Artificial Intelligence Methods in Processing and Diagnostics of the Deformed Speech Signals

IZWORSKI ANDRZEJ, TADEUSIEWICZ RYSZARD
Department of Automatics, Laboratory of Biocybernetics
AGH University of Technology
Al. Mickiewicza 30, 30-059 Kraków
POLAND
izwa@biocyb.ia.agh.edu.pl, rtad@biocyb.ia.agh.edu.pl

Abstract: - In the present work excerpts of research are presented, concerning the application of modified acoustic signal processing methods in the problem of "understanding" of selected pathologies of vocal tract. 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. The method is based on utilization of an internal model of the considered signal's generator and it is directed towards such a structure analysis of the examined sound, which enables its identification as a result of cognitive resonance. The described method allows to achieve more subtle differentiation for signal characterized by small diversification of measurable features, observed for the classes being recognized, what is the case in the problem of identification of selected pathologies considered here. The circumstances mentioned above suggest a consideration of more knowledge-based approach to the discrimination of acoustic signals, labelled here as a technique of signal understanding.

Key-Words: - Speech recognition, speech processing, speech pathology, neural networks, signal understanding, artificial intelligence

1 Introduction

In many problems of medical diagnosis, as well as planning and monitoring of therapy and rehabilitation of speech related organs, it is necessary to evaluate qualitative features of the acoustic signal of deformed speech. Tasks related to analysis and recognition of pathological acoustic signal of speech, characterizing selected pathological states, are exceptionally difficult. The difficulty results from the fact that forms of speech organ pathology, which are to be recognized (or classified) manifest themselves in various forms of speech signal deformation, often hard to predict and very inconvenient to be revealed in real, recorded speech signal of a given patient being examined. The correlation between phonetic and acoustic phenomena, observed in the temporal or spectral representation of speech signals, in general poorly correlates with morphological or pathophysiological features of the deformed speech generator. It happens that minor pathological elements (e.g. occlusion defect) strongly manifest in the speech signal, while very serious pathological changes (e.g. tumor) give only a weak and hardly readable picture of speech disturbances. Therefore it is very difficult to diagnose the condition and pathological changes of the voice tract using speech signal [1], in spite of 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). Neither is there a simple way to transfer the experience related to diagnosis of 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. What's more, it is also difficult and sometimes even impossible to indicate a proper recognition algorithm for the pathological speech signal [2]. It follows from the fact that during the identification of voice tract pathological states based on the analysis of generated deformed speech it is necessary to resort to highly specialized (atypical) methods, both for the signal parameterization as well as its categorization and classification. 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, 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 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 recognition of the speech signal i.e. the contents of the pronounced sentences.

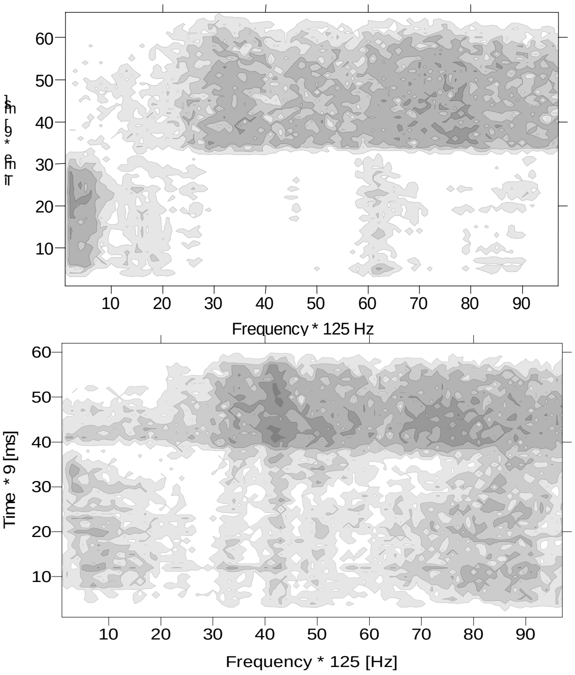

Fig. 1. Spectrogram of the "AS" statement for the reference group and the group of patients

It seems worth explaining, why the authors make an attempt to introduce a new term of automated understanding of voice tract pathologies and recommend a new realization methodology of that automated understanding on the basis of speech signal analysis, instead of confining to the better known problem of recognition and automated classification of pathological speech. As it is known the understanding in general differs from recognition by the fact, that it is very strongly based on the knowledge. In the meaning of the term "automated understanding" discussed here it 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.

Such a new approach is necessary because several previous attempts, made by the authors of the present work, directed towards the construction of a system recognizing types of speech organs pathologies, in spite of unquestionable successes, did not lead to final solutions. The reason seems to be the fact, that every attempt of a simple recognition of speech pathology must be based on the evaluation of some measure of difference between the specific utterance of a given patient and some standard of the correct speech. Alas such a simple recognition concept, superbly working in recognition of the utterance contents, or in verification of the speaker, does not meet the expectations in the attempts of recognition and classification of forms of speech pathology. The reason lies in the great changeability and diversity of speech. It concerns both regular and pathological speech [3]. Every person speaks in 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 e.g. in various days, can be very different. Particularly troublesome is the considerable diversity of correct speech, as it is rather difficult to refer to (in the measurement of degree of speech pathology) a set of speech samples, in which the temporal, spectral and parametric features exhibit huge dispersion of values. It is almost a rule, that various samples of correct speech signal exhibit a greater variety of measurable acoustic parameters, that the measurable differences of the same parameter between these samples and the registered samples of speech, which is obviously pathological (see Fig. 1)

All this is the reason that one cannot confine to models of pathological speech signal recognition in a space based on the set of its features, but in every case one should try to understand, how did such a phonetic or acoustic phenomenon occur. It means, that the diagnostic system must contain an internal model of the signal generator, based on the knowledge about 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 human perception of various informations from the environment.

2 The Material of the 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 (the standard group), 25 persons with a correct pronunciation
- the group of patients (75 persons) treated by the following surgery methods
  - hemilaryngectomia (28 persons)
  - chordectomia (17 persons), enlargata (6 persons)
  - laryngectomia subtotalis (14 persons)
  - laryngectomia fronto-lateralis (5 persons)

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) distinctions 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.

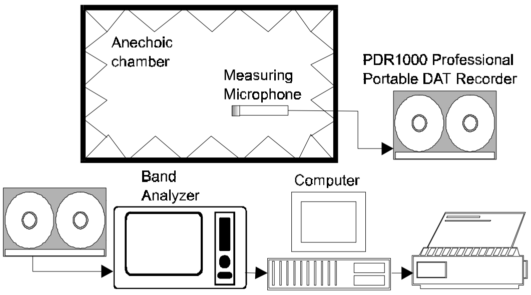

Fig.2 The measurement set-up

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 analyzed using professional, thoroughly tested acoustic analyzers. The block diagram of the measurement set-up has been presented in Fig.2.

After preliminary processing of the registered signal the result is a multispectrum, digitized in time, frequency and amplitude by the acoustic analyzer.

3. The Concept of Research

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 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:

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

where: f_i - the averaged level values in the i-th frequency band, with $\Delta f = 125$Hz

$$<F_1, F_2, F_3, M_0, M_1, M_2> = X_2 \qquad (2)$$

where: F_1,F_2,F_3 - formants, M_0,M_1,M_2 - spectral moments

$$< M_0,\ M_1, M_2, Cw, Cp, J, S> \ = \ X_3 \qquad (3)$$

where:

Cw - 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.

Cp - 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 concept described in the introduction has been presented in Fig.3.

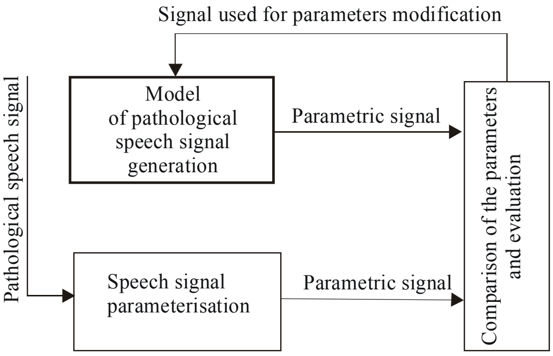
Signal used for parameters modification
Pathological speech signal
Model of pathological speech signal generation
Parametric signal
Speech signal parameterisation
Parametric signal
Comparison of the parameters and evaluation

Fig.3 A simplified diagram of the model concept

The model of signal generation represents all the knowledge about the pathological speech signal. The products of the model are spectra of the signal. The actual signal of pathological speech (obtained from a particular patient) after its transformation into the vector of features is compared with a transformed output signal of the model.

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 [7,8,9,10]. 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., the schematic diagram of which is presented in Fig.4.

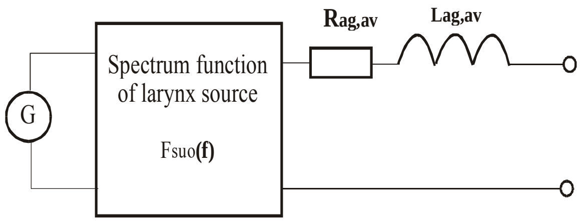

Fig.4 Simplified diagram of the larynx

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

$$F_{sou}(f) = \frac{1}{\left(\frac{f}{F_0}\right)^2} \qquad (4)$$

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

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. In Figs. 5 and Fig. 6 the spectrum of the I vowel speech signal has been presented for the actual utterance.

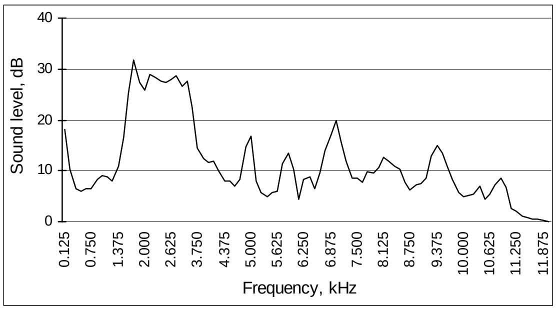

Fig.5. Spectrum of utterance of vowel I - correct pronunciation

The signal obtained from the model has been presented in Figs 7 and 8.

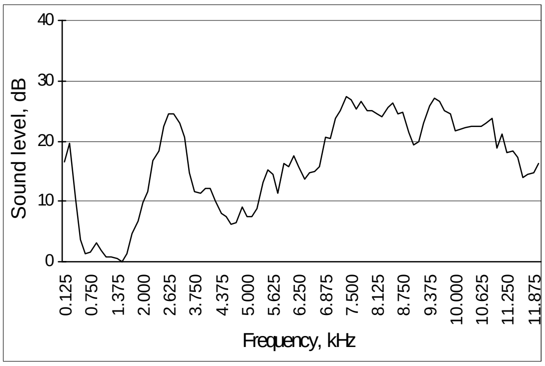

Fig.6.Spectrum of utterance of vowel I – pathological pronunciation

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 speech signal recorded for a particular patient and the signal created by the generation model (in the form of the spectrum) are processed to the form of vectors of features and then compared (using the artificial neural networks) with respect to their similarity. The result of the evaluation is used for elaboration of such correction of the respective model parameters, which result in the greatest similarity of both signals.

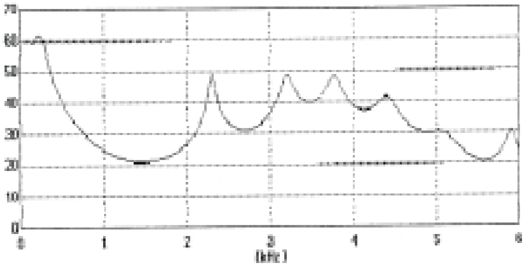

Fig. 7. Simulated spectrum of I vowel - correct

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

With such a solution, essentially different from solutions employed in the presently constructed diagnostic systems, the speech signal, received from a given patient, is recorded and analyzed (usually by use of neural networks [4]), and then it is confronted (most often in the plane of properly selected temporal and spectral multispectral features) with the reference signals, formed by the internal models mentioned above (the generators of pathological speech specific for known pathology forms). The process of adjustment of parameters of the registered signal, obtained from a given patient, and the signals obtained from the internal generators, leads first to the selection of this generator for which the strongest cognitive resonance is observed (in the process of automated understanding of the speech deformation source it is equivalent to the stage of formulation of a diagnostic hypothesis), and in the next stage of the perception modeling to the process of adjustment of internal generator parameters, executed for tuning it to the parameters of the observed signal, which leads to formulation of more exact (final) diagnostic hypothesis.

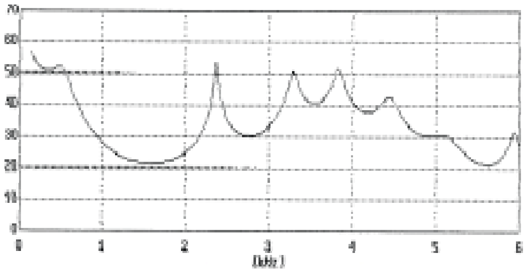

Fig. 8. Simulated spectrum of I vowel - pathological

The described concept includes a series of elements unquestionably difficult in practical realization. For the traditional way of solving diagnostic problems the answer is frequently found more easily. However longtime experience of the authors in the problems related to analysis, evaluation and classification of pathological speech signals have proved, that for this task really a new approach is required. It is because for recognition of pathological speech the standard signals processing and classification methods, used in semantic speech recognition (comprehension of the utterance contents) or voice recognition (speaker identification), totally fail [3]. These methods include the spectral analysis (sometimes executed with application of the wavelet transformation technique, popular recently), discrimination analysis, linear prediction coefficients or cepstral coefficients. They cannot satisfactorily describe the pathological speech, because of its phonetic and acoustic structure, dissimilar with respect to the correct speech signal, and also because the goal of the recognition process is totally different for that case [4]. At the same time the amount of needs related to technical assistance in diagnostic, prognostic check up tasks, performed by the physicians dealing with speech pathology, constantly grows. Successful attempts of construction of picture recognition systems [5], [6] indicate that the proposed way may be effective.

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. Such an attempt of construction of a new, fully original, special method of understanding is the concept described in the present work. It refers to solving of the problems considered here, making use of proper representation of the knowledge regarding the studied articulation process and the consequences of its deformation (in the form of adaptively adjusted models). The studies of the new method have just started, and it cannot be told whether this technique will be able to solve all the problems and to overcome all the difficulties. In general it is known, that in the tasks of acoustic signal (picture) analysis and recognition the unification of methods and standarization of algorithms has always encountered serious problems. The main source of those difficulties is the fact, that in almost every task of signal analysis, different features and different parameters, closely related to the specificity of task being solved, have to be extracted, and they are used for finding answers to different questions. Similarly in the tasks of acoustic signals recognition the criteria and goals of their classification can be very different - even for the same signals. Because of that the proposed method will have to be considerably modified, in application to various specific tasks. The adaptation will affect both the methods of preliminary processing of acoustic signals, which obviously have to be oriented for the specific features of every identification task considered, and the techniques of internal modeling of the generation processes for various forms of speech pathologies. Also the techniques of appointing of the optimal model have to be specific, because as mentioned before, the tasks of pathological speech analysis are special by the fact that no shape of standard signal, to which a reference or relation could be made, can be found.

References:

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

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

[3] Tadeusiewicz R., Izworski A., Wszołek W., (1997), Pathological Speech Evaluation Using the Artificial Intelligence Methods, Proceedings of “World Congress on Medical Physics and Biomedical Engineering”, September 14-19, 1997, Nice, France

[4] Tadeusiewicz R., Wszołek W., Izworski A., Application of Neural Networks in Diagnosis of Pathological Speech, Proceedings of NC’98, “International ICSC/IFAC Symposium on Neural Computation”, Vienna, Austria, 1998, September 23-25

[5] Leś Z., Tadeusiewicz R.: Shape Understanding System - Generating Exemplars Of The Polygon Class, in Hamza M.H., Sarfraz E. (eds.): Computer Graphics and Imaging, IASTED/ACTA Press, Anaheim, Calgary, Zurich, 2000, pp. 139-144

[6] 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

[7] Fant G.: Acoustic theory of speech production, s’-Gravenhage, Mouton and Co. 1960

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

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

[10] 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 was supported by AGH grant 10.10.120.39