Intelligent processing of stuttered speech

Andrzej Czyzewski[1], Andrzej Kaczmarek[1] and Bozena Kostek[1,2]

[1)] *Gdansk University of Technology, Sound & Vision Engineering Dept., Gdansk, Poland*

[2)] *Institute of Physiology & Pathology of Hearing, Warsaw, Poland*

Abstract: The process of counting stuttering events could be carried out more objectively through the automatic detection of stop-gaps, syllable repetitions and vowel prolongations. The alternative would be based on the subjective evaluations of speech fluency and may be dependent on a subjective evaluation method. Meanwhile, the automatic detection of intervocalic intervals, stop-gaps, voice onset time and vowel durations may depend on the speaker and the rules derived for a single speaker might be unreliable when trying to consider them as universal ones. This implies that learning algorithms having strong generalization capabilities could be applied to solve the problem. Nevertheless, such a system requires vectors of parameters, which characterize the distinctive features in a subject's speech patterns. In addition, an appropriate selection of the parameters and feature vectors while learning may augment the performance of an automatic detection system.

The paper reports on automatic recognition of stuttered speech in normal and frequency altered feedback speech. It presents several methods of analyzing stuttered speech and describes attempts to establish those parameters that represent stuttering event. It also reports results of some experiments on automatic detection of speech disorder events that were based on both rough sets and artificial neural networks.

INTRODUCTION

Stuttering is the subject of interest of researchers from many various domains like speech physiology & pathology, psychology, acoustics, signal analysis. Therefore, this area is basically an interdisciplinary field of science. One of the main problems still unsolved in the domain of speech fluency disorders is an objective and an automatic way of judgement of patient performance before and after speech therapy sessions and an assessment of gains made after intervention. Generally, classification of speech disorders is considered as a very difficult and complex problem, however some typical artifacts associated with stuttering are commonly recognized. Stuttering is a poorly understood communication disorder with 1% global prevalence. Analysis of stuttered speech includes (a) mean duration of sound/syllable repetition and sound prolongation, (b) mean number of repeated units per instance of sound/syllable and whole-word repetition, and (c) various related measures of the frequency of all between- and within-word speech dysfluencies (Zebrowski, 1991). Other research concludes that stuttering episodes affect the intensity-time profile of the speech in their vicinity and that listeners can use this acoustic information to recognize the presence and type of the stuttering (Howell and Wingfield, 1990). Despite the fact that some researchers have several attempts to use objective methods to evaluate patients' progress in speech therapy (Howell, Hamilton & Kyriacopoulos 1986; Howell & Vause 1986; Kuniszyk-Jozkowiak, 1996; Robb and Blomgren, 1997; Howell, Au-Yeung et. al. 1997; Howell, Sacking & Glenn 1997a and 1997b; Michaelis *et al.*, 1998; Archibald and de Nil, 1999; Howell *et al.*, 1999), these methods do not easily allow for automatic assessment of the severity of stuttering. As a result, still, the most common way to do that is employing a speech therapist as an expert to count manually dysfluencies in patient's speech (Kalinowski *et al.*, 1995; Howell *et al.*, 1998; Howell *et al.*, 1999;). It should be mentioned that experts employed to categorize stuttered events are usually people not only having phonetic or linguistic training, but also could be well experienced in such tasks as speech production and analysis (Howell *et al.*, 1998). A panel of experts would usually be employed, therefore reducing the level of inter-rater reliability.

The most common way to assess dysfluency is to transcribe the recorded speech and to locate occurrences of repetitions, syllable/word injections, prolongation, etc. speech (Howell *et al.*, 1998). In some research, audiovisual cues are also used in order to better classify occurrence of stuttering. On the other hand, good results are achieved using motor measures (Archibald and de Nil, 1999; Howell *et al.*, 1995). The kinematic measures that are most often examined are movements of lips, jaw and tongue. In a study carried out by Archibald and de Nil, it was indicated that movement of jaw is the best measure of relationships between speech motor deficiencies and stuttering severity, allowing easy detection of differences between fluent and stuttering speakers (Archibald and de Nil, 1999). On the other hand, from an acoustic point of view, it is possible to analyze an electric signal representing disordered speech. The results of such an analysis can provide information on the process of articulation, and consequently can form the basis for an in-depth diagnosis of the patient. Other studies have been concerned with speech signal features (Kaczmarek and Skorka, 1997; Howell and Wingfield, 1990; Kuniszyk-Jozkowiak, 1995, 1996). The speech segment durations measured are intervocalic intervals, stop-gaps, voice onset time, and vowel durations. Speech envelopes have been analyzed by Howell and Wingfield (Howell and Wingfield, 1990) resulting in some parameters that were later used in an automatic search for lexical dysfluency (Howell *et al.*, 1998). Recently, Kuniszyk-Jozkowiak compared areas under the speech envelopes of utterances of fluent and stuttering speakers (Kuniszyk-Jozkowiak, 1996).

In order to assist in therapy for stutterers, electronic devices have been designed and constructed based on auditory feedback. The most often used therapy of this type is Delayed Auditory Feedback (DAF) (Lee, 1950). The effects of using DAF have been studied thoroughly in many research institutes worldwide, and can be briefly described as a simultaneous reduction of stuttering and slowing down of speech (Kuniszyk-Jozkowiak, 1995; Lee, 1950; Kalinowski *et al.*, 1995).

Among more recent devices of this type is one based on a signal processor called a DSA (Digital Speech Aid) (Czyzewski *et al.*, 1993; Czyzewski *et al.*, 1994; Roland-Mieszkowski *et al.*, 1995). One of the algorithms used in this device is signal transposition in the spectrum domain (FAF – Frequency Altered Feedback). For many patients, this method has shown a high reduction of stuttering, especially in the case of text reading (Howell 1987, Czyzewski and Skorka, 1996). The small digital speech aid (see Fig. 18) was exploited in those experiments (Roland-Mieszkowski, Czyzewski & Kostek 1995). The research was designed to test various algorithms and measure their effectiveness. Stutterers' speech was therefre recorded and analyzed in detail. Consequently, changes in the vocal pitch in disordered speech before and after therapy were observed (Howell and Williams 1988 & 1992, Kaczmarek and Skorka, 1997). It was found that the frequency of the vocal tone of speech corrected using the FAF method is lower than before the correction. The difference of frequency ranges from about 5 to 10%. Furthermore, speech corrected in this way does not give the impression of being slowed down. Additional acoustic analyses were also performed in order to look for other significant changes in disordered speech features. The results provided the impetus to conduct studies on the possibility of automatic detection of speech dysfluencies by the computer as a possible diagnostic tool.

So far, however, studies on the vocal pitch (also called formant F_0) and on higher formants have been but a fraction of speech disorders research (Howell & Williams, 1992; Czyzewski and Skorka, 1996; Kaczmarek and Skorka, 1997; Robb and Blomgren, 1997). For this reason, they still have not been used diagnostically in speech pathology and therapy. It seems, though, that vocal pitch and higher formants make up an important element of objective studies on speech in which stuttering occurred. They can provide information on the articulation mechanisms present in stuttering, which so far have only been researched using other methods.

The fact is that in stuttered speech, contractions of articulation muscles cause changes in the speech articulation system that are visible in the results of spectral and cepstral analysis. These observations were used as a basis for the concept of analysis of stuttered speech presented here. The analysis performs segmentation of the speech signal and parametrization of the segments obtained. The parameters are constituted by the frequency of the vocal tone and the frequencies and amplitudes of the formants. Feature vectors containing sequences of parameter values were then subjected to correlation analysis and then used a training material for some intelligent algorithms as rough sets and neural networks.

As a result of the correlation analysis, one can obtain information about the behavior of formants based on the spectrum of the signal of stuttered speech signal, in addition to information about the mutual relations between the sequences of parameters obtained in the course of defective articulation. These relations can be expressed in the form of a correlation matrix and a matrix of coefficients of t-Student statistics, which indicate the significance of the obtained values of correlation coefficients. The relationships obtained in this way provide characteristics of the particular cases of speech disorders, and after some generalization over the whole researched population they can be used for further analyses that facilitate diagnosis in the case of speech dysfluency. In turn, the application of intelligent algorithms allows for the automatic detection of stuttering artifacts.

Also presented in the paper is an example of an analysis of frequency altered feedback speech obtained using the DSA electronic speech corrector (Czyzewski *et al.*, 1993; Czyzewski *et al.*, 1994).

I. METHODOLOGY OF ANALYSIS OF VOCAL TONE IN STUTTERED SPEECH

The presence of vocal tone is a substantial feature of voiced sounds of speech. This tone determines the excitation of the vocal tract. The voicing is caused by vocal cord vibrations. The shape of the vocal tract determines the resonances in it, and consequently the timbre of the resulting sound. Moreover, changes of its shape are determined by articulation. The vocal pitch constitutes the pitch of the speech signal. In stuttered speech, there are spontaneous, sometimes periodic muscle cramps, causing blocking or other disturbances of speech. This may be easily observed by analyzing the speech signal only in the time domain in order to get some data for statistical analysis (Kuniszyk-Jozkowiak, 1996; Robb and Blomgren, 1997). Analysis of the stuttered speech signal may also be useful for assessing different types of stuttered speech correction systems. In this case, analysis in the frequency domain is recommended. Cepstral analysis and cepstral smoothing may be applied to detect vowels, and assist in the study of such speech disorders that contain a syllable or vowel repetition and vowel prolongation. The analysis of the vocal tone in dysfluent speech is recommended especially with regard to estimating its quantitative changes.

Investigation of the changes in pitch of the vocal pitch was based on a modified cepstral analysis, which consists of the following steps:

1. compression of the dynamics of the spectrum,
2. reduction of the spectrum band for cepstral analysis,
3. estimation of the frequency of the vocal tone.

These modifications lead to a better discernibility of the cepstral maximum that is the consequence of the presence of vocal pitch in the analyzed speech.

Compression of the dynamics of the spectrum is made by imposing an envelope on maxima and minima of the spectrum according to the following formulae proposed by the authors:

$$G_i = \begin{cases} A_i & for \quad A_i > G_{i-1} \\ G_{i-1} + \left[1 - exp\left(-\frac{r_f}{c_f}\right)\right] \cdot \left(A_i - G_{i-1}\right) & for \quad A_i \le G_{i-1} \end{cases} \tag{1}$$

$$D_i = \begin{cases} A_i & for \quad A_i < D_{i-1} \\ D_{i-1} + \left[1 - exp\left(-\frac{r_f}{c_f}\right)\right] \cdot \left(A_i - D_{i-1}\right) & for \quad A_i \ge D_{i-1} \end{cases} \tag{2}$$

where:

G_i - values of the upper envelope,

D_i - values of the lower envelope,

A_i - values of the spectrum logarithm,

i - numbers of spectrum coefficients (i = 0, 1, 2 etc.)

r_f – frequency analysis resolution

c_f – spectral components integration constant (r_f/c_f ratio was set to –1/10)

The obtained G_i and D_i are later used to normalize the spectrum according to the formula:

$$N_i = \frac{A_{max} - A_{min}}{G_i - D_{i-1}}\left(A_i - D_{i1}\right) + A_{min} \tag{3}$$

where:

A_{max} - maximum of the spectrum logarithm,

A_{min} - minimum of the spectrum logarithm.

An example showing the effects of these operations is presented in Figs. 1-4. Calculating the cepstrum is the next step. The results may be presented as plots, for example such as those shown in Figs. 5 and 6 which show the difference between both cepstral analyses in the normalization process. The cosine transform was used for these analyses.

The vocal tone frequency is estimated as a center of gravity of a fragment of the calculated cepstrum:

$$\hat{f} = \frac{1}{r_c} \cdot \frac{\sum_{i=m}^{n} W_i}{\sum_{i=m}^{n} i \cdot W_i} \tag{4}$$

where:

$\hat{f}$ - estimated frequency,

r_c - resolution of cepstral analysis,

W_i - cepstrum coefficient number *i*,

m, n - numbers of the cepstral coefficient which contain the maximum related to the vocal pitch.

In this work, an assumption was made that *m=k-1* and *n=k+1*, where *k* is the number of the maximum cepstral coefficient. Using the sequence of values calculated from the successive frames of the speech signal, one can obtain results showing the evolution of the estimated frequency. An example of such an operation is shown in Fig. 7.

Besides the vocal tone a vital part of stuttered speech analysis is formant tracking procedure which is described in the next paragraph.

II. METHODOLOGY OF ANALYSES OF FORMANTS IN STUTTERED SPEECH

For the formant analysis monophonic digital recordings, with a 16-bit resolution at the sampling frequency 22.05 kHz, were used in the experiments. The analyzed signal was pre-emphasized with 6 dB/octave without the compression of spectrum dynamics. Computation of the formants was made using the cepstral method of spectrum smoothing based on Cosine Transform. The following parameters were used:

- length of frame equals 1024 samples,
- length of overlap equals 583 (the step of analysis is therefore 441 samples that corresponds to 20 ms signal portion),
- frequency band used to smooth the cepstrum = 5512.5 Hz (256 samples of spectrum module coefficients),
- Hamming's window,
- smoothing order in the cosine transform was set as equal to 12.

The frequencies of the particular formants were computed by interpolating second order polynomials for a cepstral smoothed spectrum. This method searches for local maxima of the smoothed spectrum and computes the value on the frequency axis for which the interpolated function reaches maximum. This method is illustrated in Fig. 8.

The general formula for x_{max} is derived from the square function, with the assumption that for $x=k$, the measurement data reach a local maximum. Therefore, using a shift to the left in the coordinate system one can express k as:

$$x_{max} = -\frac{b}{2a} = \frac{y(k-1) - y(k+1)}{2\left(y(k+1) - 2y(k) + y(k-1)\right)} \tag{5}$$

where:

a, b – coefficients at x^2 and x of the polynomial

$y(k)$ – cepstral smoothed spectral components

Due to the shift of the system and considering the resolution of the spectral analysis, one can finally express the formula for the frequency of the analyzed formant:

$$f_{max} = r_f \cdot \left(k + \frac{y(k-1) - y(k+1)}{2\left(y(k+1) - 2y(k) + y(k-1)\right)} \right) \tag{6}$$

It needs to be stressed that the stuttered speech cases of prolonged speech signals are best suited for the evolutionary analysis of the shape of the vocal tract. This is due to the fact that in these cases one is dealing with a relatively long section of time during which the examined parameters change relatively slowly, in turn allowing for easier objective measurements of their evolution. This is the reason why these particular utterances were selected for the purpose of the carried out study.

As a result of experiments, a database was obtained which contained sets of vectors of dysfluent speech parameters, including parameters of excitation, such as the frequency of the vocal tone and frequencies and amplitudes of the particular formants that represent the vocal tract structure. Each of the vectors consists of 7 parameters, with the particular elements of each vector as follows:

1. frequency of the vocal tone,
2. amplitude of the first formant,
3. frequency of the first formant,
4. amplitude of the second formant,
5. frequency of the second formant,
6. amplitude of the third formant,
7. frequency of the third formant.

An example of an analysis result, presented graphically, for a stuttered utterance of the "vowel prolongation" type is shown in Fig. 9, and the time sequence of the frequencies of the vocal tone and the frequencies of the first formant are shown in Fig. 10. As is seen from these figures, formant frequencies of the vowel are not stable during its prolongation revealing easily discernible pitch modulation effects.

III. SIGNAL ANALYSIS OF "VOWEL PROLONGATION" AND "VOWEL REPETITION" TYPE CASES

Vovel prolongation is one of the many features that characterizes stuttered speech. Due to the relatively long duration of voiced vowel articulation, prolonged utterances are easier to analyze in an evolutionary way in this case using cepstral analysis. The result is a sequence of cepstral analysis coefficients which are characterized by a distinct maximum for higher cepstral coefficients, which are characterized using the fundamental frequency, i.e. the frequency of the vocal tone. Examples of results of such analysis are shown graphically in Fig. 11.

Cases of "vowel repetition" or "syllable repetition" are more difficult to analyze than the above stated cases of vowel prolongation. This is due mainly to the fact that the duration of their articulation is much shorter, and for this reason the amount of information obtained from the analysis of single recordings of these speech disorders is much smaller. However, this problem is worth solving because of their frequency of occurrence which is significantly larger than of "vowel prolongation" cases in speech disorders related to stuttering. Consequently, using a longer sequence of analyses, one can obtain results that provide speech profile of the patients, which will support diagnosis and therapy. An example of the analysis of a typical case of syllable repetition in terms of the pitch of the vocal tone is shown in Fig. 12.

In addition, a thorough examination of the frequency of the vocal tone and the frequencies and amplitudes of the particular formants was made. Single repetitions were characterized by a single vector of parameters according to the concept presented above. An example of a result of such analysis for a speech disorder of the "repetition" type is shown graphically in Fig. 13 (see also Table III in the paragraph 7). For the purpose of comparison, Fig. 14 also shows the result of the analysis of a word uttered using the DSA electronic speech corrector. The algorithm exploited was, Frequency Altered Feedback (FAF), providing spectral transposition of signals in the frequency domain by a reduction of 1/4 of an octave (Czyzewski *et al.*, 1993).

IV. CORRELATION ANALYSIS

Analyses of speech with the "prolongation" type of disorder allow a quantitative formulation of particular utterances in terms of the degree of correlation between the obtained parameters and the actual performance of articulation muscles, which directly affect the speech signal. In stuttered speech, one can observe oscillations of formant parameters both in amplitudes and frequencies, and sometimes even the disappearance of formants (as illustrated in Figs. 13 and 14). Moreover, additional formants frequently become present in frequency ranges that are not related to the uttered vowel. Despite these disturbances, one can perform correlation analysis for parameters related to this type of speech disorder, obtaining a matrix of the correlation which includes normalized correlation coefficients for all parameters of the analyzed database. The normalized correlation coefficients were computed using the formula:

$$r_{pq} = \frac{\sum_{k=1}^{n}\left(P_k - \overline{P}\right)\cdot\left(Q_k - \overline{Q}\right)}{\sqrt{\sum_{k=1}^{n}\left(P_k - \overline{P}\right)^2}\cdot\sqrt{\sum_{k=1}^{n}\left(Q_k - \overline{Q}\right)^2}} \tag{7}$$

where:

n - number of vectors of parameters,

$\overline{P}$, $\overline{Q}$ - arithmetic means for parameters P_k and Q_k, k=1..n.

On the basis of the formula (7) coefficients of correlation matrices were calculated. To prove the significance of the obtained correlation, t-Student distribution coefficients were additionally computed at n-2 degrees of freedom according to the formula:

$$t_{pq} = \frac{r_{pq}}{\sqrt{1 - r_{pq}^{\;2}}}\cdot\sqrt{n-2} \tag{8}$$

Examples of such an analysis, which illustrate the relation between correlated parameters of stuttered speech (the case of vowel "prolongation" for which cepstrally smoothed spectrum was presented in Fig. 9) are shown in Fig. 15. The largest value of the correlation coefficient is equal to - 0.800 and the value of the coefficient computed using formula (8) is equal to -6.391, thus it signifies strong correlation.

V. COMPARISON OF FORMANTS FOR STUTTERED AND CORRECTED SPEECH

Parameters similar to those discussed above may also be computed for speech disorders of the "vowel repetition" or "syllable repetition" type. For purposes of comparison, similar analyses were made for corrected speech using two different methods of speech signal modification in the auditory feedback loop – DAF (Delayed Auditory Feedback) and FAF (Frequency Altered Feedback).

The purpose of the studies was to determine the type and size of the effects introduced by the speech corrector on the formants in the vowel spectrum, which is the consequence of the way speech is articulated. The results of the presented studies may be a contribution to an automatic diagnosis of speech dysfluency.

Stutterers are a heterogenous population, and therefore present with different speech profiles. For example among stutterers are so called clonic and tonic cases. Clonic (oscillatory) stuttering is usually related to speech elements repetitions and tonic (stationary) stuttering with prolongations. The subjects employed in testing represented both types of stuttering with medium intensity.

In speech disorders, one can observe oscillations of formant parameters in terms of amplitudes and frequencies. In fluent speech, on the other hand, (within syllables) these oscillations are much smaller (refer to Figs. 16 and 17). Moreover, in stuttered speech additional formants frequently become present in frequency ranges that are not typical for the uttered vowel (compare Figs. 16 a-e and Figs. 17 a-h). One can also observe the fading of formants typical for a given vowel. Naturally, this is the effect of an uncontrolled block of articulation muscles, which affects the shape of the vocal tract. The result is that the articulated speech is usually not very comprehensible. In some situations, one can even hear a vowel that is clearly different from the one that was supposed to be uttered. It needs to be noted, however, that the formants distribution in speech is an individual feature and may be used to identify speakers.

The above observations illustrate the scope of interpretation problems that come up when attempting to draw generalized conclusions. Examples of results from the above analyses for speech correction of the FAF type are shown in Fig. 16, and of the DAF type in Fig. 17. The settings on the DSA speech corrector were as follows: time delay ranged from 50 to 100 ms for the DAF type of feedback, and spectrum shift was set to 1/4 of an octave downwards for the FAF type of feedback.

VI. AUTOMATIC DETECTION OF STUTTERING ARTEFACTS

The automatic classification of data needs some preprocessing stages, such as parametrization and discretization. The first of these was already described in previous paragraphs. This procedure is aimed at reducing the amount of data associated with digital sound samples, and it results in feature vectors. Parameters obtained as a result of this process can directly feed the inputs of classification systems, such as artificial neural nets, even if they consist of real numbers. However, other classification algorithms which are based on rules (i.e. rough set-based classifiers) demand the discretization at the data preprocessing stage. During the learning phase, a number of rules are produced, on the basis of which the testing phase is then performed. The generated rules are of the following form:

(*param*_1)=(*value*_1) and...and (*param*_*k*)=(*value*_*k*) => (*class*_*i*) (9)

The produced rules contain parameter values and their number should be reasonably limited. Meanwhile, the number of rules generated on the basis of real valued parameters will be very large and may contain very specific values. For this reason, the discretization (quantization) process is needed (for further details on quantization see ***Appendix***).

A number of experiments related to automatic detection of speech disorders have been carried out. For this purpose, two learning algorithms were employed. The rough sets based algorithm was prepared and exploited at the TU of Gdansk (Czyzewski, 1996; Kostek, 1999). The idea of rough set was introduced in early 80s by Pawlak. From that time, the notion of rough set has been exploited in many domains and provided an effective tool for extracting knowledge from databases. (Pawlak, 1982; Slowinski, 1992; Polkowski and Skowron, 1998). Some principles of the rough set theory and on discretization processes are presented in ***Appendix***. Additionally, for comparative purposes, neural networks were applied to automatic detection of stop-gaps. In all cases, leave-one-out tests were carried out in order to check the effectiveness of tested systems. Exemplary experiment results are presented below.

A. Automatic detection of stops-gaps

Sound files were chosen which contained fluent and disordered speech patterns (with stop-gaps). The sampling frequency was 22.05 kHz and the length was equal to 25 000 samples (1.13 s of duration). There were six examples of fluent speech and six ones of speech with stop-gaps. Thus, the patterns belonged to two classes: FLUENT_SPEECH and STOP_GAP. The examples were taken from both clonic and tonic medium-intensity stutterers' utterances. The recordings were divided into segments of the size of 5 000 samples (0.22s) or of 2 500 samples (0.11s). Next, parametrization took place and evaluated parameters were stored in one of two pattern sets: *Detection of stop-gaps* (5) with 5-position vectors or *Detection of stop-gaps* (10) with 10-position vectors, respectively. Both files included 12 vectors. During experiments, various settings of the implemented rough set system were used, the range of quantization (see ***Appendix***) from 2 to 10 was tested.

B. Discerning vowel prolongations

The algorithm was based on the selection of parameters describing vowels. Sound files containing individual vowels from various speakers were prepared for the analysis. The length of the files was equal to 1024 samples and the sampling frequency was 22.05 kHz. Two different pattern sets were prepared. The first one (*Discerning vowel prolongations* (1)) included patterns for six Polish vowels edited from a medium-intensive tonic stutterer voice. There were six examples of utterance for each vowel which resulted in 36 feature vectors constituted with 6-parameters. The second pattern set (*Discerning vowel prolongations* (6)) was composed of parameters for five Polish vowels taken twice from six different clonic speakers' utterances. Thus, the set consisted of 60 vectors constituted with six parameters.

C. Detection of syllable repetitions

In the case of syllable repetitions, sound files were of the length of 25000 samples and the sampling frequency was equal to 22.05 kHz. The recordings belonged to two classes: FLUENT_SPEECH and REPETITION. The 10 examples for the class REPETITION were edited from two medium-intensive clonic stutterers' voices. In turn voices and 10 examples for the class FLUENT_SPEECH were obtained from two speakers. Hence, the pattern set (called *Detection of syllable repetitions*) contained 20 vectors. During the tests, rough set system settings were selected as follows: the value of quantization was from the range of 5 to 20 and the value of neutral point was from 0.6 to 0.9 (see ***Appendix***).

D. Application of rough sets to recognition tasks

Table I contains data concerning decision rules generated by the rough set system. The following abbreviations and symbols were used:

q – number of quantization intervals,

L_{max} - maximum rule length,

μ_{RSmin} - minimum rough-measure,

n_{μ} - neutral point,

C/I/D - number of correctly classified vectors/number of incorrectly classified vectors/ number of vectors without decision,

[%] - percentage ratio of correct decisions.

TABLE I. Results of automatic detection of some stuttering events by rough sets.

Pattern set	Parameters				Best result		Certain rules	
	q	L_{max}	μ_{RSmin}	$n\mu$	C/I/D	[%]	C/I/D	[%]
Detection of stop-gaps (5)	2	2	0.50	0.60	10/2/0	83.33	8/2/2	66.67
Detection of stop-gaps (10)	6	2	0.50	0.60	11/1/0	91.67	11/1/0	91.67
Discerning vowel prolongations (1)	7	6	0.50	0.60	35/1/0	97.22	35/1/0	97.22
Discerning vowel prolongations (6)	6	2	0.50	0.60	58/2/0	96.67	56/2/2	93.33
Detection of syllable repetitions	7	2	0.55	0.85	18/2/0	90.00	13/6/1	65.00

As is seen from Table 1, the recognition scores depend on the system settings. This is especially visible in the case of stop-gaps detection. When the number of quantization intervals was small, then the recognition scores were much smaller than in the case of division into 6 intervals. The division into two intervals apparently does not ensure the appropriate description of the data. Scores in the remaining cases are greater than 90%.

E. Application of neural networks to recognition tasks

Parameters related to stop-gaps disorder were also fed to a neural net. These parameters were obtained from pattern sets: *Detection of stop-gaps* (5) and *Detection of stop-gaps* (10). The structure of feedforward neural network with one hidden layer was chosen. The number of neurons in the input layer depends on the type of the set of parameters (5 or 10) and the size of the output layer equals two neural units (2 classes). The net was trained by the error backpropagation method. In all experiments, the unipolar function was set as the neuron activation function.

For each set of patterns, two structures of the network were chosen. In the first case, the number of neurons in the hidden layer was arbitrarily set to 5 and in the second one, the number of hidden neurons was equal to 10. To check the networks performance, the leave-one-out method was applied. For statistical purposes, each neural net for every set was trained and tested 5 times.

The training parameters for different types of networks are grouped and presented in Table II, as well as results of recognition tests. The values refer to correct responses of the network. To denote a structure of a neural net, the following convention was used: No. of_input_neurons / No._of_hidden_neurons / No. of_output_neurons.

The following abbreviations and symbols are used:

C/I - number of correctly classified vectors/number of incorrectly classified vectors

[%] - percentage ratio of correct decisions.

TABLE II. Training parameters and results of recognition tests for different structures of neural net.

Network structure	Training parameters			Results	
	Learning ratio (η)	Momentum ratio (α)	E_{max}	C/I	[%]
5/5/2	0.5	0.3	0.01	45/15	76.3
5/10/2	0.5	0.3	0.01	45/15	78.1
10/5/2	0.6	0.4	0.01	42/18	71.21
10/10/2	0.6	0.4	0.01	40/20	67.37

As seen from Table II, in the case of neural networks applied to the task of recognition of artifacts of stuttering events in the speech samples, obtained average scores were equal to 73,25%. This implies that the rough set-based algorithm performance was better while recognizing stuttering events in speech signal.

7. RESULTS

Since certain characteristics of the studied cases can be meaningful for the diagnosis of speech disorders, thus the results of the proposed analyses can be useful for the evaluation of the algorithms used in speech therapy. The presented results provide individual examples from a series of analyses made by the authors. They are typical for the analyzed cases, and allow some conclusions to be formulated:

1. In the studied cases of vowel prolongation, one can observe variations of formant frequencies within a single articulation. These variations go up and down the frequency scale and constitute a sort of pseudo-periodical frequency modulation with an ascending deviation. As this deviation has an ascending tendency especially at the end of an articulation, the frequency of the vocal tone is always higher than it was at the beginning;

2. In cases of shorter periods of time of vowel prolongation, the number of oscillations of frequencies is smaller, and can even reach the value of 1;

3. The value of the mean frequency deviation seems to be an individual feature;

4. For the purposes of enriching the parametrization suitable to the representation of this type of disorder, one should express the value of the deviation quantitatively;

5. A similar character (ascending) is shown in the behavior of frequencies of the vocal tone for speech disorders related to repetitions. The value of this change seems to be an individual feature, similar to the cases of vowel prolongation;

6. The described values also depend on the type of vowel in question;

7. For various patients, different correlation levels were observed between the frequency of the vocal tone and formant parameters. However, when the correlation was significant, the coefficient was always negative. For example, this means that as the vocal tone frequency rose, the frequency or amplitude of formants was declining;

8. Many cases of disorders of the "prolongation" type are characterized by strong oscillations of the amplitude of some of the formants, sometimes bringing about momentary fading. At the same time, additional formants may appear in frequency ranges that are not related to the spectrum of the uttered vowel. This is the natural effect of a block of the articulation muscles that affects the shape of the vocal tract. Resulting speech is almost incomprehensible;

9. For speech disorders of the "repetition" type, the above effects are present in greater intensity. This is understandable because speech signal in this case is characterized by faster changes and greater dynamics;

10. Despite the speech correction spectrum, formants of the analyzed vowel usually do not fit into the ranges recognized as typical for the given vowel;

11. Formants of disordered speech show fading with oscillations in the frequency domain. These oscillations occur around a mean frequency the value of which is usually far from the frequency occurring after electronic speech correction and are from the range of frequencies typical for the given formant;

12. The description and behavior of formants in disordered speech can be presented in the form of table (TABLE III) (using the example presented before in Fig. 13 – disordered speech; repetition of the "a" vowel):

TABLE III. Behavior of formants in disordered speech.

analyzed sound	F1 [kHz]	F2 [kHz]	F3 [kHz]
repetition 1	1.09	none	2.6
repetition 2	none	1.33	2.63
repetition 3	1.01	none	2.74
repetition 4	0.66	1.46	3.35
speech after FAF correction	0.75	1.32	2.60
mean frequency ranges	0.7 - 0.9	1.4 - 1.75	2.15 - 2.45

The last row in the Table 3 includes values related to the fluent speech patterns of the same speaker.

13. The above observations lead to some possibilities of using this type of research for purposes of diagnosis. The possibility of determining the degree of correction of the formant frequency can be represented by the variance:

$$r = f_n - f_{sr} \tag{10}$$

where:

f_n - value of formant frequencies after correction,

f_{sr} - mean value of formant frequencies of disordered speech;

14. Due to high oscillations of the values of formant frequencies, it seems that for the purpose of working on this type of data it is advisable to use algorithms based on learning algorithms.

15. For a complete description of formant behavior, one needs to add the mutual relations between the amplitudes (levels) of the particular formants.

16. When comparing the results of experiments designed to detect stuttering

events, better scores were obtained using rough set-based system than the one based on artificial neural networks. The obtained results were dependent upon the chosen representation of speech patterns and settings of the system.

8. CONCLUSIONS

Three most characteristic speech disorders related to stuttering were analyzed in this paper. They can be automatically discerned in continuous speech, however each type required a different approach in the detection procedures. The common point of all proposed methods is the application of learning algorithms based on rough sets or on neural networks. Results obtained suggest that it would be possible to automatically discern and to count most stuttered speech events in order to perform objective studies of the effectiveness of electronic speech correctors. The signal analyses of stuttered speech assisted by intelligent algorithms can bring valuable information concerning speech disorder and they provide also means of assessing the effectiveness of electronic speech aids.

Acknowledgments

The research was subsidized by the Foundation for Polish Science.

REFERENCES

1. Archibald, L. de Nil, L. F. (**1999**). "The Relationship between stuttering severity and kinesthetic acuity for jaw movements in adults who stutter", J. Fluency Disorders, **24**, 1, 25-42.
2. Czyzewski, A., Kostek, B. and Roland-Mieszkowski M., (**1993**). "Digital Speech Aid Based on the Modification of the Auditory Feedback Loop", Poster paper presented at the Eurospeech'93, 3rd Eurospeech Congress, Berlin, Germany.
3. Czyzewski, A., Kostek, B. and Roland-Mieszkowski M. (**1994**). "Digital Speech Aid to decrease stuttering - clinical results and patient's reaction", First World Congress of Fluency Disorders, Chapter 11, 526-529, Munich, Germany.
4. Czyzewski, A. and Skorka, P. (**1996**). "Modification of the Auditory Loop-Effects on Hearing and Speech Production. 100th Audio Engineering Society Convention, preprint No. 4148 (C-1), Copenhagen, Denmark.
5. Czyzewski, A. (**1996**). "Speaker-Independent Recognition of Digits - Experiments with Neural Networks, Fuzzy Logic and Rough Sets. Intelligent Automation and Soft Computing - an International Journal, **2**, 2, 133-146 .
6. Howell, P., Hamilton, A. & Kyriacopoulos, A. **(1986)**. Automatic detection of repetitions and prolongations in stuttered speech. Speech Input/Output: Techniques and Applications, 252-256. IEE Publications.
7. Howell, P. & Vause, L. **(1986)**. Acoustic analysis and perception of vowels in stuttered speech. Journal of the Acoustical Society of America, **79**, 1571-1579.
8. Howell, P., EI-Yaniv, N. & Powell, D.J. **(1987)**. Factors affecting fluency in stutterers when speaking under altered auditory feedback. In H. Peters and W. Hulstijn (Eds.), Speech Motor Dynamics in Stuttering, 361-369. Springer Press: New York.
9. Howell, P. & Williams, M. **(1988)**. The contribution of the excitatory source to the perception of neutral vowels in stuttered speech. Journal of the Acoustical Society of America, **84**, 80-89.
10. Howell, P. and Wingfield, T. (**1990**). "Perceptual and acoustic evidence for reduced fluency speech in the vicinity of stuttering episodes", Lang. Speech **33**, 1, 31-46.

11. Howell, P. & Williams, M. **(1992)**. Acoustic analysis and perception of vowels in children's and teenagers' stuttered speech. Journal of the Acoustical Society of America, **91**, 1697-1706.
12. Howell, P. **(1995)**. The acoustic properties of stuttered speech. In C. W. Starkweather and H. F. M. Peters (Eds). Proceedings of the First World Congress on Fluency Disorders, **Vol. II** (pp. 48-50), Nijmegen, The Netherlands: University Press Nijmegen.
13. Howell, P., Staveley, A., Sackin, S. and Rustin, L. (**1995**). "Comparison of speech motor development in stutterers and fluent speakers between 7 and 12 years old", J. Fluency Disorders, **20**, 1, 243-255.
14. Howell, P., Au-Yeung, J., Sackin, S., Glenn K., & Rustin, L. **(1997)**. Detection of supralexical dysfluencies in a text read by child stutterers. Journal of Fluency Disorders, **22**, 299-307.
15. Howell, P., Sackin, S. & Glenn, K. **(1997a)**. Development of A Two-Stage Procedure for the Automatic Recognition of Dysfluencies in the Speech of Children Who Stutter: I. Psychometric Procedures Appropriate for Selection of Training Material for Lexical Dysfluency Classifiers. Journal of Speech, Language and Hearing Research, **40**, 1073-1084
16. Howell, P., Sackin, S. & Glenn, K. **(1997b)**. Development of A Two-Stage Procedure for the Automatic Recognition of Dysfluencies in the Speech of Children Who Stutter: II. ANN Recognition of Repetitions and Prolongations with Supplied Word Segment Markers Journal of Speech, Language and Hearing Research, **40**, 1085-1096.
17. Howell, P., Staveley, A., Sackin, S. and Rustin, L. (**1998**). "Methods of interval selection, presence of noise and their effects on detectability of repetitions and prolongations", J. Acoust. Soc. Amer., **104**, 6, 3558-3567.
18. Howell, P., Au-Yeung, J. and Pilgrim, L. (**1999**). "Utterance rate and linguistic properties as determinants of lexical dysfluencies in children who stutter", J. Acoust. Soc. Amer., **105**, 1, 481-490.
19. Kaczmarek, A. and Skorka, P. (**1997**). "Investigation of the Pitch of the Vocal Tone for Stuttered Speech", 102nd Audio Engineering Society Convention, preprint No. 4486 (L7), Munich, Germany.

20. Kalinowski, J., Armson J. and Stuart A. (**1995**). "Effect of Normal and Fast Articulatory Rates on Stuttering", J. Fluency Disorders, **20**, 3, 293-302, 1995.
21. Kostek, B. (**1999**). "Soft Computing in Acoustics, Applications of Neural Networks, Fuzzy Logic and Rough Sets to Musical Acoustics, Studies in Fuzziness and Soft Computing", Physica Verlag, Heilderberg, New York.
22. Kuniszyk-Jozkowiak, A. (**1995**). "The statistical analysis of speech envelopes in stutterers and non-stutterers", J. Fluency Disorders, **20**, 1, 11-23.
23. Kuniszyk-Jozkowiak, W. (**1996**). "A comparison of speech envelopes of stutterers and nonstutterers", J. Acoust. Soc. Amer., **100**, 2, 1105-1110.
24. Michaelis, D., Frohlich, M. and Strube, H. W. (**1998**). "Selection and combination of acoustic features for the description of pathologic voices", J. Acoust. Soc. Amer., **103**, 3, 1628-1639.
25. Lee B. (**1950**). "Effects of delayed speech feedback", J. Acoust. Soc. Amer., **22**, 824-826.
26. Robb, M. and Blomgren, M. (**1997**). "Analysis of F2 Transitions in the Speech of Stutterers and Non-stutterers", J. Fluency Disorders, **22**, 1, 1-16.
27. Pawlak, Z. (**1982**). "Rough sets", J. Computer and Information Science, 11, 5, 341-356.
28. Polkowski, L. and Skowron A. (Eds.) (**1998**). "Rough Sets in Knowledge Discovery 1: Methodology and Applications", Physica-Verlag, Heidelberg and New York.
29. Slowinski, R. (ed.) (**1992**). "Intelligent Decision Support. Handbook on Applications and Advances of the Rough Sets Theory". Kluwer Academic Publisher, Dordrecht/Boston/London.
30. Roland-Mieszkowski, M., Czyzewski, A. and Kostek, B. (**1995**). "DSA (Digital Speech Aid) - a new device to decrease or eliminate stuttering. First World Congress of Fluency Disorders", Chapter 11, Munich, 535-539, Munich, Germany.
31. Zebrowski, P.M. (**1991**). "Duration of the speech disfluencies of beginning stutterers", J. Speech-Hear-Res., **34**, 3.

Appendix. ***Basic notions of the rough set theory***

The rough set theory and its basic concepts were proposed by Pawlak in the early 80's (Pawlak, 1982). Since then, this theory has been introduced to different domains of science, including the domain of audio acoustics (Czyzewski, 1996; Kostek, 1999). The rough set concept is a new mathematical tool to reason about imperfect knowledge. Rough set theory is based on the assumption that elements that exhibit the same information are indiscernible and form blocks that can be understood as elementary granules of knowledge. These granules are called elementary sets. Any union of elementary sets is called a crisp set. Due to the granularity of knowledge, rough sets cannot be defined finally by available knowledge. That is why with every rough set we associate two crisp sets, called its lower and its upper approximation.

The universe *U*, defined as a collection of objects stands at the top of the rough set hierarchy. On the other hand, a basic entity is placed at the bottom of this hierarchy. Between them, the approximation space *AS* is defined. Several important concepts are connected with such notions as an upper approximation $\overline{R}(S)$, a lower approximation $\underline{R}(S)$, and a boundary region *BR*. Intuitively, the lower approximation of a set consists of all elements that surely belong to the set, whereas the upper approximation of the set constitutes of all elements that possibly belong to the set. The boundary region is the difference of the upper and the lower approximation. It consists of all elements that cannot be classified uniquely to the set or its complement, by employing available knowledge. The concept set is a subset of the entire set of elements representing knowledge, so that this subset contains only elements fulfilling the desired relation.

Consequently, the rough set approximates a given concept from below and from above, using both lower and upper approximations. It means that a set in this theory is not defined straightforwardly but it is defined in terms of its lower and its upper approximations. Thus, a standard set *S* can be approximated in approximation space AS by the pair $\underline{R}(S), \overline{R}(S)$, called the rough set (see Fig. A1).

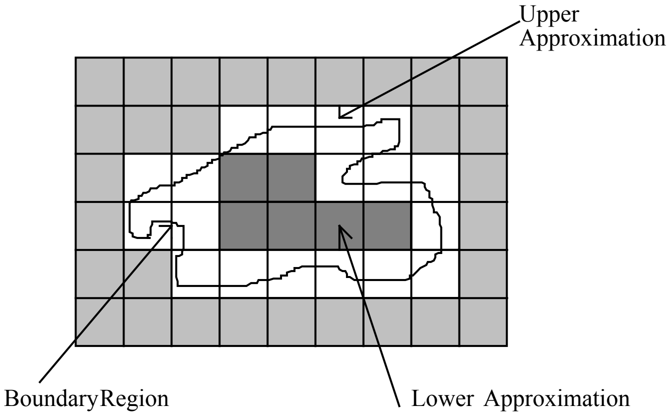


Fig. A1. Basic structure of rough sets

Three other properties of rough sets defined in terms of attribute values are: *dependencies, reduct* and *core* (Pawlak, 1982).

An inductive learning system based on a rough set theory consists of a learning component for the automatic rule derivation from training samples and of an inference system used for decision-making at the recognition stage. Rules are in the following format:

if <condition1>and<condition2>...and<condition *n*> then <decision> (A1)

Conditions are based on attributes. Attributes should be equal to concrete values or should belong to certain ranges. The decision always uses a single rule matching approach. There are two kinds of rules, certain and uncertain (possible) ones. The certain rules are derived from the lower approximation of the concept set and the possible rules are derived from the upper approximation of the concept set.

The knowledge base in the rough set method can be conveniently represented in the form of a decision table, in which rows represent objects and columns represent attributes. In the last column decision attributes are collected.

An important parameter of rules, reflecting their quality is the rough measure μ_{RS}, defined as follows:

$$\mu_{RS} = \frac{|X \cap Y|}{|Y|} \tag{A2}$$

where: X is the set determined by the given concept (concept set) and Y is the set of all examples described by the rule. The rough measure of certain rules is always equal to 1, whereas the rough measure of possible rules fulfills the term: $0<\mu_{RS}<1$.

An additional parameter is defined which allows one to optimize the rule generation process. This parameter was called the rule strength r and is defined as follows:

$$r = c(\mu_{RS} - n_{\mu}) \tag{A3}$$

where: c - number of cases conforming to the rule, and n_{μ} - the neutral point of the rough measure is one of several parameters of the rule generation system to be set by its operator. This parameter allows one to regulate the effect of possible (uncertain) rules on the process of decision making.

The knowledge base represented by rules in the rough set system allows one to control the processes burdened with vagueness. More details concerning the rough set concepts are documented elswhere (Pawlak 1982; Slowinski, 1992).

Discretization

The discretization process can be performed in two ways, using quantization and clusterization methods:

˜ the parameter domain can be divided into subintervals and each parameter value belonging to the same subinterval will take the same value (quantization process);

- parameter values can be clustered together into a few groups, forming intervals, and each group of values will be considered as one value (clusterization process).

Various attempts have been made in computing practice to process real-value data with several methods classified as global and local. The discretization methods are considered global if they are applied to the entire set of parameter values. On the other hand, some methods are limited to one parameter domain. Several discretization schemes were reviewed by Kostek (1999) among them: *Equal*

Interval Width Method, Equal Frequency per Interval Method, *Minimal Class Entropy Method, Hierarchical Cluster Analysis*, etc.

Generally speaking, The quantization process can be performed using various algorithmic approaches. The division of the parameter domain into subintervals is defined as follows:

Let *A* be a real value parameter and let the interval [*a*, *b*] be its domain. The division $\prod_A$ on [*a*, *b*] is defined as the set of *k* subintervals:

$$\prod_A = \{[a_0, a_1), [a_1, a_2), \ldots, [a_{k-1}, a_k]\} \quad \text{(A4)}$$

where: $a_0 = a, a_{i-1} < a_i,\ i = 1,\ldots,k,\ a_k = b$

This approach to quantization is based on calculating division points a_i. After quantization, the parameter value is transformed into the number of the subinterval to which this value belongs.

FIGURE CAPTIONS

FIG. 1. Graphical presentation of the logarithm of the spectrum for a fragment of disordered speech – Polish vowel "*o*" before normalization.

FIG. 2. Graphical presentation of the upper envelope of the spectrum shown in Fig. 1.

FIG. 3. Graphical presentation of the lower envelope of the spectrum shown in Fig. 1.

FIG. 4. Graphical presentation of the logarithm of the spectrum for a fragment of disordered speech – Polish vowel "*o*" after normalization.

FIG. 5. Graphical presentation of the cepstrum for a fragment of disordered speech – Polish vowel "*o*" before normalization.

FIG. 6. Graphical presentation of the cepstrum for a fragment of disordered speech – Polish vowel "*o*" after normalization.

FIG. 7. Graphical presentation of the evolution of the vocal tone frequency for a fragment of disordered speech – Polish vowel "*o*" (vowel prolongation case). The cepstrum was calculated on the basis of the Cosine Transform.

FIG. 8. Graphical presentation of the method of square interpolation for the formant estimation.

FIG. 9. Graphical presentation of the evolution of formants for a fragment of stuttered speech which contains prolongation – Polish vowel "*a*". The diagram presents the relation: spectral amplitude - time -frequency.

FIG. 10. Graphical presentation of the results of disordered speech signal analysis – Polish vowel "*a*" (vowel prolongation case). The diagram presents changes of the vocal tone frequency F_0 (above) versus time and changes of the frequency of formant F_1 (below).

FIG. 11. Graphical presentation of results of analysis of disordered speech - Polish vowel "*i*" (vowel prolongation case):

a. time-domain plot; b. evolution of cepstrum characteristics; c. changes of vocal tone frequency.

FIG. 12. Graphical presentation of the results of disordered speech analysis – Polish syllable "*ba*" (repetition case - syllable pronounced twice):

a. time-domain plot; b. evolution of cepstrum characteristics; c. changes of vocal tone frequency.

FIG. 13. Graphical presentation of formant parameter changes for disordered speech – Polish vowel "*a*" (vowel repetition case). Cepstrally smoothed spectra are shown. Left: first repetition; Right: second repetition. Vocal tone frequency and formant frequencies are listed – see also Table III.

FIG. 14. Graphical presentation of formant parameter changes for corrected speech using the DSA – Polish word "*zamieηila*". Cepstrally smoothed spectra are shown. Left: first vowel "*a*"; Right: second vowel "*a*". Vocal tone frequency and formant frequencies are listed.

FIG. 15. Graphical presentation of the relation between the vocal tone frequency (axis x) and formant F1 frequency (axis y) of disordered speech – Polish vowel "*a*" (vowel prolongation case). The correlation coefficient is negative in this case.

FIG. 16. Graphical presentation of cepstrally smoothed spectra for disordered speech – Polish vowel "*a*" (vowel repetition case). Formant levels in [dB] and frequencies of formants in [kHz] are listed:

(a) first repetition; (b) second repetition; (c) third repetition; (d) fourth repetition;

(e) electronically corrected vowel "*a*" (with FAF method).

FIG. 17. Graphical presentation of cepstrally smoothed spectra for disordered speech – Polish vowel "*a*" (vowel repetition case). Formant levels in [dB] and frequencies in [kHz] are listed

(a) first repetition; (b) second repetition; (c) third repetition; (d) fourth repetition; (e) fifth repetition; (f) sixth repetition; (g) seventh repetition; (e) electronically corrected vowel "*a*" (with DAF method).

FIG. 18. Digital speech aid – the device introduced by Czyzewski & Roland-Mieszkowski

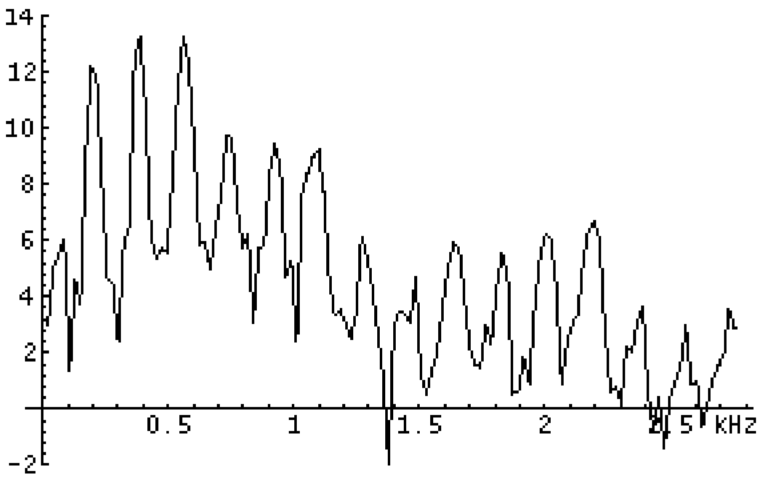


FIG. 1. Graphical presentation of the logarithm of the spectrum for a fragment of disordered speech – Polish vowel "*o*" before normalization.

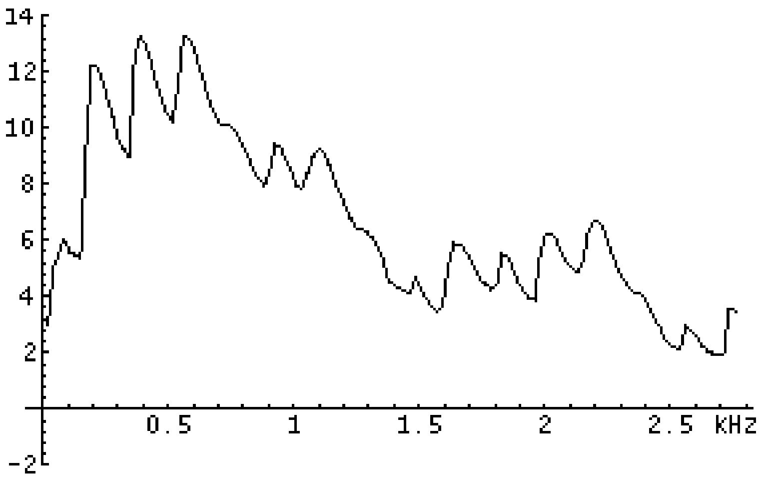


FIG. 2. Graphical presentation of the upper envelope of the spectrum shown in Fig. 1.

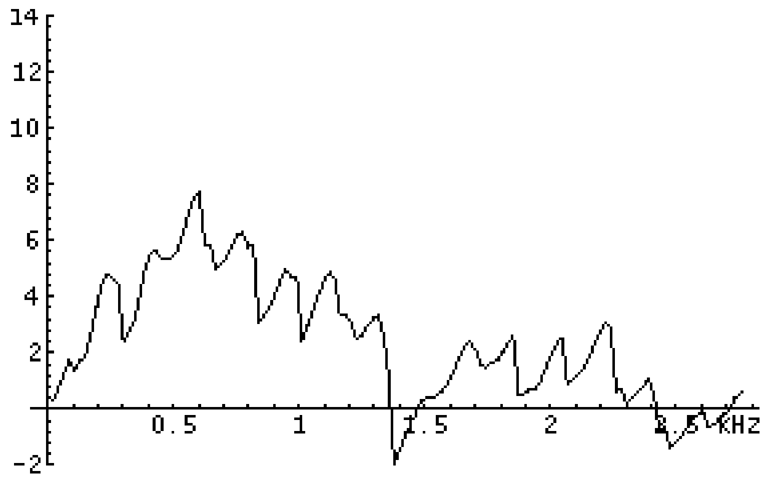


FIG. 3. Graphical presentation of the lower envelope of the spectrum shown in Fig. 1.

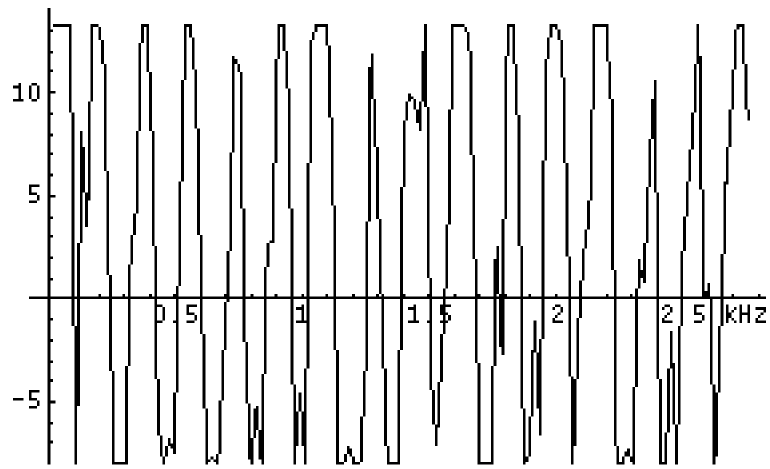


FIG. 4. Graphical presentation of the logarithm of the spectrum for a fragment of disordered speech – Polish vowel "*o*" after normalization.

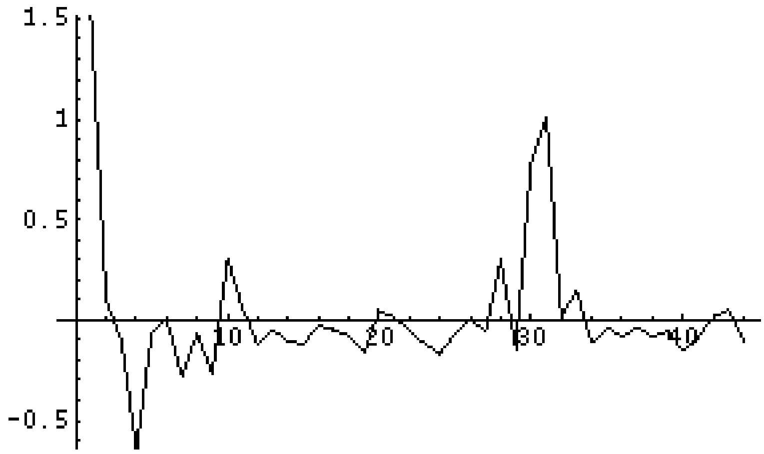


FIG. 5. Graphical presentation of the cepstrum for a fragment of disordered speech – Polish vowel "*o*" before normalization.

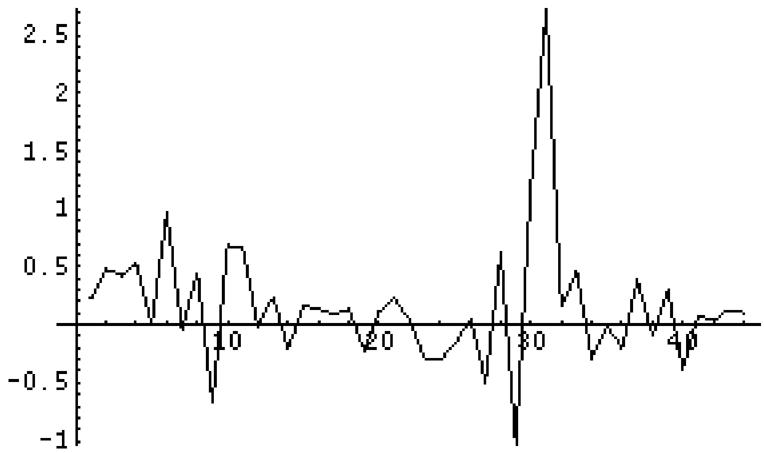


FIG. 6. Graphical presentation of the cepstrum for a fragment of disordered speech – Polish vowel "*o*" after normalization.

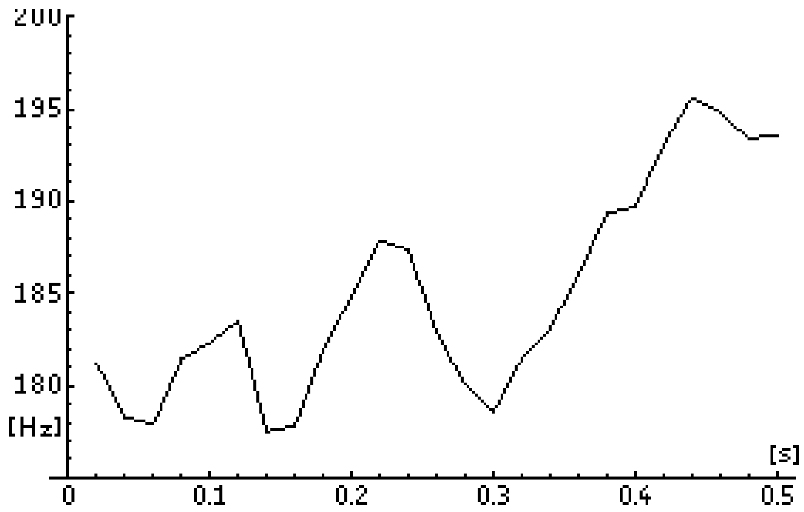


FIG. 7. Graphical presentation of the evolution of the vocal tone frequency for a fragment of disordered speech – Polish vowel "*o*" (vowel prolongation case). The cepstrum was calculated on the basis of the Cosine Transform.

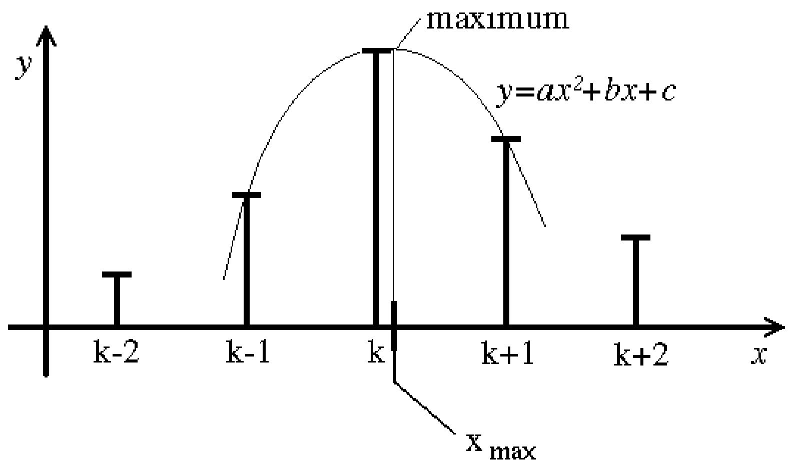


FIG. 8. Graphical presentation of the method of square interpolation for the formant estimation.

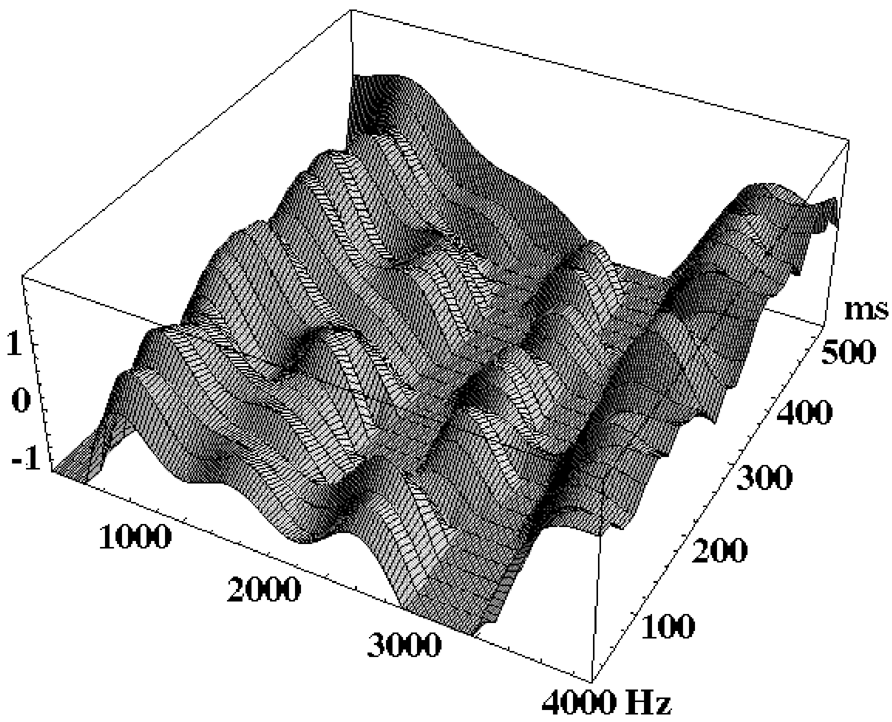


FIG. 9. Graphical presentation of the evolution of formants for a fragment of stuttered speech which contains prolongation – Polish vowel "*a*". The diagram presents the relation: spectral amplitude - time -frequency.

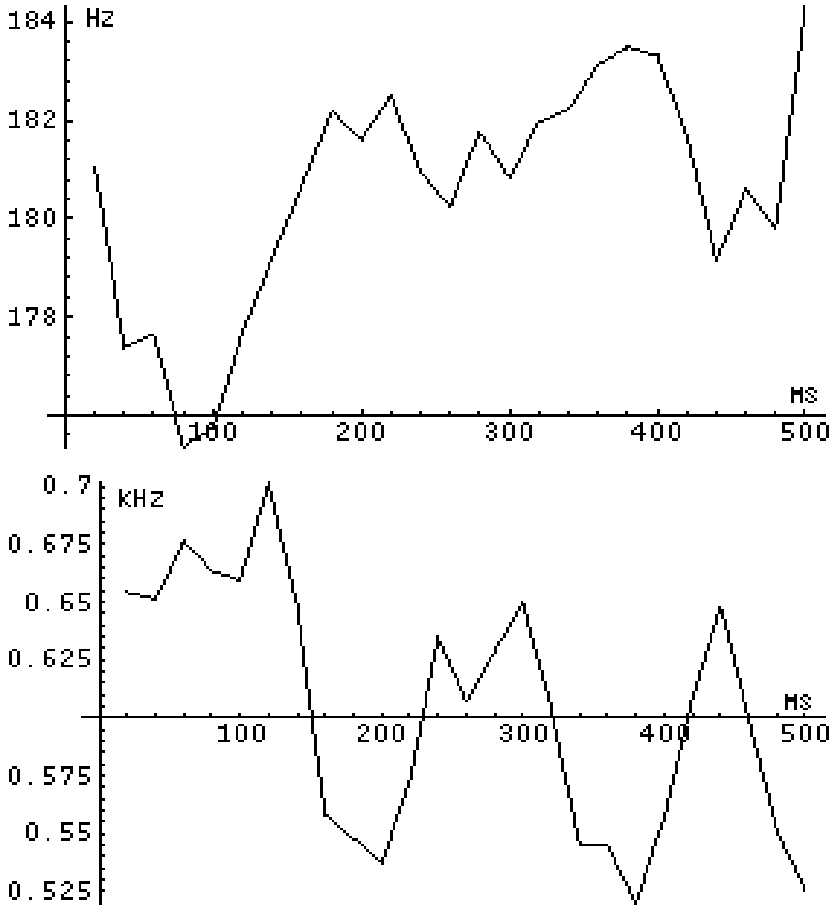
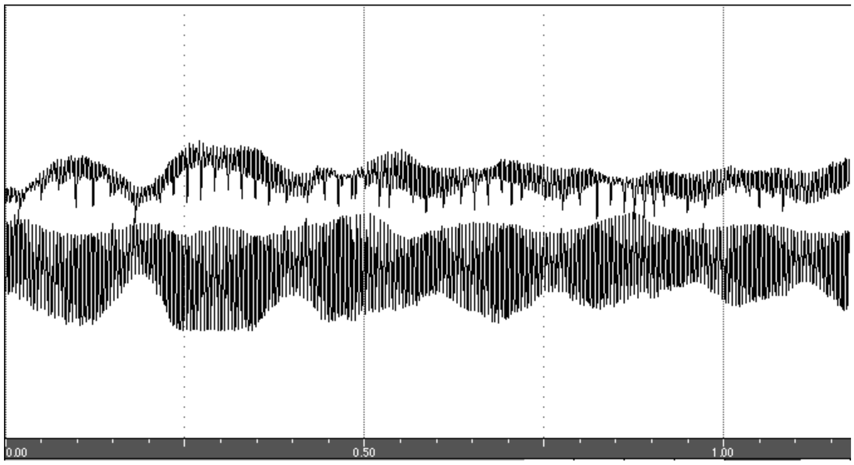
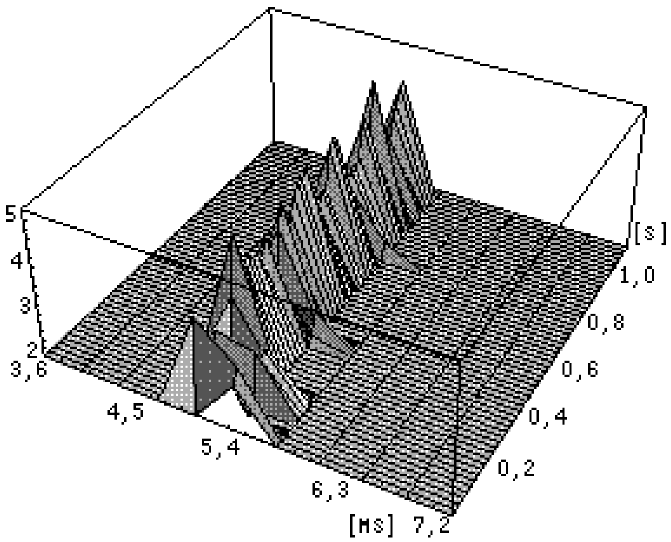


FIG. 10. Graphical presentation of the results of disordered speech signal analysis – Polish vowel "*a*" (vowel prolongation case). The diagram presents changes of the vocal tone frequency F_0 (above) versus time and changes of the frequency of formant F_1 (below).

a.



b.



c.

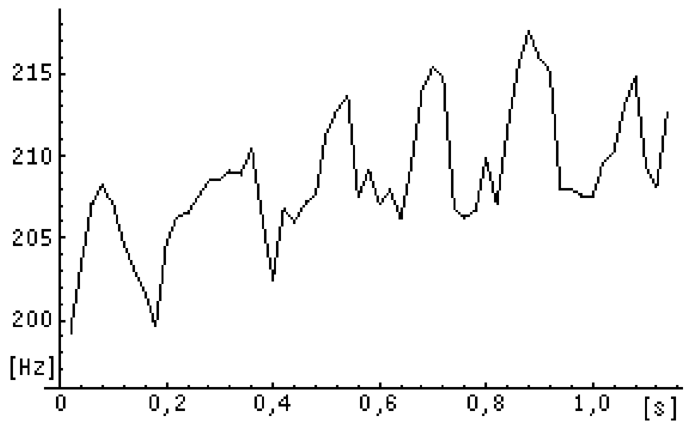
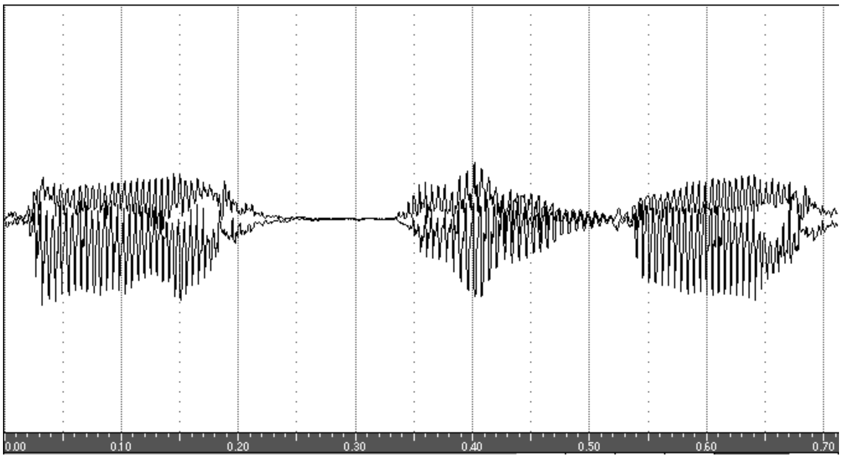


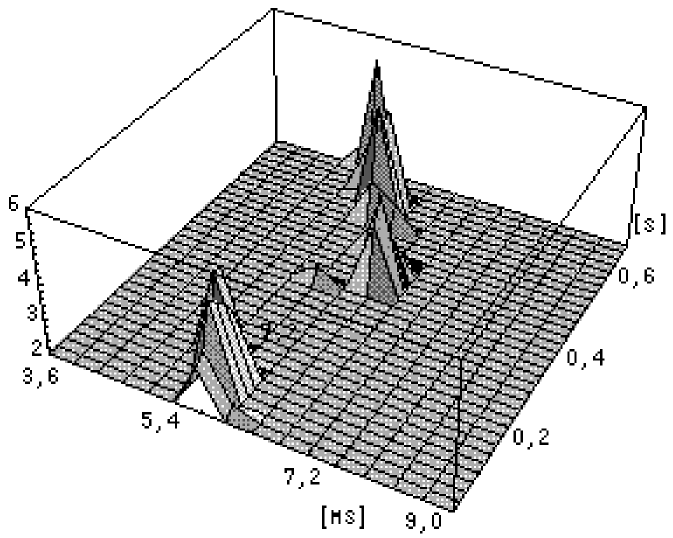
FIG. 11. Graphical presentation of results of analysis of disordered speech - Polish vowel "*I*" (vowel prolongation case):

a. time-domain plot; b. evolution of cepstrum characteristics; c. changes of vocal tone frequency.

a.



b.



c.

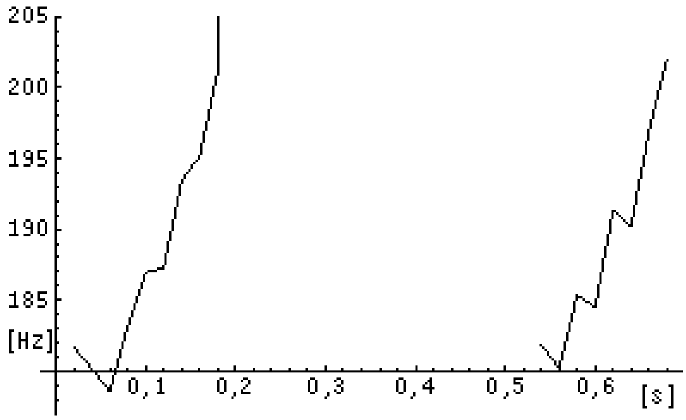


FIG. 12. Graphical presentation of the results of disordered speech analysis – Polish syllable "*ba*" (repetition case - syllable pronounced twice):

a. time-domain plot; b. evolution of cepstrum characteristics; c. changes of vocal tone frequency

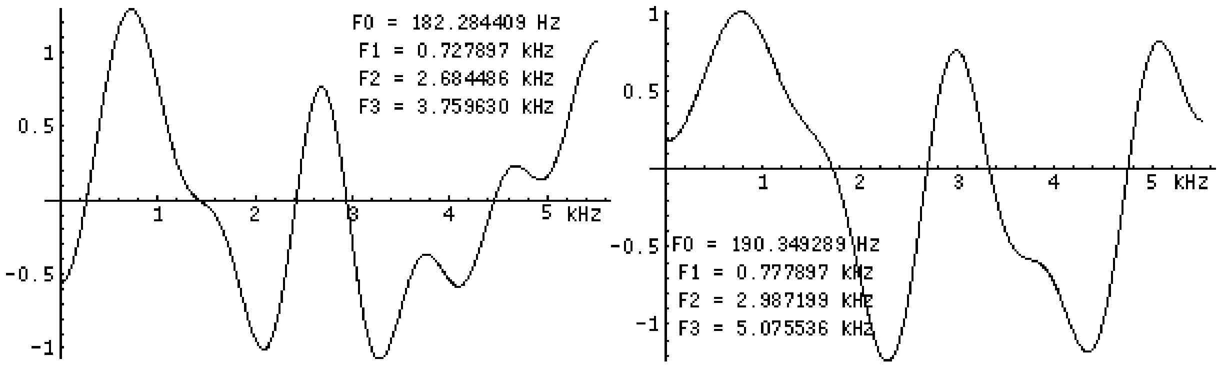


FIG. 13. Graphical presentation of formant parameter changes for disordered speech – Polish vowel "*a*" (vowel repetition case). Cepstrally smoothed spectra are shown. Left: first repetition; Right: second repetition. Vocal tone frequency and formant frequencies are listed – see also Table III.

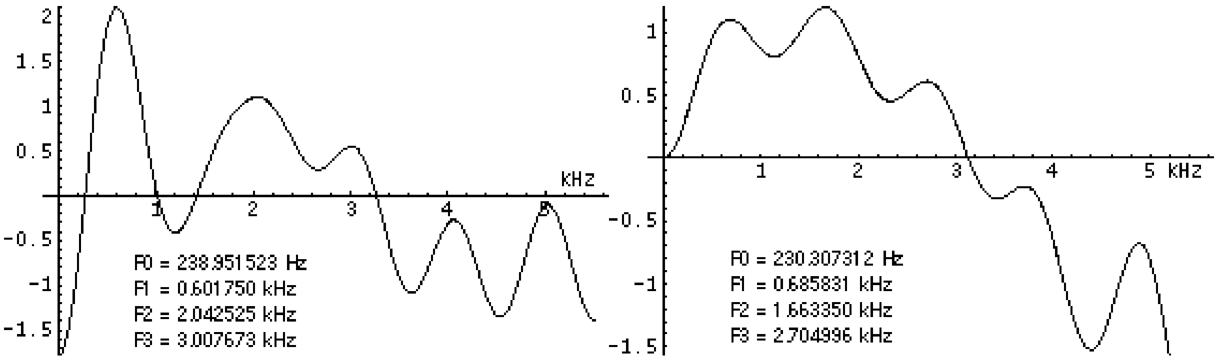


FIG. 14. Graphical presentation of formant parameter changes for corrected speech using the DSA – Polish word "*zamieηila*". Cepstrally smoothed spectra are shown. Left: first vowel "*a*"; Right: second vowel "*a*". Vocal tone frequency and formant frequencies are listed.

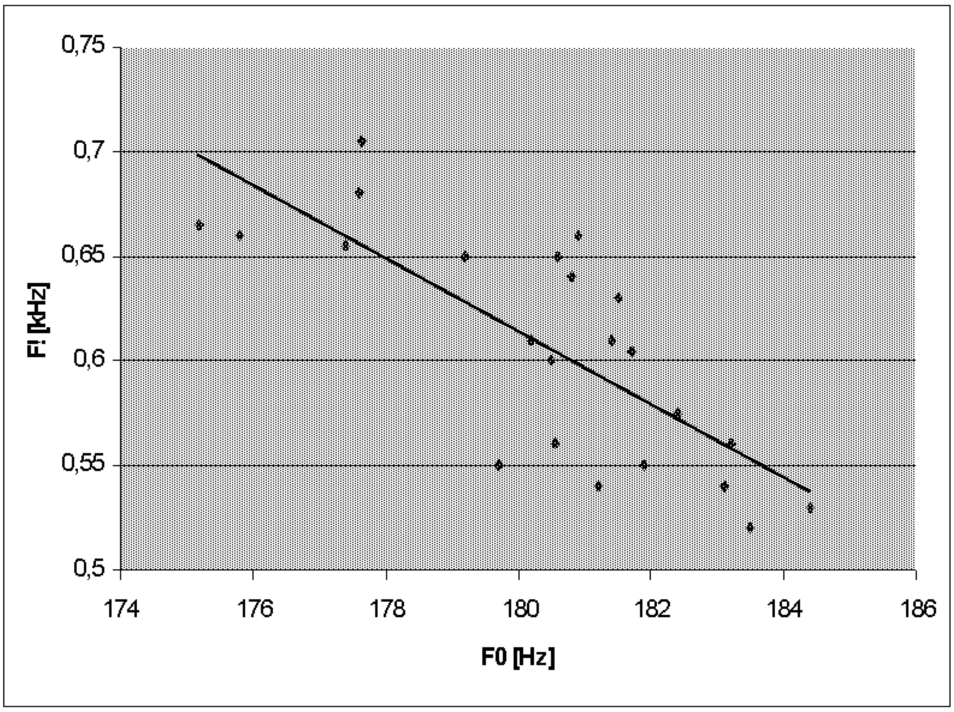
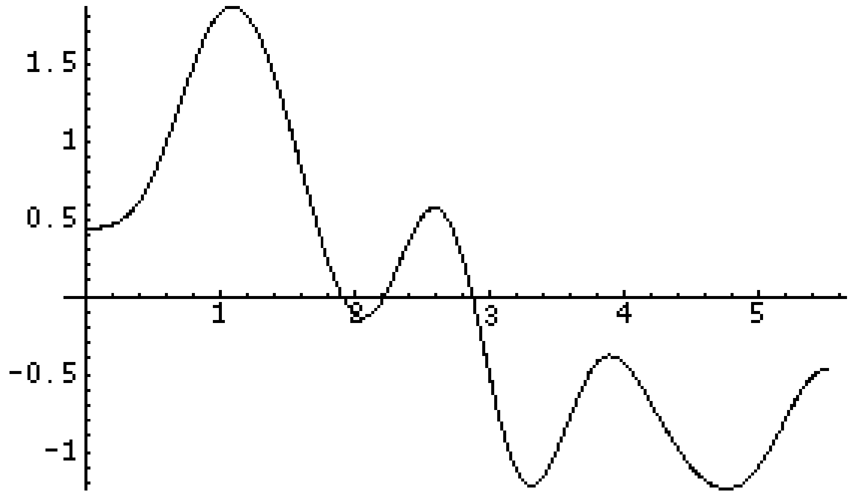


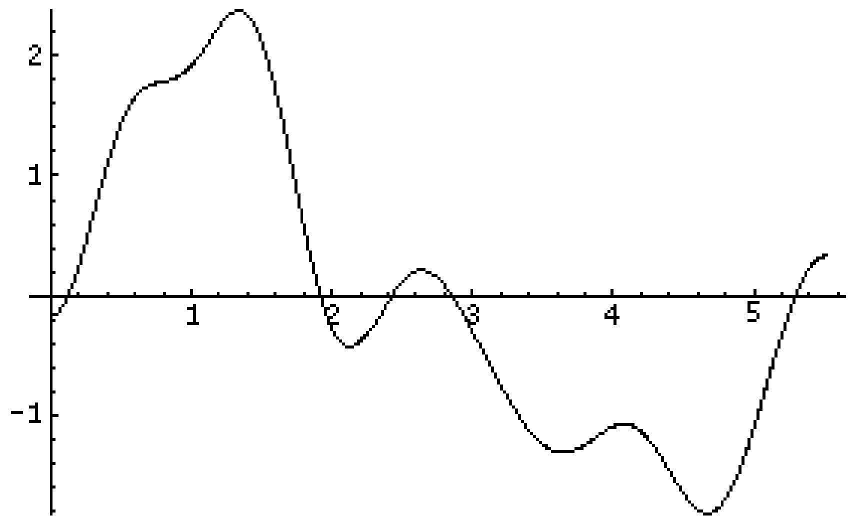
FIG. 15. Relation between the vocal tone frequency F_0 and formant frequency F_1 of disordered speech – Polish vowel "*a*" (vowel prolongation case). The correlation coefficient is negative in this case and equals -6.391.

(a)



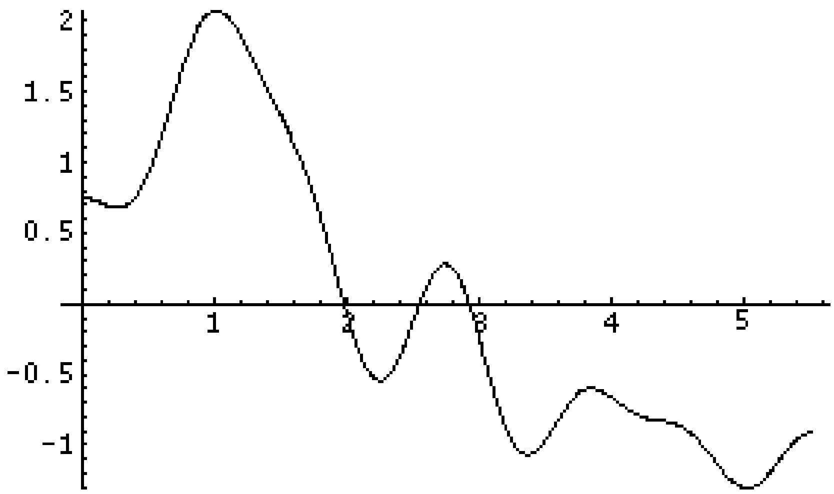
L1 = 1.867490 F1 = 1.094157 kHz L2 = 0.570181 F2 = 2.595860 kHz
L3 = -0.381293 F3 = 3.901403 kHz

(b)



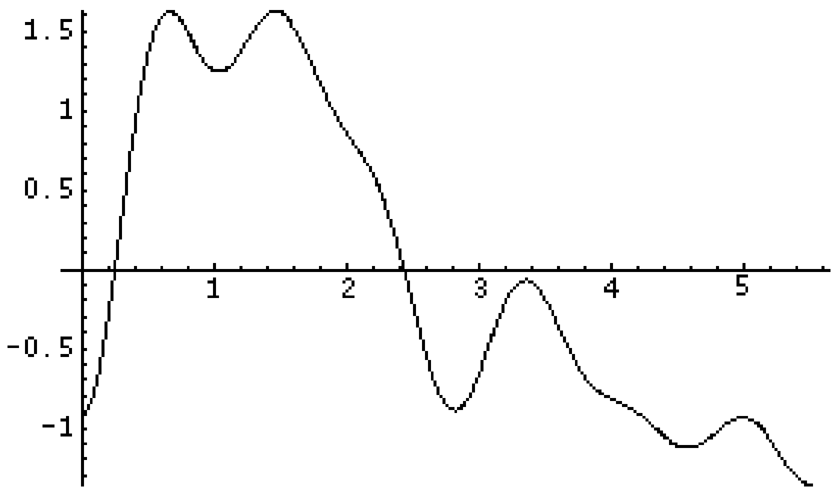
L1 = 2.378656 F1 = 1.333182 kHz L2 = 0.207837 F2 = 2.633388 kHz
L3 = -1.074840 F3 = 4.073639 kHz

(c)



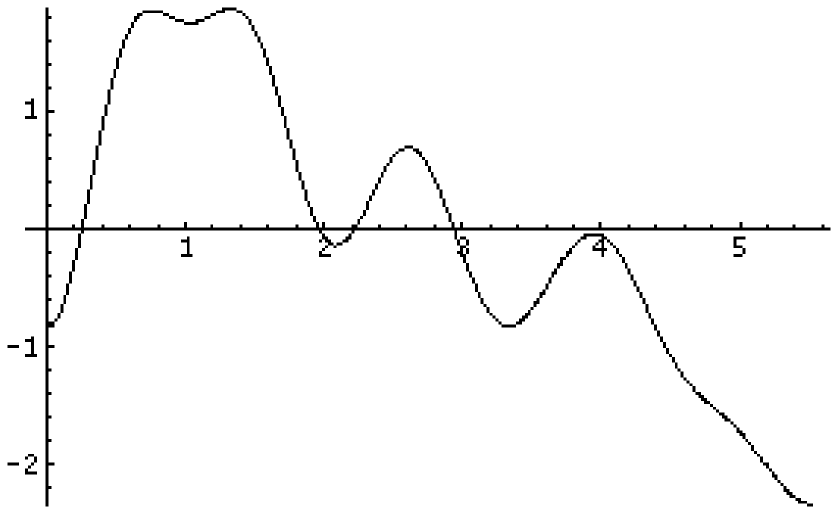
L1 = 2.073406 F1 = 1.018925 kHz L2 = 0.278272 F2 = 2.746737 kHz
L3 = -0.598770 F3 = 3.849827 kHz

(d)



L1 = 1.624132 F1 = 0.662184 kHz L2 = 1.630457 F2 = 1.462814 kHz
L3 = -0.070226 F3 = 3.352183 kHz L4 = -0.935523 F4 = 4.990126 kHz

(e)



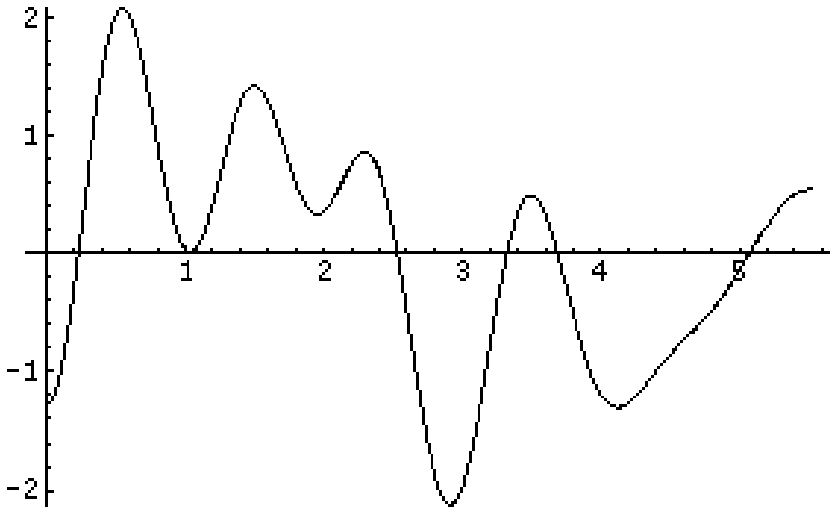
L1 = 1.853511	F1 = 0.754539 kHz	L2 = 1.866170	F2 = 1.326586 kHz
L3 = 0.695308	F3 = 2.603917 kHz	L4 = -0.038948	F4 = 3.945079 kHz

FIG. 16. Graphical presentation of cepstrally smoothed spectra for disordered speech – Polish vowel "*a*" (vowel repetition case). Formant levels in [dB] and frequencies of formants in [kHz] are listed:

(a) first repetition; (b) second repetition; (c) third repetition; (d) fourth repetition;

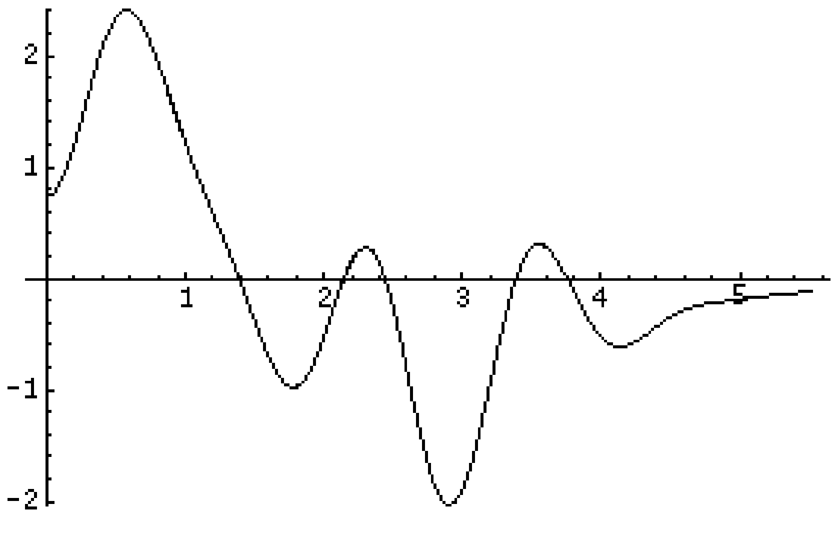
(e) electronically corrected vowel "*a*" (with FAF method).

(a)



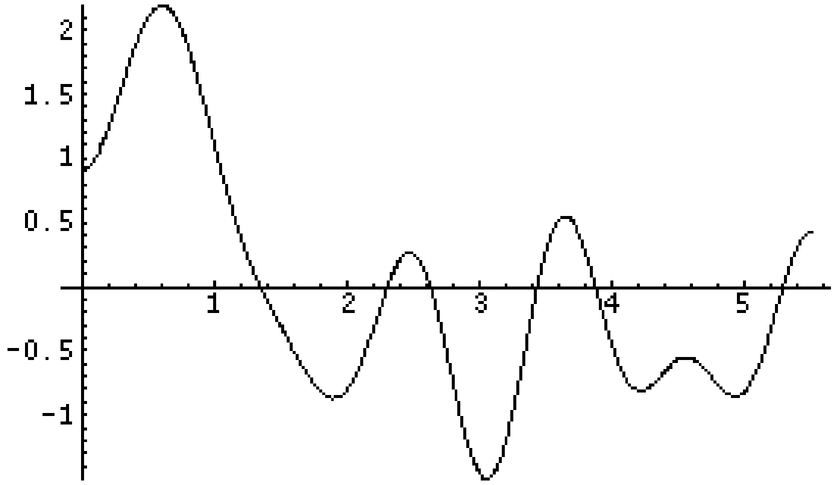
L1 = 2.073615 F1 = 0.547536 kHz L2 = 1.420132 F2 = 1.499503 kHz
L3 = 0.849996 F3 = 2.295185 kHz L4 = 0.492743 F4 = 3.496861 kHz

(b)



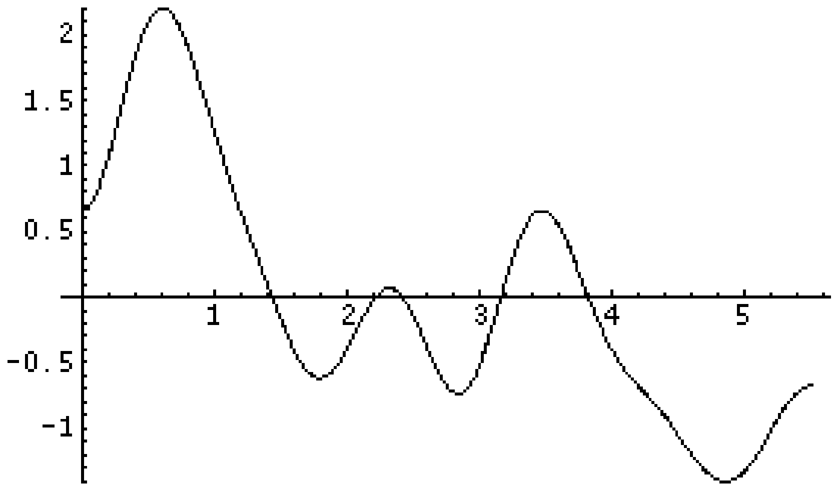
L1 = 2.406654 F1 = 0.576670 kHz L2 = 0.274573 F2 = 2.298181 kHz
L3 = 0.309148 F3 = 3.555708 kHz L4 = 0.000000 F4 = 5.544800 kHz

(c)



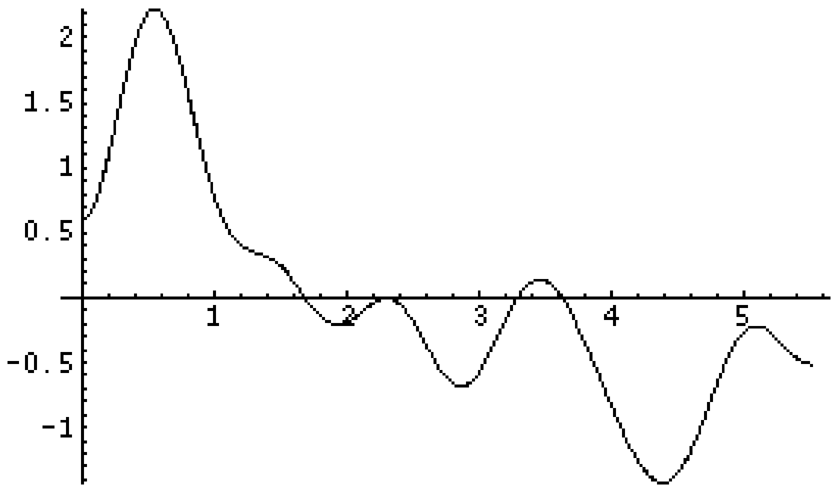
L1 = 2.191803	F1 = 0.605000 kHz	L2 = 0.265227	F2 = 2.478196 kHz
L3 = 0.544154	F3 = 3.649057 kHz	L4 = -0.551388	F4 = 4.561376 kHz

(d)



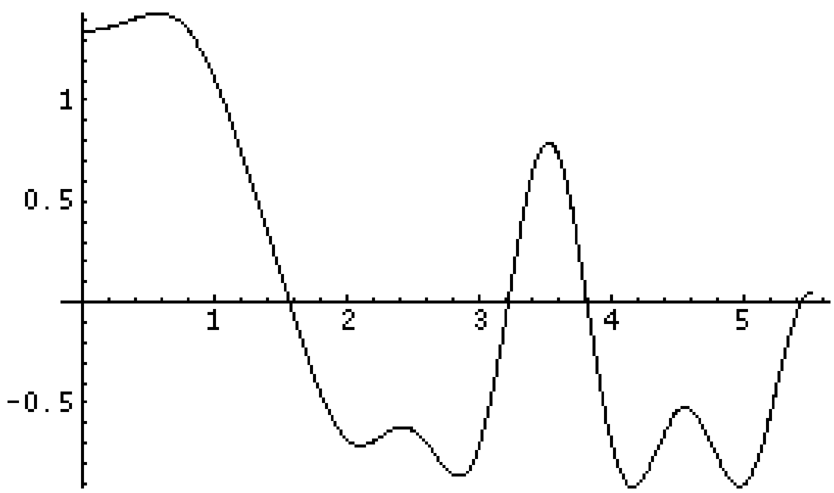
L1 = 2.202443	F1 = 0.608633 kHz	L2 = 0.071304	F2 = 2.321937 kHz
L3 = 0.662271	F3 = 3.473420 kHz		

(e)



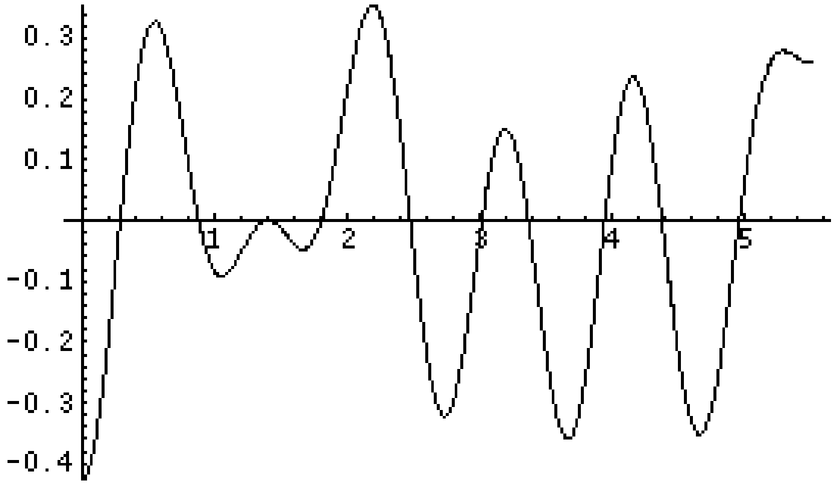
L1 = 2.217035 F1 = 0.551158 kHz L2 = -0.003053 F2 = 2.303142 kHz
L3 = 0.137483 F3 = 3.459164 kHz L4 = -0.216584 F4 = 5.095179 kHz

(f)



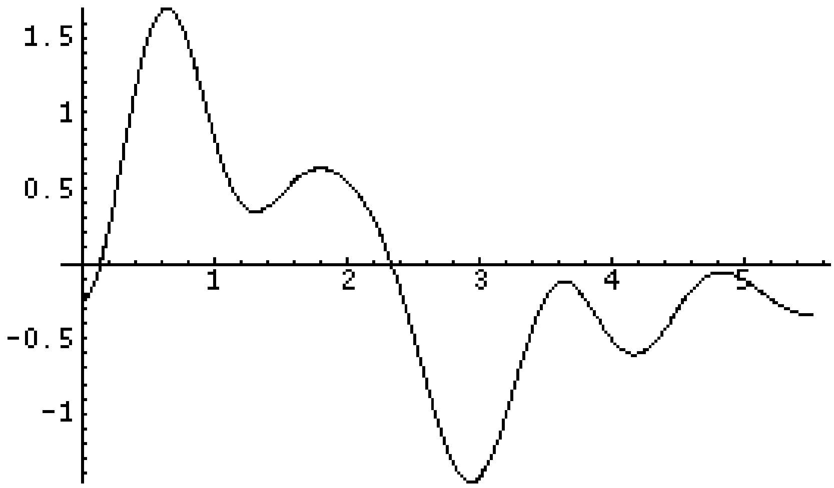
L1 = 1.430661 F1 = 0.578986 kHz L2 = -0.619051 F2 = 2.426605 kHz
L3 = 0.792556 F3 = 3.525269 kHz L4 = -0.522849 F4 = 4.559778 kHz

(g)



L1 = 0.326151	F1 = 0.549942 kHz	L2 = 0.000379	F2 = 1.401725 kHz
L3 = 0.353586	F3 = 2.194561 kHz	L4 = 0.151260	F4 = 3.197909 kHz

(h)



L1 = 1.696113	F1 = 0.641047 kHz	L2 = 0.631089	F2 = 1.806168 kHz
L3 = -0.119042	F3 = 3.639956 kHz	L4 = -0.060941	F4 = 4.832399 kHz

FIG. 17. Graphical presentation of cepstrally smoothed spectra for disordered speech – Polish vowel "*a*" (vowel repetition case). Formant levels in [dB] and frequencies in [kHz] are listed

(a) first repetition; (b) second repetition; (c) third repetition; (d) fourth repetition; (e) fifth repetition; (f) sixth repetition; (g) seventh repetition; (e) electronically corrected vowel "*a*" (with DAF method).