Analysis and classification of the pathological speech using artificial intelligence methods

Wieslaw Wszolek[1], Ryszard Tadeusiewicz[2], Andrzej Izworski[2], Tadeusz Wszolek[1]
[1]Department of Mechanics and Vibroacoustics, [2]Department of Automatics
University of Mining and Metallurgy,
Al. Mickiewicza 30, 30-059 Kraków, Poland

Abstract: - In the present work we introduce a new approach to pathological speech processing methods and recognition of various speech pathologies. We no longer attempt to recognise forms of pathological speech deformations, because it is impossible to show any compact template of a normal speech signal (as reference) and it is also impossible to show a standard form of any deformation. The presented concept of the research scheme is based on the technique of advanced acoustic signal analysis and it refers to the analysis of artificial neural networks functioning in the task of recognition of selected types of vocal tract pathologies. It is recommended here that the simple process of signal recognition should be replaced by a more advanced method of its analysis, called the process of automated understanding of the signal. Selected excerpts of research are presented concerning the application of (modified) acoustic signal processing methods and (in particular) the neural network techniques specially designed for solving the problem of "understanding" selected pathologies of vocal tract.

Key-Words – artificial intelligence, computerised signal processing, speech analysis, applications of neural networks,

1. Introduction

Tasks related to the analysis and recognition of a pathological acoustic signal of speech, characterising selected pathological states, are exceptionally difficult. It happens that minor pathological elements (e.g. occlusion defect) strongly manifest in the speech signal, while very serious pathological changes (e.g. tumour) give only a weak and hardly readable picture of speech disturbances. In spite of the existence of multiple examples of successful automated speech recognition in the semantic (recognition of the utterance contents for e.g. voice control of machines and devices) or personal aspect (verification and identification of persons by using their speech samples), it remains very difficult to diagnose the condition and pathological changes of the voice tract using a speech signal [1]. Additionally, there is no simple way to transfer the experience related to diagnosis of the technological system, because the problems of pathological speech diagnosis are specific by the fact that for such tasks it is very difficult to find an appropriate rule for the preliminary signal analysis [2]. On the basis of the statement that for the cases of analysis of speech pathology forms and sources discussed here, the well-known methods of automated signal recognition cannot be applied. Thus, the authors propose in the present paper a completely new approach, based on the concept of automated understanding. Because of possible multiple meanings of that phrase it should be stressed that the meaning used in the presented work concerns itself with the automated understanding of the nature and character of the pathological speech signal deformations. The exact meaning of the term understanding has no connections with the frequently discussed problem of semantic understanding of the speech signal (i.e. the contents of the pronounced sentences). As it is known the understanding in general differs from recognition by the fact that it is very strongly based on knowledge. To clarify, the term "automated understanding" discussed here denotes such a deformed speech signal analysis, which is oriented towards revealing the sources of the observed signal forms and not towards bare analysis of these forms and diagnostic deduction based on their typology. Every person speaks in a somewhat different way, various (with respect to the contents or speed) utterances of the same person reveal various phonetic and acoustic features of his/her speech signal, and even various registrations of the same utterance recorded from the same person but in various days, can be very different. It is almost a rule, that various samples of a regular speech signal exhibit a greater variety of measurable acoustic parameters, than the measurable differences of the same parameter between these samples and the registered samples of speech, which is obviously pathological.

All this is the reason that one cannot solely relay on models of pathological speech signal recognition in a space based on the set of its features, but that in every case one should try to understand, how such a phonetic or acoustic phenomenon occurred. This means, that the diagnostic system must contain an internal model of the signal generator, based on knowledge of the speech signal and the ways of its generation - in regular and pathological conditions. It should be noticed that such a way of signal analysis closely reflects the contemporary views on the essence of the human perception of various informations from the environment. The perception concept sketched here is widely known, but the authors' original contribution was its application as a standard for a construction of the system of automated understanding of selected voice system pathologies, treated as objects which are being diagnosed by the analysis of the acoustic speech signal, deformed by these pathologies. The described concept includes a series of elements unquestionably difficult in practical realisation. For in the traditional way of solving diagnostic problems, the answer is frequently found more easily. However the authors' long-time experience in the problems related to analysis, evaluation, and the classification of pathological speech signals have proved that for this task a really a new approach is required.

2. The material of study

The studies of the speech articulation have been carried out for persons treated for the larynx cancer (men after various types of operations). Depending on the stage of the tumour, various types of partial larynx surgery have been applied. In the recorded and studied material the following cases have been present: subtotal larynx remove (laryngetctomia subtotalis), unilateral vertical laryngectomy (hemilaryngectomia). Remove of cord vocalise with arytenoid cartage (chordectomia enlargata) and fronto-lateral laryngectomy (laryngectomia fronto-lateralis).

The final acoustic material has been collected from 95 persons divided into two groups:

- ❑ The reference group, 25 persons with a correct pronunciation
- ❑ The group of patients
  - Hemilaryngectomy, includes 28 patients treated by Hemilaryngectomy
  - Subtotal Laryngectomy (near total),includes 14 patients treated by Subtotal Laryngectomy
  - Cordectomy (classical cordectomy), includes 17 patients treated by Cordectomy
  - Enlarged Cordectomy (includes aretenoid cartilage), includes 6 patients treated by Enlarged Cordectomy
  - Fronto-Lateral Laryngectomy, includes 5 patients treated by Laryngectomy Fronto-Lateral

Both the patients and the persons from the reference group pronounced the same text (three times), which consisted of: vowels (A, U, E, I), words containing vowels. The selection of phrases and sets of words pronounced by the examined persons has been based on morphological and functional analysis of the expected (for a given pathology) dysfunction of speech organs, what resulted in collection of research material including sets of words selected with respect to their phonetic features in order to carry the maximum amount of information. In order to receive undisturbed results, ensuring a precise and sometimes even very subtle evaluation of the quality and usefulness of specific sets of input parameters, it was necessary to collect signal samples of very high quality. This is why all the acoustic studies have been carried out in an anechoic chamber, the samples have been registered using professional recording equipment and analysed using professional, thoroughly tested acoustic analysers. The block diagram of the measurement setup has been presented in Fig.1.

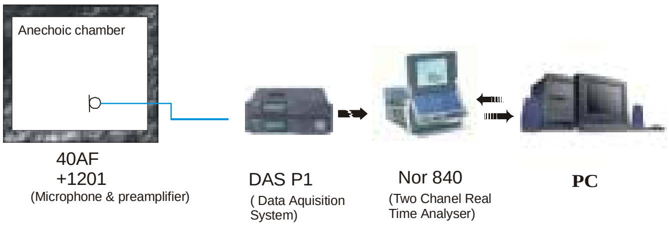
Anechoic chamber
40AF
+1201
(Microphone & preamplifier)
DAS P1
(Data Aquisition System)
Nor 840
(Two Chanel Real Time Analyser)
PC

Fig.1 The measurement setup

After registration, the signal samples were prepared in may ways, using a wide collection of signal processing methods (both in time and spectral domain), to finally obtain key elements for building algorithms realising automatic understanding of the nature of speech deformation.

3. The concept of new methodology

The research task of the present work is the evaluation of origins of speech signal deformations after larynx surgery treatment. One of the important problems encountered during the elaboration of the collected samples was the reduction of the very large information file, the source of which was the analyzed acoustic speech signal (e.g. in the form of dynamic spectra), to the space of features with reduced number of dimensions but information contents sufficient and useful from the diagnostic point of view. In the further signal processing stage the dynamic spectra W(j,k) has been transformed to several variants of feature vectors.

The above mentioned features have been selected during the long-time studies concerning the evaluation of the speech deformation level and the search for features combining the following three advantages:

- ❑ are insensitive to the content of the statement and personal features of the speaker's voice
- ❑ exhibit great sensitivity for distinguishing between various forms of the same type of pathology and in

classification of various stages of development for a given pathology

- □ are easy to determine from the registered speech signal samples and exhibit the required numerical stability (are insensitive to small errors in the signal measurement)

The authors have selected and studies several feature vectors, for which the respective spaces could be satisfactorily metricized, and which are presented below:

$$\mathbf{<f_1, f_2, \ldots ,f_{96}>=X_1} \qquad (1)$$

where: fi - the averaged level values in the i-th frequency band, with Δf =125Hz

$$\mathbf{<F1, F2, F3, M0, M1, M3>=X_2} \qquad (2)$$

where: F_1, F_2, F_3 - formants, M_0, M_1, M_3 - spectral moments

$$\mathbf{< M_0, M_1, M_3, C_w, C_p, J, S>=X_3} \qquad (3)$$

where: C_w - the relative power coefficient, denoting the ratio of signal power in the reference phone frequency range to the signal power in the whole frequency band of the signal.

C_p - the relative power coefficient, denoting the ratio of the signal power in the remaing frequency band to the signal power in the whole frequency band of the signal

J - Jitter (denotes the deviation of the larynx tone frequency in consecutive cycles from the average frequency of the larynx tone)

S - Shimmer , (denotes the deviation of the larynx tone amplitude in the consecutive cycles from the average amplitude of the larynx tone)

The widely known and simple concepts of sound pattern recognition satisfy their tasks in the routine recognition (comprehension) of the utterance contents or in verification of the speaking person. However they do not meet expectations in the classification of various forms of speech pathology. The reason lies in the great variability and diversity of the acoustic signal of pathological speech. As a result the attempts of finding such feature space, in which a representative description and an effective discrimination between particular forms of speech pathology would be possible, encounter very serious difficulties (problems). The perfection of the human perception, the emulation of which is difficult, particularly for complex phenomena and transient processes (and exactly such are encountered in the context of pathological speech evaluation), results from the fact, that human brain does not work out its decisions using only a precise analysis of the received signals. The human perception always follows from the interaction of two information streams: the inner one (generated by the stored knowledge resources) and the one coming from outside (as a stream of information provided by the reception organs). The idea of using the above mentioned scheme of "cognitive resonance" as a basis for the system of automated diagnosis of pathological speech means, that the diagnostic system has to be furnished with an internal model of the signal generator. Such a model has to include the knowledge about the speech signal and the ways of its generation - both in regular and pathological conditions. The parameters of the model (based on the knowledge about the speech signal) are modified by process of the input signal analysis. In fact both information streams are being processed - the one coming from the external source to the interior of the system and the one transferred from the internal generator to the comparator accomplishing the necessary comparisons and negotiations. The general concept described above has been presented in Fig.2.

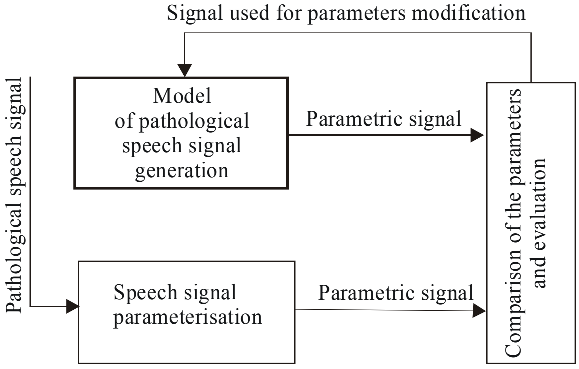
Signal used for parameters modification
Pathological speech signal
Model of pathological speech signal generation
Parametric signal
Speech signal parameterisation
Parametric signal
Comparison of the parameters and evaluation

Fig.2 A simplified diagram of the model concept

The general scheme of the described method, presented above, assumes that the knowledge about the signal, expressed in the form of properly adjusted parameter values, is confronted with the external (sensory) information, coming from the signal of actual pathological speech. This is a completely new approach, based on the attempts of penetration of the causal dependencies [relations], underlying the origins of such or another form of the studied signal deformation Therefore in contrast to the tasks of automated recognition of pathological speech (what was the object of our group's activity for many years) the new approach is called "an automated understanding" of the pathological speech signal. The key to that method is the model of the articulation process (regular or pathological), controlled by the parameters of the observed signal. Therefore in the proposed concept "understanding" denotes the matching of the general knowledge to a specific situation. The outputs of both the model and the system for processing of the pathological speech signal are selected signal parameters. On the basis of the measured similarity of these parameters the choice of optimal model and the correction of the assumed model parameters are done. The concept, described in general above, can be realised in a system presented in Fig.3.

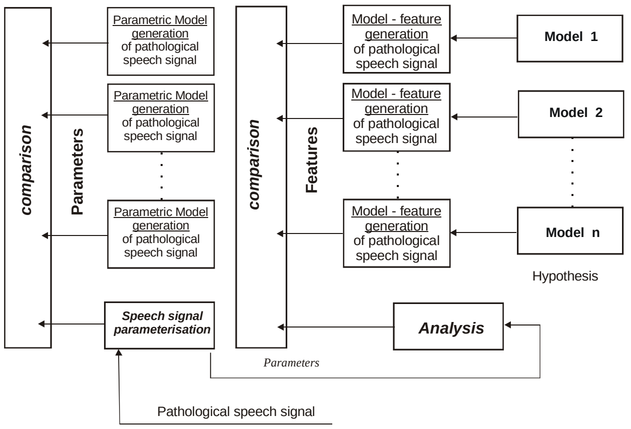
comparison
Parameters
Parametric Model generation of pathological speech signal
Parametric Model generation of pathological speech signal
Parametric Model generation of pathological speech signal
comparison
Features
Model - feature generation of pathological speech signal
Model - feature generation of pathological speech signal
Model - feature generation of pathological speech signal
Model 1
Model 2
Model n
Hypothesis
Speech signal parameterisation
Analysis
Parameters
Pathological speech signal

Fig.3 The concept's block diagram for the parallel design.

The models 1 to n have been adjusted to the previously known cases and forms of speech pathologies. The outputs provided by the generator (model) within that concept are the parameters and features of the generated signal.

4. The model of speech organs simulation

The complex process of acoustic speech signal generation can be presented in the form of a theoretical model mapping functions performed by particular organs. It is essential for the simulation model to enable the determination of the signal spectrum, based on the geometrical parameters of the vocal tract specific for the articulation of particular speech sounds. The basis for presentation of the model has been taken from the works [5, 6, 7]. In the simulation model three principal modules have been distinguished:

- the source of the acoustic wave G, characterised by impedance $Z_g(j\omega)$, refers to phonotory organ
- four-terminal network, characterised by transmittance $K(j\alpha\omega)$, refers to phonotory organ
- load impedance $Z_{lo}(j\omega)$.

which are presented in Fig.4

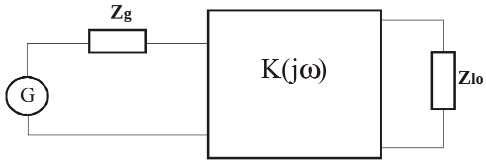

Fig.4. Model block diagram of the speech organs

In the present work a model of larynx generator has been assumed, considered as a source of signals of frequencies F_0, 2 F_0, $3F_0$ etc., where $F_0=1/T_0$, and the amplitude proportion to the up and bellow pressure glottis difference The schematic diagram of which is presented in Fig.5.

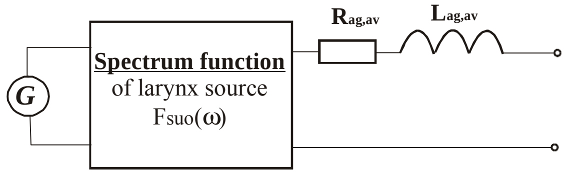

Fig.5. Model block diagram of the larynx.

The introduced notation is as follows: F_{sou}- reflects a simplified envelope of the spectral characteristic $|Ag(j\omega)|$.

$$F_{sou}(\omega) = \frac{1}{\left(\frac{\omega}{\omega_0}\right)^2} \qquad (4)$$

While the resistance R_{agav} and the source's acoustic mass L_{agav} are taken for respective of these elements for average value of the glottis section A_{gav}.

5. Results

On this stage of research the subject is limited to the comparison of pathological signal with the model created one. As a pathological vocal tract the larynx after surgery is assumed. The product of such comparison and evaluation is a signal used for modification of internal model parameters, in order to minimize the difference between the vectors of features of the actual pathological speech signal and the signal generated by the model. . The size and direction of the model modification is a measure of the speech signal deformation degree.

Figs.5 and 6 the spectrum of the A vowel speech signal has been presented for the actual utterance and the signal obtained from the model respectively

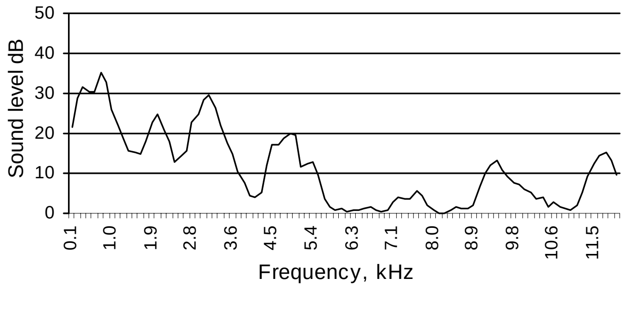

Fig.6.Spectrum of utterance of A vowel

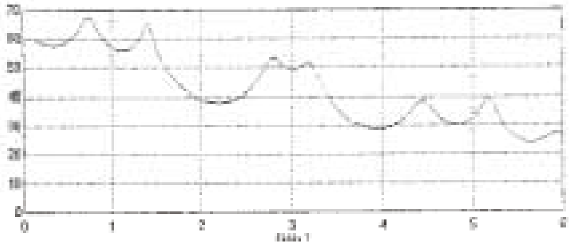

Fig. 7. Simulated spectrum of A vowel

The reference spectrum of vowel „E" is shown in Fig.8.

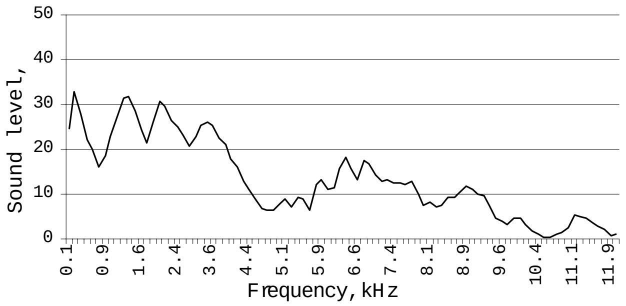

Fig.8. Spectrum of utterance of vowel E, correct pronunciation

In the followed figures the spectrums of vowel "E" created by model for reference.

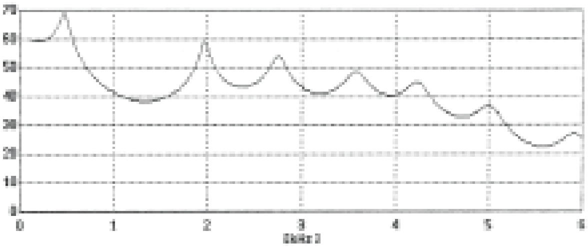

Fig. 9. Simulated spectrum of E vowel – correct

The introduced concept of signal understanding consists of introduction of quantitative factors, describing the essence of the origins of signal deformation (e.g. various pathologies of the vocal tract). The signal of pathological speech is recorded separately of each patient and after that is converted into parameter s spaces and vector features. The model creates signal in the spectrum and features domains. The comparison of these two signals in those domains is carrying out. The magnitude of changes of the selected model parameters is a measure of the signal deformation, and the information specifying which of the model parameters induced the signal change ensuring the greatest similarity determines the level of "understanding" of the deformation origins.

6. Conclusion

The described concept includes a series of elements unquestionably difficult in practical realisation. In conclusion it can be stated, that in the field of automated diagnosis of pathological speech it is necessary to construct special methods of automated understanding of the nature of processes leading to speech deformation, which could replace the presently employed methods of typical acoustic signal analysis and recognition, and which would be fully adapted to the specificity of the considered problem. Because of that the proposed method will have to be considerably modified, in application to various specific tasks.

References:

[1] Tadeusiewicz R., Wszolek W., Wszolek T, Izworski A.: *Methods of Artificial Intelligence for Signal Parameterisation Used in the Diagnosis of Technical and Biological Systems*, 4th World Multic. on Systemics, Cybern. and Informatics, July 23-26,2000 Orlando, FL,USA, Proceedings on CD.

[2] R. Tadeusiewicz, W. Wszolek, A. Izworski, T.Wszolek; *Methods of deformed speech analysis.* Proc. International Workshop Models and Analysis of vocal Emissions for Biomedical Applications, Firenze, Italy, 1-3 September 1999, pp.132-139

[3] Tadeusiewicz R., Wszolek W., Izworski A., (1998), *Application of Neural Networks in Diagnosis of Pathological Speech*, Proceedings of NC'98, "Intern. ICSC/IFAC Symposium on Neural Computation", Vienna, Austria, 1998, September 23-25

[4] Ogiela M. R., Tadeusiewicz R.: Automatic *Understanding of Selected Diseases on The Base of Structural Analysis of Medical Images*, Proceedings of ICASSP 200, Salt Lake City, 2001

[5] Fant G.: *Vocal tract wall effects, losses and resonance bandwidths*, Quart. Progr. Rep. Speech Trans. Lab. In Stockholm, STR-QPSR, 2-3/1972, 28-52.

[6] Flanagan J.L.*: Speech analysis, synthesis and perception.* Springer-Verlag, Berlin-Heidelberg-New York, 1965

[7] Kacprowski J.: *An acoustic model of the vocal tract for the diagnostic of cleft palate*. Speech analysis end synthesis (ed. by W.Jassem), vol.5, 165-183, Warsaw, 1981.

This paper is supported by AGH grant No 10.10.120.39