Recognition and Feature Extraction of Motor Imagery EEG Signals Based on Complementary Ensemble Empirical Mode Decomposition

Qidong Huang, Min Tang

ABSTRACT
In view of the characteristics of high nonlinearity and fractional stationarity of motor imagery EEG signals, an improved Complementary Ensemble Empirical Mode Decomposition (CEEMD) method is proposed in this paper. The method does not need to select the basis function in advance and has the characteristics of high self-adaptability. In this paper, Hilbert transform, mutual information, sensitive factor and approximate entropy are used to obtain the time-frequency characteristics, recognition accuracy and other parameters of motor imagery EEG signals in time-frequency domain, and compared with other methods. Using mutual information and sensitive factors, the IMF component with useful information of the original signals can be effectively identified by CEEMD decomposition, and the selected IMF component can be reconstructed and identified by common space model and approximate entropy. The results show that the recognition rate of EEG signal by using the combination of approximate entropy and time-frequency feature is better than that by using time-frequency feature vector alone. Compared with other algorithms, the proposed algorithm has the highest classification accuracy of 84.1%, among 80.9% for ANN algorithm, 79.6% for WT algorithm and 75.8% for EMD-HT algorithm. It indicates that the method in this paper can distinguish the motor imagery tasks and proves its effectiveness and superiority in the extraction and classification of motor EEG signals.

Key Words: EEG Signals, Cerebral Nerve, Motor Imagery, Hand and Left Hand Grip Movement Recognition, Brain-Computer Interface, CEEMD-HT, Feature Extraction

DOI Number: 10.14704/nq.2018.16.6.1642

Introduction
The brain is the highest center of the nervous system of the human body, and controls the emotion, thinking and motor ability of the human body (Wolpaw et al., 2002; Millán et al., 2010; Dong, 2009). Brain cells are conducting electrical activity at every moment, and the curve that electrical activity develops as time changes is electroencephalogram (EEG) signal. In case of neurologic diseases such as frozen human disease, paralysis, and stroke, the nerve and muscle channels in the human body are destroy, and the neural signals of the brain cannot be transmitted (Wang et al., 2013; Cincotti et al., 2007). These patients lose control of some or all of their autonomic muscles, which eventually leads to their inability to communicate with the outside world (Birbaumer and Cohen, 2007; Machado et al., 2013). The purpose of the brain-computer interface (BCI) is to provide a new concept to help patients recover their ability to move (Rozado et al., 2015; Zou, 2016).

Corresponding author: Qidong Huang
Address: Suihua University, Suihua 152061, China.
e-mail: 415159248@qq.com

Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
Received: 6 March 2018; Accepted: 7 May 2018
The purpose of BCI is to establish a direct communication path between the brain and the external environment, and the instructions of the brain can be transmitted to the external device only through the established path instead of through neurons and muscles (Nguyen et al., 2015; Yang and Wu, 2013; Bajaj and Pachori, 2012). At present, the brain signals which have been discovered and accepted by the academic circle mainly include VEP signal (controlling brain visual cortex), P300, SMR (motor synchronization), SCP signal (brain signal voltage change), etc. (Mak et al., 2012; Ortiz-Rosario and Adeli, 2013; Blankertz et al., 2010).

Motor imagery belongs to a kind of mental activity of brain signal, the actual motion of human body produces related brain signals with imaginary motion, and then causes the change of related potential of cerebral cortex (Bajaj and Pachori, 2013; Pachori and Bajaj, 2011). BCI based on motor imagery is a research hotspot in the field of brain nerves (Djemili et al., 2013; Rodríguez-Bermúdez et al., 2011). For example, Mousavi et al. constructed the virtual environment based on motor imagery BCI, and extracted the characteristic energy of brain signals (Mousavi et al., 2011); Kus et al. established a multi-classification cursor control system for motor imagery BCI, with the accuracy of information transmission exceeding 75% (Kus et al., 2012); and Hasan et al. established an adaptive motor imagery BCI system using sensory motion control (Hasan and Gan, 2012). BCI based on motor imagery can realize event-related synchronization (ERS) phenomenon of thought and actual motion, which is the characteristic of motor imagery BCI system (Bajaj and Pachori, 2013; Rodríguez-Bermúdez et al., 2013; Tang and Chen, 2016; Elbeltagy et al., 2018; Luo and Jia, 2016).

The EEG signals based on motor imagery have typical non-stationary and non-linear characteristics (Hsu, 2015). At present, the conventional EEG signal extraction method is to use scalp electrodes to obtain EEG signals, and this method has such defects as weak signal, and large background noise. Extracting useful information feature of brain signals has become the difficult point of studying motor imagery EEG signals at present (Hwang, Kwon and Im, 2009; Fu et al., 2014; Yeh et al., 2010). In previous studies, linear regression model, wavelet packet, Fourier transform, pow spectrum and other methods were mainly adopted for feature extraction of EEG signals, and all of the above methods have greater limitations and defects and were often unable to collect useful signals ((Hsu, 2010; Gong et al., 2013; Hsu, 2013). HHT algorithm is a successful algorithm which can extract the features of motor imagery BCI signal at present. It has higher resolution in both time domain and frequency domain due to such features as self-adaptability and no-apriority (Zhang et al., 2013; Higashi and Tanaka, 2012).

For the features of high nonlinearity and fractional stationarity of motor imagery EEG signals, an improved complementary ensemble empirical mode decomposition (CEEMD) method is proposed in this paper. CEEMD does not need to select the basis function in advance and is highly adaptive. By means of Hilbert transform, mutual information, sensitive factor and approximate entropy, the parameters such as time-frequency features and recognition accuracy of motor imagery EEG signals in time-frequency domain are obtained and compared with other methods.

**Test processes and data processing methods**

**Test processes**

There are 9 subjects, 5 men and 4 women, at the age of 24.8 years on average, without serious diseases in the past and related EEG and BCI test experience. The test environment is Windows64bit system with 16GB memory and 2.3GHz main frequency. The simulation software is Matlab.

Figure 1 illustrates the motor imagery EEG signal acquisition process. The subject sits on the stool and a display screen is hung 1 meter in front. The prompt tone is played at the 2nd second when the test starts, and at the 3rd second, a non-directional arrow on the screen prompts the subject should imagine the movement direction of his left hand and right hand according to the arrow direction. The sampling frequency of EEG signals is 64 Hz, and the test data is composed of 1/3 training sample and 2/3 test sample.

![Figure 1. Motor Imagery EEG Signal Acquisition Process](image-url)
It has been known from previous studies that the main frequency bands of motor imagery EEG signals are mainly distributed in 8-12 Hz and 15-30 Hz. Therefore, after the EEG signals are collected, they can be band-pass filtered in advance, and the processed signals can be linearly corrected. Data are collected via C3, C4 and Cz electrodes.

**Ceeemd—ht**

Empirical Mode Decomposition (EMD) can decompose the original signals into a sum of several Intrinsic Mode Functions (IMFs), and the decomposed IMFs can have linear or nonlinear characteristics. Its essence is to "screen" the signals and separate the IMF of different components. EMD decomposition is complete, that's, the sum of IMFs and residual quantity \( r_n(t) \) obtained by EMD decomposition of original traffic data are the same as the original data, and there is no energy loss in the decomposition process.

The original EMD decomposition has the problem of modal aliasing, that’s, any IMF contains other traffic data with different time scales, which makes the subsequent sub-band analysis difficult. Based on EMD decomposition, Yeh proposed a noise-assisted EMD decomposition method called CEEMD, and the specific steps are as follows:

1. Addition of white noise \( n(t) \) in pairs to original traffic data \( m(t) \)

\[
\begin{align*}
    m^+_i(t) &= m(t) + n_i(t) \\
    m^-_i(t) &= m(t) - n_i(t)
\end{align*}
\]

2. EMD is conducted respectively for \( m^+_i(t) \) and \( m^-_i(t) \) to obtain IMF \( IMF_{i^+} \) and IMF \( IMF_{i^-} \),

\[
IMF_i = \left[ IMF_{i^+} + IMF_{i^-} \right] / 2
\]

3. EMD is conducted for \( x(t) \) with additional \( L-1 \) white noise, and there are \( L \) groups of different modal components. The sum and the arithmetic mean of components of the same order in group \( L \) are obtained to find the final components of each order in accordance with CEEMD requirements

\[
c_j(t) = \frac{1}{L} \sum_{k=1}^{L} c_{jk}(t)
\]

\( c(t) \) is IMF. Compared with EMD, CEEMD decomposition method can completely solve the noise residue by adding noise in pairs, and can effectively suppress the mode aliasing phenomenon. The separation degree of the original traffic data in each frequency band is better, and the local self-correlation in the traffic data can be better suppressed.

The typical EEG signals are decomposed with CEEMD to obtain IMF components of each order as shown in Figure 2.

![CEEMD decomposition results of brain signals](image)

A Hilbert transform is performed on the decomposed IMF, then

\[
H(\omega,t) = \text{Re} \sum_{i=1}^{n} C_i(t)e^{i\theta_i(t)} = \text{Re} \sum_{i=1}^{n} C_i(t)e^{i\theta_i(t)dt}
\]

\( \text{Re} \) represents the real part. By integrating \( H(\omega,t) \) to the time, the Hilbert marginal energy spectrum is obtained:

\[
ES(\omega) = \int_{0}^{T} H^2(\omega,t)dt
\]

The marginal spectrum can show the superposition of all amplitudes of the signals at each frequency. The instantaneous energy of the signals:

\[
IE(t) = \int_{\omega} H^2(\omega,t)d\omega
\]

---

---

Huang Q., Recognition and Feature Extraction of Motor Imagery EEG Signals Based on Complementary Ensemble Empirical Mode Decomposition
IE expresses the relationship between energy and time. There are differences in instantaneous energy and marginal spectrum for different subjects. Therefore, the obtained IMF of previous orders is analyzed in sub-frequency bands, that's, segmenting one IMF into several sub-frequency bands to obtain the average amplitude of data in each sub-frequency band as the spectrum characteristic of the frequency band. For energy, the energy mean and variance are obtained. Take the first-order IMF as example, the video feature vector $F_1$ is

$$
F_1 = \left\{ C_{3\text{IE}}, C_{3\text{IE}}, C_{3\text{IE}}, \ldots, C_{4\text{IE}}, C_{4\text{IE}}, \ldots, C_{4\text{IE}}} \right\}
$$

(7)

C3 and C4 are data acquisition electrodes, and are divided into 11 bands.

The approximate entropy is used to measure the irregularity of the EEG signals and describe their regularity. The approximate entropy $Ap\text{En}$ can be expressed as

$$
Ap\text{En}(m, r, N) = \Phi^m(r) - \Phi^{m+1}(r)
$$

(8)

$$
\Phi^m(r) = (N - m + 1)^{-1} \sum_{i=1}^{N-m+1} \ln C_i^m(r)
$$

(9)

N is the length of sequence; $r$ is 0.2 times the standard deviation of the original signals; $m$ is the number of dimensions.

The approximate entropy feature vector $F_2$ of any IMF of the signals can be expressed as:

$$
F_2 = \{ C_{3\text{ApEn}}, C_{3\text{ApEn}}, \ldots, C_{4\text{ApEn}}, C_{4\text{ApEn}}, \ldots, C_{4\text{ApEn}} \}
$$

(10)

**Mutual information and IMF component selection**

Mutual information can be used to represent the correlation between two variables. Set the entropy of two functions, $X$ and $Y$ as $H(X)$ and $H(Y)$ respectively, and the joint entropy as $H(X, Y)$, then their expressions are

$$
H(X) = -\sum_{x \in X} p_x(x) \log p_x(x)
$$

(11)

$$
H(Y) = -\sum_{y \in Y} p_y(y) \log p_y(y)
$$

(12)

$H(X, Y) = -\sum_{x \in X, y \in Y} p(x, y) \log p(x, y)$

(13)

$p(x, y)$ is joint probability density, $p_x(x)$ and $p_y(y)$ is marginal probability density, then mutual information of $X$ and $Y$ is

$$
I(X, Y) = H(X) + H(Y) - H(X, Y)
$$

(14)

The mutual information is used to determine whether the IMF components of the decomposed IMF of each order contain useful information about the original signals. The decomposed IMF component of the $i$ order is $c_{x_i}(t)$, and the original signal is $x(t)$, then the mutual information $A_i$ between two is

$$
A_i = I(c_{x_i}(t), x(t))
$$

(15)

The noise signal $n(t)$ is decomposed through CEEMD, and the mutual information $B_i$ between $c_{n_i}(t)$ and $n(t)$ of the IMF component of the $i$ order is

$$
B_i = I(c_{n_i}(t), n(t))
$$

(16)

The mutual information $E_i$ between $c_{x_i}(t)$ and $c_{n_i}(t)$ is calculated according to above two equations:

$$
E_i = I(c_{x_i}(t), c_{n_i}(t))
$$

(17)

The sensitive factor $R_i$ of IMF component of $x(t)$ of the $i$ order is calculated:

$$
R_i = \exp \left[ \frac{(A_i + B_i)}{2} - E_i \right]
$$

(18)

The greater $R_i$ is, and the greater $A_i$ and $B_i$ values are, then the smaller $E_i$ is, and the more the useful information of original signals the IMF component contains.

In conclusion, the process of feature extraction of the motor imaginary EEG signals proposed in this paper is shown in Figure 3.
**Figure 3.** Process of feature extraction of the motor imaginary EEG signals

**Test Results and Analysis**

In order to verify the accuracy of feature extraction of motor imagery EEG signals based on CEEMD-HT proposed in this paper, wavelet transform and CEEMD-HT method are respectively used to carry out feature extraction of EEG signals, with the extraction results are shown in Figure 4.

![Figure 4. Accuracy of feature extraction of motor imagery EEG signals based on wavelet transform and CEEMD-HT method](image)

As can be seen from the figure, when wavelet transform is used to extract features, db8 wavelet base is selected to decompose raw data into three layers, and the signals are decomposed into four sub-bands: 0-8Hz, 8-16Hz, 16-32Hz and 32-64Hz. The average correct rate of feature extraction in 4-7s is 72.4% with the highest correct rate of 80.9%. The average correct rate of feature extraction in 7-9s is 51.8% with the highest correct rate of 63.6%.

When CEEMD-HT method is used to extract features, the average correct rate of feature extraction is 80.1% in 4-7s with the highest correct rate of 87.5%. The average correct rate of feature extraction in 7-9s is 70.8% with the highest correct rate of 81.6%. The comparison between the two methods shows that CEEMD-HT-based method has a better effect on extracting features of EEG signals from left and right hand motor imagery EEG signals.

![Figure 5. Curve of subjects' recognition rate](image)

Table 1 shows the recognition rate of 9 subjects in the ascending stage (4-7s) and the optimal stage (7-9s). Among them, 4-7s is the initial period of feedback of motor imagery EEG signals, and is the transition period for the subjects to form motor imagery EEG signals. The 7-9s signal is the mature stage of the motor
imagery EEG signals, and the signals are basically stable. As can be seen from Figure 5 and Table 1, the recognition rate of motor imagery EEG signals using the combination of approximate entropy and time-frequency features is obviously superior to that of time-frequency feature vectors alone.

Figure 7. Average instantaneous energy spectra of motor imagery EEG signals in left hand (upper) and right hand (lower) of the subject

Figure 6. Average instantaneous energy spectra of motor imagery EEG signals in left hand (upper) and right hand (lower) of the subject

Table 1. Recognition rate of 9 subjects in different periods

<table>
<thead>
<tr>
<th>The subjects</th>
<th>Feature combination</th>
<th>Recognition rate rising period average recognition rate(%)</th>
<th>The best time average recognition rate(%)</th>
<th>Optimal time of the highest recognition rate(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NO.1</td>
<td>F_1</td>
<td>69.31</td>
<td>84.33</td>
<td>86.11</td>
</tr>
<tr>
<td></td>
<td>F_1 + F_2</td>
<td>72.59</td>
<td>85.97</td>
<td>86.27</td>
</tr>
<tr>
<td>NO.2</td>
<td>F_1</td>
<td>68.06</td>
<td>77.02</td>
<td>77.36</td>
</tr>
<tr>
<td></td>
<td>F_1 + F_2</td>
<td>73.22</td>
<td>77.37</td>
<td>75.84</td>
</tr>
<tr>
<td>NO.3</td>
<td>F_1</td>
<td>79.37</td>
<td>88.49</td>
<td>88.59</td>
</tr>
<tr>
<td></td>
<td>F_1 + F_2</td>
<td>80.35</td>
<td>86.53</td>
<td>88.43</td>
</tr>
<tr>
<td>NO.4</td>
<td>F_1</td>
<td>65.79</td>
<td>85.95</td>
<td>78.63</td>
</tr>
<tr>
<td></td>
<td>F_1 + F_2</td>
<td>63.37</td>
<td>78.66</td>
<td>78.47</td>
</tr>
<tr>
<td>NO.5</td>
<td>F_1</td>
<td>64.27</td>
<td>78.39</td>
<td>80.75</td>
</tr>
<tr>
<td></td>
<td>F_1 + F_2</td>
<td>66.68</td>
<td>77.71</td>
<td>85.58</td>
</tr>
<tr>
<td>NO.6</td>
<td>F_1</td>
<td>73.66</td>
<td>93.04</td>
<td>96.41</td>
</tr>
<tr>
<td></td>
<td>F_1 + F_2</td>
<td>79.48</td>
<td>95.45</td>
<td>95.75</td>
</tr>
<tr>
<td>NO.7</td>
<td>F_1</td>
<td>69.52</td>
<td>82.74</td>
<td>80.27</td>
</tr>
<tr>
<td></td>
<td>F_1 + F_2</td>
<td>79.83</td>
<td>79.55</td>
<td>80.44</td>
</tr>
<tr>
<td>NO.8</td>
<td>F_1</td>
<td>66.29</td>
<td>79.48</td>
<td>78.46</td>
</tr>
<tr>
<td></td>
<td>F_1 + F_2</td>
<td>69.05</td>
<td>76.84</td>
<td>78.18</td>
</tr>
<tr>
<td>NO.9</td>
<td>F_1</td>
<td>62.49</td>
<td>70.06</td>
<td>74.61</td>
</tr>
<tr>
<td></td>
<td>F_1 + F_2</td>
<td>67.28</td>
<td>74.18</td>
<td>74.83</td>
</tr>
</tbody>
</table>

Figure 8. Sensitive factors of motor imagery EEG signals of four testers

It can be seen from the figure that the Ri values of the four tested subjects are different, but the Ri values of the first three orders’ IMF are larger than 0.3, while the Ri values of the fourth-seventh orders’ IMF are less, ranging 0.25-0.3. The Ri value of IMF2 is the largest, which has a significant advantage over other IMF orders. Since the size of the Ri value reflects the amount of
useful information, the useful information of the original signals is mainly contained in the first 3 IMF orders. Since some high-frequency ambient noise may be included in the first-order IMF, the IMF 1 is filtered.

The reconstructed EEG signals of the first subject is selected for feature extraction. The feature vector dimension is 4, and the feature vector of 150 × 4 is obtained after calculation. The first 100 feature vectors of the subjects are used to draw the curve, with the results shown in Figure 9, where the first class represents the feature value of the right-hand imaginary movement, and the second class represents the feature value of the left-hand imaginary movement. As can be seen from the figure, the overall distinction of the feature value is better.

![Figure 9. Curve of feature value extracted from the subjects](image)

**Table 2.** Comparison of classification accuracy and standard deviation of motor imagery EEG signals of 9 subjects

<table>
<thead>
<tr>
<th></th>
<th>WT</th>
<th>EMD-HT</th>
<th>ANN</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>NO.1</td>
<td>76.9±4.9</td>
<td>76.2±4.9</td>
<td>82.3±3.3</td>
<td>86.8±4.2</td>
</tr>
<tr>
<td>NO.2</td>
<td>73.5±4.2</td>
<td>76.5±4.1</td>
<td>80.6±3.7</td>
<td>84.5±4.0</td>
</tr>
<tr>
<td>NO.3</td>
<td>89.2±4.6</td>
<td>75.3±4.4</td>
<td>81.2±3.4</td>
<td>90.2±3.6</td>
</tr>
<tr>
<td>NO.4</td>
<td>82.3±4.3</td>
<td>73.4±4.7</td>
<td>76.3±4.1</td>
<td>85.4±3.3</td>
</tr>
<tr>
<td>NO.5</td>
<td>81.1±4.7</td>
<td>74.2±4.6</td>
<td>84.2±3.2</td>
<td>81.9±3.4</td>
</tr>
<tr>
<td>NO.6</td>
<td>79.2±4.2</td>
<td>79.1±4.2</td>
<td>83.7±3.6</td>
<td>84.2±3.1</td>
</tr>
<tr>
<td>NO.7</td>
<td>78.5±4.1</td>
<td>80.2±4.6</td>
<td>78.9±4.2</td>
<td>80.6±3.5</td>
</tr>
<tr>
<td>NO.8</td>
<td>76.9±4.3</td>
<td>75.7±4.9</td>
<td>79.6±3.1</td>
<td>80.6±3.6</td>
</tr>
<tr>
<td>NO.9</td>
<td>78.8±4.4</td>
<td>71.6±4.8</td>
<td>81.3±3.8</td>
<td>82.7±3.7</td>
</tr>
<tr>
<td>Average</td>
<td>79.6±4.3</td>
<td>75.8±4.6</td>
<td>80.9±3.6</td>
<td>84.1±3.6</td>
</tr>
</tbody>
</table>

The algorithm proposed hereby is compared with wavelet transform (WT), EMD-HT decomposition and artificial neural network (ANN), and the classification accuracy and standard deviation of motor imagery EEG signals are calculated by four methods. The calculation results are shown in Table 2.

It can be seen from the table that body differences among different subjects have an effect on classification accuracy. For example, the classification effect of the ANN of the 4th subject and the method proposed hereby is better than that of WT algorithm, while the WT and EMD-HHT algorithms of the 7th subject are better than that of ANN algorithm. From the average, the classification accuracy of the algorithm proposed hereby is the highest, reaching 84.1%, ANN is 80.9%, WT and EMD-HT are only 79.6% and 75.8% respectively. It is shown that the method in this paper can distinguish the motor imagery tasks, and proves its effectiveness and superiority in the extraction and classification of motor imagery EEG signals.

**Conclusions**

According to the characteristics of high nonlinearity and fractional stationarity of motor imagery EEG signals, an improved complementary empirical mode decomposition (CEEMD) method is proposed in this paper, and CEEMD does not need to select the basis function in advance with high adaption. By means of Hilbert transform, mutual information, sensitive factor and approximate entropy, the parameters such as time-frequency feature and recognition accuracy of motor imagery EEG signals in time-frequency domain are obtained and compared with other methods. The conclusions of the study are as follows:

(1) With mutual information and sensitive factors, the IMF component containing useful information of the original signals decomposed via CEEMD can be recognized effectively. With common space model and approximate entropy, the selected IMF component can be reconstructed and recognized. The results show that the recognition rate of EEG signals using the combination of approximate entropy and time-frequency feature is better than that using time-frequency feature vector alone.

(2) Compared with other algorithms, the proposed algorithm has the highest classification accuracy, reaching 84.1%, while the classification accuracy is 80.9% for ANN algorithm, 79.6% for WT algorithm and 75.8% for EMD-HT algorithm respectively. It is shown that the method proposed in this paper can better distinguish the motor imagery tasks, and proves its effectiveness.
and superiority in the extraction and classification of motor imagery EEG signals.

References


