A Novel Approach for Baseline Correction in ^{1}H-MRS Signals Based on Ensemble Empirical Mode Decomposition

Mohammad Ali Parto Dezfouli, Mohsen Parto Dezfouli, and Hamidreza Saligheh Rad

Abstract— **Proton magnetic resonance spectroscopy (^{1}H-MRS) is a non-invasive diagnostic tool for measuring biochemical changes in the human body. Acquired ^{1}H-MRS signals may be corrupted due to a wideband baseline signal generated by macromolecules. Recently, several methods have been developed for the correction of such baseline signals, however most of them are not able to estimate baseline in complex overlapped signal. In this study, a novel automatic baseline correction method is proposed for ^{1}H-MRS spectra based on ensemble empirical mode decomposition (EEMD). This investigation was applied on both the simulated data and the in-vivo ^{1}H-MRS of human brain signals. Results justify the efficiency of the proposed method to remove the baseline from ^{1}H-MRS signals.**

I. Introduction

Proton magnetic resonance spectroscopy (^{1}H-MRS) is widely used to detect metabolites' concentration variation, which can indicate a diseased condition in human tissue. ^{1}H-MRS should be quantified well to show the metabolic variations and exact metabolites detection. However, existence of a background signal, called baseline signal, originated from macromolecules and lipids, hinders the performance of ^{1}H-MRS quantification methods. The background signals are characterized by broad spectral lines, resulting to overlap with metabolites' spectra in the frequency domain

Recently, a number of studies have been carried out to explore a proper method to remove the baseline artifact. Two different techniques were implemented to correct the baseline signal; reconstruction of the first points of the free induction decay (FID) [1], and approximation of the baseline in the frequency domain [2]. Most of such techniques are composed of two essential steps: 1) baseline recognition step in which the signal-free regions of the spectrum are detected using some thresholding mechanisms [3, 4], and 2) implementation based on iterative methods. The performance of such baseline correction methods are limited in the complex spectra with overlapping peaks. In a recent investigation, an automatic baseline correction method was implemented based on a combination of baseline recognition and an iterative baseline modeling method [5]. It provided good results for the complex MRS spectra with overlapping peaks, but it is sensitive to noise factor and cannot follow up the corrected baseline when the trend of baseline is similar to metabolites.

In this study, ensemble empirical mode decomposition (EEMD) was employed to estimate the baseline of ^{1}H-MRS signals. The method which was optimized for the applications in the human brain, is able to estimate the most of the baseline points hidden under the overlapping signals and noise.

Mohammad Ali Parto Dezfouli is with the Department of Biomedical Engineering and Medical Physics, Tehran University of Medical Sciences, Tehran, Iran and Quantitative MR Imaging and Spectroscopy Group, Research Center for Molecular and Cellular Imaging, Tehran, Iran (e-mail: a.parto@razi.tums.ac.ir).

Mohsen Parto Dezfouli is with the School of Electrical Engineering, faculty of Biomedical Engineering, Iran University of Science and Technology, Tehran, Iran (e-mail: parto@elec.iust.ac.ir).

H. Saligheh Rad is with the Department of Biomedical Engineering and Medical Physics, Tehran University of Medical Sciences, Tehran, Iran and Quantitative MR Imaging and Spectroscopy Group, Research Center for Molecular and Cellular Imaging Tehran, Iran (email: h-salighehrad@tums.ac.ir).

II. Materials and Method

In order to estimate the baseline, the signal was decompose to several sub signals based on ensemble empirical mode decomposition (EEMD) algorithm. Then the sub signals are selected and linear combined to estimate the baseline.

A. Ensemble Empirical Mode Decomposition (EEMD)

The EMD is an iterative method that can decompose any complex data set into some oscillations called intrinsic mode functions (IMF). An IMF represents an oscillatory mode of the main signal, that the summation of them get the original signal [6]. Each IMF was calculated in an iterative procedure called sifting procedure as follow:

1. Determine the maxima and minima envelopes with a spline interpolation.
2. Calculate the mean of two envelopes and subtract it from the signal.

$$d_{(1)}(t) = s(t) - m_1(t) \qquad (1)$$

where $s(t)$ is the original signal and the parameter $m_1(t)$ is the mean of the calculated envelopes at first iteration. $d_{(1)}(t)$ is considered as the modified data. The sifting procedure repeat k times and $d_{(k)}(t)$ is calculated.

3. Consider $d_{(1)}(t)$ as the modified data and similarly the steps one and two repeat k times.

$$d_{(k)}(t) = d_{(k-1)}(t) - m_{(k-1)}(t) \qquad (2)$$

where $d_k(t)$ is the k-th data and $m_{(k-1)}(t)$ is the (k-1)-th mean of the calculated envelopes at (k-1)-th iteration. The value of k is determined according to a stopping criterion that usually defined as:

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

$$SD = \left\| \frac{d_{(k+1)}(t) - d_{(k)}(t)}{d_{(k)}(t)} \right\| \qquad (3)$$

where $\|.\|$ denotes the Euclidean distance. The parameter $d_{(k)}(t)$ consider as the first IMF when the value of *SD* becomes smaller than a threshold.

4. When the criterion *SD* becomes smaller than a threshold, the procedure stops and $d_{(k)}(t)$ is determined the first IMF.

$$IMF_i(t) = d_{i,(k)}(t) \,|\, \{SD \le Threshold\ \} \qquad (4)$$

5. The residual $r_1(t)$ is calculated by subtracting IMF_1 from signal $s(t)$ and the procedure are repeated for $r_1(t)$ to find next IMF.

$$r_1(t) = s(t) - IMF_1(t) \qquad (5)$$

6. This repetition stops when $r_1(t)$ becomes a monotonic function and cannot extract more IMFs from it.

$$r_i(t) = IMF_{i-1}(t) - IMF_i(t) \qquad (6)$$

There is a mode-mixing problem in the EMD method [6] , which causes a serious aliasing in the time-frequency distribution in an IMF. In order to increase the reliability an improved EMD method is used which called EEMD [7]. EEMD is an iterative noise-assisted data analysis method that that repeatedly performs sifting process. It takes the ensemble means of IMFs by adding a white noise to the original signal and calculation the intrinsic oscillatory functions (IOF).

The IOFs calculation procedure can be obtained as follow:

1. Set the number of ensemble *M*
2. Construct the noise and add it to signal
3. Apply the EMD method on the noised-added signal and calculate IMFs: $IMF_i = [IMF_{1,k} \ldots IMF_{i,k}]^T$
4. Repeat the steps 2 and 3 while *i* be equal to *M*
5. Calculate the ensemble means of *M* trials for each IMF to achieve IOFs:

$$IOF_i = \frac{1}{M} \sum_{i=1}^{M} IMF_i \qquad (7)$$

B. Signal Model

^{1}H-MRS signal $s(t)$ is contained three main terms, which can be modeled by:

$$s(t) = \mathrm{met(t)} + b(t) + n(t) \qquad (8)$$

The first term; *met(t)* represents the metabolites' signal whose model function is known. The function *met(t)* can be defined as:

$$met(t) = \sum_{k=1}^{K} a_k e^{(i\varphi_k)} e^{(-d_k t + 2\pi i f_k t)} m_k(t) \qquad (9)$$

where K is the number of metabolites (k = 1, ..., K), $m_k(t)$ the profile of metabolite, a_k the amplitude, φ_k the phase shift, d_k the damping correction, f_k the frequency shift and $j = \sqrt{-1}$. The second term; *b(t)* represents the baseline signal which its model function is not known exactly. The third term; *n(t)* denotes the white Gaussian-distributed noise..

C. ^{1}H-MRS Simulated signal generation

In order to evaluate the proposed method we used the simulated ^{1}H-MRS signals, which were generated using MATLAB. Nine different metabolites were combined with a Gaussian noise and a baseline were added to it. The parameters phase shifts φ_k, damping perturbations d_k, and frequency shifts f_k were set within the following values: $-\pi \le \varphi_k \le \pi$, 0ppm $\le d_k \le$ 0.1ppm, -0.1 ppm $\le f_k \le$ 0.1 ppm.

The profiles were multiplied by an amplitude value to change the concentration of each metabolite mimicking the normal and tumours brain (Table. I). Baseline was generated and added to signal with interest signal-to-baseline ratio (SBR) [8]. In addition, white Gaussian noise is added to signal with interest signal-to-noise ratio (SNR) to obtain *s(t)* in (8). Finally, the *S(f)* was calculated by applying Fourier Transform (FT) to *s(t)*.

$$\Im\{s(t)\} = S(f) \qquad (10)$$

TABLE I. METABOLITES AND THEIR CONCENTRATION [9]; ABSOLUTE CONCENTRATIONS AND CONCENTRATION RATIOS OF UNTREATED PEDIATRIC BRAIN TUMORS

Metabolite	***Amplitude (mean ± standard deviation)***
Alanine	1.7±1.1
Choline	3.4±2.0
Creatine	3.5±2.6
γ-Amino butyric acid (GABA)	2.0±1.0
Glutamine	5.0±3.5
Lactate	3.0±2.5
Myo-Inositol	8.2±5.8
N-Acetyl-L-Aspartic Acid (NAA)	1.3±1.1
Taurine	2.5±1.0

D. Signal Acquisition

Proton MRS imaging experiments were performed on a 1.5T Siemens Avanto MRI/MRS system in the room temperature using point resolved spectroscopy (PRESS) [10] pulse sequence with manufacture's built-in auto-shimming on the volume-of-interest, chemical shift selective suppression (CHESS) [11] to suppressed water and 3D imaging parameters as follows: TE/TR = 30/1500ms, voxel size = 8mm × 8mm × 8mm, NEX = 1, frequency bandwidth = 1000Hz, number of data points = 512 and 16×16×16 array chemical shift imaging (CSI) grid.

E. Baseline Estimation

Baseline is a low-frequency phenomenon at the frequency-domain. Hence, it is expected that the major baseline components are located in the high-order IOFs. The baseline can be estimated by subtracting the last several IOFs from the signal. EEMD was applied to the simulated signal and found that the summation of some last IOFs, follow the baseline trend plus an residual signal (*R(f)*). The *R(f)* is calculated by the half distance between extrema envelopes.

$$R(f) = \frac{Env_{\max}(f) - Env_{\min}(f)}{2} \qquad (11)$$

where $Env_{max}(f)$ and $Env_{min}(f)$ are the maximum and minimum envelope of the signal which were obtained by finding the peaks and performing interpolation. There is significant different across the full width half maximum (FWHM) of the metabolites and baseline. The FWHM of the metabolites in ^{1}H-MRS is less than 0.2ppm [12]. The threshold set to 0.4ppm based on a priori knowledge and experimentally can be tuned according to the baseline behavior.

The peaks of all IOFs were calculated, then the curve fitting was performed to fit Lorentzian to each peak. The Lorentzian function parameters define as:

$$L(f) = \frac{P_1}{(f - P_2)^2 + P_3} + C \qquad (12)$$

The nonlinear curve-fitting problem was solved in the least-squares sense. The Lorentzian FWHM calculation is actually straightforward and can be read off from the equation (12).

$$FWHM = 2\sqrt{P_3} \qquad (13)$$

There are fewer baseline components in the early IOFs, but more metabolite components. Therefore, the attenuation correction was applied to IOFs to estimate the baseline by linear weighted combination of selected IOFs.

$$B(f) = \sum_{i=1}^{I} a_i C_i(f) \qquad (14)$$

where B(f) is the estimated baseline by combination of IOFs, I is the number of selected IOFs, $C_i(f)$ is the i-th IOF and a_i is the attenuation correction which should be obtained by analyzing the simulated signal.

The following steps constitute the proposed baseline estimation procedure:

1. *R(f)* was calculated from (11).
2. EEMD was applied to *R(f)* to obtain IOFs.
3. FWHM of all the peaks were calculated in all IOFs from (12) and (13).
4. Threshold was applied to FWHM of all the peaks in each IOF to select candidate IOFs for baseline estimation.
5. Selected IOFs were linearly combined to estimate the baseline (14).

III. Results

Two types of experiments are presented. First, several different simulated signals are generated to evaluate the performance of the proposed EEMD-based method. The *S(f)* is processed to obtain an enhanced reconstructed *Š(f)*, which is free from baseline. Second, the ^{1}H-MRS of normal brain was used to calculate the correlation of baseline in different regions. There are correlations, between the baseline spectra for different tissues.

A. Simulated ^{1}H-MRS signals

For the simulated signal, the quantitative evaluation is assessed by the relative signal-to-error ratio (RSER):

$$RSER = 10\log_{10}\left(\frac{\sum_{f=0}^{L-1} S^2(f)}{\sum_{f=0}^{L-1}\left(S(f) - \tilde{S}(f)\right)^2}\right) \qquad (15)$$

The first IOF is contained the highest frequency component and the last IOF is contained the lowest frequency component. The proposed method is compared to the automatic baseline correction method which employed continuous wavelet transform (CWT) [5]. Fig. 1 shows the spectrum of the estimated baseline performing by the two methods. RSER of the proposed method was 17.95dB and 14.54dB for the baseline estimation method in [5]. Fig. 2 shows the results of the RSERs in different SNR calculated in 100 experiments.

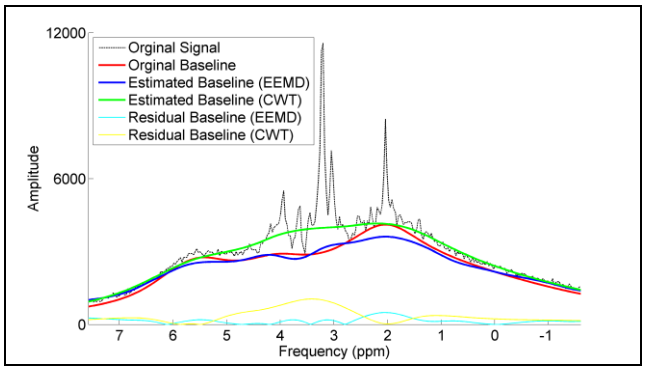

Figure 1: Baseline estimation result on simulated data; sample spectrum of baseline estimation.

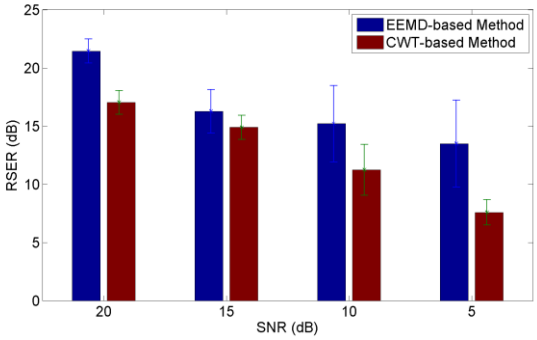

Figure 2: RSERs of proposed method and iterative method which used CWT, under different SNR.

B. In-vivo 1H-MRS of human brain

The proposed method was applied on ^{1}H-MRS of normal human brain. All the voxels' spectrums were recorded and then the correlation coefficient of the selected voxel from white matter (WM) and gray matter (GM) compared to all of the voxel at the same slice. Fig. 3 shows the result of the correlation coefficient map between selected voxel and other voxels. The map shows there is a strong correlation between

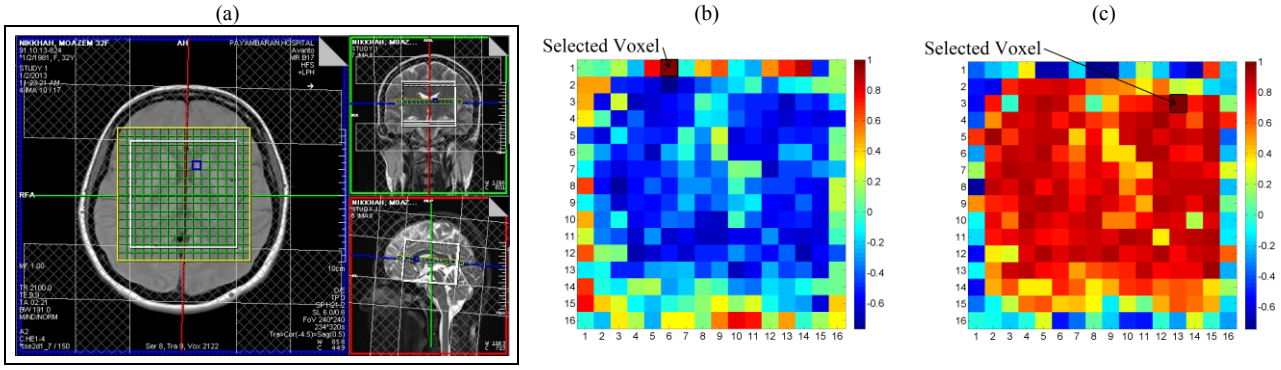

Figure 3: Baseline estimation result on normal brains; (a) T2-weighted anatomical image of the brain (b) baseline estimation correlation coefficient map between selected voxel from GM and all other voxels at the same slice (c) baseline estimation correlation coefficient map between selected voxel from WM and all other voxels at the same slice

the same regions. The map shows there is strong correlation between the same regions. Fig 4 shows the correlation coefficient map of the baseline estimation results on the tumor brain. The baseline estimation processing takes on average less than 1 second for each voxel (about 1 hour for full 3D 16×16×16 voxels MRSI with 512 data points) with implementation on a PC, Intel Core i7 and 4 GB of memory using the MATLAB v 7.14 under Microsoft Windows 7.

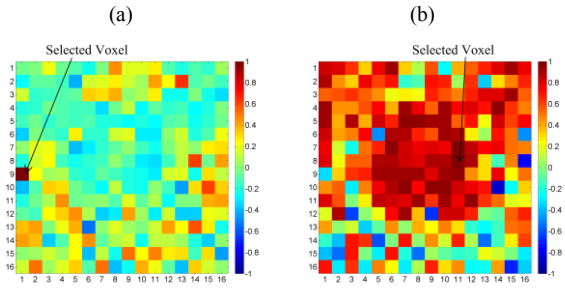

Figure 4: Baseline estimation results on tumor brain; (a) baseline estimation correlation coefficient map between selected voxel from non-tumor region and all other voxels at the same slice and (b) baseline estimation correlation coefficient map between selected voxel from tumor region and all other voxels at the same slice

IV. Conclusion

A novel method for ^{1}H-MRS baseline correction based on the EEMD is presented. Baseline estimation and removal is achieved through the development of EEMD-based methods with prior knowledge on ^{1}H-MRS signal. The techniques developed by using ^{1}H-MRS metabolites information to modify the IOFs and linear combine them as estimated baseline. Simulated and *in-vivo* results indicate that the EEMD is an effective tool to estimate the baseline in ^{1}H-MRS signal. The techniques can be optimized by tuning the parameters of the IOFs combination with analyzing of much more data.

Acknowledgment

All MR Imaging and Spectroscopy were performed at imaging center, Payambaran Hospital, Tehran, Iran.

References

[1] C. Tang, "An analysis of baseline distortion and offset in NMR spectra," *Journal of Magnetic Resonance, Series A,* vol. 109, pp. 232-240, 1994.

[2] Y. Hiltunen, M. Ala-Korpela, J. Jokisaari, S. Eskelinen, K. Kiviniitty, M. Savolainen, *et al.*, "A lineshape fitting model for 1H NMR spectra of human blood plasma," *Magnetic resonance in medicine,* vol. 21, pp. 222-232, 1991.

[3] J. Carlos Cobas, M. A. Bernstein, M. Martín-Pastor, and P. G. Tahoces, "A new general-purpose fully automatic baseline-correction procedure for 1D and 2D NMR data," *Journal of Magnetic Resonance,* vol. 183, pp. 145-151, 2006.

[4] S. Golotvin and A. Williams, "Improved baseline recognition and modeling of FT NMR spectra," *Journal of Magnetic Resonance,* vol. 146, pp. 122-125, 2000.

[5] Q. Bao, J. Feng, F. Chen, W. Mao, Z. Liu, K. Liu, *et al.*, "A new automatic baseline correction method based on iterative method," *Journal of Magnetic Resonance,* vol. 218, pp. 35-43, 2012.

[6] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, *et al.*, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences,* vol. 454, pp. 903-995, 1998.

[7] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Advances in adaptive data analysis,* vol. 1, pp. 1-41, 2009.

[8] U. Seeger, U. Klose, I. Mader, W. Grodd, and T. Nägele, "Parameterized evaluation of macromolecules and lipids in proton MR spectroscopy of brain diseases," *Magnetic resonance in Medicine,* vol. 49, pp. 19-28, 2003.

[9] A. Panigrahy, M. Krieger, I. Gonzalez-Gomez, X. Liu, J. McComb, J. Finlay, *et al.*, "Quantitative short echo time 1H-MR spectroscopy of untreated pediatric brain tumors: preoperative diagnosis and characterization," *American journal of neuroradiology,* vol. 27, pp. 560-572, 2006.

[10] P. A. Bottomley, "Selective volume method for performing localized NMR spectroscopy," ed: Google Patents, 1984.

[11] A. Haase, J. Frahm, W. Hanicke, and D. Matthaei, "1H NMR chemical shift selective (CHESS) imaging," *Physics in medicine and biology,* vol. 30, p. 341, 1985.

[12] P. Pouwels, M. Steenweg, F. Barkhof, and M. Van der Knaap, "Absolute metabolite quantification in human brain using short echo-time CSI and a phased-array headcoil," in *Proceedings of International Society for Magnetic Resonance in Medicine*, 2010.