AUTOMATED UNDERSTANDING OF SELECTED VOICE TRACT PATHOLOGIES BASED ON THE SPEECH SIGNAL ANALYSIS

Wiesław Wszołek[1], Ryszard Tadeusiewicz[2], Andrzej Izworski[2], Tadeusz Wszołek[1]
[1]Department of Mechanics and Vibroacoustics, [2]Department of Automatics
University of Mining and Metallurgy, Kraków, Poland

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, labeled here as a technique of signal understanding.**

Keywords **- Speech recognition, speech processing, speech pathology, neural networks, signal understanding**

I. 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 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. 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 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 information from the environment.

Report Documentation Page

Report Date	Report Type	Dates Covered (from... to)
25OCT2001	N/A	-

Title and Subtitle Automated Understanding of Selected Voice Tract Pathologies Based on the Speech Signal Analysis	**Contract Number**
	Grant Number
	Program Element Number
Author(s)	**Project Number**
	Task Number
	Work Unit Number
Performing Organization Name(s) and Address(es) Department of Mechanics and Vibroacoustics, Department of Automatics University of Mining and Metallurgy, Kraków, Poland	**Performing Organization Report Number**
Sponsoring/Monitoring Agency Name(s) and Address(es) US Army Research Development & Standardization Group (UK) PSC 802 Box 15 FPO AE 09499-1500	**Sponsor/Monitor's Acronym(s)**
	Sponsor/Monitor's Report Number(s)

Distribution/Availability Statement
Approved for public release, distribution unlimited

Supplementary Notes
Papers from the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, October 25-28, 2001, held in Istanbul, Turkey. See also ADM001351 for entire conference on CD-ROM.

Abstract

Subject Terms

Report Classification unclassified	**Classification of this page** unclassified
Classification of Abstract unclassified	**Limitation of Abstract** UU

Number of Pages
4

II. 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) disfunctions 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 a professional recording equipment and analyzed using professional, thoroughly tested acoustic analyzers.

The block diagram of the measurement setup has been presented in Fig.1.

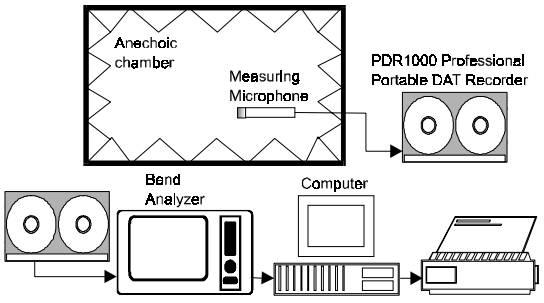

Fig.1 The measurement setup

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

III. 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 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:

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

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

$$<F_1, F_2, F_3, M_0, M_1, M_3>=X_2 \quad (2)$$

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

$$< M_0, M_1, M_3, Cw, Cp, J, S>=X_3 \quad (3)$$

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 remaining 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.2.

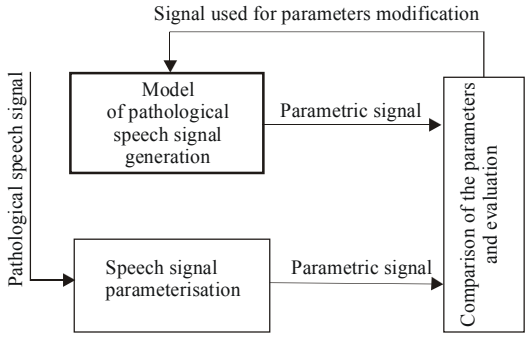

Fig.2 A simplified diagram of the model concept

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.

IV. 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 simulation model three principal modules have been distinguished:

- the source of the acoustic wave G, characterized by impedance Zg,
- - four-terminal network, characterized by transmittance K(jω),
- load impedance Z_{lo}

which are presented in Fig.3

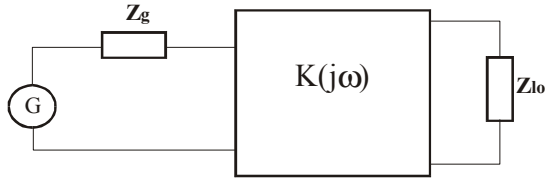

Fig.3. Model block diagram of the speech organs

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., the schematic diagram of which is presented in Fig.4

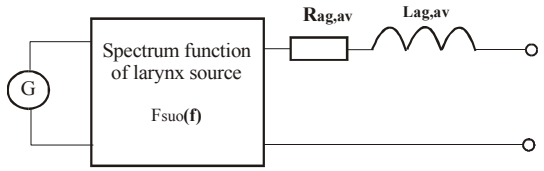

Fig.4 Simplified diagram of the larynx

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}.

V. 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 6 the spectrum of the I vowel speech signal has been presented for the actual utterance.

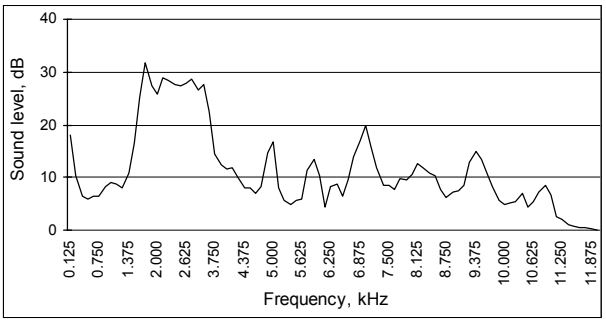

Fig.5. Spectrum of utterance of vowel I - correct pronunciation

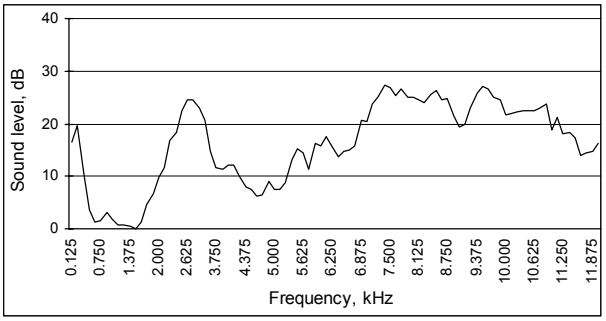

Fig.6.Spectrum of utterance of vowel I – pathological pronunciation

The signals obtained from the model has been presented in figs 6 and 7

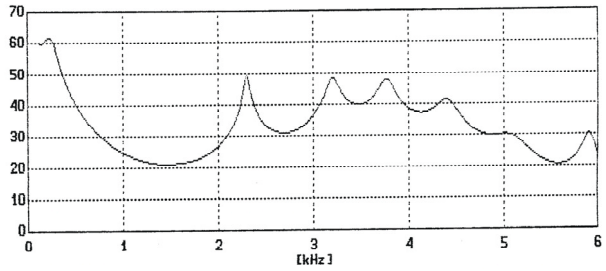

Fig. 6. Simulated spectrum of I vowel - correct

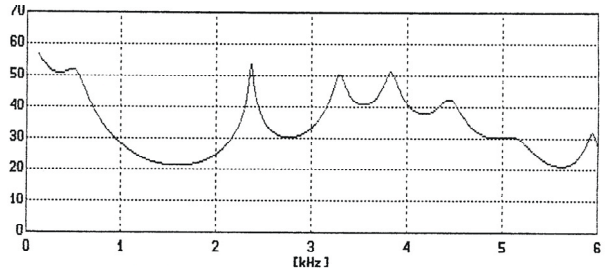

Fig. 7. Simulated spectrum of I vowel - pathological

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. 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.

V. Conclusion

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. 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. 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 standarisation of algorithms has always encountered serious problems.

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 Multiconference on Systemics, Cybernetics and Informatics, July 23-26,2000 Orlando, FL,USA, Proceedings on CD.

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

[3] Tadeusiewicz R., Izworski A., Wszolek 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.