Features Extraction from NIRS Data using Extreme Decomposition

Yan Jiang
Department of Software, Shenyang University of Technology, Shenyang, China
Email: panjiang1@163.com

Xin Jin, Xin Li
Shenyang HeXing Testing Equipment Co., Ltd., China
Department of information science and technology, Shenyang University of Technology, Shenyang, China
Email: sealovesoft@163.com; 644876180@qq.com

***Abstract*—the main aim of BCI builds a communicating bridge between brain and peripheral devices. NIRS is dependent on changes of blood flow, as it measures oxygenated and deoxygenated hemoglobin's in the super-facial layers of the human cortex. We are able to detect HbO and HbR of imaged movement and movement on the surface of the brain with NIRS. If we want to achieve the control of external devices with HbO and HbR, the change of these data must be analyzed, and be extracted. In this paper, we present a new method to achieve in the analysis of the data of HbO and HbR, realize the removal of high frequency and achieve preliminary extraction the characteristics of the data. Secondary analysis of the extracted feature points could reduce the number of feature points. Designing a new compensated interpolation algorithm achieve completely new feature points to replace the original feature points to represent the data .The interpolated data curves response the change of original data, and realize the removal of high frequency to smooth the output curve.**

***Index Terms*—interpolation algorithm, time series analysis, near-infrared spectroscopy, oxygenated hemoglobin**

I. INTRODUCTION

The main aim of Brain Computer Interface (BCI) builds a communicating bridge between brain and peripheral devices [1]-[3]. One of the essential conditions to want better development and popularizing application of the BCI system is finding a kind of signal which could reflect different mental state of brain and could be extracted and classified in real time or short term. Electroencephalogram (EEG) is a non-invasive technology of brain activity, with high resolution, reliability, the amount of information, visual images of features, so it becomes one of the best choices for BCI [4-6]. The Nuclear Information and Resource Service(NIRS) is dependent on changes of blood flow, as it measures oxygenated and deoxygenated hemoglobin's ([HbO] and [HbR]) in the super-facial layers of the human cortex. We are able to detect HbO and HbR of imaged movement and movement on the surface of the brain with NIRS. We could judge simple task they want to do with these data.

Detected through equipment brain oxygen data change frequency is higher. The difference is mainly reflected in the continuous movement, blood oxygen concentration curve may show large fluctuations, such a change is a manifestation of the oxygen concentration data measured by brain. But such a signal is not suitable as a direct control of the output signal. We hope that the change trend of the data is represented by a smooth curve through signal processing methods for the next processing data.

In the field of signal processing, the empirical mode decomposition (EMD) [7] - [10] has been recognized as the driving signal decomposition method of effective data, and has been widely applied to multiscale signal analysis. EMD method is to remove the average of superior envelope and inferior envelope in the source data, and it will inevitably affect the true value to the original data [11-16]. In the new algorithm, we ensure the effective characteristics of the data at the same time, and use a smaller number of feature points to complete the description of the source data. Through calculation we can get the new data replacing feature points. These data not only describe the characteristics of the source data, but also facilitate interpolation algorithm to interpolate. The source data is described as a smooth curve by interpolated data we need. The curves can describe the basic characteristics of the source data, and output stable data.

II. METHOD

We give a general flowchart of the Features Extraction from NIRS (FEFN) algorithm showing in Fig.1. The maxima and the minima are extracted from input time series by given the same time slice. We can get one maxima and one minima in one time slice. Before getting extreme, we remove noise data first of all. All of the maxima and the minima form maxima set and minima set. They are processed by an amalgamating algorithm and get rid of some redundant data from the maxima set and the minima set. Three cases of the adjacent nodes are processed with different function. Details will be covered in the subsequence sections.

doi:10.4304/jcp.9.7.1671-1677

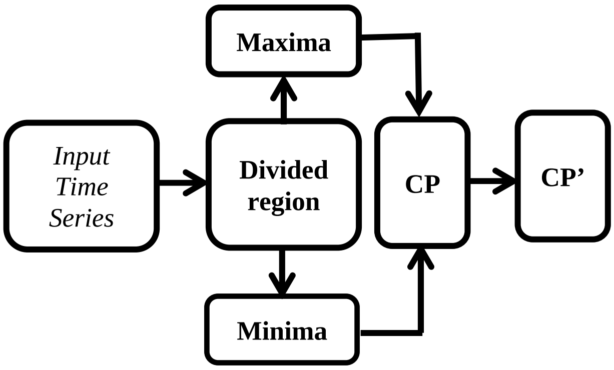


Figure 1. General Flowchart of the algorithm

A. FEFN Basics

Let {x[n],n=1,…,N*M} be an N*M time series which we will refer to as x[n].T is the time series of one whole experiment, Let { Ti, i=1,…,M} be an M-element time series which is the M equal segments. The value of M can affect the result of an operation, so we will discusses in detail in the subsequence sections. There are N elements in one segment. Let{(L[i],x[L[i]]),i=1,…,M} be an array with M elements which is maximum value in i segment. Let{(S[i],x[S[i]]),i=1,…,M} be an array with M elements which is minimum value in i segment.

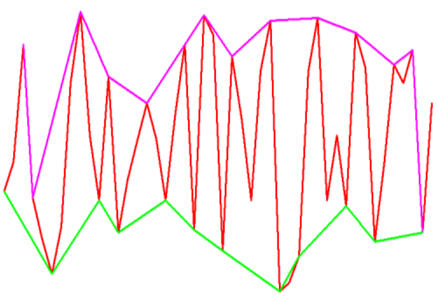
Maxima Envelope: The maxima envelope of a time series is Line segment passes through all of its maxima.(see the purpose one in Fig. 2)

Minima Envelope: The minima envelope of a time series is Line segment passes through all of its minima. (see the green one in Fig. 1)

M'value: The value of M is can affect the result of the maxima Envelope and the minima envelope. Polylines with the different M's value are in Fig.2

Cross Polyline: Starting points and end points of the cross polyline come from the set of maxima Envelope and minima Envelope alternately. Firstly we take out a node from maxima envelope as the starting node, and take out a node from minima envelope as the end node, and drawing a straight line joining these two points. The end node from minima envelope becomes a start node, and end node is the second node in maxima envelope then. These nodes from maxima envelope and minima Envelope alternately are taken out and are joined together. The implementation process of joining lines is two steps. One is MaxtoMin line, the other is MintoMax line. Cross Polyline (CP) is the set of MaxtoMin and MintoMax. Tcp is the number of every segment as in (1), and fcp is reciprocal of Tcp as in (2).

(a)



(b)

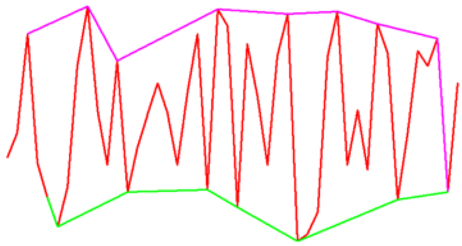


Figure 2. Comparison between different value of M. For both (a) and (b), the red curve is original time series, and its superior and inferior envelopes are depicted as pink and green dashed curves. The value of M in (a) is larger than in(b).

$$\mathrm{Tcp} = \mathrm{N/M} \tag{1}$$

$$\mathrm{fcp} = \mathrm{M/N} \tag{2}$$

Theoretically, we divide time series into M equal divisions. All data from CP curves are the maximum and minimum values which were selected from M parts. These data could present features of each part's data. Each part is an independent part of data processing, and the interpolation operation will be implemented in each part at last. We could not retain all data in CP curves in solving practical problems. We need analysis data in CP curves, and eliminate data which could not present features of original data information. After the modification of CP' curves could keep the primary data features using few data.

In short, the objective of algorithm is removing the noise data which disturbs data presenting, and forming CP curves which we defined after divided time series. With the relevant algorithms, some useful data was reserved, and useless data was removal. The improved CP' curves instead of intrinsic curves

B. CP Process

The main process of FEFN forms CP curves and CP' curves from time series data. Firstly, we divide time series into M region, and form CP curves. By analyzing the data of CP, We realize data removing and merging.

M's value: the algorithm idea is choosing the maximum and minimum in every region, and presenting data character with these data. There are the maximum and minimum in every region, and connect these data to generate CP curves. Of course, there are some anomaly data in those data. The selection of extreme is greatly affected by the number of anomaly data. The method of removing anomaly data is removing several maximum and minimum, and the more accurate method is removing the middle node in three consecutive nodes which crossing angle is small. Y_{i-1}, Y_i, Y_{i+1} is value of three consecutive nodes. Because of the same sample time, there are the same the abscissa value. The value is t0. We could calculate the angle α with Y_{i-1}, Y_i, Y_{i+1} and t_0. If the angle α is smaller than a given threshold, Y_i is considered as an anomaly data which we call noise data and will be removed. An approximate method used to find out noise data could be chosen .the value of $\Delta Y_i/\Delta Y_{i+1}$or $\Delta Y_{i+1}/\Delta Y_i$ is larger than a given threshold, Y_i is considered as an anomaly data as in (3) and (4).

$$\Delta Y_i=|\Delta Y_i-\Delta Y_{i-1}| \tag{3}$$

$$\Delta Y_{i+1}=|\Delta Y_{i+1}-\Delta Y_i| \tag{4}$$

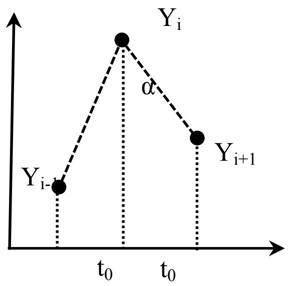


Figure 3. The positional relationship of the three consecutive points

We begin to define the value of M after removing noise data. The principle of selecting M's value is satisfied with such conditions.

1. For the convenience of the interpolation operation, extreme value distribution in each region should be comparatively distributed with the mean distance. The distance between two consecutive maxima or minima should be close to Tcp. If we are able to determine the M'value with the basis of the analysis and comparison of previous data, its calculation efficiency is higher.

2. In the process of determining the value of M, it should try to ensure the maximum and minimum values appear alternately.

Because of data's own characteristics, it is very difficult to meet the above requirements. If the value of M is the close to or meets the need of requirement, there is a high efficiency in the back of the computing process. The value of M should be greater than 1/2 of the cycle of changes in the value of the brain signal in general case. If the value of M is not larger than the cycle of the signal changes in the brain of 1/2, some data may be lost.

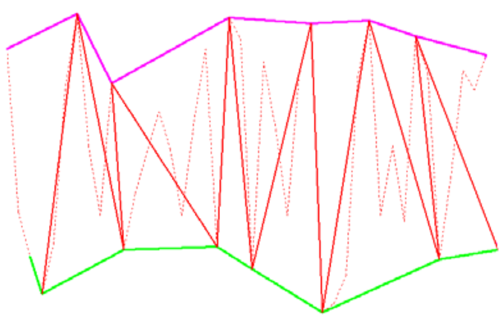


Figure 4. CP line

C. Interpolation&CP' Process

After determining the value of M, We divide time series into M segments. Maximum value in each segment will be put into L[i], and minimum value will be put into S[i]. Variable iis the NO. of segment, and the values of i are chosen between 1 and M. under the perfect conditions, Connection of the data in the L and S may be formed CP.

In the three points, if there is a maximum near the point of maximum value, the subsequent far greater than a maximum value, we can remove the intermediate value. Reducing the data in Lensure that this data does not exist in the L. The S array data is processed in the same way. In the CP of the maximum value and minimum value curve, we remove the peak around the protrusion, and the extreme point of attachment is retained.

After getting a simplified L and S array, we can select the value in turn. In the L and S data, the points will be connected into CP which meets the conditions. The method is that the maximum value point appears in the individual changes will be retained, and others points will be removed.

After getting the line which is simplified, we will form smooth curve with interpolation method. In the common interpolation algorithm, we can construct interpolation points according to the points we provided, and then formed a relatively smooth curve. But the interpolation curve can't meet our requirements. Because the interpolation data cannot guarantee form in accordance with the original discount trend, and it does not guarantee that the extreme we give is appeared in the curve peaks and troughs, as shown in Fig 5. Therefore, we need to improve the interpolation algorithm.

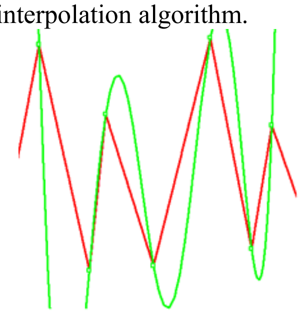


Figure 5. Interpolated curve

In the process of traditional interpolation algorithm, when the number of interpolation points and operations are too many, so the insertion value is uncertainty. It is said that although we can get the given value, there will be a great deviation between "fact" and the value in the vicinity, this kind of phenomenon is also called "the Runge phenomenon". The solution is piecewise interpolation polynomial in low degree, to reduce the interpolation point. According to the features of the data we need interpolation, we need to be closed interval segmentation, to make the relatively small number of interpolation points in each cell and then to interpolate between each cell.

After determining the number of interpolation points, according to the experiment data, the effect between the two interpolation points in about 5 is relatively satisfactory, but more than 10 will be more serious, so if the interpolation points are not too many, every interval with 5 interpolations can get good effect. But to ensure that each contact more closely, make the interpolated curve more smooth, we should let the interval contains several identical interpolation point in the interval distribution .The number of selected the interpolation points for N, The interpolation points for each interval contains m.

Our approach is in accordance with the sliding window, the window contains the M points, each window sliding down the instrument point. This two time interpolation calculation of a point M-1 is repeated last used interpolation point. The algorithm only needs to compute at point N-m+1, the end of the computation. The smoothness of the m value determination will affect the curve.

In the interpolation process, we can obtain a smooth curve. However, the curve is not necessarily according to data changes. Drawing line graph we can use three points to finish a crest or trough description. But when we use the same three points to describe smooth curves, it is unable to draw the desired curve. Therefore we need to interpolate some supplemental point, which can guide the interpolation algorithm meet the requirements during the interpolation process. Derivation formulas are as in (5), (6) and (7).

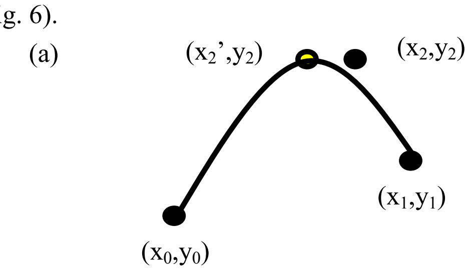
$$\begin{cases} y_0 = a * (x_0 - a_0)^2 + y_2 \\ y_1 = a * (x_1 - a_0)^2 + y_2 \end{cases} \qquad (5)$$

$$a0 = \frac{x_1 - x_0}{\pm \sqrt[2]{\frac{y_0 - y_2}{y_1 - y_2}} - 1} + x_1 \qquad (6)$$

$$a = \frac{y_0 - y_2}{\sqrt[2]{x_0 - a_0}} \qquad (7)$$

In order to obtain a smooth curve segments, we can define each segment as a parabola. Supposing a parabola through two points (x_0,y_0) and (x_1,y_1), we cannot determine it pass through point (x_2,y_2).We can use the similar point (x'_2,y_2) to replace point (x_2,y_2),and the abscissa of (x'_2,y_2) is the same as abscissa of vertices which could be calculated. According to the longitudinal coordinate the coordinates of two points and a point, we could get the coordinate of new vertices.

In the process of solving equations, we can get expressions for a and a_0. We analyze that the auxiliary points have the effects on the procession of interpolation S. The five points we selected are assumed as the feature points. If we want the interpolated curve to pass through these five points, and these five points are continuous peaks of the curve, the curve will pass through the intermediate points of the connection of each two adjacent points. The intermediate points will be added into original peaks to calculate auxiliary points' position. We use the auxiliary points to replace original the characteristic points to realize interpolation calculation. We find six intermediate points as characteristic points in the connectivity, and figure out 12 new points as the auxiliary points to complete the interpolation calculation (Fig. 6).



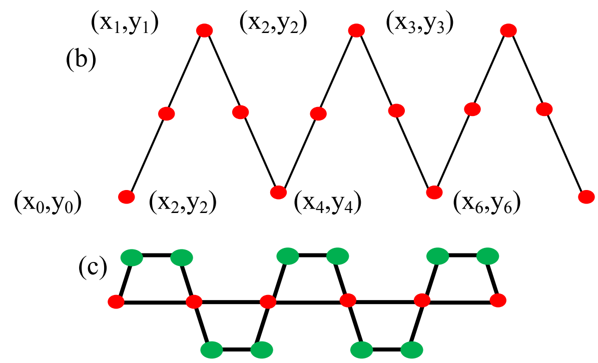


Figure 6. (a) is using (x_2',y_2) to replace (x_2,y_2).Calculate new points with (x_0,y_0), (x_1,y_1) and (x_2',y_2).For both (b) and (c), the redpoints are characteristic points, and the green points are auxiliary points.

Because the expression needs to prescribe in the calculation process, the a_0 may appear two values in the process of calculation. We can judge the two values which is consistent with our design curves. With analysis of the data, a_0 is the abscissa of parabola apex, and it is the value of the x_2'. a_0 value should be between x_0 and x_1. In this way we can choose one of the two values as a curve equation of value. However, in different circumstances, every time we need to compare the relationship of three values to get the curve of a_0 value. From the first set of data, we can see that there are large fluctuations in the original data curve. By analyzing the data, there is the difference result in brain surface oxygen concentration. In the continuous movement of finger case, there is larger fluctuation in the brain blood oxygen concentration. We hope that with the new method, the output of oxygen concentration curve is smooth.

III. Result

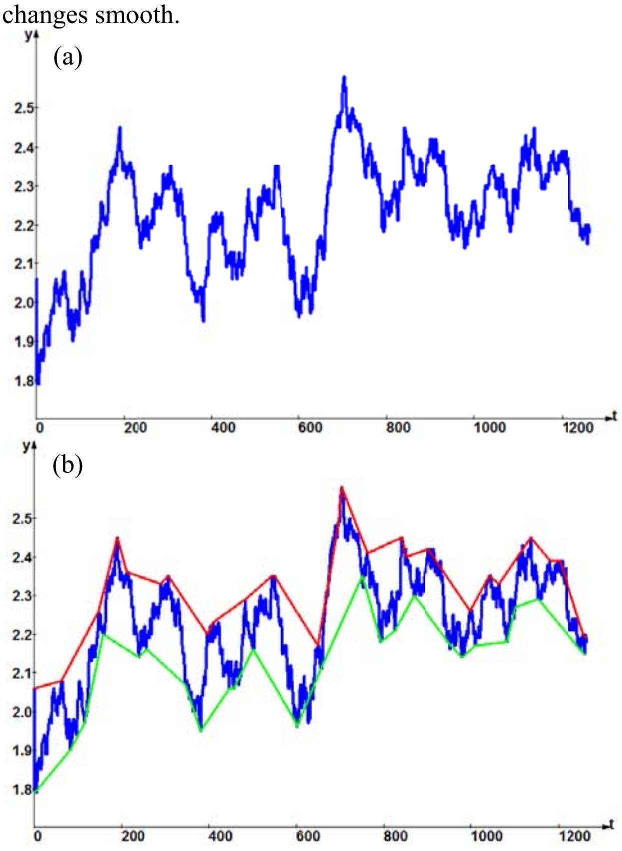
Ten experiments which have been popularly used in BCI are designed and implemented to revalue and validate the proposed ENDS. During these tasks simultaneous measurements of NIRS were performed. The NIRS-System was equipped with 37 optical fibers (15 sources with wavelengths of 850 nm and 760 nm, 22 detectors convolving to 37 measurement channels). Frontal, motor and parietal areas of the head were covered. The sampling frequency was fNIRS= 10 Hz. We choose 2 representative experiments with large fluctuations as the object we analysis.

The first study is based on non-continuous fingers motions on one side, and 20 sets of data of the sample divided into two groups which were selected as training set. The subjects were seated in a comfortable chair with armrests and were instructed to relax their arms. The experiment has 5 parts. Before the experiment, the subjects have 10 seconds' rest. There are 20 seconds' fingers motion and 30 seconds' rest in every part. The subjects were finger movement on one side in each group. This enables us to situate the NIRS channel positions according to the standard 10 – 20 system.

Fig. 7(a) is a graph of the initial data. There are two large fluctuations between the second and third peaks appeared. We start to process this data set. In the first step we should determine the value of M. According to the preliminary analysis, the sampling frequency is 10Hz, and the period of action is 20 seconds, so we determine 50 data in every group. With 50 sample data in one group, we find out the maximum value of each stored in the L array and the minimum value of a stored into a S array. Then the values of all L[i] and S[i] are connected to form performance curve. one is maximum curve, and the other is minimum curve.

According to the rules defined above, we should find the points which need to deal with "bulge" in the two curves, and remove the "bulge" point, as shown in Fig. 7(b).Based on the two curves which has removed the "bulge" points, according to the time sequence in the L array and S array, we find the maximum value and the minimum value of the single direction, and the maximum and minimum values are connected to form CP in turn.

Through we can see that the basic characteristics of the original curve data from the CP line, in order to get the interpolation curve which can after a maximum value curve and the minimum value, we need to insert the calculated value in CP. According to the three continuous points CP1, CP2, CP3 in CP line, we could calculate the four values as the first trapezoid used to realize the interpolation. We use the CP_{i-1}, CP_i, CP_{i+1} to calculate f4 values as the interpolation data, as shown in Fig. 7(c).Fig. 7 (d) depicts the interpolation curve. The comparative result between original data and the interpolation data is in Fig. 7(e). According to the interpolation algorithm, the data reflect the change trend of the distance, and the data changes smooth.



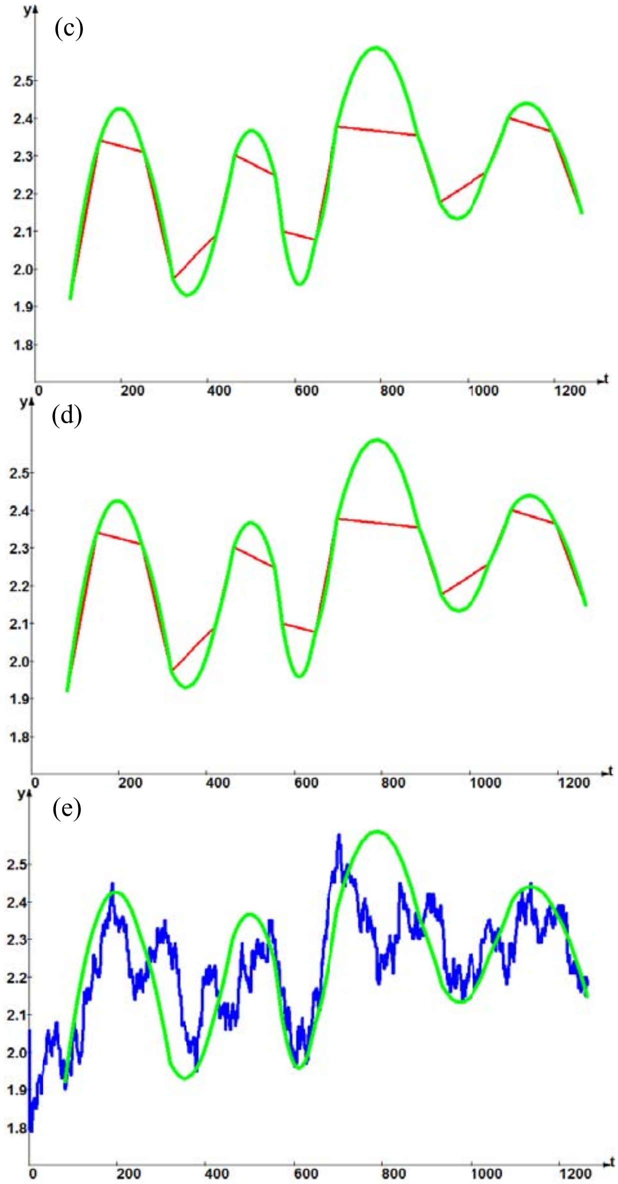


Figure 7. The processing of the first group of data.(a) the blue curve is original time series.(b)the superior and inferior envelopes are depicted as red and green dashed curves.(c)the cp line is depicted as blue dashed curve. Calculated nodes are connected into the red dashed curve.(d)Interpolation curve is depicted as green curve.(e)Composition between original data and interpolation curve.

With the curve equation, we can calculate coordinates of two points on both sides of the top point. With these two points instead of vertex, it can depict the smooth curve vertex with interpolation algorithm. That is to say, we draw the peaks and valleys with four points of trapezoidal, and realize interpolation with trapezoidal. Curve after interpolation doesn't exceed the range between the maximum value and the minimum value of the original data. Finally we use CP ' to instead CP. The interpolation based on CP' of data eventually forms a smooth curve.

In the second experiment data, fluctuation phenomenon is very common and obvious. When we get feature points, we should distinguish between fluctuation which should be repaired and not. The processing of the second group data is shown in Fig. 8.

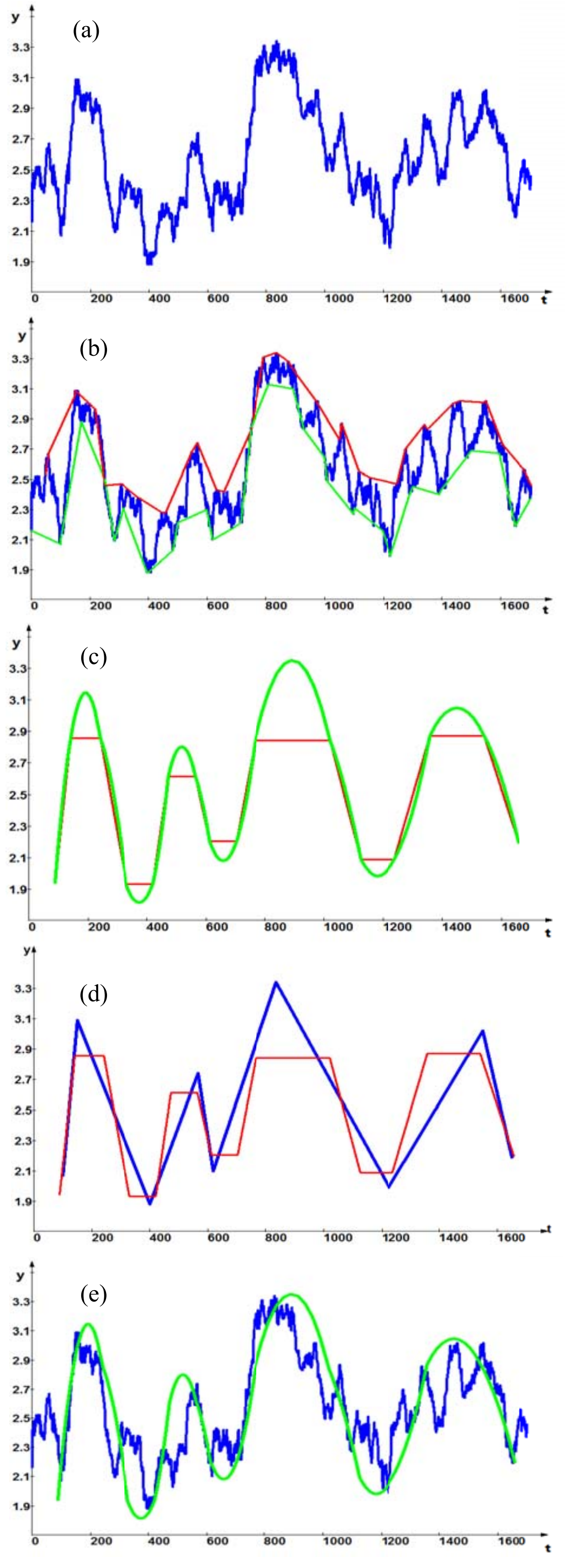


Figure 8. The processing of the second group of data.(a) the blue curve is original time series.(b)the superior and inferior envelopes are depicted as red and green dashed curves.(c)the cp line is depicted as blue dashed curve. Calculated nodes are connected into the red dashed curve.(d)Interpolation curve is depicted as green curve.(e)Composition between original data and interpolation curve.

IV. Conclusions

We presented a novel NIRS data extraction algorithm that iteratively extracts the eigenvalue curves of time series data. The FEFN framework divides the whole time series into some regions and extracts maximum value and minimum value in every region.

With extreme processing, we implemented the extraction of eigenvalue of the original data, and generated CP curves. In order to ensure data after interpolation is agreement with features of original data information, we use a group of calculated data which was called CP' to replace data of CP. The experimental results show data curve based on CP ' data after interpolation is in qualitative agreement with original data. The smoothed data is in qualitative agreement with original data, and could output stable control signals. This algorithm is mainly for testing crowd 'data whose oxygenated hemoglobin has greater fluctuation with continuous motion.

In this paper we focused on the presentation of the FEFN framework, and how to extract the feature points of the data. Based on the maximum and minimum values we represent the variation characteristics of data. To calculate the new data point with the feature points which can reflect the characteristics to replace the original extreme, it should be suitable for interpolation algorithm of interpolation. So the data curve is smoother after interpolation and it is suitable as a control signal output. The new algorithm is based on time sequence segmentation, and the whole data is divided into the same part. We extract characteristic values for each part, and extract the maximum value and the minimum value. So the data points each region are replaced by two feature points. That is to say, if the data is divided into M parts, then the initial data will be reduced to 2*M times. In computing, we only analysis 2*M data, so the algorithm time complexity is reduced substantially. In order to guarantee the output data is continuous, we design the auxiliary point to realize interpolation. Not only can ensure the data continuously and smoothly, but also can ensure the output curve to reflect the changes in the source data.

Acknowledgment

The authors would like to thank the Intelligent Robotics Laboratory, Kochi University of Technology for their support to conduct NIRS experiments. The authors also thank for the support of Natural science Foundation of Liaoning Province under Grant 2013020032. The authors also thank for the support of the Education Department of Liaoning Province under Grant L2011021.

References

[1] Flandrin, Patrick, and Paulo Goncalves, "Empirical mode decompositions as data-driven wavelet-like expansions", International Journal of Wavelets, Multiresolution and Information Processing, Vol. 2, NO. 4. 2004, pp. 477-496.

[2] M. Alamgir, M. Grosse-Wentrup, and Y. Altun, "Multitask learning for Brain-Computer Interfaces", International

Conference on Artificial Intelligence and Statistics (AISTATS 2010), May 13-15, 2010, pp 17–24.

[3] B. Blankertz, C. Sannelli, S. Halder, E. M. Hammer, A. Kübler, K. R. Müller, G. Curio, and T. Dickhaus, "Neurophysiological predictor of SMR-based BCI Performance", Neuro Image, Vol. 51, NO. 4, 2010, pp. 1303–1309.

[4] S. Fazli, J. Mehnert, G. Curio, A. Villringer, K.-R. Müller, J. Steinbrink, and B. Blankertz, "Enhanced performance by a hybrid NIRS-EEG Brain Computer Interface", Neuro Image, Vol. 59, NO. 1, 2012, pp. 519-529.

[5] T. Tsubone, T. Muroga, and Y. Wada, "Application to robot control using brain function measurement by near-infrared spectroscopy", Proceedings of the 29th Annual International Conference of the IEEE EMBS, August 23-26, 2007, pp. 542–534.

[6] Daly, Janis J., and Jonathan R. Wolpaw, "Brain–computer interfaces in neurological rehabilitation", Lancet neurology, Vol. 7, NO. 11, 2008, pp. 1032-1043.

[7] Fan Deng, Dajiang Zhu, JingleiLv, Lei Guo, and Tianming Liu, "FMRI signal analysis using empirical mean curve decomposition", IEEE transactions on biomedical engineering, Vol. 60, NO. 1, January 2013, pp. 42-54.

[8] M. De Luca, C.F. Beckmann, N. De Stefano, P.M. Matthews, and S. M. Smith, "fMRI resting state networks define distinct modes of long-distance interactions in the human brain", Neuroimage, Vol. 29, NO. 4, 2006, pp. 1359-1367.

[9] D. J. Heeger and D. Ress, "What does fmri tell us about neuronal activity?",Nature Rev. Neurosci , vol. 3, no. 2, pp. 142–151, Feb. 2002.

[10] N. K. Logothetis, "What we can do and what we cannot do with fmri", Nature, vol. 453, no. 7197, pp. 869–878, Jun. 2008.

[11] G. Buzski and A. Draguhn, "Neuronal oscillations in cortical networks", Science, vol. 304, no. 5679, pp. 1926–1929, Jun. 2004.

[12] Xiangjun CHEN, Zhanfeng GAO, "Data Processing Based on Wavelet Analysis in Structure Health MonitoringSystem", Journal of computers, vol. 6, no. 12, pp. 2686-2691.

[13] Xinming Zhang, Lin Yan. A Fast Image Thresholding Method Based on Chaos Optimization and Recursive Algorithm for Two-Dimensional Tsallis Entropy. Journal of computers, vol. 5, no. 7, pp. 1054-1061.

[14] S. Waldert, H. Preissl, E. Demandt, C. Braun, N. Birbaumer, A. Aertsen, and C. Mehring, "Hand movement direction decoded from MEG and EEG", Journal of Neuroscience, Vol. 28, NO. 4, 2008, pp. 1000–1010.

[15] Lebedev, Mikhail A., and Miguel AL Nicolelis, "Brain-machine interfaces: past, present and future", TRENDS in Neurosciences, Vol. 29, NO. 9, 2008, pp. 536-546.

Yan Jiang received the Ph.D. degree in Institute of information science and technology from Northeastern University, in 2009. Currently, he is an Associate Professor at Shenyang University of Technology, China. His interests are in XML and intelligent control.

Xin Jin received the Bachelor's degree in Shenyang University of Technology, in 2010. Currently, he is an Engineer at Shenyang Hexing Testing Equipment Co., Ltd, China. His interests are in mechanical control and computer network.

Xin Li received the bachelor's degree in computer science and technology from Shenyang University of technology, in 2013. Currently, she is a graduate student at Shenyang University of Technology. Her interests are database and the data processing.