Detecting movement-related EEG change by wavelet decomposition-based neural networks trained with single thumb movement

Chih-Wei Chen a, Chou-Ching K. Lin b, Ming-Shaung Ju a,*

a Department of Mechanical Engineering, National Cheng Kung University, Tainan, Taiwan 701
b Department of Neurology, National Cheng Kung University Hospital, Tainan, Taiwan

Accepted 9 December 2006

Abstract

Objective: The main goal of this study was to develop a real-time detection algorithm of movement-related EEG changes for the naive subjects with a very small amount of training data. Such an algorithm is vital for the realization of brain–computer interface.

Methods: The target algorithm developed in this study was based on the wavelet decomposition neural network (WDNN). Surface Laplacian EEG was recorded at central cortical areas and processed with wavelet decomposition (WD) for feature extraction and neural network for pattern recognition. The new algorithm was compared with other three methods, namely, threshold-based WD and short-time Fourier transform (STFT), and Fourier transform neural network (FTNN), for performance. The trainings of all algorithms were based, respectively, on the changes of μ and β rhythms before and after voluntary movements. In order to investigate whether WDNN could adapt to the nonstationarity of EEG or not, we also compared two training modes, namely, fixed and updated weight. The significances of the success rates were tested by ANOVA (analysis of variance) and verified by ROC (receiver operating characteristic) analysis.

Results: The experimental data showed that (1) success rates of movement detection were acceptable even when the training set was reduced to a single trial data, (2) WDNN performed better than WD or STFT without optimized thresholds and (3) when weights were updated and thresholds were optimized, WDNN still performed better than WD, while FTNN had a marginal advantage over STFT.

Conclusions: We developed a detection algorithm based on WDNN with the training set being reduced to a single trial data. The overall performance of this algorithm was better than the conventional methods as such.

Significance: μ wave suppression could be detected more precisely by the wavelet decomposition with neural network than the conventional algorithms such as STFT and WD. The size of training data could be reduced to a single trial and the success rates were up to 75–80%.

Keywords: Brain–computer interface; Wavelet decomposition; Neural network

1. Introduction

μ and β rhythms (8–14 and 16–24 Hz, respectively) recorded from the primary sensorimotor cortical area are highly related to voluntary movements. Their characteristics are suppression before movements (event-related desynchronization, ERD) and enhancement after movements (event-related synchronization, ERS) (Pfurtscheller and Berghold, 1989; Müller-Gerking et al., 2000; Pineda et al., 2000; Delorme and Makeig, 2003). Because of the low signal-to-noise ratio of EEG, surface Laplacian operation was a standard technique for removing distant artifacts (Oostendorp and Adriaan, 1996). Pure movement attempts or imagination without actual movement could...
also result in similar phenomena. It was shown that subjects could be trained to regulate their μ or β waves after some training (Neuper et al., 1999; Babiloni et al., 2001), so these brain rhythms can be employed as the control sources of brain–computer interface (BCI).

BCI provides an artificial communication channel between the human brain and computers (Wolpaw et al., 2000; Guger et al., 2000). It has been developed for locked-in or stroke patients, whose cortices are relatively intact, to help them communicate with others or assist their daily lives (Donchin et al., 2000; Kennedy et al., 2000; Hinterberger et al., 2003). These patients can be trained to control devices such as computer cursors, spelling systems, or prostheses by means of motor imagination (Pfurtscheller and Neuper, 2001). For those BCI systems that use the dynamics of oscillatory activity as the control source, EEG processing is one of the essential core technologies.

The conventional surface EEG is a summation contributed from the activated sources of the brain, which have poor spatial resolution. Surface Laplacian operation has been considered as an operator that estimates the local current density, which emits from the cortical sources through the skull and scalp (Bradshaw and Wikswo, 2001 and Srinivasan et al., 1996). The estimation of surface Laplacian operation utilized in this study is actually the so-called Hjorth (1975) method, which is a method of local spatial enhancement of EEG. The poor spatial resolution can be improved from 6–10 to 2–3 cm (Nunez, 2000). Several studies (Pfurtscheller and Neuper, 1997) showed that unilateral (real or imaginary) hand movement resulted in a contralateral ERD close to the primary hand motor areas (C3 and C4). The distance between C3 and C4, depending on the size of a subject’s head, is about 7–9 cm. Therefore, the improvement of spatial resolution offered by the surface Laplacian operation makes it possible to discriminate the local current density at C3 and C4. In addition to the local surface Laplacian operation, another global method based on the spline interpolation techniques was also suggested to compare the differences (Tandonnet et al., 2005), and an empirical evidence showed that both methods provided equivalent improvement of spatial resolution. Moreover, the authors also concluded that the local Laplacian operation method was more suitable for studying particular cortical areas.

Feature extraction is a key issue in EEG processing. Short-time Fourier transform (STFT) is a conventional approach to estimate the activity of the EEG in different frequency bands. The problem with STFT is that fast transients are missed. On the other hand, wavelet decomposition (WD) provides a multi-level time-frequency decomposition of functions or signals, which allows simultaneous use of long-time interval for low-frequency information and short-time interval for high-frequency information. Anil et al. (Rao and Jones, 2000) developed a wavelet-based multi-sensor denoising algorithm to extract EEG out of noises. The results concluded that the wavelet-based algorithm performed better than discrete Fourier transform and Wiener filter.

Neural network is a good method in prediction, pattern recognition, and function approximation for processing chaotic or nonstationary signals (Pfurtscheller et al., 1996; Sun and Slabassi, 2000). Haselsteiner and Pfurtscheller (2000) used a neural network of multi-layer perceptrons (MLP) as the EEG classifier, and the error rates were compared with those of finite impulse response (FIR) MLP. It was concluded that the FIR MLP had better performance and a lower error rate. Bostanov (2004) introduced the method of t-CWT, which was based on the continuous wavelet transform (CWT) and Student’s t-statistic, to process EEG data sets provided by other research groups. The authors concluded that this method resulted in lower error rates and higher precision just like the powerful methods as support vector machine (SVM).

The main goal of this study was to test whether it was possible to train the wavelet decomposition neural network (WDNN) algorithm by the data of one single-trial thumb movement in an untrained subject for detecting μ wave suppression in later movements. Two training processes, namely fixed weight and updated weight weightings, were designed to investigate whether re-training was beneficial for coping with the nonstationarity of EEG. For comparison, we also designed Fourier transform neural network (FTNN), which used ERD and ERS patterns as the training sets. The success rates of these WDNNs and FTNNs were compared with those of the wavelet decomposition and conventional STFT methods.

2. Methods

2.1. Experimental setup

Experiments were performed in an electrically shielded room to avoid electrical contamination. EEG was amplified 10,000× and filtered by a 3–100 Hz analog band-pass filter in a commercial EEG recorder. Seven channels of signals were recorded in the experiments, including five channels of EEG, one channel of right thumb extensor electromyography (EMG) and one channel of cue signal.

In this study, the positioning of EEG electrodes was based on the standard international 10–20 system. One EEG electrode was placed on C3, the other four were placed 2 cm to the right, left, front, and rear of C3, respectively, and A1 (left earlobe) was chosen as the reference. Two electrodes were placed on the forearm to record one channel of right thumb extensor EMG. EMG was also amplified and filtered by the same EEG machine. The cue was produced by a custom-made buzzer with the ‘ beep’ sound. All channels of signals were sampled and digitized...
by a digital signal processing card (D1104, dSPACE, www.dspaceinc.com) at 256 Hz.

### 2.2. Experimental procedures

A subject lay on a deck chair in a supine position with the right upper limb well-supported and was instructed to be relaxed and keep eyes open as long as possible during the experiments. The right forearm was in the neutral position with the fingers in the natural flexed position. The length of one round was 30 s, including two right thumb swings separated by an interval of 10 s. Each movement was cued by a beep. The subject was asked to swing the right thumb up promptly and let it fall down naturally at about 2 s after a beep was heard. The time-delay between right thumb up promptly and let it fall down naturally was cued by a beep. The subject was asked to swing the swings separated by an interval of 10 s. Each movement of movement attempts.

There were 2 min of rest between two consecutive rounds. Twenty-one successive rounds including 42 thumb movements were collected in one experiment. The data were stored for later off-line analyses. The subject was requested not to think about his hand, fingers or anything about movements during the trial. Ten subjects in total participated in this study, and all of them were normal and healthy young males without history of neuromuscular diseases, and none had any previous experience with this study or other similar studies.

### 2.3. Data processing

#### 2.3.1. Preprocessing

The onset of right thumb movement was defined by the EMG linear envelope, which was computed by the following steps: (1) high-pass filtering by a fourth-order Butterworth filter with a cutoff frequency at 64 Hz, (2) taking the absolute values, and (3) low-pass-filtering by a fourth-order Butterworth filter with a cutoff frequency at 3 Hz.

EEG signals were preprocessed with Laplacian operation (Hjorth, 1975) by the formula,

\[
S = E_1 - \left( \frac{E_1 + E_3 + E_4 + E_5}{4} \right),
\]

where \(S\) is the processed EEG signal, \(E_1\) is the EEG signal recorded on C3, and \(E_2-E_5\) are the other four EEG signals recorded at the points surrounding C3.

Movement detection was a pattern-based procedure that sequentially performed feature extraction and classification. A pattern contained features of 5 s EEG data. We used wavelet decomposition (WD) and short-time Fourier transform (STFT) for feature extraction and thresholding and neural network (NN) for pattern classification.

#### 2.3.2. Wavelet decomposition

We chose Daubechies 8th order wavelet, an orthonormal wavelet, to extract \(\mu\) (8–16 Hz) and \(\beta\) (16–24 Hz) wave components from the preprocessed EEG signals. One level of wavelet decomposition separated the original signal into two complementary halves, i.e., approximation and detail. The approximation was the high-scale, low-frequency component, and the detail was the low-scale, high-frequency component. In each half, the number of samples was reduced by half because of dyadic downsampling and the frequency range was also reduced by half. EEG signals were sampled and digitized at 256 Hz and, thus, the frequency content was 0–128 Hz. Four-level decomposition was required to decompose the features corresponding to \(\mu\) (8–16 Hz) and \(\beta\) (16–24 Hz) bands (128/2^4) and the sample number became 16 (256/2^4). The steps to derive a pattern from 5 s EEG data by WD were (1) cutting the 5 s train of EEG into successive 1 s segments with 0.75 s overlapping between consecutive segments, (2) taking WD and deriving the 16 coefficients representing \(\mu\) and \(\beta\) waves for each of 17 successive segments of EEG, and (3) calculating the mean-squared values of the coefficients for these 17 segments as the features of \(\mu\) and \(\beta\) waves. A pattern consisted of 17 mean-squared features. These three steps were repeated to derive a sequence of 5 s patterns with 1 s shift between adjacent patterns.

#### 2.3.3. Short-time Fourier transform

A rectangular window with 1 s length was chosen, and there was a 0.25-s shift between windows. The features of \(\mu\) and \(\beta\) waves were derived by (1) taking Fourier transform on each successive window for 5 s (17 windows in total), (2) calculating the mean-squared values of Fourier coefficients of 8–16 and 16–24 Hz, respectively, and, then, (3) computing the relative percentage of bandpower by the formula

\[
RP = \frac{(P - P_0)}{P_0} \times 100\% ,
\]

where \(P\) are the results after steps (1) and (2), \(P_0\) is the first value of \(P\), and \(RP\) is the relative power in percentage. These three steps were repeated to derive successive 5 s patterns with 1 s shift between adjacent patterns.

The methods that combined feature extraction by STFT and WD based on \(\mu\) and \(\beta\) waves and thresholding classification were denoted as STFT_\(\mu\), STFT_\(\beta\), WD_\(\mu\), and WD_\(\beta\), respectively, and those combining a NN classifier were denoted as FTNN_\(\mu\), FTNN_\(\beta\), WDNN_\(\mu\), and WDNN_\(\beta\), respectively.

#### 2.3.4. Thresholding

The values of threshold defined for the two feature extraction methods were different. The thresholds for \(\mu\) and \(\beta\) patterns derived from WD were 4 and 5, respectively. The thresholds for \(\mu\) and \(\beta\) patterns derived from STFT were \(-30\%\) and \(35\%\), respectively. The definitions of thresholds for WD and STFT were different because they were calculated in different ways and were in different scales. Thresholds in STFT were relative parameters that were derived from Eq. (2), while thresholds in WD were absolute values without normalization.
The decision rules of successful detection by WD$_{\mu}$ and WD$_{\beta}$ were different. For WD$_{\mu}$, the minimum feature of a pattern must exist between the 2nd and 4th s (5th–13th features) and must be smaller than 5, and for WD$_{\beta}$, the maximum feature of a pattern must exist between the 3rd and 5th s (9th–17th features) and must be larger than 4.

The criteria of successful movement detection by STFT$_{\mu}$ and STFT$_{\beta}$ were different. For STFT$_{\mu}$, the minimum feature of a pattern must exist between the 2nd and 4th seconds (5th–13th features) and must be smaller than $-30\%$. For STFT$_{\beta}$, the maximum feature of a pattern must exist between the 3rd and 5th s (9th–17th features) and must be larger than $35\%$.

Success rates were different under different choices of thresholds for WD$_{\mu}$, WD$_{\beta}$, STFT$_{\mu}$ and STFT$_{\beta}$. The optimal threshold was defined by minimizing the difference between specificity and sensitivity (Wayne, 1999). The thresholds of WD$_{\beta}$ and WD$_{\mu}$ were tuned from 2 to 10 and 1 to 9, respectively, and the thresholds of STFT$_{\mu}$ and STFT$_{\beta}$ were tuned from $-50$ to $-10\%$ and $10$–$50\%$, respectively. The optimal thresholds were determined and the optimized success rates were denoted as WD$_{\mu}^{\beta}$, WD$_{\beta}^{\mu}$, STFT$_{\mu}^{\beta}$ and STFT$_{\beta}^{\mu}$.

### 2.3.5. Neural network

We used a three-layered feed-forward neural network (NN) as the other classifier. The input layer had 16 neurons, the hidden layer had eight neurons and the output layer had one neuron. A sigmoid function was chosen as the neural excitation function,

$$G(x) = \frac{2}{1 + e^{-x}} - 1.$$  (3)

The weights of NN were tuned by the back-propagation learning algorithm. The learning objective was to minimize the sum of squared errors between the output and the target with a learning rate of 0.01.

WDNN that we used in this study was a combination of WD and NN classifier. The input pattern of WDNN was derived by computing the wavelet coefficients of $\mu$ and $\beta$ waves for 17 successive EEG segments (5 s in length, the same as in the step 1 of WD), and arranging these wavelet coefficients into a $16 \times 17$ matrix. The input patterns based on $\mu$ and $\beta$ waves, $c_\mu$ and $c_\beta$, were fed into the 3-layer NN and single-valued outputs, $y_{\text{out}}$, were obtained by the formula

$$y_{\text{out}} = G(W_{23} * G(W_{12} * G(c))),$$  (4)

where $G$ is the activation function shown in Eq. (3) with output range of $[-1, 1]$. $W_i$ are the weights from the $i$th to $j$th layer, 1st layer is the input layer, 2nd layer is the hidden layer, and the 3rd layer is the output layer, and $c$ is the input pattern ($c_\mu$ or $c_\beta$), the structure of WDNN is shown in Fig. 1.

The beginning time of training data, $t_b$, was determined by the cue-given time, and the end time of training data, $t_e$, was 5 s after $t_b$ (Fig. 2(a)),

$$t_e = t_b + 5.$$  (5)

The training data was extracted from $t_b$ to $t_e$ (in time course) from the EEG of one thumb movement (a $16 \times 17$ matrix).

For WDNN$_{\mu}$, the training target ($y_{At}$, Fig. 2(b)) was determined by the following rule,

$$y_{At} = -1, \text{ for } t \in [t_b, t_m + 1],$$
$$= +1, \text{ for } t \in (t_m + 1, t_e].$$  (6)

where $t_m$ is the time of movement onset defined by the peak time of EMG envelope. $y_{At}$ was defined based on the trend of $\mu$ wave suppression. Similarly, the training target of WDNN$_{\beta}$ ($y_{Bt}$, Fig. 2(c)) was determined by:

$$y_{Bt} = -1, \text{ for } t \in [t_b, t_m],$$
$$= +1, \text{ for } t \in (t_m, t_m + 1],$$
$$= -1, \text{ for } t \in (t_m + 1, t_e].$$  (7)

$y_{Bt}$ was defined based on the characteristics of $\beta$ wave post-movement enhancement.

To detect movements by WDNNs, the decomposed wavelet coefficients were fed into a WDNN$_{\mu}$ or WDNN$_{\beta}$ with well-trained weights, respectively. Since the interval of each piece of EEG was 1/4 s, WDNN gave one output value per 1/4 s. The outputs of WDNN could not be
wise, it was defined to be ‘+1’. Other-ness, if they were smaller than −0.7 and larger than 0.7, respectively. Outputs of WDNN were assigned values of −1 and +1, if they were between −0.7 and 0.7. If the sum of the reassigned values in 1 s interval was positive, the final output was defined to be ‘+1’. On the contrary, negative sum meant no movement was detected in that 1 s interval if the final output was ‘+1’; on the contrary, output was defined to be ‘−1’. Movement was detected in a 1-s interval if the final output was ‘+1’; on the contrary, negative sum meant no movement was detected in that 1 s interval. When the detection result matched the EMG result, it was a successful detection.

Two training strategies, i.e., single-section training (fixed weights) and re-training (updated weights) at regular intervals, were adopted to investigate the effects of re-training of neural network on success rate. In fixed-weight training strategy, only one out of 42 thumb movement EEG data were assigned value of 0 if they were between −0.7 and 0.7. If the sum of the reassigned values in 1 s interval was positive, the final output was defined to be ‘+1’. Otherwise, it was defined to be ‘−1’. Movement was detected in a 1-s interval if the final output was ‘+1’; on the contrary, negative sum meant no movement was detected in that 1 s interval. When the detection result matched the EMG result, it was a successful detection.

In summary, only one detection result was calculated per second for all the methods mentioned above, and the calculation was based on the information in the previous 5 s period. The real-time processor produced a series of ‘positive’ and ‘negative’ judgments by different criteria of these methods. But the ‘true’ or ‘false’ had to be determined after experiments by comparing these results with EMG recordings. A ‘movement interval (T_m)’ of 5 s length was defined by the peak of EMG envelope (shown in Fig. 6c), and the beginning and stopping times were rounded to the nearest integers. If the algorithm produced at least one positive judgment within T_m, the final judgment was a true positive (TP), otherwise it was a false negative (FN). If the algorithm produced positive judgments outside T_m, false positives (FP) were counted. Negative judgments outside T_m were counted as true negatives (TN). The success rate was defined as the ratio of the number of successful detections to the number of actual movements. In other words, it was identical to TP.

### 2.4. Statistical analyses

In total, 12 methods were used for movement detection, i.e., WD_F, WD_B, STFT_F, STFT_B (WD_F, WD_B, STFT_F, STFT_B), WDNN_F, WDNN_B, WDNN_F, WDNN_B, FTNN_F, FTNN_U, FTNN_F, FTNN_U, FTNN_F, FTNN_U. Twelve success rates were computed for each subject. The significance of differences in success rates among these methods was tested by analysis of variance (ANOVA) and Tukey’s honestly significant difference (HSD) post hoc test. The latter was used to verify which methods had honestly different success rates from others (Wayne, 1999).

Though TP was the simplest and straightforward index to determine which method was better, other indices (FN, FP and TN) were also important. Considering that TP might not be sufficient to assess the performance of each method mentioned above, ROC (Receiver Operating
Characteristic analysis, which considered not only TP but also FP, was introduced to compare the performances of these methods. The performances of all the methods were compared by calculating the area under ROC curves, which were plotted by FP versus TP under different choices of thresholds. A good classifier should show high TP and low FP, resulting in a larger area under the ROC curve.

3. Results

3.1. $\mu$ and $\beta$ wave extraction by wavelet decomposition

Fig. 3 shows the averages of the patterns derived from WD and STFT in subject S8 as an example of WD and STFT processing results. Fig. 3(a) is the 5 s EEG after Laplacian operation, and the corresponding EMG envelope in Fig. 3(b) represents one right thumb movement at the 3rd s. The solid lines shown in Fig. 3(c and d) are the patterns of STFT $\mu$ and STFT $\beta$, and the dashed lines are two patterns taken randomly at the time of no movement. STFT $\mu$ reached 30% suppression at 2.5 seconds, and STFT $\beta$ reached 40% rebound at 4 seconds. The solid lines shown in Fig. 3(e and f) are the patterns of WD $\mu$ and WD $\beta$, and the dashed lines are two patterns taken within the time period of no movement. As mentioned before, every point of the WD pattern was the mean-squared value of WD coefficients in the relevant frequency band ($\mu$ or $\beta$ wave). WD $\mu$ and WD $\beta$ patterns were similar to STFT $\mu$ and STFT $\beta$ patterns, respectively.

3.2. Movement detection by WD $\mu$ and WD $\beta$

Fig. 4(a–d) shows an example of how movement detection was determined. Fig. 4(a and b) are patterns of WD $\mu$ and Fig. 4(c and d) are of WD $\beta$. The horizontal dashed lines represent different thresholds. For Fig. 4(a), the detection was successful, i.e., there was a thumb movement, for all chosen thresholds because the minimum value between the 2nd and 4th s was smaller than the smallest threshold. In Fig. 4(b), the detection was unsuccessful because the minimum value was outside the range between the 2nd and 4th seconds. In Fig. 4(c), the detection result was successful for all chosen thresholds because the maximum

---

**Fig. 3.** EEG, EMG envelope, and averaged patterns of STFT $\mu$, STFT $\beta$, WD $\mu$, and WD $\beta$ of subject S8. (a) One of 5 s EEG data after Laplacian operation. The cue was given at 1 s. (b) The corresponding EMG envelope of right thumb extensor. Movement onset was at 3 s. (c, d) Show the averaged STFT $\mu$ and STFT $\beta$ patterns with thumb movements (solid lines), respectively, compared with STFT $\mu$ and STFT $\beta$ patterns without thumb movements (dashed lines). (e, f) Show the averaged WD $\mu$ and WD $\beta$ patterns with thumb movements (solid lines), respectively, compared with randomly selected WD $\mu$ and WD $\beta$ patterns without thumb movements (dashed lines).
value was between the 3rd and 5th s. In Fig. 4(d), the detection result was a failure because the maximum value was outside the range between the 3rd and 5th s.

3.3. Movement detection by STFT

Fig. 4(e–h) shows typical results of movement detection in STFT\(_\mu\) and STFT\(_\beta\) patterns of the same subject as in Fig. 4(a–d). Fig. 4(e, f) are patterns of STFT\(_\mu\) and Fig. 4(g and h) are patterns of STFT\(_\beta\). The horizontal dashed lines represent different thresholds. In Fig. 4(e), the minimum value (the marked point) of the pattern was between the 2nd and 4th s and smaller than \(-30\%\), so the detection was successful. In Fig. 4(f), the detection result of this pattern was unsuccessful because the minimum value (the marked point) was not between the 2nd and 4th s. In Fig. 4(g), the maximum of the pattern value existed between 3rd and 5th s and the detection was successful. In Fig. 4(h), the detection result was a failure because the maximum value was outside the period between 3rd and 5th s.
3.4. Success rates versus thresholds

The success rates of STFT\textsubscript{$\mu$}, STFT\textsubscript{$\beta$}, WD\textsubscript{$\mu$} and WD\textsubscript{$\beta$} with constant thresholds, ranging from 42.86 to 66.67\%, are shown in Table 1 (column 1–4). As mentioned before, the success rates of movement detection by the thresholding methods were dependent on the chosen thresholds. Fig. 5 shows the results of sensitivity and specificity versus different thresholds in subject S8, where Fig. 5(a–d) are results of STFT\textsubscript{$\mu$} and STFT\textsubscript{$\beta$}, WD\textsubscript{$\mu$}, WD\textsubscript{$\beta$}, respectively. The optimized thresholds were $-20\%$, $25\%$, $5\%$, and $4\%$, respectively, and the success rates (sensitivities) for this subject under the optimal thresholds were 66.67\%, 78.57\%, 73.81\%, and 61.90\%, respectively. The success rates of these four methods with optimized thresholds in all subjects are shown in Table 1 (column 5–8). S10 had the largest improvement in success rate of WD\textsubscript{$\beta$} for 26.19\%, S1 had the largest improvement in success rate of WD\textsubscript{$\beta$} for 28.57\%, S5, S8 and S10 all had improvement of STFT\textsubscript{$\mu$} for 16.67\%, and S8 had improvement of STFT\textsubscript{$\beta$} for 21.43\%.

3.5. Movement detection by WDNN

Fig. 6 presents an example result of movement detection in one round by WDNN\textsubscript{$\mu$}. Fig. 6(a) shows the original outputs of WDNN\textsubscript{$\mu$}. The dashed lines are the thresholds of +0.7 and −0.7. Fig. 6(b) shows the results after re-assignment. Fig. 6(c) shows the results after the logical operation. Fig. 6(d) shows the final results. The vertical dashed lines mark the interval of 2 s before and after the movement onset determined by the EMG envelope. Both the movements were detected correctly. The success rates of WDNN\textsubscript{F}, WDNN\textsubscript{U}, WDNN\textsubscript{$\mu$} and WDNN\textsubscript{$\beta$} in all subjects are shown in the 1–4 columns of Table 2.

3.6. Movement detection by FTNN

Fig. 7 shows the results of movement detection in one trial by FTNN\textsubscript{$\mu$}. Fig. 7(a) depicts the original outputs of FTNN\textsubscript{$\mu$}, and the dashed line is the threshold of +0.7. After re-assignment, the binary outputs are shown in Fig. 7(b). The final results (Fig. 7(c)) show that only the first movement was detected correctly. The success rates of FTNN\textsubscript{F}, FTNN\textsubscript{U}, FTNN\textsubscript{$\mu$} and FTNN\textsubscript{$\beta$} in all subjects are shown in the 5–8 columns of Table 2. The means and standard deviations of success rates of all methods are also plotted in Fig. 8.

3.7. Success rates and statistical analyses

We used one-way ANOVA to test whether the differences among the success rates of 12 methods were significant, and Tukey’s HSD post hoc test to evaluate among which pairs the differences between methods were honestly signifi-

Table 1
The success rates (%) of WD\textsubscript{$\mu$}, WD\textsubscript{$\beta$}, STFT\textsubscript{$\mu$} and STFT\textsubscript{$\beta$} with constant and optimized (with * superscript) thresholds

<table>
<thead>
<tr>
<th></th>
<th>WD\textsubscript{$\mu$}</th>
<th>WD\textsubscript{$\beta$}</th>
<th>STFT\textsubscript{$\mu$}</th>
<th>STFT\textsubscript{$\beta$}</th>
<th>WD\textsubscript{$\mu$}*</th>
<th>WD\textsubscript{$\beta$}*</th>
<th>STFT\textsubscript{$\mu$}*</th>
<th>STFT\textsubscript{$\beta$}*</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>57.14</td>
<td>47.62</td>
<td>57.14</td>
<td>54.76</td>
<td>71.43</td>
<td>76.19</td>
<td>61.90</td>
<td>59.52</td>
</tr>
<tr>
<td>S2</td>
<td>47.62</td>
<td>42.86</td>
<td>52.38</td>
<td>45.24</td>
<td>61.90</td>
<td>57.14</td>
<td>52.38</td>
<td>54.76</td>
</tr>
<tr>
<td>S3</td>
<td>52.38</td>
<td>50.00</td>
<td>42.86</td>
<td>52.38</td>
<td>71.43</td>
<td>69.05</td>
<td>57.14</td>
<td>64.29</td>
</tr>
<tr>
<td>S4</td>
<td>61.90</td>
<td>54.76</td>
<td>47.62</td>
<td>54.76</td>
<td>73.81</td>
<td>76.19</td>
<td>61.90</td>
<td>61.90</td>
</tr>
<tr>
<td>S5</td>
<td>57.14</td>
<td>57.14</td>
<td>45.24</td>
<td>61.90</td>
<td>78.57</td>
<td>83.33</td>
<td>61.90</td>
<td>69.05</td>
</tr>
<tr>
<td>S6</td>
<td>47.62</td>
<td>50.00</td>
<td>50.00</td>
<td>52.38</td>
<td>71.43</td>
<td>73.81</td>
<td>64.29</td>
<td>64.29</td>
</tr>
<tr>
<td>S7</td>
<td>54.76</td>
<td>50.00</td>
<td>54.76</td>
<td>64.29</td>
<td>71.43</td>
<td>69.05</td>
<td>66.67</td>
<td>69.05</td>
</tr>
<tr>
<td>S8</td>
<td>61.90</td>
<td>52.38</td>
<td>50.00</td>
<td>57.14</td>
<td>73.81</td>
<td>61.90</td>
<td>66.67</td>
<td>78.57</td>
</tr>
<tr>
<td>S9</td>
<td>66.67</td>
<td>52.38</td>
<td>59.52</td>
<td>50.00</td>
<td>76.19</td>
<td>73.81</td>
<td>69.05</td>
<td>59.52</td>
</tr>
<tr>
<td>S10</td>
<td>47.62</td>
<td>50.00</td>
<td>54.76</td>
<td>40.48</td>
<td>73.81</td>
<td>66.67</td>
<td>71.43</td>
<td>54.76</td>
</tr>
<tr>
<td><strong>Means ± SD</strong></td>
<td><strong>55.46 ± 6.74</strong></td>
<td><strong>50.71 ± 3.89</strong></td>
<td><strong>51.43 ± 5.29</strong></td>
<td><strong>53.33 ± 7.12</strong></td>
<td><strong>72.38 ± 4.38</strong></td>
<td><strong>70.71 ± 7.62</strong></td>
<td><strong>63.33 ± 5.64</strong></td>
<td><strong>63.57 ± 7.28</strong></td>
</tr>
</tbody>
</table>
As mentioned before, there were 12 methods and 10 subjects. The differences of results by methods listed in column 1–4 of Tables 1 (in which heuristic thresholds were used in WDNN, WDNN, STFT, and STFT) and 2 were significant, $F(9, 108) = 7.15, p < 0.05$. The results of post hoc test were, for μ wave; (1) WDNN > STFT, (2) Table 2

The success rates (%) of WDNN, WDNN, WDNN, STFT, FTNN, FTNN, FTNN, and FTNN

<table>
<thead>
<tr>
<th></th>
<th>WDNN</th>
<th>WDNN</th>
<th>WDNN</th>
<th>WDNN</th>
<th>FTNN</th>
<th>FTNN</th>
<th>FTNN</th>
<th>FTNN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>U</td>
<td>F</td>
<td>U</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S1</td>
<td>73.81</td>
<td>88.10</td>
<td>76.19</td>
<td>92.86</td>
<td>66.67</td>
<td>83.33</td>
<td>50.00</td>
<td>64.29</td>
</tr>
<tr>
<td>S2</td>
<td>57.14</td>
<td>85.71</td>
<td>69.05</td>
<td>83.33</td>
<td>73.81</td>
<td>54.76</td>
<td>73.81</td>
<td>66.67</td>
</tr>
<tr>
<td>S3</td>
<td>66.67</td>
<td>97.62</td>
<td>76.19</td>
<td>83.33</td>
<td>73.81</td>
<td>78.57</td>
<td>73.81</td>
<td>71.43</td>
</tr>
<tr>
<td>S4</td>
<td>73.81</td>
<td>83.33</td>
<td>80.95</td>
<td>88.10</td>
<td>69.05</td>
<td>78.57</td>
<td>54.76</td>
<td>78.57</td>
</tr>
<tr>
<td>S5</td>
<td>73.81</td>
<td>83.33</td>
<td>90.48</td>
<td>83.33</td>
<td>73.81</td>
<td>78.57</td>
<td>57.14</td>
<td>57.14</td>
</tr>
<tr>
<td>S6</td>
<td>52.38</td>
<td>88.10</td>
<td>78.57</td>
<td>83.33</td>
<td>73.81</td>
<td>78.57</td>
<td>57.14</td>
<td>57.14</td>
</tr>
<tr>
<td>S7</td>
<td>61.90</td>
<td>92.86</td>
<td>73.81</td>
<td>83.33</td>
<td>73.81</td>
<td>78.57</td>
<td>57.14</td>
<td>57.14</td>
</tr>
<tr>
<td>S8</td>
<td>69.05</td>
<td>83.33</td>
<td>90.48</td>
<td>80.95</td>
<td>73.81</td>
<td>71.43</td>
<td>66.67</td>
<td>66.67</td>
</tr>
<tr>
<td>S9</td>
<td>66.67</td>
<td>88.10</td>
<td>83.33</td>
<td>83.33</td>
<td>73.81</td>
<td>71.43</td>
<td>66.67</td>
<td>66.67</td>
</tr>
<tr>
<td>S10</td>
<td>59.52</td>
<td>85.71</td>
<td>76.19</td>
<td>83.33</td>
<td>69.05</td>
<td>64.29</td>
<td>64.29</td>
<td>64.29</td>
</tr>
</tbody>
</table>

Means ± SD 65.48 ± 7.55 87.14 ± 5.29 76.19 ± 7.61 86.92 ± 5.30 70.24 ± 5.05 69.29 ± 8.51 63.81 ± 10.64 67.38 ± 10.59

Fig. 6. Movement detection by WDNN. (a) The original outputs of WDNN. Time interval between adjacent points was 1/4 s and the horizontal dashed lines indicate the tolerance thresholds of +0.7 and −0.7. (b) WDNN outputs after thresholding. (c) Logical judgment in each second. If the sum of outputs in each second was larger than 0, the result was +1. Otherwise, the result was −1. (d) Two movement intervals as determined by EMG envelope. If there was at least one ‘+1’ in the movement interval, the movement was defined as being detected successfully.
WDNN_μ > WD_β and STFT_μ, (3) FTNN_μ for wave, (1) WDNN_β > WD_β and STFT_β, (2) FTNN_β > WD_β. In summary, the algorithms with neural network showed higher success rates.

The differences of results by methods listed in column 5–8 of Table 1 (in which optimized thresholds were used in WD_μ, WD_β, STFT_μ and STFT_β) and 2 were significant, F(9,108) = 5.35, p < 0.05. Then the results of post hoc test were, (1) WDNN_μ > WDNN_β and STFT_μ and (2) WDNN_β > STFT_β.

We also performed ROC analyses. The solid lines of Fig. 8(a) show the ROC curves of WDNN_μ with fixed weights for the 10 subjects. Each triangle indicates FP vs. TP under each threshold, and the dashed line indicates the reference. The thresholds were chosen from 0.2 to 0.9, the increment was 0.07, and the averaged area under ROC curves was 0.76. Fig. 8(b) show the ROC curves of STFT_μ with different thresholds for the 10 subjects. The thresholds were chosen from −0.5 to −0.1, the increment was 0.04, and the averaged area under ROC curves was 0.65. In summary, the averaged areas under ROC curves of the methods in our study were 0.67 for STFT_μ, 0.67 for STFT_β, 0.68 for WD_μ, 0.66 for WD_β, 0.66 for FTNN_μ, 0.72 for FTNN_β, 0.67 for FTNN_μ, 0.71 for FTNN_β, 0.76 for WDNN_μ, 0.79 for WDNN_β, 0.72 for WDNN_μ and 0.81 for WDNN_β. In summary, the results of ROC analyses were similar to those of ANOVA.

4. Discussion

4.1. Design of 2 s delay from the cue to the movement onset

In this study we asked the subjects to swing their thumbs at about 2 s after the cue. The purpose of the designed delay was to keep the movement preparation for at least 2 s, so that the WDNN could detect the attempts of movements. It was not necessary to keep the length of delay between the cue and the movement very precise, because the task was motivated by the subject himself and the sound of cue was just a prompt to prepare the movement. To prevent the subjects from predicting or memorizing the next cue, the first cue (T_1) was given in the interval of 5–10 s, and the second cue (T_2) was given in 20–25 s randomly.
Right thumb movement was chosen to be the movement task in this study, because the cortical area for the thumb was much larger and closer to the scalp, so that the potentials generated from the area could be recorded more easily than the corresponding potentials provoked by movements of other fingers or limbs, and that was why surface Laplacian operation was applied at C3. The reason for choosing the right thumb was that all the subjects were right-handed, although there was no evidence indicating the right-handed people could provoke EEG changes more easily.

4.2. Effect of the sound cue on motion potential

As mentioned before, we expected that $\mu$ wave should be suppressed in the period from the cue to actual movement. Another research compared the movement intentions of left/right index fingers and foot (Peters et al., 2001). The cue was also given by a short beep. In order to ensure that the suppression of $\mu$ wave was not caused by the cue sound, we also conducted a validating experiment in which the cue sound was given but the subject was asked not to make any movement. Two subjects (S3 and S4) with the highest success rates were selected for this experiment. We conducted 20 trials for each subject and STFT of $\mu$ wave band was computed to determine whether $\mu$ wave was affected by the sound. We found that there were only 3 and 2 trials for S3 and S4, respectively, that $\mu$ wave was affected by sound cues, and all of them occurred in the first five trials. When the subjects had been accustomed to just hearing the sound but making no attempt of movements in the next 15 trials, the $\mu$ wave suppression disappeared. From these results, it was concluded that the cue sound alone was not a major determinant of $\mu$ wave suppression in the main experiments.

4.3. Training of WDNN and FTNN

In most supervised learning procedures of artificial neural networks, the size of training data is usually larger than that of testing data, because a larger training set makes the network recognize more features of the sampled population. The more training data, the better the network resulted, and the more time spent for network learning. In this study, we tried to make the training set smaller and save the training time as much as possible, and to keep the classification results at an acceptable percentage of 70%. We found that the data of single trial movement was enough to train the networks for each subject. The classification algorithm that needs small learning data has a practical significance.

Though several studies with similar methods showed comparable or higher success rates, we found these results were compromised by some difficulties. For example, some algorithms were very sophisticated (Kennedy et al., 2000; Bostanov, 2004), or the features had to be extracted from many channels of EEG (Pfurtscheller et al., 1996; Bernhard et al., 2004; Tkacz et al., 2002), or more than 50% of training data were needed for network training (Haselsteiner and Pfurtscheller, 2000). Sophisticated algorithms and large amount of training data can result in higher detection
rates. Yet, they are difficult to be realized in real-time computation. Domider et al. (2001) analyzed ECG by the neural networks with wavelet decomposition, and the advantages of WDNN were revealed by comparing with the networks without wavelet decomposition. Graumann et al. (2004) proposed wavelet-packet analysis to detect the event-related potentials of four different tasks and a genetic algorithm was introduced to select the best feature combinations of every subject. The results were satisfactory but the algorithms were so complicated that they could only be performed in pseudo-online analysis.

Though the size of training data was much smaller than the testing data, the results of post hoc tests indicated that adding neural networks could improve the success rates. The success rate of adding a neural network was comparable to the rate of the corresponding method with optimized threshold. The advantage of adding a neural network was that only the data of one trial was needed. On the other hand, to obtain an optimal threshold requires computing the success rates of all EEG data under all possible thresholds.

In this study, we trained the networks by back-propagation algorithm. This method tuned the weights by the principle of least mean-squared-error (MSE). Learning procedure was stopped when MSE was reduced to less than 0.1 and the change of MSE to less than 0.05. The training error usually converged within 1500 epochs. Taking one training course as an example, the MSE of the training epochs converged and was smaller than 0.1 at the 782\textsuperscript{nd} epoch when the training procedure was stopped. To prevent the NN from being overtrained, the termination criteria of training should not be too stringent. Therefore, the testing outputs were not exactly equal to the training targets, and the tolerance thresholds (±0.7) shown in Fig. 6(a) and Fig. 7(a) were needed to modify the original outputs of WDNNs and FTNNs.

4.4. The effects of updating weights

Because of the nonstationarity of EEG, retraining procedures were designed to update the weights of WDNNs and FTNNs per 10 movement detections. For WDNN\textsubscript{U}, WDNN\textsubscript{F}, FTNN\textsubscript{U}, and FTNN\textsubscript{F}, we used both fixed and updated weights. It was supposed that the WDNNs and FTNNs with updated weights could adapt to the nonstationary EEG, so the success rates of WDNN\textsubscript{U} and FTNN\textsubscript{U} should be higher than those of WDNN\textsubscript{F} and FTNN\textsubscript{F}. Yet, there was no significant difference between the corresponding pairs by post hoc test. To verify the effect of weight updating on success rates, the results shown in columns 1–4 and 5–8 of Table 2 were tested by ANOVA and post HSD separately. In column 1–4 of Table 2, both the results of WDNN\textsubscript{U} and WDNN\textsubscript{F} were better than WDNN\textsubscript{F}. One the other hand, there was no significant difference for the results in column 5–8 of Table 2. The above analyses indicated that updating the weights of WDNN, but not FTNN, could improve the performance. Even though updating weights was only marginally advantageous, it was conceivable that updating weight would become indispensable for long-term use. For real-time systems, a self-tuning algorithm will be necessary to update weights automatically.

4.5. Comparing our methods to Fisher’s linear discriminant analysis (FLDA)

For comparison, we also performed FLDA, a linear operator that reduced patterns from multi-dimension to one-dimension. According to the algorithms for two-cluster discriminant (Richard and Peter, 1973), a project matrix (denoted as \( w \)) must be calculated to project patterns from high-dimension to one-dimension.

We noticed that more than 60 patterns were required to make the project matrix \( w \) converge. This meant that 30 patterns of \( \mu \) wave with movements and another 30 patterns without movements were needed to derive \( w \). We emphasized again that, in our method, only one pattern was taken as the training input for WDNN. In contrast, FLDA needed at least 30 patterns to obtain the classifier. Nevertheless, the rates of TP for FLDA, as shown in the following, were not superior. The success rates of all subjects by FLDA were 66.7\%, 44.16\%, 75\%, 58.3\%, 83.3\%, 79.16\%, 83.3\%, 58.3\%, 50\%, and 79.16\%, and the mean and standard deviations were 67.73 \( \pm \) 14.32\%. Comparing these means and standard deviations as shown in Tables 1 and 2, we found that the success rates of conventional methods (Table 1) were comparable to or worse than FLDA, while the methods using wavelet decomposition and NN (Table 2) had higher success rates than FLDA.

4.6. States of subjects

All subjects participating in this study had no previous experience with BCI. Thus, the effects of adaptation and learning could be excluded. The purpose of choosing first-time subjects was to ensure that the classification algorithm developed in this study was suitable for users without prior learning. It was well known that rigorous training for a period of time could increase the success rates. For clinical applications, many potential users have mental and emotional disabilities, which may inhibit them from tolerating protracted learning.

In this study, actual movements were chosen to be the tasks. But in the practical application of BCI, the users are the patients with different levels of muscle weakness and the commands have to be generated by thoughts without movements. In future studies, we will test the applicability of the detection algorithms developed in this study to the imaginary movements.

5. Conclusion

In this study, we developed an EEG processing algorithm to identify the movement-related EEG changes for
the first time subjects by using only the data of single thumb movement. The algorithm, based on WDNN, had a better performance than the conventional short-time Fourier transformation method with or without combining a neural network. Its advantage was also revealed that the size of required training data was much smaller than a linear discriminant classifier. We found that re-training of the network at short regular intervals only marginally improved the detection rate. WDNN, by eliminating the need for protracted training with a large amount of data, has a great potential for clinical and practical applications.

References


