Measurement of Speech Signal Patterns under Borderline Mental Disorders

Alan Alimuradov, Alexander Tychkov, Andrey Kuzmin, Pyotr Churakov, Alexey Ageykin, Galina Vishnevskaya
Penza State University
Penza, Russia
alansapfir@yandex.ru, tychkov-a@mail.ru, flickerlight@inbox.ru, snpbi-pgu@mail.ru, keokushinkai@yandex.ru, galina.vishnevskaya.62@mail.ru

Abstract—**An algorithm for pitch frequency measurement for pattern detecting systems of borderline mental disorders is developed. The essence of the algorithm is decomposition of a speech signal into frequency components using an adaptive method for analyzing of non-stationary signals, improved complete ensemble empirical mode decomposition with adaptive noise, and isolating the component containing pitch. A block diagramfor the developed algorithm and a detailed mathematical description are presented. A research of the algorithm using the formed verified signal base of healthy patients, and male and female patients with psychogenic disorders, aged from 18 to 60, is conducted. The research results are evaluated in comparison with the known algorithms for pitch frequency measurement, realized on the basis of the autocorrelation function and its modifications, the robust algorithm for pitch tracking, and the sawtooth waveform inspired pitch estimation. In accordance with the results of the study, the developed algorithm for pitch frequency measurement provides an accuracy increasein determination of borderline mental disorders: for the error of the first kind, on the average, it is more accurate by 10.7%, and for the second type error by 4.7%.**

I. INTRODUCTION

Currently, an assessment of human mental health is a socially significant problem for every state, since it is directly related with the formation of a healthy lifestyle of the population. According to the World Health Organization, current socially significant diseases, which are the main cause of temporary disability, invalidity and mortality, negatively affecting the socioeconomic factors of the state development, are directly related to the mental health of the population [1].

An assessment of borderline mental disorders is of particular importance in those branches of human activity that involve an increased risk to human life and the risk of economic consequences (operators of control systems with a high degree of responsibility: pilots, astronauts, servicemen, airport dispatchers, dispatchers of hazardous production facilities, e.g., nuclear power plants, thermal power plants, chemical industry facilities, etc.).

Recently, the research in the field of mental (psycho-emotional) state assessment has been actively supported by international funds and grants of organizations: Remote Assessment of Disease and Relaps in Central Nervous System Disorders, RADAR CNS (#115902), Foundation/Grant Organization: EU H2020 / EFPIA Innovative Medicines Initiative (IMI); Emotion Sensitive Assistance System, EmotAsS (#16SV7213), Foundation/Grant Organization: BMBF IKT2020-Grant (Sozial- und emotionssensitive Systeme für eine optimierte Mensch-Technik-Interaktion); Promoting Early Diagnosis of Rett Syndrome through Speech-Language Pathology, (#16430), Foundation/Grant Organization: Österreichische Nationalbank (OeNB) Jubiläumsfonds.

II. RELATED WORK

Currently, various experimental and statistical techniques and differentiation of signal processing methods on accessible recording channels of the human body reactions are used for the detection of borderline mental disorders. Methods for evaluation, implemented on the basis of video data reflecting mimic and gestural changes [2], [3]; signals reflecting parameters of physiological activity of a human body (electroencephalography, electrocardiography, electromyography, etc.) [4], [5], [6]; biochemical blood parameters [7], [8]; parameters of handwriting and keyboard writing of texts [9], [10]; parameters of oculography (eye tracking) [11], [12] are of particular interest.

An essential shortcoming limiting a wide practical application of these methods is the obligatory condition of contact recording/sampling/writing, which certainly affects the mental state, which it is no longer possible to effectively evaluate. The most promising and adaptive (in real time and free activity) is the method based on the analysis of speech signals (SS) [13], [14], [15].

III. MATERIALS AND METHODS

A.Informative parameters of speech signals reflecting the borderline mental disorders

The importance of SS analysis for the purpose of diagnosing the nervous systemdisturbances is noted in [16], where the authors showed that the grouping of certain informative parameters reflects the allegedly underlying pathology.

The type and degree of severity of mental disorders are coded into certain informative parameters of SS, called patterns. A review analysis [17], [18], [19], [20], [21], [22] in the field of speech formation, psychology and psycholinguistics has revealed that speech characteristics capable of serving as patterns of psychogenic states (manifested at the level of voice segments, syllables, words and whole sentences associated with the geometric shape and change dynamics of the speech apparatus), can be divided into three main groups: spectral-temporal, cepstral, and amplitude-frequency.

Each group of patterns is designed to describe certain aspects of SS, and finds its application in the detection of borderline mental disorders.

B. Pitch frequency

Speech represents a non-stationary acoustic signal of complex shape, which amplitude and frequency characteristics are rapidly changing in time. Speech consists of voiced and unvoiced sections, being formed, respectively, as a result of periodic and non-periodic oscillations of the vocal cords. Periodic oscillations of the vocal cords are called the pitch (P). The oscillation frequency of the cords is an important informative parameter of speech, called the pitch frequency (PF). From an acoustic point of view, the PF is the first component of the formant frequencies (the harmonic sieve) of speech. Besides PF, the vocal characteristics of speech also are: the P intensity, the dynamics of the P intensity change, the dynamics of the PF change, the PF deviation, and the harmonics intensity ratio to the P intensity.

A peculiarity of the PF in disorders is that the intervals of impulses repetition of the vocal cords are continuously changing within a considerable range. In many cases, the duration of voiced sections of speech is low; a significant part is occupied by transient processes.

C. Approaches for the pitch frequency measurement

The task of the PF measurementconsists ofthe P contour spotting, marking of the P periods, and the PF measuring. At present, a large number of methods for PF measuring are known which in general can be classified as methods in time, frequency, and frequency-time domains.

In the time domain [23], [24], the measurement is carried out by analyzing peaks' distribution, transitions through zero, correlation (autocorrelation, measured and normalized autocorrelation) of the signal oscillogram. Temporal methods are the most accurate, but require careful filtering and setting (work is done only with vocalized sections) of the original SS. The main disadvantage is high sensitivity to the noise level in the signal.

In the frequency domain [25], [26], the measurement is carried out by using the maximum energy values of the spectrum (peaks), and comparing them with frequencies that are multiples of the PF. The main disadvantage of frequency methods is the presence in the considered frequency band the second or third harmonics with higher energy, in addition to the PF.

In the frequency-time approaches [27], [28], the measurement is carried out by analyzing the intended P contour, highlighting the instantaneous maxima of the individual harmonics, and dividing the signal into voiced and unvoiced sections. The disadvantage of the frequency-time methods is high probability of obtaining the instantaneous maximum of energy in the unvoiced sections due to the presence of noise in the SS.

A wide practical application belongs to the approaches implemented on the basis of the autocorrelation function and its modifications ("YIN") [29], robust algorithm for pitch tracking (RAPT) [30], and the sawtooth waveform inspired pitch estimation (SWIPE) [31]. The popularity of these algorithms is due to high functionality, low percentage of gross errors, and the availability of freely distributed software implementations.

Nevertheless, given the irregularity of the organs motor of the speech apparatus under borderline mental disorders, the possibilities of these algorithms are substantially limited. The limitation is due to the use of inefficient and non-adaptive methods for processing of complex non-stationary SS, leading to low accuracy and large errors in the PF measurements.

In this paper, the development of an algorithm for the PF measurement for pattern detecting systems of borderline mental disorders is proposed. This study is the development of previously published papers of the authors [32], [33].

D. Methods for empirical mode decomposition

The research of the SS processing methods has revealed the perspective of using the adaptive technology for analysis of non-stationary signals: empirical mode decomposition (EMD) [34].

The EMD [34] is an adaptive method for analyzing non-stationary signals stemming from nonlinear systems. The EMD produces 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 a mode, or intrinsic mode functions (IMF). An analytical expression of theEMD is as follows:

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

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

As a result of the SS decomposition using the EMD, one IMF may haveoscillatory functions incommensurate in amplitude and frequency scales, or vice versa, commensurate oscillatory functions can appear in different modes. This phenomenon is called 'mode mixing'. To alleviate it, a new method was proposed: the ensembleempirical mode decomposition (EEMD)[35]. The essence of the method is in addition of white noise to the original signal to create new extremes:

$$x_j(n) = x(n) + w_j(n) \tag{2}$$

where $x_j(n)$ are noise copies of the original signal, $w_j(n)$ is the white noise realizations with zero mean unit variance.

An analytical expression of the EEMD is as follows:

where j=1, 2, ..., J is the amount of white noise realizations.

Thus, as a result of the decomposition, more regular IMFs with commensurate scales of oscillatory functions are obtained. The main disadvantage of the EEMD is that the decomposition is not complete. Each noise copy of the original signal $x_j(n)$ is decomposed independently of other realizations, and the residue $r_{ji}(n)=r_{ji-1}(n)-IMF_{ji}(n)$ is computed for each of them at each stage, without connection between different implementations. In addition, residual white noise is observed in the IMF, and various implementations of noise copies can generate different amounts of IMFs, which makes final averaging more difficult.

Another type of the EMDis complementary ensemble empirical mode decomposition (CEEMD) [36] which more qualitatively solves the problem of residual noise, using additional (that is, adding and subtracting) pairs of noise with direct and inverse amplitude values. Nevertheless, the averaging problem remains unresolved, since different noise copies of the original signal can produce different amounts of IMFs.

$$\begin{bmatrix} x_j(n) \\ x_j^*(n) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \times \begin{bmatrix} x(n) \\ w_j(n) \end{bmatrix} \tag{4}$$

where $x_j(n)$ is a noisy signal with white noise, $x_j*(n)$ is a noisy signal with white noise with inverse amplitude values.

Taking into account these shortcomings, a new method called the completeEEMD with adaptive noise (CEEMDAN) is proposed in [37]. The main idea of the method is the addition of controlled noise to the original signal to create new extremes. The first IMF is extracted using the EEMDmethod, averaging the first modes of the signal with white noise:

$$\overline{IMF_1}(n) = \frac{1}{J}\sum_{j=1}^{J} IMF_{j1}(n) = \widetilde{IMF_1}(n) \tag{5}$$

Then the first residue is calculated, independent of the noise realization:

$$r_1(n) = x(n) - \widetilde{IMF_1}(n) \tag{6}$$

For further IMF extracting, a specific noise is addedto the current first residue. This noise is an IMF of white noise, obtained by the EMD method.

Despite the above mentioned advantages of the CEEMDAN method, the authors note its disadvantages in [38]:

- IMFs contain a residual noise.
- Informative modes about the signal during the decomposition are extracted later that using the EEMD with some parasitic modes in the early stages of the decomposition.

In the same paper, the authors solve the noted shortcomings, and propose an improved complete EEMD with adaptive noise.

The essence of reducing the residual noise is the use of

$$x_j(n) = \sum_{i=1}^{I} IMF_{ji}(n) + r_{ji}(n) \tag{3}$$

local mean values instead of IMF values.

The EEMD method independently decomposes each realization of the signal with noise, therefore at the first stage of each decomposition implementation there is one local mean value and one mode.

Given that the true mode can be defined as the difference between the current residual and the averaged value of its local averaged value, we get the following expression:

$$E_1(x(n)) = x(n) - M(x(n)) \tag{7}$$

where M is the operator creating the local average value of the applied signal.

Then, for the first IMFs obtained by the EEMD and the CEEMDAN, the expression will have the following form:

$$\begin{aligned} \widetilde{IMF_1}(n) &= \langle E_1(x_j(n)) \rangle = \langle x_j(n) - M(x_j(n)) \rangle \\ &= \langle x_j(n) \rangle - \langle M(x_j(n)) \rangle \end{aligned} \tag{8}$$

where $\langle . \rangle$ is the averaging action.

Estimating only the local average value and subtracting it from the original signal, we obtain the following expression:equation is centered.

$$\widetilde{IMF_1}(n) = x(n) - \langle M(x_j(n)) \rangle \tag{9}$$

The essence of the elimination of emerging parasitic IMFs in the early stages of decomposition consists in the reduction in the overlap of the scale-energy spaces of the first two modes.

Here is an algorithm and mathematical description of the improved CEEMDAN method.

- Step 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 (8), the first residue $r_1(n) = \langle M(x_j(n)) \rangle$is calculated.
- Step 2. At the first stage, the first mode $IMF_I(n)=x(n)-r_I(n)$ for i=1 is calculated.
- Step 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 $\widetilde{IMF_2}(n) = r_1(n) - r_2(n) = r_1(n) - \langle M(r_1(n) + \beta_1 E_2(w_j(n))) \rangle$is determined.
- Step 4. In the subsequent stages, the i-th residue $r_i(n) = \langle M\left(r_{i-1}(n) + \beta_{i-1} E_i(w_j(n))\right) \rangle$for i=3, ..,I is calculated.
- Step 5. The i-th mode is calculated: $\widetilde{IMF_i}(n) = r_{i-1}(n) - r_i(n)$.
- Step 6. Go to Step 4 for the next value i.

The constants β_i=ε_i std ($r_i(n)$) are selected to obtain the desired signal-to-noise ratio between the added noise and the residue to which the noise is added. Note, that usingthe EEMD, the signal-to-noise ratio between the added noise and the residue is increased by an order of *i*. This is due to the fact that the noise energy in the *i*-th residue, *i*>1, is only a low noise energy added at the beginning of the algorithm operation. To emulate this behavior, in this paper we will set β_0 so that ε_0 is directly opposite to the desired signal-to-noise ratio between the first added noise and the analyzed signal: if we express the signal-to-noise ratio as a standard deviation factor, then we have β_0=$\varepsilon_0$std ($x(n)$)/std($E_1(w_j(n))$). To obtain noise realizations with smaller amplitudes for the last stages of decomposition, we will perceive noise in the remaining IMFs as a result of its pre-processing by the EMD, i.e. without their normalization by standard deviation.

IV. DESCRIPTION OF THE ALGORITHM

The authors have developed analgorithm for the PF measurement to detect the patternsof borderline mental disorders. A block diagram for the algorithm (blocks 1-8) is shown in Fig. 1.

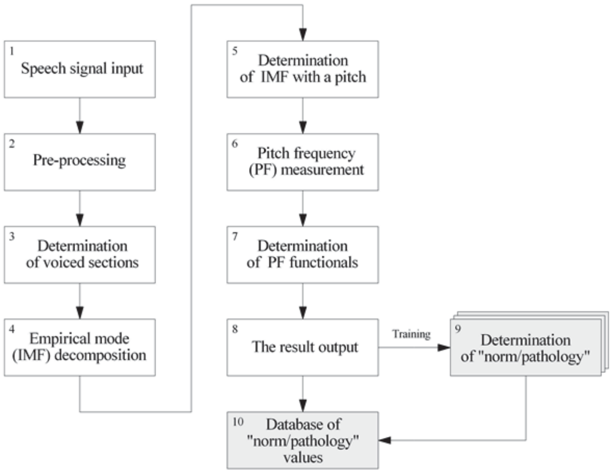


Fig. 1.A block diagram for the algorithm for the PF measurement to detect the patterns of borderline mental disorders

The algorithm decomposes the SS into frequency components, allocates the component containing the P, and measures the PF.

The algorithm works as follows (Fig. 2):

- Processing in block 2 is performed for the SS in the full time interval.
- Processing in block 3 is performed in a sliding window mode with a duration of 15 ms, and an overlap of 7.5 ms.
- Processing in blocks 4-8 is performed in a sliding window mode with a duration of 20 ms (variations are possible),and a 10 ms overlap.

Let's consider each stage of processing in more detail.

1) Input of a speech signal: The input is carried out with the following parameters: sampling frequency is 8000 Hz, quantization capacity is 16 bits.

2) Pre-processing: The removal of a constant component (displacement of the signal relative to zero by a certain constant value), which usually occurs in the analog-to-digital converter, is carried out. In order to remove the constant component or, in other words, to equalize the signal with respect to zero, the arithmetic mean of all signal samples is determined, and subtracted from the original signal.

The next stage of pre-processing is the SS filtering by the fourth-order Chebyshev high-pass filter to remove frequencies below 130 Hz, which include the main rumble, crackle and other noises in this range. Filtering with a frequency cut at 130 Hz does not affect the useful information in the signal [39].

At the end, the correction of natural distortions of the spectrum (minus 6 dB per octave) is carried out, arising in the human speech apparatus during the speech [40].

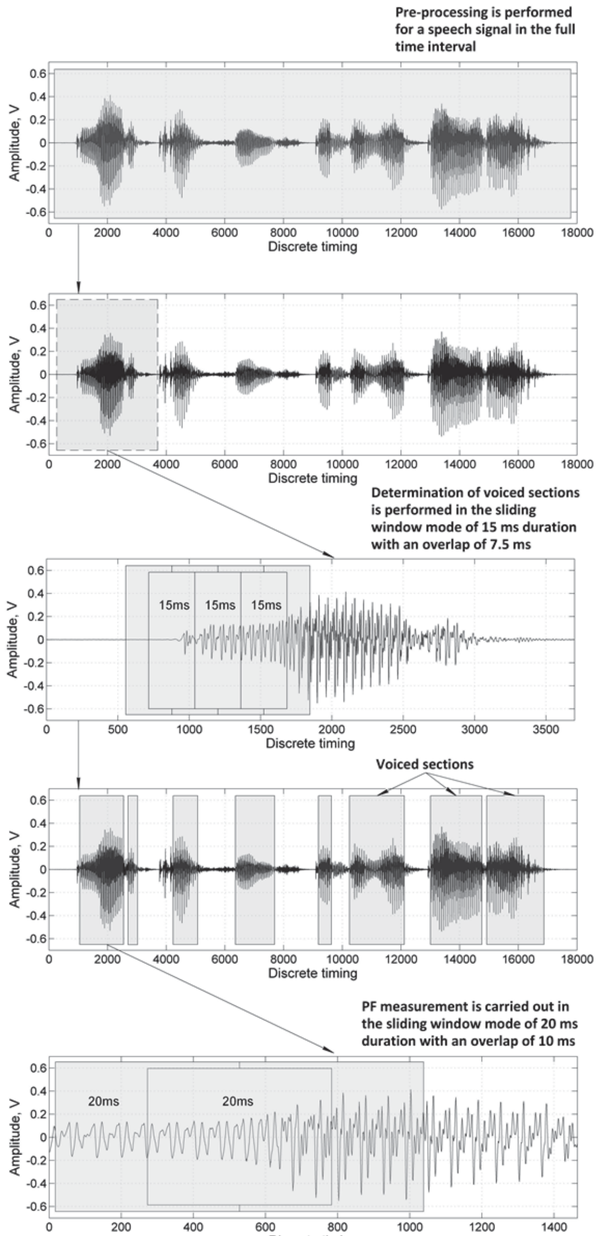


Fig. 2.The algorithm operation

The SS are passed through a correction filter with a transfer function:

$$W(z) = \sum_{k=0}^{m} a_k \, z^{-k} \tag{10}$$

where a_k are constant coefficients, m is an integer (m> 0), k is the coefficient number. Most often m=1, and the transfer function has the form: $W(z) = a_0 - a_1 z^{-1}$

3) Determination of voiced sections: Voiced sections are isolated from the SS using the cluster analysis in the space of the calculated values of Zero-Crossing Rate (ZCR), Autocorrelation Function(ACR), and energy/power (PWR) [41]. The vocalization segments were identified as segments with the highest PWR value and the lowest values of ZCR and ACR. The solutions were smoothed by the fifth-order median filter using the following rule: the voiced segments shorter than 30 ms were classified as segments that did not contain vocalization (formed without the participation of vocal cords); but segments that did not contain vocalization, shorter than 20 ms, were classified as vocalized due to the physiological aspect of speech formation [41].

4) Empirical mode decomposition: On the basis of a detailed analysis of the advantages and disadvantages of various types of decomposition, and taking into account the SS specificity under borderline mental disorders, the authors decided to use the improved CEEMDAN to decompose the signal into frequency components [39].

The result of decomposition of the voiced SS section by the improved CEEMDAN is shown in Fig. 3. Decomposition parameters: the standard deviation of noise is 0.2 mV, the number of realizations is 500, and the maximum permissible number of sifting iterations is 5000.

As it can be seen from Fig. 3, the voiced SS section is decomposed into nine IMFs. The first two modes contain the main noise present in the original signal. The sixth mode and the subsequent ones are low-frequency, and correspond to the trend present in the signal. Valuable high-frequency information associated with the closure of vocal cords appears with the third to fifth IMF.

5) Determination of the IMF with a pitch: Informative modes (IMF3-IMF5) have more energy than the trend modes. The amplitude distribution of IMFs is well described by the short-term energy function. In the developed algorithm, the logarithm of energy is used to compress the signal amplitude in a large dynamic range, maximizing the operation of the algorithm to the work of the human auditory apparatus:

$$LE_i = log_2 \sum_{n=1}^{N} (IMF_i(n))^2 \tag{11}$$

where LE_i is the logarithm of IMF energy.

The process of determination of the IMF with a pitch consists in sequent modulo calculating of the value difference of energy logarithms between the current and subsequent modes:

$$d = |LE_i - LE_{i+1}| \tag{12}$$

where d is the difference between the logarithms energy values of the current and subsequent IMF.

As a result, from the sequence of the obtained d values, the greater of them corresponds to a sharp decrease of energy between the informative IMF containing the P and the trend one [42], [43]. Fig. 4 shows a graphical interpretation of the mode determining process containing the pitch.

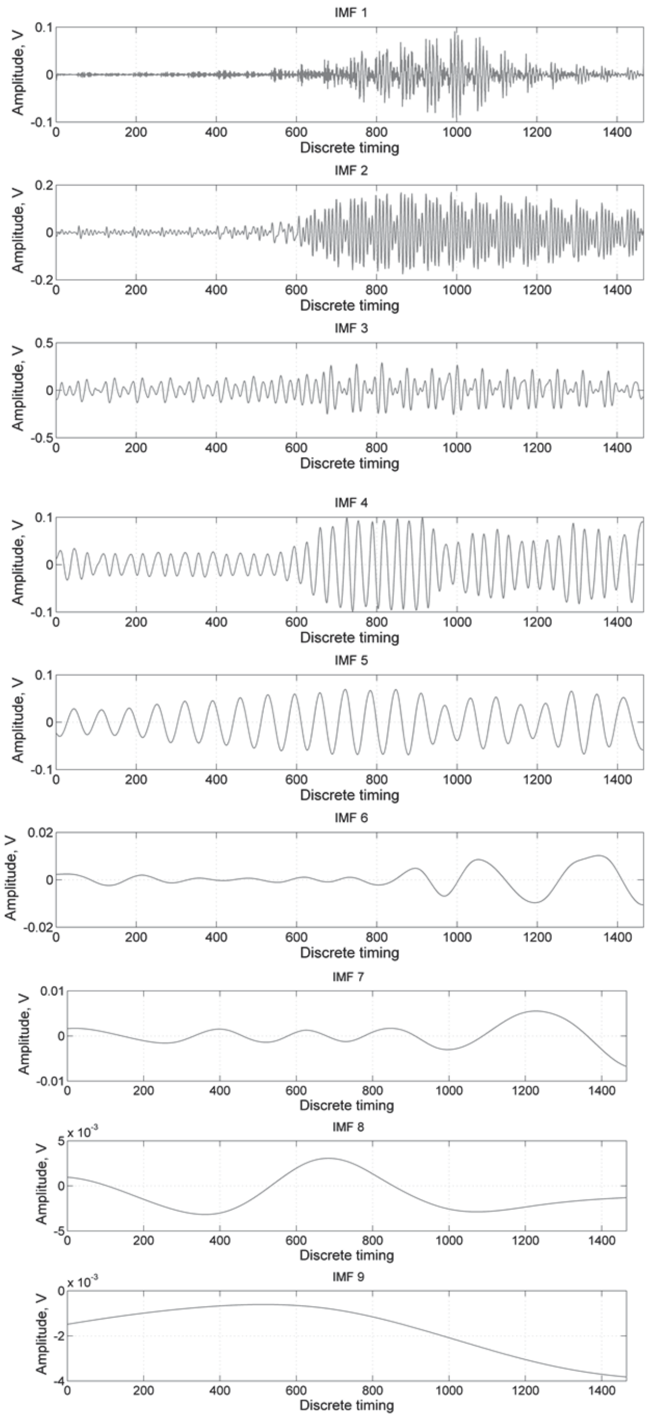
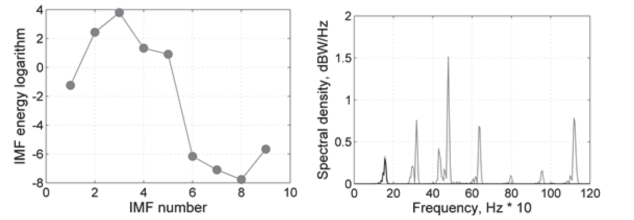


Fig. 3.Decomposition of the voiced SS section by the improved CEEMDAN method

According to the determination rule, it follows from Fig. 4a that the fifth IMF can contain the pitch. The analysis of the spectral distribution of the fifth IMF and the original SS confirms the correctness of the mode determination containing the pitch: the only harmonic PF component of the fifth IMF corresponds to the first component of the harmonic sieve of the original SS (Fig. 4b).



a. Logarithms of the IMF energy

b. Spectral distribution of the original SS and the fifth IMF

Fig. 4.Determination of IMF containing the pitch

Fig. 5 shows the oscillograms of the original signal of the fifth IMF with the pitch. As a measurement unit of the power spectral density, a composite value of dBW/Hz with a reference level of 1 W/Hz (power allocated in the frequency band of 1 Hz wide) was selected.

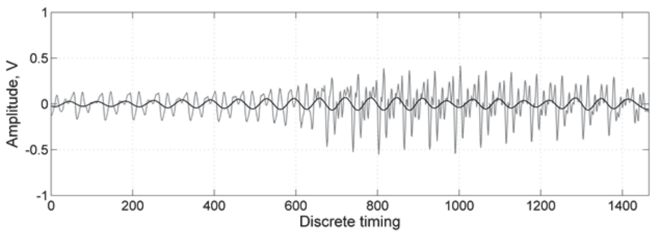


Fig. 5.Oscillograms of the original signal (black color) of the fifth IMF with the pitch (grey color)

6) The pitch frequency measurement: It is carried out using the measuring function of the signal instantaneous energy, the Teager operator, which has simplicity, efficiency, and good susceptibility to the SS change:

$$T(n) = \left(IMF_{i,PF}(n)\right)^2 - IMF_{i,PF}(n-1) \times IMF_{i,PF}(n+1) \quad (13)$$

where $T(n)$ is the function of the Teager operator; IMF_i, $PF(n)$ is the IMF containing the pitch.

Fig. 6 shows the oscillogram and the function of the Teager operator of the fifth IMF.

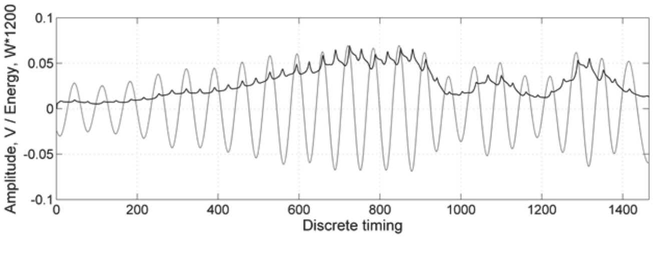


Fig. 6.Oscillogram (black color) and the function of the Teager operator (grey color) of the fifth IMF

To measure frequency, closely spaced maxima and the Teager operator function sare used, between which the difference in discrete time readings is determined; the P period is calculated in seconds, and the PF is in Hz:

$$P_0 = \frac{T_{max}(n+2) - T_{max}(n)}{f_d} \quad (14)$$

$$f_0 = \frac{1}{P_0} \quad (15)$$

where P_0 is the pitch, f_0 is the PF; $T_{max}(n)$, $T_{max}(n+1)$ are the maxima of the Teager operator function; f_d is the sample rate.

*7) Determination of the pitch frequency:*To expand the information space on the P frequency, the following values are determined:

- PH mean value in Hz:

$$f_{0,\text{mean}} = \frac{1}{P}\sum_{p=1}^{P} f_{0,p} \quad (16)$$

where p=1, 2, ..., P is the period number of the PF.

- The maximum max (f_0) and the minimum min (f_0) of the PF values, in Hz.
- Standard contour deviation of the PF:

$$SD_{f_0} = \frac{1}{P-1}\sum_{p=1}^{P}\left(f_{0,p} - f_{0,\text{mean}}\right)^2 \quad (17)$$

- A range of background frequencies:

$$PFR = 12 \times \frac{\log\left(\frac{\max(f_0)}{\min(f_0)}\right)}{\log_2} \quad (18)$$

- Mean absolute jitter value:

$$MAJ = \frac{1}{P-2}\sum_{p=P-1}^{1}\left|f_{0,p-1} - f_{0,p}\right| \quad (19)$$

- A jitter:

$$J = \frac{MAJ}{f_{0,\text{mean}}} \quad (20)$$

- An average relative perturbation of the PF, smoothed over tree P periods:

$$RAP = \frac{\frac{1}{P-2}\sum_{p=2}^{P-1}\left|\frac{f_{0,p+1}+f_{0,p}+f_{0,p-1}}{3} - f_{0,p}\right|}{f_{0,\text{mean}}} \times 100 \quad (21)$$

- The PF perturbation coefficient, smoothed over five P periods:

$$PPQ = \frac{\frac{1}{P-4}\sum_{p=3}^{P-2}\left|\frac{\sum_{k=p-2}^{p+2} f_{0,k}}{5} - f_{0,i}\right|}{f_{0,\text{mean}}} \times 100 \quad (22)$$

8) Output of the results: At this stage of the algorithm, the vector formation of the received PF patterns and its values in a convenient for the further the norm/pathology definition is carried out (Fig. 1 blocks 9, 10 in grey color).

V. Investigation of the Algorithm

A.Description of the speech signal database

To conduct the research of the developed algorithm, a group of patients, and a verified signal base have been formed with the support of K.R. Evgrafov Regional Psychiatric Hospital (Penza, Russian Federation), and Penza State University. The group of patients has been formed according to the clinical disorder state of the following diagnostic headings of the International Classification of Diseases ICD-10: F48.0, F45.3, F43.2, and F41.2.

The group of patients with psychogenic disorders included 100 males and females, aged from 18 to 60, with clearly expressed symptoms. The control group of 100 patients, without signs of borderline mental disorders (conditionally healthy) was also formed. The average age in the experimental group of patients with borderline mental disorders was 40.2 years, and 35.4 years in the control group.

In both groups, women were predominated (75%), agedfrom 40 to 59, and men aged from 50 to 59. The majority of patients were employed (90.8%), among which were employees of enterprises and organizations (65.0%). Smaller shares fell on workers (14.2%), creative workers (12.5%), and a small amount were pupils, students and non-employed. The majority of patients had higher or incomplete higher education (69.2%).

B.The research results

To evaluate the algorithm efficiency, errors of the first and second kind were used. In this study, the task wasto determine the patterns of borderline mental disorders, so the first kind error would be a false assignment of the ‘normal’ SS status, spoken by a person with a psychogenic disorder, and a second-kind error would be a false assignment of the ‘pathology’ SS status, pronounced by a healthy person. The results of the study of the developed algorithm are evaluated in comparison with the PF measurement algorithms, the software implementation of which is available for free access: “YIN”, RAPT, and SWIPE (Table 1).

VI. Conclusion

In accordance with the results obtained, it follows that the developed algorithm by the authors provides an accuracy improvement of the PF measurement: for the error of the first kind,it is more accurate by 8% than for the RAPT algorithm; by 13% than for the “YIN” algorithm, and by 11%, than for the SWIPE algorithm; for the second-kind error, it is more accurate by 4%, 7%, and 3%, respectively.

These results allows to conclude that the developed algorithm based on the improved CEEMDAN method can be successfully used in pattern detecting systems of borderline mental disorders, and introduced into psychiatric clinical practice.

TABLE I. Results of determination of borderline mental disorders

Predictable result	**Determination result**		**Errors of the first and second kind, %**	
	Pathology	Norm		
Robust Algorithm for Pitch Tracking (RAPT)				
Pathology	84 pers.	16 pers.	1st	16
Norm	8 pers.	92 pers.	2nd	8
Algorithm based on autocorrelation function (“YIN”)				
Pathology	79 pers.	21 pers.	1st	21
Norm	11 pers.	89 pers.	2nd	11
Sawtooth Waveform Inspired Pitch Estimation(SWIPE)				
Pathology	81 pers.	19 pers.	1st	19
Norm	7 pers.	93 pers.	2nd	7
Developed algorithm				
Pathology	92 pers.	8 pers.	1st	8
Norm	4 pers.	96 pers.	2nd	4

Acknowledgment

The authors are grateful to the Russian Science Foundation for the financial support of the project “Search for hidden patterns of borderline mental disorders, and the development of a rapid assessment system of human mental health”, No. 17-71-20029.

References

[1] A.Yu. Tychkov, A.V. Ageykin, A.K. Alimuradov, V.B. Kalistratov, and S.Yu. Mitroshina, “Modern ideas about the specifics and problems of diagnosis of borderline mental disorders”, *Mental Health,* vol. 5, 2017, pp. 69-75.

[2] Koelstra, M. Pantic, and I. Patras, “A dynamic texture-based approach to recognition of facial actions and their temporal models”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 32, Nov. 2010, pp. 1940-1954.

[3] A.S. Bobkov, “Models and methods for determining the emotional state of a person based on automated processing and analysis of video information”, PhD thesis, Institute for Systems Analysis of Russian Academy of Sciences, Moscow, 2013. (In Russian)

[4] O.O. Kislova, “Psychophysiological preconditions for successful recognition of emotional speech expression”, PhD thesis, Institute of Higher Nervous Activity and Neurophysiology of Russian Academy of Sciences, Moscow, 2009. (In Russian)

[5] F. Agrafioti, “ECG in biometric recognition: time dependency and application challenges”, PhD thesis, University of Toronto, 2011.

[6] C. Kappeler-Setz, F. Gravenhorst, J. Schumm, B. Arnrich, and G.Tröster, “Towards long term monitoring of electrodermal activity in daily life”, *Personal and Ubiquitous Computing*, vol. 17(2), Feb. 2013, pp. 261-271.

[7] L.A. Farwell, “Brain fingerprinting: a comprehensive tutorial review of detection of concealed information with event-related brain potentials”, *CognitiveNeurodynamics,* vol. 6, 2012, pp. 115-154.

[8] C. Quan and F. Ren, “Sentence emotion analysis and recognition based on emotion words using Ren-CECps”, *International Journal of Advanced Intelligence*, vol. 2(1), Jul. 2010, pp.105-117.

[9] A.V. Zaboleeva-Zotova, Yu.A. Orlova, V.L. Rozaliev, and A.S. Bobkov, “A task of creating a system of automated recognition of emotions”, *in Proc.of the International Conference on Open Semantic Technology for Intelligent Systems (OSTIS-2012)*, 2012, pp. 347-350.

[10] L. Likforman-Sulem, A. Esposito, M. Faundez-Zanuy, S. Clémençon, and G. Cordasco, “EMOTHAW: A novel database for emotional state recognition from handwriting and drawing”, *IEEE Trans. on Human-Machine Systems*, vol.47(2), Apr. 2017, pp. 273-284.

[11] V.A. Barabanschikov and A.V. Zhegallo, “Methods of eye tracking in psychology: educational program”, *Experimental Psychology*, vol. 7(1), 2014, pp. 132-137.

[12] C.M. Loughland, L.M. Williams, and E. Gordon, “Visual scanpaths to positive and negative facial emotions in an outpatient schizophrenia sample”, *Schizophrenia Research,* vol. 55(1-2), May 2002, pp. 159-170.

[13] B. W. Schuller and A. M. Batliner, *Computational Paralinguistics: Emotion, Affect and Personality in Speech and Language Processing*. New York: Wiley, 2013.

[14] A.G. Davydov, V.V. Kiselev, D.S. Kochetkov, and A.V. Tkachenya, “A choice for the optimal set of informative signs for classification of the speaker’s emotional state by voice”, *in Proc.of theInternational Conference on Computational Linguistics and Intellectual Technologies (Dialogue-2012)*, vol. 1(11), 2011, pp. 122-128. (In Russian)

[15] J.R. Williamson, T.F. Quatieri, B.S. Helfer, G. Ciccarelli, and D.D. Mehta, “Vocal and facial biomarkers of depression based on motor incoordination and timing”, *in Proc. of the 4th International Workshop on Audio/Visual Emotional Challenge(AVEC)*, 2014, pp. 65-72.

[16] F.L. Darley, A.E. Aronson, and J.R. Brown, “Differential diagnostic patterns of dysarthria”, *Journal of Speech Language and Hearing Research*, vol. 12, 1969, 47-57.

[17] M.K. Dorry, “Emotion identification from spontaneous communication”, Master thesis, Addis Ababa University College of Natural Sciences, Addis Ababa, Ethiopia, 2016.

[18] M. Schwenzer, E. Zattarin, M. Grozinger, and K. Mathiak, “Impaired pitch identification as a potential marker for depression”, *BMC Psychiatry*, vol. 12(1), pp. 32-37.

[19] M.U. Morist, “Emotional speech synthesis for a radio dj: corpus design and expression modeling”, Master thesis, MTG-UPF, Barcelona, 2010.

[20] Th. Yingthawornsuk, “Analysis of speech based on spectral entropy in detecting depressed among control subjects”, *inProc. of the International Conference on Computer and Communication Technologies (ICCCT’2012)*,May 26-27, 2012.

[21] B. Yu, Th.F. Quatieri, J.R. Williamson, and J.C. Mundt, “Cognitive impairment prediction in the elderly based on vocal biomarkers”, *in Proc. of the 16th Annual Conferenceof the International Speech Communication Association*, 2015, pp. 3734-3738.

[22] S. Grossberg, *The Adaptive Self-Organization of Serial Order in Behavior: Speech, Language, and Motor Control.* In E.C. Schwab and H.C. Nusbaum (Eds.), *Pattern Recognition by Humans and Machines: Speech Perception.* New York: Academic Press, 1986.

[23] M. Staudacher, V. Steixner, A. Griessner, and C. Zierhofer, “Fast fundamental frequency determination via adaptive autocorrelation”, *EURASIP Journal on Audio, Speech, and Music Processing*, Dec. 2016, pp. 8.

[24] S. Kumar and M. Jangra, “Efficient time domain fundamental frequency estimation (pitch estimation)”, *International Journal of Electronics &Communication Technology*,vol. 6(3), Jul.-Sept. 2015.

[25] S. Eun-Joo and H. Kwang-Seok, “Feature extraction method for predicting depression by frequency domain analysis”, *in Proc. of the International Conference on Civil, Materials and Environmental sciences (CMES 2015)*, pp. 600-603.

[26] Y. Haraa, M. Matsumotob, and K. Miyoshi, “Method for estimating pitch independently from power spectrum envelope for speech and musical signal”, *Journal of Temporal Design in Architecture and the Environment*, vol. 9(1), 2009, pp. 121-124.

[27] C. Shahnaz, W.-P. Zhu, and M.O. Ahmad, “A spectro-temporal algorithm for pitch frequency estimation from noisy observations,” *in Proc. of IEEE International Symposium on Circuits and Systems (ISCAS2008)*, May 18-21, 2008, pp. 1704-1707.

[28] A.P. Klapuri, “Multiple fundamental frequency estimation based on harmonicity and spectral smoothness”, *IEEE Trans. on Speech and Audio Processing*, vol. 11(6), Nov. 2003, pp. 804-816.

[29] A. Cheveigne and H. Kawahara, “YIN, a fundamental frequency estimator for speech and music”, *TheJournal of the Acoustical Society of America*, vol. 111(4), Apr. 2002, pp. 1917-1930.

[30] D. Talkin, *A Robust Algorithm for Pitch Tracking (RAPT).* In W.B. Kleijn and K.K. Paliwan (Eds.), *Speech Coding & Synthesis*, Elsevier Science, 1995.

[31] A. Camacho and J.G. Harris, “A sawtooth waveform inspired pitch estimator for speech and music”, *TheJournal of the Acoustical Society of America*,vol. 123(4), Sep. 2008, pp. 1638-1652.

[32] A.K. Alimuradov, “Research of frequency-selective properties of empirical mode decomposition methods for speech signals’ pitch frequency estimation”, *in Proc. of the International Conference on Engineering and Telecommunication (EnT)*, Nov. 18-19, 2015, pp. 77-79.

[33] A.K. Alimuradov, A.Yu. Tychkov, A.V. Ageykin, P.P. Churakov, Yu.S. Kvitka and A.P. Zaretskiy, “Speech/pause detection algorithm based on the adaptive method of complementary decomposition and energy assessment of intrinsic mode functions,” *in Proc. of the XXth IEEE International Conference on Soft Computing and Measurements (SCM),* May 24-26, 2017, pp. 610-613.

[34] N.E. Huang, Sh. Zheng, and R.L. Steven, “The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis,” *inProc. of the Royal Society of London,* vol. A454, 1998, pp. 903-995.

[35] W. Zhaohua and N.E. Huang, “Ensemble empirical mode decomposition: A noise-assisted data analysis method”, *Advances in Adaptive Data Analysis*, vol. 1(1), 2009, pp. 1-41.

[36] J.-R. Yeh, J.-S. Shieh, and N.E. Huang, “Complementary ensemble empirical mode decomposition: A novel noise enhanced data analysis method”, *Advancesin Adaptive Data Analysis*, vol. 2(2), 2010, pp. 135-156.

[37] M.E. Torres, M.A. Colominas, G. Schlotthauer, and P. Flandrin, “A complete Ensemble Empirical Mode decomposition with adaptive noise”, *in Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP-11)*, pp. 4144-4147.

[38] M.A. Colominasa, G. Schlotthauera, and M.E. Torres, “Improved complete ensemble EMD: A suitable tool for biomedical signal processing”, *Biomedical Signal Processing and Control,* vol.14, 2014, pp. 19-29.

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

[40] G.K. Fant, *Acoustic Theory of Speech Formation*. Moscow: Nauka, 1964. (In Russian)

[41] J. Hlavnička, R. Čmejla, T. Tykalová, K. Šonka, E. Růžička, and J. Rusz, “Automated analysis of connected speech reveals early biomarkers of Parkinson’s disease in patients with rapid eye movement sleep behaviour disorder”, *Scientific Reports*, vol.7(12), 2017, p. 13.

[42] A.K. Alimuradov, “An algorithm for measurement of the pitch frequency of speech signals based on complementary ensemble decomposition into empirical modes”, *Measurement Techniques*, vol. 59(12), Mar. 2017, pp. 1316-1323.