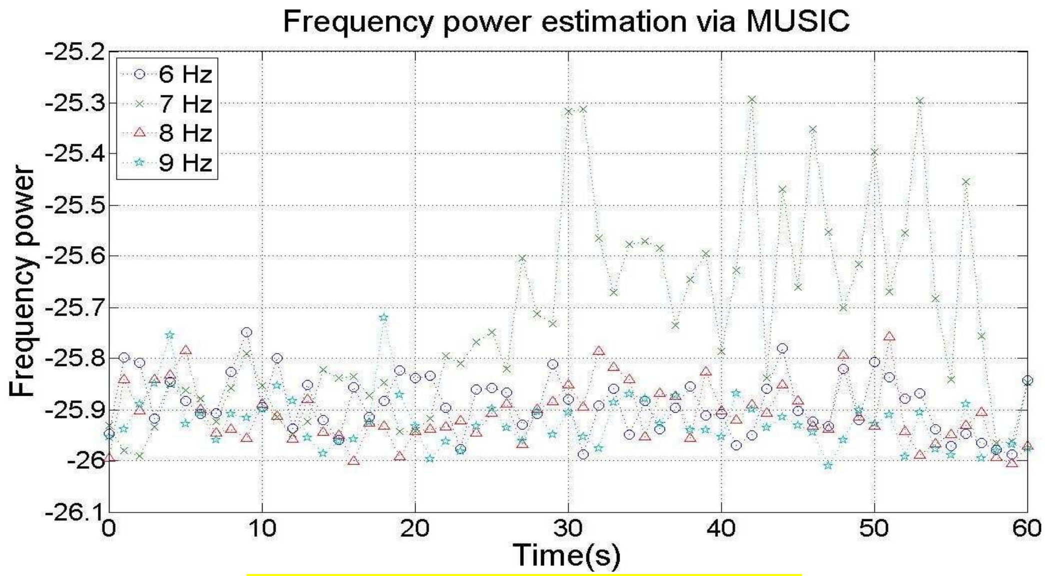


Fig. 6 Frequency power estimation corresponding to dataset A03

In the third session, the use of covers on stimuli, the distance of neighboring stimuli were changed to compare the different recognition results. The four stimulus frequencies used in Table 3 were 6.5, 7.5, 8.5 and 9.5 Hz, respectively.

Table 1 Recognition rates under different conditions during the first session in the synchronous mode[24]

Conditions *		1	2	3
A01	MUSIC	0.8571	0.9024	0.7073
	CCA	0.7619	0.8293	0.6829
A03	MUSIC	0.8571	0.878	0.9268
	CCA	0.7619	0.8293	0.9268
A05	MUSIC	0.8571	0.878	0.9268
	CCA	0.7143	0.878	0.8537
A07	MUSIC	0.8571	0.8537	0.9268
	CCA	0.8095	0.8293	0.9024

* Condition 1: data length=400, overlapping=0; number of channels=16;
Condition 2: data length=400, overlapping =50%; number of channels=16;
Condition 3: data length=400, overlapping =50%; number of channels=4.

In the first session of the asynchronous mode, subjects looked at stimuli following predefined number sequences '1-2-3-4', '1-3-2-4' and '1-4-2-3'. Once a target was recognized, subjects turned to look at another one immediately. It is different from the experiments in the synchronous mode, in which subjects looked at each stimulus for a fixed period of time.

Table 2 Recognition rates under different conditions during the second session in the synchronous mode[24]

Conditions*		1	2	3	4
A02	MUSIC	0.6607	0.5495	0.7928	0.8908
	CCA	0.5	0.5135	0.6937	0.8829
A04	MUSIC	0.6429	0.5315	0.7838	0.8288
	CCA	0.6429	0.6126	0.8108	0.8198
A06	MUSIC	0.7679	0.5856	0.8378	0.8829
	CCA	0.6429	0.6667	0.8559	0.9099
A08	MUSIC	0.75	0.6937	0.9279	0.8649
	CCA	0.75	0.5676	0.8288	0.8198

* Condition 1: data length=400, overlapping =0; number of channels=16; 2 consecutive decisions (overlapping: 1) made one final target choice;
Condition 2: data length=400, overlapping =50%; number of channels=16; 2 consecutive decisions (overlapping: 1) made one final target choice;
Condition 3: data length=400, overlapping =50%; number of channels=16; 3 consecutive decisions (overlapping: 2) made one final target choice;
Condition 4: data length=400, overlapping =50%; number of channels=4; 3 consecutive decisions (overlapping: 2) made one final target choice.

Table 3 Recognition rates during the third session in the synchronous mode[24]

Conditions	MUSIC	CCA
Far distance (8 cm), no use of covers	0.8025	0.8148
Far distance (8 cm), use of covers	0.8148	0.8025
Near distance (4 cm), no use of covers	0.7901	0.7531
Near distance (4 cm), use of covers	0.8148	0.8148

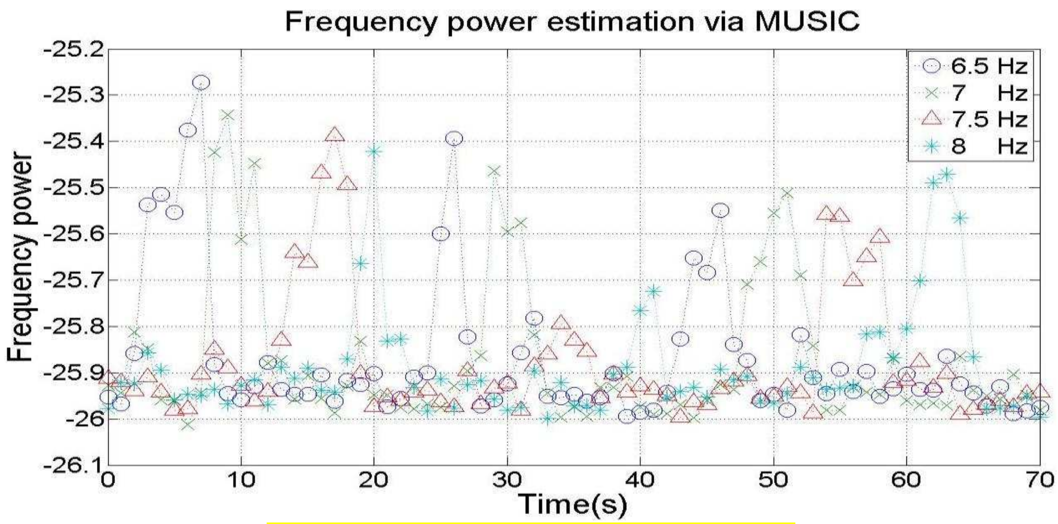


Fig. 7 Frequency power estimation in the asynchronous mode

Taking one number sequence '1-2-3-4'as an example, the frequency power at four stimulus frequencies was illustrated in Fig. 7. Subjects repeated looking at stimuli orderly three times. The frequencies used were 6.5, 7, 7.5 and 8 Hz, respectively. It means the frequency resolution was reduced to 0.5 Hz, while the above experimental results were obtained with frequency resolution of 1 Hz. As seen in Fig. 7, frequencies with an interval of 0.5 Hz can be clearly distinguished using the MUSIC method just like the simulated data.

Five subjects participated in experiments. Each of them finished looking at all stimuli corresponding to three number sequences for three times. The experimental time consumption for all subjects to complete different tasks is listed in the following Table 4.

Table 4 Time consumption (seconds) during the first session in the asynchronous experimental mode

Subjects	Number sequences (repeating three times for each)		
	'1-2-3-4'	'1-3-2-4'	'1-4-2-3'
Subject 1	57	57	57
Subject 2	58	59	63
Subject 3	57	51	55
Subject 4	57	58	61
Subject 5	60	62	59

Table 5 Recognition accuracy when subjects looked at stimuli randomly

Subjects	Accuracy	Total time (seconds)
Subject 1	1.0000	100
Subject 2	1.0000	103
Subject 3	0.9523	100
Subject 4	1.0000	101
Subject 5	0.9048	102

In the second session of the asynchronous mode, subjects looked at whichever stimulus as they would like to for about 100 seconds. This situation is more like the real world applications without external interference. The first session was to estimate the time consumption time of target recognition, while this one was to estimate the recognition accuracy in the asynchronous mode as seen in Table 5.

3.3 SSVEP-based Virtual Keyboard

Fig. 4 illustrates the whole procedure of spelling the word "BRAIN". For five different subjects, the time consumption of spelling "BRAIN" or "BCI TEST" was recorded as shown in Table 6.

Table 6 shows the spelling speed for different subjects were not the same even for the same word. As for "BRAIN", it takes 11.6 to 17 seconds to output a letter or a symbol. In terms of "BCI TEST", the average time is 15 to 15.88 seconds. Taking the word "BRAIN" as an example, the steps of choosing these five letters starting from the keyboard center are 2, 1, 2, 1 and 1, respectively. The total step is 7. Considering the five "confirm" commands, 12 steps are needed to spell the word "BRAIN". It means each step takes about 4.83 to 7.08 seconds. This time is longer than that of the asynchronous mode experiments without device control. It is partly because subjects need time to choose which the next target is when spelling a specific word.

Table 6 Time consumption for different subjects of spelling different words

Subjects	Words	Time (Seconds)
Subject 1	BRAIN	58
Subject 2	BRAIN	68
Subject 3	BCI TEST	120
Subject 4	BCI TEST	127
Subject 5	BRAIN	85

4 Discussion

4.1 Enhancement of Frequency Resolution

For stimulated data, the MUSIC method was used to estimate the power at two nearby frequencies. As for FFT, power values can be computed only at points where the frequency is an integer multiple of the frequency resolution, such as, 0, 0.5, 1, 1.5 Hz and so on. In Fig. 5(b), it is obvious that two maxima appear at around 5.3 and 5.8 Hz, which are the real harmonic components of the stimulated data. That means the MUSIC method can achieve high frequency resolution. In addition, MUSIC decomposes data into signal subspace and noise subspace. The noise can be removed to some extent when estimating the power spectrum.

4.2 One-target Recognition

For the real SSVEP signals, we compared the experimental results of MUSIC and another typically used method CCA. As shown in Tables 1-3, the MUSIC method achieved higher recognition accuracy in most cases. Exceptions existed, as the numbers in red color in these tables illustrated that CCA performed better than MUSIC sometimes.

In Table 1, one subject looked at only one stimulus for 40 seconds. All segmented data length was 2 seconds. The accuracy was better when 50% data overlapping was used for 16-channel SSVEP signals. Also, the results showed that 4-channel (P3, P4, O1 and O2) signals could be utilized for detecting the target stimulus except dataset A1. This is because not all frequencies evoke strong SSVEPs when applied to different individuals. It is proved that more channels do not definitely produce better results, because SSVEP signals are not evenly distributed on the brain surface. Some channels might contain more noise rather than SSVEP related signals. In addition, reducing the number of channels shortens the preparation time before experiments, which is important for implementing a real BCI system.

4.3 Idle State Detection

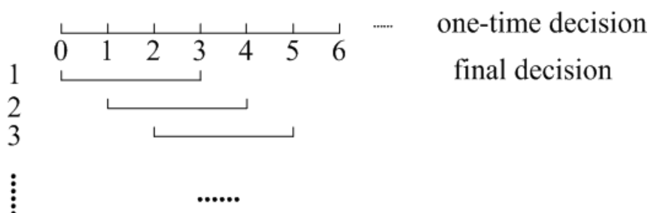


Fig. 8 A dwell time used for final decisions

The frequency with the maximum power value was recognized as the target one in Table 1. However, in the idle state, when subjects didn't focus on any stimulus, this method could cause false positives. Idle state detection becomes another key issue. There are two common ways to solve this problem. A dwell time or a threshold can be employed for idle state detection. The accuracy in Table 2 was calculated by using a dwell time. A same target was recognized for several consecutive times, and then a final decision was made as seen in Fig. 8. Fig. 8 shows that three consecutive decisions make a final decision, which means if three consecutive one-time decisions are identical, and then a final decision is made to confirm a target. The overlapping percentage is 2/3 in this example. That means there is a 'decision window' moving along the time axis. It can be seen in Table 2 that under condition 4, where the data length was 400 points with 50% overlapping and the number of channels was 4, better performance was achieved than other conditions.

For comparison, the threshold method was also employed. The threshold was obtained from training data. For each stimulus frequency, a period time of data in the work state and the idle state was acquired. The threshold was adjusted to maximize the total recognition rate, which was defined as value of the sum of true positives and true negatives divided by the total training number.

As seen in Table 7, the idle state could be better detected by using a threshold. If the threshold and the dwell time were used together, the recognition accuracy can be even up to 100%, such as dataset A06. It is noticed that using a threshold can achieve better results, but it is more time consuming compared with using a dwell time. Preparation time is needed to acquire training data for all stimulus frequencies. Recognition accuracy was higher than 83% when using a dwell time in experiments. And there is no extra time needed in advance. That is an advantage of using a dwell time. However, a longer dwell time leads to delay of recognition. In real applications, the choice of the dwell time should be considered based on requirements of the detection speed and the detection accuracy.

Table 7 Recognition rates with a threshold during the second session in the synchronous mode

Conditions*	A02	A04	A06	A08
1	0.8468	0.8468	0.8378	0.8739
2	0.8559	0.7658	0.8739	0.8739
3	0.9640	0.9820	1.0000	0.9820

*Condition 1: data length=400, overlapping =50%; number of channels=4; 3 consecutive decisions (overlapping: 2) made one final target choice;

Condition 2: data length=400, overlapping =50%; number of channels=4; a threshold made one final target choice;

Condition 3: data length=400, overlapping =50%; number of channels=4; a threshold and 2 consecutive decisions (overlapping: 1) made one final target choice.

4.4 Multi-target Recognition

Table 8 Recognition rates under different conditions during all sessions in the synchronous mode[24]

A01	MUSIC	0.7869	A06	MUSIC	0.8378
	CCA	0.7705		CCA	0.7748
A02	MUSIC	0.8468	A07	MUSIC	0.9180
	CCA	0.8649		CCA	0.8852
A03	MUSIC	0.7705	A08	MUSIC	0.8739
	CCA	0.7377		CCA	0.8198
A04	MUSIC	0.8468	B01	MUSIC	0.716
	CCA	0.8649		CCA	0.7284
A05	MUSIC	0.7869	B02	MUSIC	0.6914
	CCA	0.6885		CCA	0.6914

Table 3 reflects the recognition accuracy when subjects looked at each stimulus consecutively when all stimuli were flickering. The influence of stimulation design on recognition results was tested in experiments. If two neighboring stimuli were too close, it was not easy for subjects to focus on one stimulus. This explained the poor results. Through experiments, a distance of 4 cm seemed reasonable. The distance made the stimulation panel not too big, and the mutual influence of nearby stimuli could be reduced. Also, a layer of thin paper was covered on the stimuli, which made the light more centralized. Recognition rates shown in Table 3 proved that it did improve the recognition results.

The idle state data and the work state data were analyzed alone with or without the use of a dwell time as discussed above. In real applications, all data should be processed under the same condition. Table 8 shows the recognition results of one subject during all three sessions. Three consecutive decisions with 2/3 overlapping made a final target choice. It is shown in Table 8 that the recognition rates in the work state were reduced compared with the situation when no dwell time was used. Subjects might have difficulty in keeping focusing on one stimulus for 40 seconds. This affected the overall recognition accuracy. For the third session, the results became even worse. This is mainly caused by the transition from looking at one stimulus to looking at another, as seen in Fig. 9 showing the stimulus frequency power values of data B02 in Table 8. It took time for subjects to get used to another stimulus. The response time made the results worse. In addition, 80 seconds was quite a long time for subjects to keep focused during this session. The results could be affected if subjects' attention was distracted.

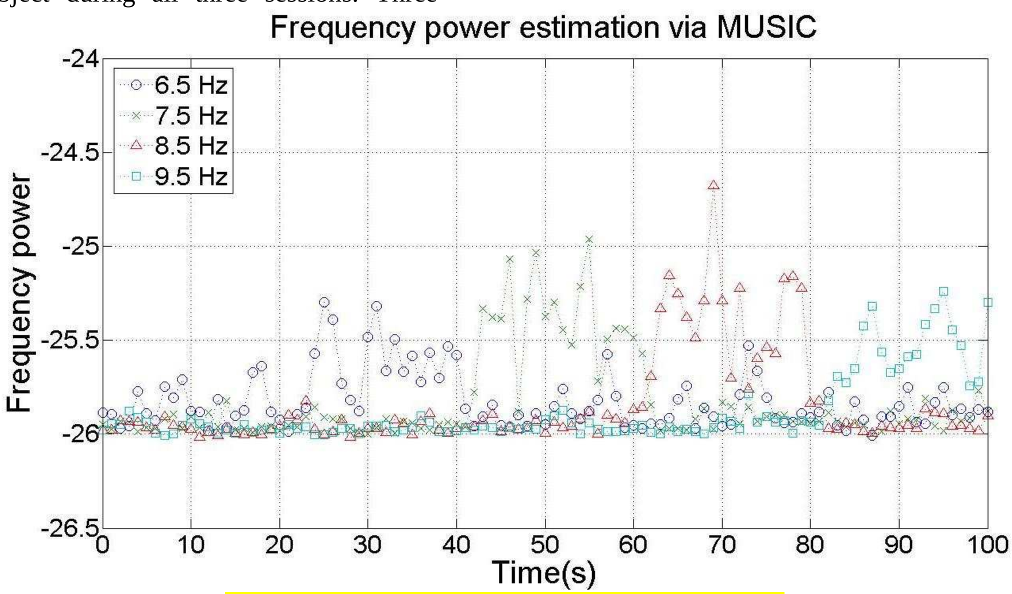


Fig. 9 Frequency power estimation corresponding to dataset B02

4.5 Asynchronous Mode Target Recognition

In the asynchronous mode, from both Fig. 7 and Table 4, it can be estimated that it took about 5 seconds to finish one-target recognition. But the consumption time is different for different subjects or sequences. Actually the original data were processed every one second, but subjects couldn't move eyes from one stimulus to another instantly. The response time existed inevitably when subjects were informed to look at the next stimulus. At most times, one response potential was evoked following a former potential quickly. However, for each potential, it lasted for a period of time. Also, a dwell time caused the delay of one final decision, which made the whole procedure longer.

In the second session, the accuracy when subjects looked at stimuli as they wanted for about 100 seconds was tested. The performance was quite good as seen in Table 5. Three of them achieved accuracy of 100%. For this experiment, not only the target stimulus should be recognized, but also the idle state had to be detected. For all experiments in the asynchronous mode, a dwell time was used to reduce false positives. In addition, the idle state detection was conducted in this asynchronous mode. All stimuli were flickering and the subject looked other places rather than the stimulation panel for 100 seconds. However, in the previous synchronous mode, only one of the stimuli was on and the subject looked at other stimuli areas even they were not flickering. This test in the asynchronous mode means subjects are completely in an idle state without any external interference. The recognition accuracy for this experiment is shown in the following Table 9.

Table 9 Idle state detection in the asynchronous mode

Subjects	Accuracy	Total time (seconds)
Subject 1	0.9303	100
Subject 2	0.9055	100
Subject 3	0.9950	100
Subject 4	0.8905	100
Subject 5	0.9876	100

It can be seen from Table 9 that the accuracy is better in this experiment than that in the synchronous mode for idle state detection. The subjects kept in an idle state for a period of 100 seconds, and the detection accuracy for Subject 3 was even up to 99.50%, which is a satisfactory result. For in this experiment, although all stimuli were flickering, the subject looked at other areas rather than the stimuli. While for the previous one, the flickering stimulus had influence when subjects looked at other stimuli areas. Also, no predefined time duration was set in the asynchronous mode. All these contributed to good results.

4.6 Number of Eigenvectors Constituting the Signal Subspace

Table 10 Recognition rates when p is a constant or adjustable

Data	p: constant	p: adjustable
A01	0.5372	0.5372
A02	0.9095	0.9457
A03	0.6529	0.7438
A04	0.9231	0.9819
A05	0.6033	0.6529
A06	0.9548	0.9593
A07	0.8430	0.8512
A08	0.9186	0.9276

It was mentioned that a segmentation ratio was proposed to confirm the number of eigenvectors to be used for constituting the signal subspace of the MUSIC method in our research. This value can be set to a constant, too. Let p represent this value. We did experiments to compare the recognition results as shown in Table 10 by using the constant method and the proposed method.

Table 10 reflects that the proposed method produced better results. It is more flexible than the constant method. However, the segmentation ratio is set to be 0.01 in our research. This ratio was obtained after conducting some previous data analysis, too.

4.7 Control of a Virtual Keyboard

Fig. 10 illustrates the power values corresponding to five stimuli. The five frequencies are 6.5, 7, 7.5, 8 and 8.5 Hz, which are mapped to five numbers 1, 2, 3, 4 and 5. The numbers are used to represent five commands: up, confirm, left, down and right.

As seen in Fig. 10, the frequency corresponding to the maximum power value is different during the spelling process. The number sequence is “4-5-2-4-2-3-3-2-1-2-3-2” when mapping the frequencies to numbers. This sequence is translated to commands “up-right-confirm-down-confirm-left-left-confirm-up-confirm-left-confirm”. According to these commands, five letters B, R, A, I and N are selected. It takes about 60 seconds to finish the whole procedure, which means each step consumes about 5 seconds. It is noticed that a control strategy was used during the spelling procedure. Only if one recognition number (corresponding to one frequency target) lasts for a certain period of time, is a control command confirmed. This explains the recognition time is longer than that of the asynchronous mode experiments without device control, too. If the commands change too quickly, the cursor moves quickly. It is hard for subjects to focus on one letter or symbol. Also, the subjects need time to decide the move path of the cursor. This is why the control strategy is necessary.

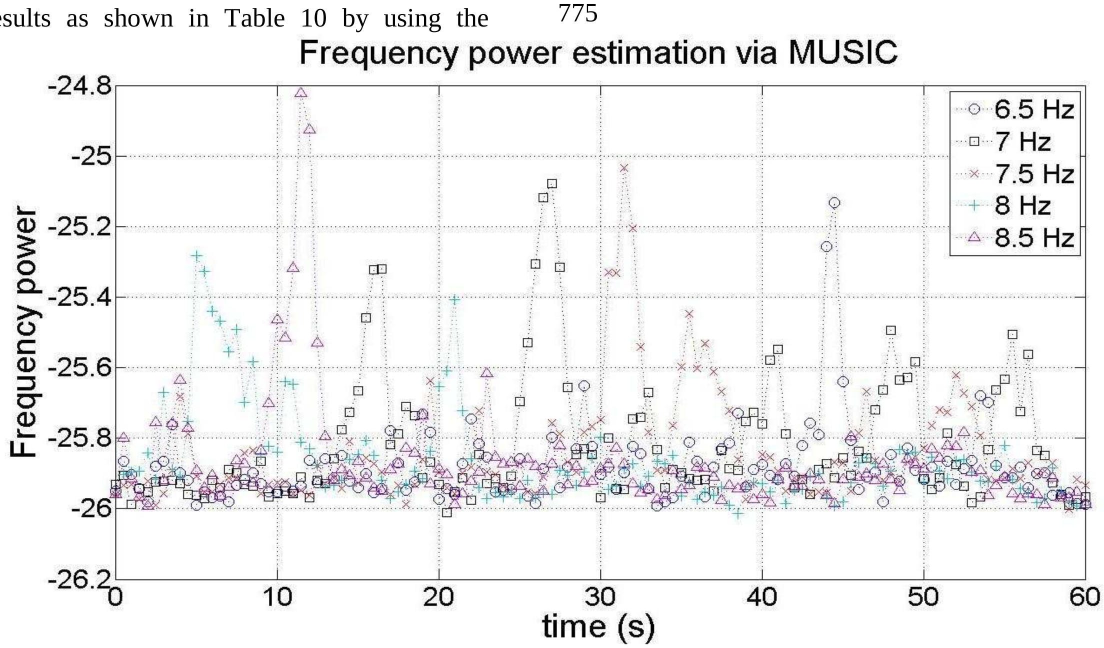


Fig. 10 Frequency power values during the process of spelling “BRAIN”

5 Conclusions

SSVEP-based BCI systems have great potential in real world applications, and the signal processing algorithm is of great importance. In this paper, a MUSIC-based method was proposed for multi-dimensional signal processing and a segmentation ratio for determining the number of eigenvectors constituting the signal subspace was suggested.

The experimental results with simulated data proved that this method could provide high frequency resolution. Also, the multi-channel(dimensional) signals have the anti-noise ability, so the power spectrum estimation is more accurate.

The method was verified with real SSVEP signals acquired in an unshielded environment. The original data were segmented into 2-second epochs with a sampling frequency of 200 Hz. Four channels were enough for feature extraction. The basic idea was to compute the frequency power at stimulus frequencies. Then the frequency with the maximum power value was recognized as the target one. The results showed targets could be well recognized either one stimulus was on or all stimuli were on, compared with a typical multi-channel signal processing method CCA. It was feasible to detect the idle state using a dwell time in our experiments. A disadvantage of this method is that it caused time delay for each decision. An alternative way is to use a threshold, which was obtained with training data. In our research, the threshold was gradually adjusted to maximize the overall recognition accuracy of the training data. Though it didn't cause recognition delay, the preliminary time was longer.

Finally, the proposed method was implemented in a real time application of keyboard control. Different subjects could use the virtual keyboard successfully.

Future work includes improving the algorithm to recognize the target within less time and with higher accuracy. Other factors like design of the stimuli, the subjects comfort, all should be taken into account to develop practical applications for disabled people.

ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation (Grant No. 51475342). The authors would like to thank all subjects for their participating in the experiments.

REFERENCES

1. Wolpaw JR, Birbaumer N, Heetderks WJ, McFarland DJ, Peckham PH, Schalk G, Donchin E, Quatrano LA, Robinson CJ, Vaughan TM (2000) Brain-computer interface technology: A review of the first international meeting. IEEE Trans. Rehabil. Eng. 8 (2):164-173
2. Mason SG, Birch GE (2003) A general framework for brain-computer interface design. IEEE Trans. Neural Syst. Rehabil. Eng. 11 (1):70-85
3. Wang Y WR, Gao X (2006) A practical VEP-based brain-computer interface. IEEE Trans. Neural Syst. Rehabil. Eng. 14 (2):234-240
4. Picton T (1990) Human brain electrophysiology: evoked potentials and evoked magnetic fields in science and medicine. J. Clin. Neurophysiol. 7:450-451
5. Bin GY, Gao XR, Yan Z, Hong B, Gao SK (2009) An online multi-channel SSVEP-based brain-computer interface using a canonical correlation analysis method. J. Neural Eng. 6 (4):1-6(046002)
6. Liu Q, Chen K, Ai Q, Xie SQ (2014) Review: Recent Development of Signal Processing Algorithms for SSVEP-based Brain Computer Interfaces. J. Med. Biol. Eng. 34 (4):299-309
7. Muller-Putz GR, Pfurtscheller G (2008) Control of an electrical prosthesis with an SSVEP-based BCI. IEEE Trans. Biomed. Eng. 55 (1):361-364
8. Bian Y, Li HW, Zhao L, Yang GH, Geng LQ (2011) Research on steady state visual evoked potentials based on wavelet packet technology for brain-computer interface. Proc. Eng. 15:2629-2633
9. Zhang Z, Li X, Deng Z (2010) A CWT-based SSVEP classification method for brain-computer interface system. In: Internation Conference on Intelligent Control and Information Processing, pp 43-48
10. Yan B, Li Z, Li H, Yang G, Shen H (2010) Research on brain-computer interface technology based on steady state visual evoked potentials. In: 4th International Conference on Bioinformatics and Biomedical Engineering, pp 1-4
11. Wu CH, Chang HC, Lee PL, Li KS, Sie JJ, Sun CW, Yang CY, Li PH, Deng HT, Shyu KK (2011) Frequency recognition in an SSVEP-based brain computer interface using empirical mode decomposition and refined generalized zero-crossing. J. Neurosci. Methods 196 (1):170-181
12. Wu CH, Chang HC, Lee PL (2009) Instantaneous gaze-target detection by empirical mode decomposition: application to brain computer interface. In: World Congress on Medical Physics and Biomedical Engineering, pp 215-218
13. Friman O, Luth T, Volosyak I, Graser A (2007) Spelling with steady-state visual evoked potentials. In: 3rd International IEEE/EMBS Conference on Neural Engineering, pp 354-357
14. Volosyak I (2011) SSVEP-based Bremen-BCI interface - boosting information transfer rates. J. Neural Eng. 8 (3):1-11(036020)
15. Volosyak I, Malechka T, Valbuena D, Graser A (2010) A novel calibration method for SSVEP based brain-computer interfaces. In: 18th European Signal Proc. Conf., pp 939-943
16. Cecotti H (2010) A self-paced and calibration-less SSVEP-based brain-computer interface speller. IEEE Trans. Neural Syst. Rehabil. Eng. 18 (2):127-133
17. Zhang ZM, Deng ZD (2012) A kernel canonical correlation analysis based idle-state detection method for SSVEP-based brain-computer interfaces. In: 2nd International Conference on Material and Manufacturing Technology, pp 634-640
18. Lin ZL, Zhang CS, Wu W, Gao XR (2006) Frequency recognition based on canonical correlation analysis for SSVEP-based BCIs. IEEE Trans. Biomed. Eng. 53 (12):2610-2614
19. Hakvoort G, Reuderink B, Obbink M (2011) Comparison of PSDA and CCA detection methods in a SSVEP-based BCI-system. Centre for Telematics and Information Technology, University of Twente. http://eprints.eemcs.utwente.nl/19680/01/Comparison_of_PSDA_and_CCA_detection_methods_in_a_SSVEP-b

ased_BCI-system.pdf

20. Schmidt RO (1986) Multiple emitter location and signal parameter estimation. IEEE Trans. Antennas Propag. 34 (3):276-280

21. Swami A, Mendel JM, Nikias CL (1998) Higher-order Spectral Analysis Toolbox: for Use with MATLAB: User's Guide. Mathworks, Incorporated

22. Wang Yongliang CH, Peng Yingning, Wan Qun (2004) Theories and algorithms of spatial spectrum estimation. Press of Tsinghua University

23. Hardoon D R SS, Shawe-Taylor (2004) Canonical correlation analysis: An overview with application to learning methods. Neural Comput. 16 (12):2639-2664

24. Chen K, Liu Q, Ai QS (2014) Multi-channel SSVEP pattern recognition based on MUSIC. In: 4th International Conference on Intelligent Structure and Vibration Control, pp 84-88