

INCREASING DETECTION EFFICIENCY OF BORDER MENTAL DISORDERS BASED ON ADAPTIVE DECOMPOSITION AND CEPSTRAL ANALYSIS OF SPEECH SIGNALS

Alan K. Alimuradov ¹, Alexander Yu. Tychkov ¹, Pyotr P. Churakov ²,
Alexey P. Zaretskiy ³, Ivan B. Prokhorov ³, Kirill S. Mityagin ³

¹ Research Institute for Fundamental and Applied Studies, Penza State University
40, Krasnaya str., Penza, 440026, Russia

² Data Measuring Equipment and Metrology Department, Penza State University
40, Krasnaya str., Penza, 440026, Russia

³ Department of radio engineering and cybernetics, Moscow Institute of Physics and
Technology Moscow Institute of Physics and Technology (State University),
9 Institutsky Per., Dolgoprudny, Moscow region 141700, Russia

The paper is received on May 25, 2019

Abstract. The detection accuracy of borderline mental disorders depends on correct processing of speech signals. The main reason of low accuracy and large errors in measurements is associated with the use of inefficient and non-adaptive methods for processing of non-stationary speech signals. In this paper, the authors propose a method for increasing the detection efficiency of borderline mental disorders based on adaptive decomposition technology for non-stationary signals, namely, improved complete ensemble empirical mode decomposition with adaptive noise and mel-frequency cepstral analysis. A block diagram for the method and a brief mathematical description are presented. The research results are presented, on the basis of which it was concluded that the method proposed by the authors can successfully be tested in remote monitoring systems of psychogenic disorders to accelerate the treatment process.

Keywords: speech signal, border mental disorders, mel-frequency cepstral coefficients, improved complete ensemble empirical mode decomposition with adaptive noise.

1. Introduction

According to the Federal State Statistics Service [1], more than 326.7 thousand patients having mental illness (mental disorders, behavioral disorders, alcohol and drug psychoses, etc.) have been registered over the previous five years.

A borderline mental disorder is the most common affective disorder. As a rule, the reason for borderline mental disorders is low socioeconomic status and the presence of a large number of social stresses.

For a long time, accurate detection of borderline mental disorders was not possible, and it was limited to perceptual tests and laboratory analysis. However, since the power increase in computers, experimental statistical methods have appeared [2 – 6], the most adaptive of which are the methods based on the analysis of speech signals [7].

There is an important rule in diagnosing of the nervous system disorders: pathophysiological mechanisms of the development of borderline mental disorders are based on the principles of interaction of the nervous and speech-forming systems of the organism; disorders of the nervous system activate a cascade of mechanisms that affect the functioning of motor organs of the speech apparatus. So, we can conclude that borderline mental disorders are “encoded” in certain informative parameters of speech signals.

The detection accuracy of borderline mental disorders depends on correct processing of speech signals, which is determined by the correctness of measurement of its amplitude, time, frequency and energy characteristics. The main reason of low accuracy and large errors in measurements is associated with the use of inefficient and non-adaptive methods for processing of non-stationary speech signals.

In this paper, the authors propose a method for increasing the detection efficiency of borderline mental disorders based on adaptive decomposition technology for non-stationary signals, namely, improved complete ensemble empirical mode decomposition with adaptive noise and mel-frequency cepstral analysis. The study is a development of previously published works of the authors [8, 9].

2. Materials and methods

A. Empirical mode decomposition

Empirical mode decomposition (EMD) [10] is an adaptive technology for the decomposition of non-stationary signals stemming from nonlinear systems. The EMD

provides local decomposition of a signal into fast and slow oscillatory functions. As a result of the decomposition, the original signal can be represented as a sum of amplitude and frequency modulated functions, called intrinsic mode functions (IMF), or modes. An analytical expression of the EMD is as follows:

$$x(n) = \sum_{i=1}^I IMF_i(n) + r_i(n)$$

where $x(n)$ is the original signal; $IMF_i(n)$ is the IMF; $r_i(n)$ is a residue; $i = 1, 2, \dots, I$ is the IMF number, n is discrete timing ($0 < n \leq N$, N is a number of discrete samples in the signal).

Among all the varieties of decomposition methods, the most adaptive one to biomedical and speech signals is an improved ensemble empirical mode decomposition with adaptive noise (CEEMDAN) [11]. Here is an algorithm and mathematical description of the improved CEEMDAN method:

1. Using the EMD apparatus and determining the local average values of noise copies of the original signal ($x_j(n) = x(n) + \beta_0 E_1(w_j(n))$) from the formula $\langle E_1(x_j(n)) \rangle = \langle x_j(n) \rangle - \langle M(x_j(n)) \rangle$, the first residue is determined:

$$r_1(n) = \langle M(x_j(n)) \rangle$$

where $E_i(\cdot)$ is the IMF extraction operator by the EMD method (i is the mode number); $x_j(n) = x(n) + w_j(n)$ is the noise copies of the original signal ($x(n)$ is the initial speech signal, $w_j(n)$ is the implementation of white noise with zero mean unit variance); $M(\cdot)$ is the operator creating the local average value of the applied signal; $\beta_i = \varepsilon_i \text{std}(r_i)$ is the coefficient that allows choosing of different values of signal-to-noise ratio.

2. At the first stage, the first mode for $i = 1$ is calculated:

3. The second residue is calculated as the averaged local average value of the noise copies of the first residue $r_1(n) + \beta_1 E_2(w_j(n))$, and the second mode is determined:

$$\widetilde{IMF}_2(n) = r_1(n) - r_2(n) = r_1(n) - \langle M(r_1(n) + \beta_1 E_2(w_j(n))) \rangle$$

4. At subsequent stages, the i -th residue for $i = 3, \dots, I$ is calculated:

$$r_i(n) = \langle M(r_{i-1}(n) + \beta_{i-1} E_i(w_j(n))) \rangle$$

5. The i -th mode is calculated:

$$\widetilde{IMF}_i(n) = r_{i-1}(n) - r_i(n)$$

6. Go to Step 4 for the next i value.

The constants β_i are selected to obtain the desired signal-to-noise ratio between the added noise and the residue to which the noise is added.

B. Mel-frequency cepstral analysis

Cepstral analysis has gained a wide practical popularity in the field of speech signal processing, due to the advantage of information compression on the signal during the transition into the frequency processing domain [12]. Cepstral analysis is based on the allocation of cepstral coefficients on the Mel scale, called mel-frequency cepstral coefficients (MFCC). MFCC include two basic concepts: a cepstrum, and a mel scale. A cepstrum is a discrete cosine transform of the signal amplitude spectrum in a logarithmic scale. The signal cepstrum is determined by the formula:

$$c(n) = DCT[\log(X(|x(n)|^2))]$$

where DCT is a discrete cosine transform; X is a spectral representation of the signal $x(n)$; n is discrete timing ($0 < n \leq N$, N is a number of discrete samples in the signal).

The method of MFCC obtaining is based on the model of the human hearing system functioning and uses the frequency scale in Mel, which simulates the frequency sensitivity of a human ear [12].

3. Description of the method

Fig. 1 presents a simplified block diagram for the method for detecting borderline mental disorders. The processing steps 1, 2, 5 and 6 are a classic detection method. Steps 3 and 4 were introduced by the authors to improve detection efficiency. Processing step 6 has undergone modernization. Steps 7 and 8 are not related to processing. Let us consider some of the stages.

Stage 2. During pre-processing, the following actions are performed: the removal of the constant component, the speech signal filtering by the fourth-order

Chebyshev high-pass filter to remove frequencies below 130 Hz, and segmentation into voiced and unvoiced sections.

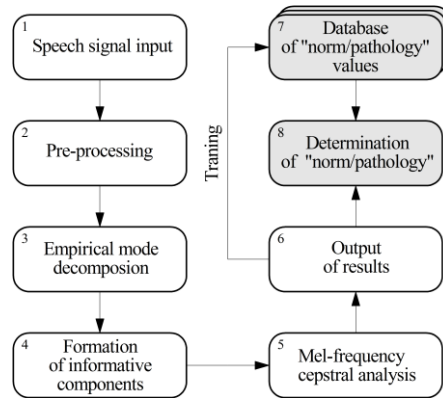


Fig. 1. Block diagram for speech/pause detection algorithm.

Stage 3. Based on the above mentioned brief analysis of the advantages and disadvantages of different types of decomposition, and taking into account the specifics of speech signals under borderline mental disorders, the authors decided to use the improved CEEMDAN, which will provide:

- adaptive decomposition, since basic functions used in the decomposition are extracted directly from the original speech signal, and allow to take into account only its own peculiarities (hidden modulation, energy concentration regions, etc.).
- minimum level of residual noise;
- absence of parasitic IMF arising in the early stages of decomposition as a result of the overlap of the scale-energy mode spaces.

Stage 4. All IMF for an absolutely arbitrary signal can be divided into two categories: informative IMF with noise and signal components; non-informative IMF with trend components. The formation of informative components consists in subtraction of informative noise and non-informative IMF from the initial signal. The purpose of formation of informative components is the collection of information reflecting violations in the functioning of speech apparatus organs due to borderline mental disorders. The formation of a set of informative components is carried out according to the formula:

$$x_{ab,i}(n) = x(n) - \left(a \times \sum_{i=0}^2 IMF_{i+1}(n) + b \times \sum_{i=0}^2 IMF_{i-1}(n) \right)$$

where $x_{ab,i}(n)$ is the informative component; $x(n)$ is the initial signal; a, b are the coefficients determining the participation of IMF in the formation of informative components.

Stage 5. As noted earlier, the MFCC are calculated as the cepstral characteristics in the method. The main stages of MFCC calculation are as follows:

- converting of the signal from the Hertz scale to the Mel scale;
- spectral transformation and determination of the signal periodogram;
- filtering of the signal periodogram with a set of mel filters;
- taking the logarithm of the signal energy in each mel filter;
- discrete cosine transform of the energy logarithm.

The result of the calculation is the MFCC ($MFCC(c)$, where $c = 1, 2, \dots; C$ is the MFCC number; C is the desired number of coefficients).

Normalization and calculation of the first and second increments of the MFCC values allows us to obtain dynamic information about the coefficients. As it is known, high frequencies are less susceptible, and the MFCC at these frequencies are less informative than the MFCC at low frequencies. Normalization of the MFCC is the multiplication of each coefficient by a number that increases with the coefficient index. Thus, the first coefficients are reduced, and the last coefficients are increased. The following formula is used for this operation:

$$MFCC_N(c) = MFCC(c) \times \left(1 + \frac{L}{2} \sin\left(\frac{\pi c}{2}\right) \right)$$

where L is an empirically chosen value, being 22.

The coefficient vector describes a fixed spectral envelope of one fragment, but it is obvious that the speech signals carry information about dynamics in the form of a slight change in coefficients over time:

$$MFCC_D(c) = \frac{\sum_{d=1}^D d(MFCC(c+d)) - MFCC(c-d)}{2 \sum_{d=1}^D d^2}$$

$$MFCC_{DD}(c) = \frac{\sum_{d=1}^D d(MFCC_D(c+d)) - MFCC_D(c-d)}{2 \sum_{d=1}^D d^2}$$

where $MFCC_D(c)$, $MFCC_DD(c)$ are the first and the second MFCC increments, $MFCC(c)$ are static MFCC, D is a typical increment value, being 2.

4. Investigation of the method

To assess the detection effectiveness of borderline mental disorders based on the proposed authors' with the support of K.R. Evgrafov Regional Psychiatric Hospital (Penza, Russian Federation) and Penza State University, a group of subjects and a verified signal base have been formed. A group of subjects included 100 males and females, aged 18 to 60 years old, with clearly expressed symptoms of borderline mental disorders. To evaluate the effectiveness of the method, errors of the first and second kind were used.

The objective of the study was to detect borderline mental disorders using the classical method and the method proposed by the authors based on the improved CEEMDAN and mel-frequency cepstral analysis.

All stages of signal processing and data analysis were performed in the mathematical modeling environment © Matlab (MathWorks). Table 1 presents the results of detection of borderline mental disorders.

Table 1. Results of detection of borderline mental disorders

Predictable result	Detection result		Errors of the first and second kind, %	
	Pathology	Norm		
<i>Classical method for detection</i>				
Pathology	78 pers.	22 pers.	1st	22
Norm	16 pers.	84 pers.	2nd	16
<i>Proposed method for detection</i>				
<i>Informative component No. 1 (subtraction of the last three IMF's)</i>				
Pathology	84 pers.	16 pers.	1st	16
Norm	9 pers.	81 pers.	2nd	9
<i>Informative component No. 2 (subtraction of the first IMF and last three IMF's)</i>				
Pathology	91 pers.	9 pers.	1st	9
Norm	6 pers.	94 pers.	2nd	6
<i>Informative component No. 3 (subtraction of the first two and last three IMF's)</i>				
Pathology	87 pers.	13 pers.	1st	13
Norm	8 pers.	82 pers.	2nd	8

5. Conclusion

In accordance with the data obtained, it can be concluded that the use of the improved CEEMDAN and mel-frequency cepstral analysis in the classical method makes it possible to improve the detection efficiency of mental disorders. The best results are achieved by subtracting the first noise and the last three trend IMFs from the initial speech signal (information component No. 2). This is explained by the fact that the first IMF contains most of the residual noise and when subtracting, the original signal is filtered. The last three trend IMFs are the result of the decomposition of the sum of the harmonic signal and the polynomial trend. Subtraction of these IMFs saves the original signal from redundant information.

Thus, the results obtained make it possible to conclude that the method proposed by the authors can be successfully tested in remote monitoring systems of psychogenic disorders to accelerate the treatment process.

6. Acknowledgment

The authors are grateful to the Council on Grants of the President of the Russian Federation for the financial support of the project “Development of methods and virtual means for adaptive noise-robust processing and detection of clinically significant parameters of medical electrical and acoustic signals in patients under borderline mental disorders”, No. MK-250.2017.8, 2017-2018.

References

1. Healthcare in Russia 2017. Statistical Bulletin. [Online] Available at: http://www.gks.ru/free_doc/doc_2017/zdrav17.pdf (In Russian)
2. S. Koelstra, M. Pantic, I. Patras. A dynamic texture-based approach to recognition of facial actions and their temporal models. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2010, Vol. 32, pp. 1940–1954.
3. C. Kappeler-Setz, F. Gravenhorst, J. Schumm, B. Arnrich, G.Tröster. Towards long term monitoring of electrodermal activity in daily life. *Personal and Ubiquitous Computing*, 2013, Vol. 17, No. 2, pp. 261–271.

4. L.A. Farwell. Brain fingerprinting: a comprehensive tutorial review of detection of concealed information with event-related brain potentials. *Cognitive Neurodynamics*, 2012, Vol. 6, pp. 115–154.
5. L. Likforman-Sulem, A. Esposito, M. Faundez-Zanuy, S. Cléménçon, G. Cordasco. EMOTHAW: A novel database for emotional state recognition from handwriting and drawing. *IEEE Trans. on Human-Machine Systems*, 2017, Vol.47, No. 2, pp. 273–284.
6. V.A. Barabanschikov, A.V. Zhegallo. Methods of eye tracking in psychology: educational program. *Experimental Psychology*, 2014, Vol. 7, No. 1, pp. 132–137.
7. J.R. Williamson, T.F. Quatieri, B.S. Helfer, G. Ciccarelli, D. Mehta. Vocal and facial biomarkers of depression based on motor incoordination and timing. *Proc. 4th International Workshop on Audio/Visual Emotional Challenge (AVEC)*, 2014, pp. 65-72.
8. A.K. Alimuradov, A.Yu. Tychkov, A.V. Ageykin, P.P. Churakov, Y.S. Kvitka, A.P. Zaretskiy. Speech/pause detection algorithm based on the adaptive method of complementary decomposition and energy assessment of intrinsic mode functions. *Proc. XX IEEE International Conference on Soft Computing and Measurements (SCM)*, May 24-26, 2017, Russia, St. Petersburg, pp. 610–613. DOI: [10.1109/SCM.2017.7970665](https://doi.org/10.1109/SCM.2017.7970665). (In Russian)
9. A.K. Alimuradov, A.Yu. Tychkov, P.P. Churakov, Yu.S. Kvitka, A.P. Zaretskiy, G.V. Vishnevskaya. Noise-robust algorithm for “speech/pause” segmentation in diagnostic systems of psychogenic states. *2016 International Conference on Engineering and Telecommunication (EnT)*, Nov. 29–30, 2016, Russia, Moscow, pp. 3–6, DOI: [10.1109/EnT.2016.009](https://doi.org/10.1109/EnT.2016.009).
10. N.E. Huang, Sh. Zheng, R.L. Steven. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Lond.*, 1998, Vol. A454, pp. 903–995.
11. M.A. Colominasa, G. Schlotthauera, M.E. Torres. Improved complete ensemble EMD: A suitable tool for biomedical signal processing. *Biomed. Signal Proces.*, 2014, Vol.14, pp. 19–29.

12.X. Huang, A. Acero, H.-W. Hon. *Spoken Language Processing. Guide to Algorithms and System Development*. Prentice Hall, 2001.

For citation:

A.A. Alimuradov, A.Yu.Tychkov, P.P.Churakov, A.P.Zaretskiy, I.B.Prokhorov, K.S.Mityagin. Increasing detection efficiency of border mental disorders based on adaptive decomposition and cepstral analysis of speech signals. *Zhurnal Radioelektroniki - Journal of Radio Electronics*. 2019. No.6. Available at <http://jre.cplire.ru/jre/jun19/6/text.pdf>