Wiesław WSZOŁEK*, Maciej KŁACZYŃSKI*

METHODS OF DEFORMED VOICE SIGNAL EVALUATION AFTER LARYNX SURGERY

SUMMARY

In the work has been shown from studies concerning the application of modified acoustic signal processing methods to the task of evaluation and classification of larynx surgery effects. The goal of the standard speech recognition studies is to reveal the semantic aspects of the pronounced text. In the tasks of medical diagnosis employing the speech signal analysis the semantic aspects are insignificant. The required signal characteristics should be as sensitive as possible to small deformations of the layers directly related to the voice functioning and the structure of vocal tract. The goal of the work is presentation of voice quality after various surgical treatments, performed in the ENT area. The research subject is the speech articulation process itself and all its pathological deformations, which determines both the used signal analysis tools as well as the techniques of the selected objects recognition, which are the forms of the particular ill person speech deformation forms in comparison to the speech of the whole sound people population. The evaluation has been carried out both for voice quality after larynx surgery as well as voice quality after surgical treatment of resonance cavities (nose, paranasal sinussis). The study was oriented towards the construction of systems based on the analysis of objectively registered acoustic signals of deformed speech.

Keywords*: speech analysis, pathological speech, speech recognition, surgical treatment*

METODY OCENY ZNIEKSZTAŁCONEGO SYGNAŁU MOWY PO OPERACJACH KRTANI

W pracy przedstawiono badania dotyczące metod przetwarzania sygnału akustycznego do oceny i klasyfikacji mowy po zabiegach w obrębie kanału głosowego. W zagadnieniach rozpoznawania mowy, problem dotyczy ujawniania semantycznych aspektów wypowiedzi. Natomiast w zagadnieniach diagnostyki medycznej przy wykorzystaniu sygnału mowy, cechy semantyczne są nieistotne. Poszukiwane cechy sygnału mowy winny być wrażliwe na małe deformacje, które mogą wystąpić w poszczególnych warstwach kanału głosowego. Celem pracy jest ocena jakości głosu po różnorodnych zabiegach chirurgicznych wykonanych w obszarze kanału głosowego. Tematem badań jest zarówno sam proces artykulacji mowy, jak i jego patologiczne deformacje. Diagnostykę narządu głosu można określić jako jednoznaczne rozpoznanie cech aktualnego stanu źródła głosu na podstawie zespołu istotnych cech akustycznych, zwartych w sygnale akustycznym. Ocena jakości głosu została przeprowadzona dla osób po chirurgicznym leczeniu krtani, nosa oraz zatok przynosowych. Badania zostały ukierunkowane na stworzenie systemu analizy umożliwiającego obiektywne rozpoznawanie deformacji sygnału mowy.

Słowa kluczowe*: analiza mowy, mowa patologiczna, rozpoznawanie mowy*

1. INTRODUCTION

In many problems of medical diagnosis, as well as in planning and monitoring of the therapy and rehabilitation of vocal organs, the evaluation of quality of the deformed speech signal is very important. The main purpose of the research projects mentioned above was to increase the accuracy of pathology detection and eliminate the most dangerous error – classification of a patient with laryngeal disease as a normal speaker (Bull 1999).

Pathological processes that affect the vocal tract in most cases cause changes in the speech production process, which can be heard as abnormal voice. These changes are often the first, isolated and therefore very important symptom in early stages of larynx pathologies. It should be emphasized, that voice problems such as hoarseness are frequently underestimated by the patients. Moreover, they often cannot be properly diagnosed on the most accessible primary health care level, since they require an expert laryngologist and expensive professional equipment. Consequently, in many cases the correct diagnosis and treatment is introduced in advanced stages of the disease, often when it is no longer possible to cure the patient. There are many diseases which can be characterized by the sequence described above, but it is evident that the most important problem in this field is early detection of larynx cancer. It is well known that at early stages the disease can be cured with minimally invasive methods, while advanced stages of cancer often require more aggressive, crippling treatment, including permanent loss of the ability to speak. The mortality rate in advance stage is also significantly higher than in early stages (Hadjitodorov *et al.* 2000). An easily accessible, low-cost and noninvasive method of laryngological prediagnosis, based on acoustic signal analysis and advanced processing of the phonological data could therefore improve the detection and treatment of larynx diseases. There have been many attempts to create a reliable computer system, which could distinguish between patients without serious vocal tract problems and those who need a consequent laryngological diagnosis and treatment. Acoustic parameters that can be extracted from the signal reflect such changes in the voice as loss of power, changes in the pitch,

* Faculty of Mechanical Engineering and Robotics, AGH University of Science and Technology, Krakow, Poland

constriction of the voice range (displacement towards lower frequency), addition of noises, etc. which are important from the medical point of view (Deller *et al.* 1993).

2. SPEECH SIGNAL GENERATED

The process of human speech generation is a complex phenomenon, comprising many topics in psychology, biology, medicine, as well as aerodynamics and acoustics. In a simplified description one can distinguish two basic layers of features that are specific for a given speaking person: the physical layer – originating from the anatomical structure of the vocal tract (the source and filters) and the psychological layer, related to the individual manner of controlling the phonation and articulation organs. In the physical layer it is necessary to distinguish the two stages of speech generation and recognize the possibility of separate definition and measurement of the parameters of source and filter. The speech signal, treated as a time-dependence of acoustic pressure (upper part of the Figure 1), exhibits a complicated time-course, reflecting the complex nature of the process of its articulation. The signal parameters are affected by the source (vibrating vocal cords, or the noise of turbulent flow of the air-stream through the straits in speech organs) as well as the dynamical properties of the vocal tract, forming the final structure of the signal. Deformation of the vocal organ, related to the larynx dysfunction, manifests itself in the change of the vocal cords vibrations parameters, what influences on F_0 (Titze 1994). Exact determination of the fundamental tone function becomes a priority in the voice gene-rator research. An accurate and frequently used method is the electrical method, EGG[1], based on the recording and analysis of the electroglottographic signal (Marasek 1997) (lower part of the Fig. 1).

In the time domain the signal can be mathematically described using a convolution of the original time dependence of the source signal $g(t)$ and the impulse response of the vocal tract $h(t)$:

$$p(t) = \int_0^t h(t-\tau)g(\tau)d\tau \tag{1}$$

Interpretation of the above-mentioned formula indicates that in the time-dependent acoustic speech signal the properties of the source and the properties of the sound forming voice channel are closely related (Jurkiewicz *et al*. 2006).

Speech acoustics provides several methods of speech signal quality evaluation, enabling a multilateral analysis with its results visualisation and their changeability process during speaking. However, the direct analysis of this type of a process is very complex and requires a lot of experience, especially in case of pathological speech analysis. Hence the methods of automation analysis processes and speech signals recognition are developed and still enriched, and the results of these examinations are presented in number of works, including this article.

Before we go to the detailed consideration it is necessary to point what new and original values this particular article implements, in relation to this extremely rich and diversified bibliography of the subject (Modrzejewski *et al*. 1999, Rabiner 1993, Reroń *et al*. 1998, Titze 1994).

Namely, the key term is the fact that the subject of research in this work is a pathological speech signal, and the crucial aim of the research is determining the nature, kind, and degree of illness changes advance, manifested by acoustic changes in the considered speech signal.

In a typical research, concerning speech recognition, the aim is mostly to disclose (through selected parameters) the

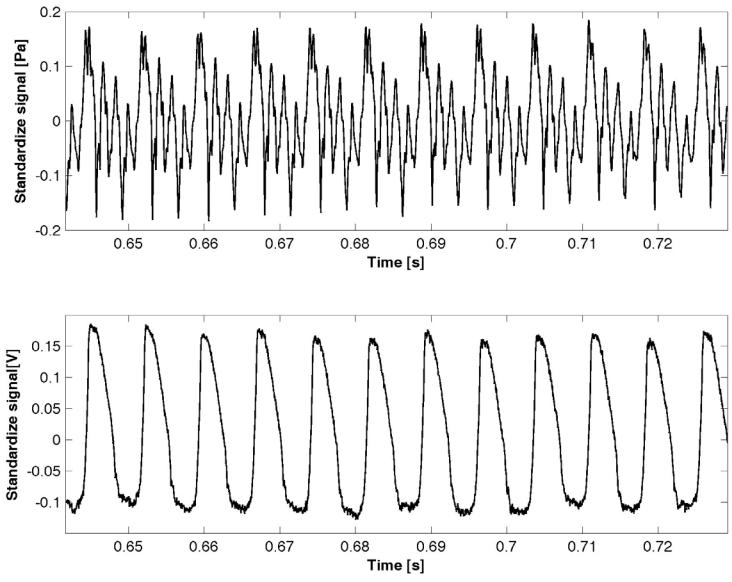

Fig. 1. Acoustic speech signal, vibrations of vocal folds

[1] Electroglottography is a noninvasive method of the glottis electrical impedance measurement. The impedance is measured between two electrodes placed on the subject's skin on the larynx level.

semantic aspects of the pronounced text, in case of medical diagnosis, based on the pathologically deformed speech signal analysis; the semantic content of the statement is insignificant (and can even be treated as a disturbance).

Selected methods of transformation and signal analysis in the task of pathological speech evaluation were presented in the article; moreover their usefulness in the selected medical issues was discussed. The wide range of various medical issues, which are presented in the work as the problem areas of the performed research are worth stressing. It proves that the acoustic techniques and methods described in the article, as well as the techniques related to transforming signals, are quite universal, although from the medical point of view quite often they can concern very different tasks (Titze 1994).

It was stated that the most important (and the most difficult!) element of research work preceding a practical using speech as the source of medically useful diagnostic and prognostic information, is finding and describing these signal parameters, which are maximally independent, both from the context and the personal features of the examined voice. Additionally searched features of the signal must be maximally sensitive to its deformations in this layer, even the small ones, which is connected with the structure and functioning of speech signal generators (larynx and strangulations, being the source of speech noise constituents) and with the voice tract structure being used during articulation (Tadeusiewicz 1988, Wszołek 2006).

During the research special attention has been focused on the analysis and description of the structure for the feature space describing the pathological speech signal, because the exact knowledge of the feature space topology (which is not easy for a direct evaluation because of the its multidimensional nature) enables the subsequent effective application of proper automated recognition methods.

3. SPEECH DATA PROCESSING

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 person's clarity of speech has been evaluated using a verbal test including the forms of signal generation and its articulation, which have been selected as carrying the greatest amount of diagnostic information. 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.

Recording of sound pressure time waveforms, $p(t)$, of speech signal was made in the anechoic chamber at the Chair of Mechanics and Vibroacoustics, AGH. The recording system employed a professional digital magnetic recorder (HHB type PDR 1000).

The examined group consisted of patients with disease changes in their vocal folds, glottis, and larynx. The database of the deformed speech signals comprised recordings from 80 patients treated in the Otolaryngology Clinic CMUJ in Cracow. Data were collected at three stages of patients medical treatment:

- first recording – before a surgery, approximately 7 to 14 days before the planned date of treatment, on the day of visit at clinic,
- second recording – during early check-up, approximately 14–30 days after the surgery,
- third recording – during late check-up, approximately 90 days after the surgery.

In addition, voices of 36 persons were recorded solely before the hospital treatment (what corresponds to first recording above). During the same period of time normal voices of 128 persons, both male and female, which no pathology of voice found, were recorded as the reference for the standard Polish language.

The product of the initial transformation of the recorded signal was a dynamic spectrum $W(i, j)$, digitized in time, frequency and amplitude by output circuit of the acoustic analyzer (Wszołek 2006). In order to standardise the research process, and in order to provide the results comparability, the same scheme of signal transformation, with the amplitude resolution quant Δs = 1 dB and with an evenly frequency digiti-zed signal in the waveband, f_d = 125 Hz, fg = 12 kHz every Δf = 125 Hz was applied. Instant spectra were determined with the usage of a digital analyser and were timely digitised (sampled) every Δt = 9 ms. The applied professional registration system provided a transfer band from 20 Hz to 20 kHz at the dynamics amounted to not less than 80 dB. The dynamic spectra obtained from the analysis have been sometimes used directly in the present study as vectors of distinctive features for analysis and evaluation of the pathological speech signal, particularly in the preliminary stage of the study, when there is a need to reveal the essence of the irregularity in the time-frequency structure of the speech signal produced by the person affected by one of the stu-died pathology forms, but because of redundancy and considerable dimension of such a feature space it has been usually transformed to feature vectors X of the following form (Szaleniec *et al.* 2005):

$$\langle f_1, f_2, f_3, ..., f_{96} \rangle = X_1 \tag{2}$$

where: f_i – averaged amplitude of j dynamic spectrum waveband

$$\langle F_1, F_2, F_3, M_0, M_1, M_2 \rangle = X_2 \tag{3}$$

where:

F_1, F_2, F_3 – forments' frequencies,
M_0, M_1, M_2 – spectra moments introduced defined following

$$M_n(m) = \sum_{i=f_d}^{i=f_g} \left|G_n(t_j, f_i)\right| [f_i]^m \tag{4}$$

$$F_{kj} \equiv f_{kj} \Leftrightarrow \left[\left(\frac{\partial G_n(t_j, f)}{\partial f} = 0\right) \wedge \left(\frac{\partial^2 G_n(t_j, f)}{\partial f^2} < 0\right) \wedge \left(f_d \le F_{kj} \le f_g\right)\right] \tag{5}$$

where:

$G_n(t_j,f)$ – frequency spectrum in the j^{th} time instant, t_j,
f_d, f_g – lower and higher limiting frequencies for frequency band in which spectrum moment was determined on the discrete frequency scale,
k – forment number k = 1, 2, 3, 4, at the j^{th} time instant, t_j.

$$\langle M_0, M_1, M_2, WS_S, WS_1, WS_2, WS_3 \rangle = X_3 \tag{6}$$

where:

WS_s – relative power coefficient determining the ratio of signal power in the pattern phoneme band (determined on the basis of statistical research of undeformed speech), to signal power in the whole band of pathological speech signal,
WS_i – relative power coefficient determining the ratio of signal power in the i-band (i = 1, 2, 3) to signal power in the whole band (selection of the released bands boundaries was one of the main research problems solved within the framework of the research presented here).

$$\langle M_0, M_1, M_2, C_W, C_P, J, S \rangle = X_4 \tag{7}$$

where:

C_w – the relative power coefficient, denoting the ratio of signal power in the reference phoneme 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 remaining frequency band to the signal power in the whole frequency band of the signal,
J – Jitter, denotes a frequency deviation of the basic tone – in consecutive periods,
S – Shimmer, denotes an amplitude deviation of the basic tone – in consecutive periods.

$$\langle M_0, M_1, M_2, F_1, F_2, F_3, F_4, AF_1, AF_2, AF_3, AF_4, WS_1, WS_2, WS_3, C_1, C_2, C_3, C_4, C_5, F0\_SR, J, S \rangle = X_5 \tag{8}$$

where:

$AF_1 - AF_4$ – forments' amplitudes values,
$F0\_SR$ – basic frequency medium value of larynx tone

$$C_n = \sqrt{\frac{2}{N}} \sum_{i=1}^{N} \log(s_i) \cdot \cos\left[\frac{\prod \cdot n}{N}(i - \frac{1}{2})\right] \tag{9}$$

where:

C_n – n^{th} cepstral coefficient,
S_i – i^{th} coefficient obtained from signal conversion by the set of filters,
N – number of filters in the set, N = 12.

In the research, the following features, comprising three following advantages were sought:

– being little sensitive to the statement content and to individual features of the speaker's voice,
– demonstrating high sensitivity while diversifying various forms of speech pathologies and while classifying various degrees of the same pathology type advancement,
– being easily determined on the basis of the recorded speech signal samples and demonstrating the desired numerical stability (are little sensitive to small errors in the signal measurement)

It seems that the sets, presented above, realize the given task in the highest quality standard.

4. SELECTED RESEARCH RESULTS

The selected results of research are presented below. Figure 2 provide visualisation of the feature vector of the deformed speech recorded form male patient diagnosed with the laryngeal cancer. Correspondingly, Figure 3 present the feature vectors of the deformed speech of the female patient diagnosed with the chronic la-ryngitis. Criteria for an objective assessment of speech signals are based on distance metric in the space of features. Corresponding calculations were performed for simple Hamming metric and standardised Euclidean metric. These metrics allow for objective ordering of certain measures, which subjectively correspond to aurally perceived differences among patient's voice and an average for normal voice.

In Figure 4 the deformation of vowels in relation to the correct speech in case of people suffering from various illness changes in the larynx area is shown (Szaleniec *et al.* 2003).

In Figure 5 there was shown the degree of a vowel deformation in relation to the correct speech in case of a person suffering from larynx cancer in various periods of time, before the surgical operation, after the operation and after a long period of rehabilitation.

Figures 4 and 5 "distance vaules from the the standard" are presented as the normalised distance value between the standard speech and pathological speech according to established metrics in the chosen space of features.

The presented results definitely confirm the opinion that the selected parameters characterise well the analysed phenomena connected with the speech signal degradation forms in the context of selected pathologies.

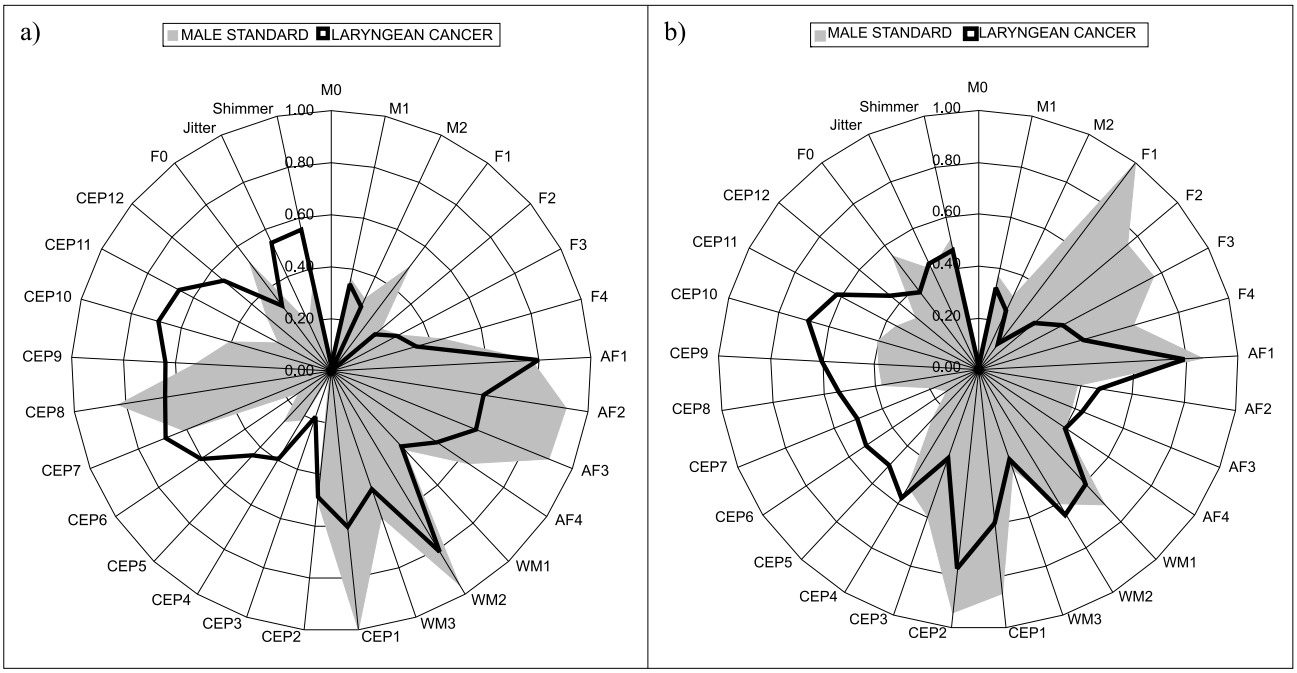

Fig. 2. Graphic interpretation of a feature vector for deformed male speech (laryngean cancer). Vowel /e/ (e.g., /test/) with prolonged phonation. Notes. M0–M2, F0–F4, AF1–AF4, WM1–WM3, CEP1–CEP12 – co-ordinates of the feature vector (a). Graphic interpretation of a feature vector for deformed male speech (laryngean cancer). Vowel /u/ (e.g., /puk/) with prolonged phonation. Notes. M0–M2, F0–F4, AF1–AF4, WM1–WM3, CEP1–CEP12 – co-ordinates of the feature vector (b)

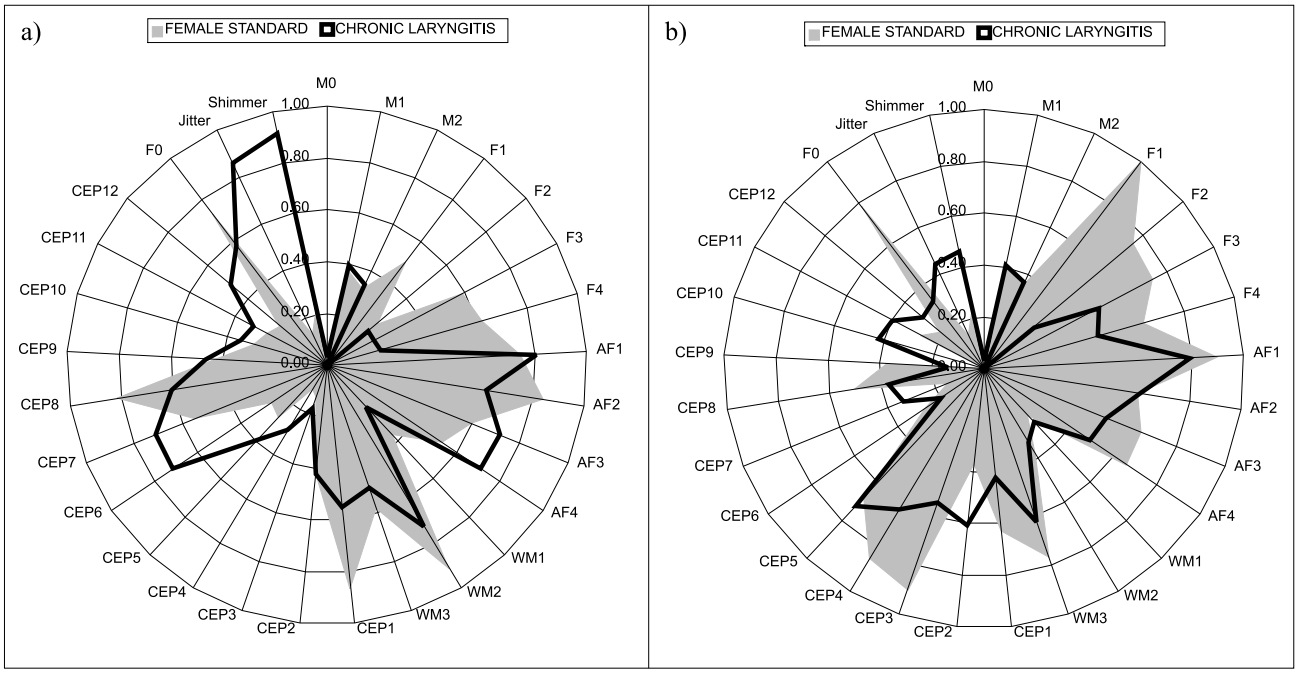

Fig. 3. Graphic interpretation of a feature vector for deformed female speech (chronic laryngitis). Vowel /a/ (e.g., /pat/) with prolonged phonation. Notes. M0–M2, F0–F4, AF1–AF4, WM1–WM3, CEP1–CEP12 – co-ordinates of the feature vector (a). Graphic interpretation of a feature vector for deformed female speech (chronic laryngitis). Vowel /i/ (e.g., /pit/) with prolonged phonation. Notes. M0–M2, F0–F4, AF1–AF4, WM1–WM3, CEP1–CEP12 – co-ordinates of the feature vector (b)

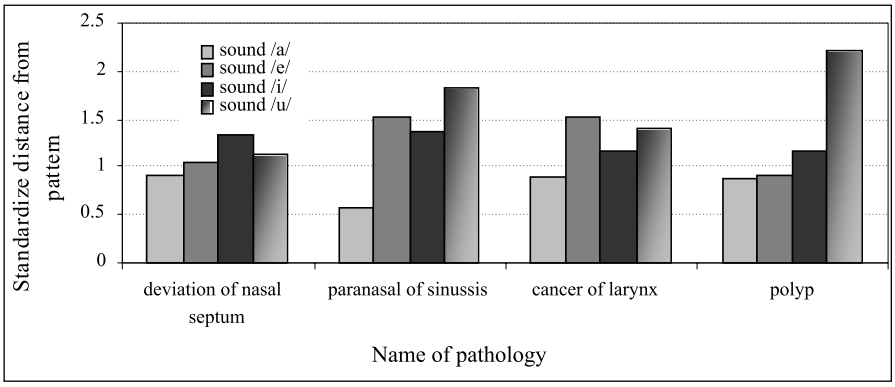

Fig. 4. Distance from standard in case of Cammber's formula – examination before operation

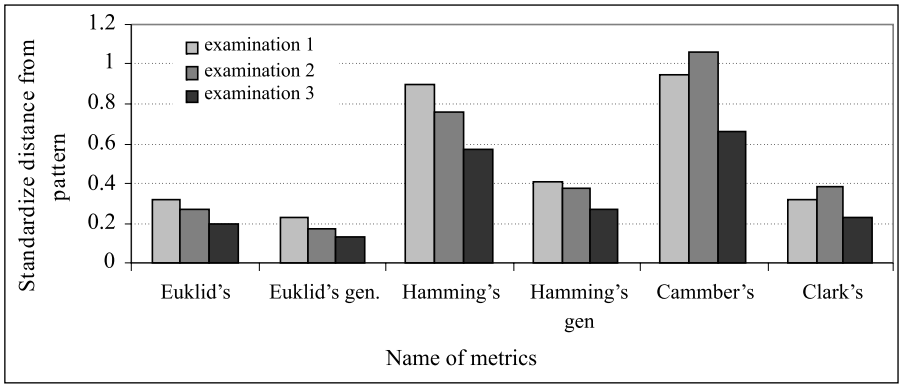

Fig. 5. Distance from standard in case of Cammber's formula, examination 1 – before operation, examination 2 – after operation, examination 3 – six months after operation

5. CONCLUSIONS

Integrated acoustical analysis of deformed pathological speech including mapping, visualisation and quantitative assessment was presented in the paper. This analysis conducted among groups of patients showed that speech pathology caused by various laryngeal diseases can be assessed using acoustical methods supported by distance metrics measuring the degree of signal deformation. Assessment executed before the surgery and during recovery period corresponds to changes in vocal tract, may be used for checking the decrease in speech deformation, and may be a useful tool to control prosthetic restoration and rehabilitation.

The obtained parameters (co-ordinates of a feature vector), was useful for development of vibroacoustic model of the diseases of the human vocal tract. In this model, the feature vectors of the deformed speech signal are presented as graphs or tables. Information presented in such a way can be used by phoniatrists, laryngologists and logopedics as additional objective tool for estimating the degree of speech deformation. Physicians can use visual representation of acoustical analysis to assess changes in speech signal at different stages of medical treatment.

The presented methodology can be directly applied in the monitoring examinations for patients after larynx surgery as well as the operations of the nose and paranasal sinussis and intubation in general anaesthesia.

The results of the present study seem to be helpful in proper qualification of patients for the specific types of operations, in order to provide at the same time the maximal therapeutic effect and minimal speech deformation degree after the operation.

References

Bull Ph.D. 1999, *Lecture Notes on Diseases of the Ear, Nose and Throat*. Gdansk, Via Medica.

Deller J.R. Jr., Proakis J.G., Hansen J.H.L. 1993, *Discrete-Time Processing of Speech Signals*, Macmillan Publishing Company, New York.

Engel Z., Kłaczyński M., Wszołek W. 2007, *A Vibroacoustic Model of Selected Human Larynx Diseases*, International Journal of Occupational Safety and Ergonomics (JOSE), Vol. 13, No. 4, pp. 367–379.

Hadjitodorov S., Mitev P., 2002, *A computer system for acoustic analysis of pathological voices and laryngeal diseases screening*. Medical Engineering & Physics 24, pp. 419–429.

Hadjitodorov S., Boyanov B., Teston B. 2000, *Laryngeal Pathology Detection by Means of Class-Specific Neural Maps*. IEEE Trans Inf Technol Biomed, 4 (1), pp. 69–73.

Jurkiewicz D., Dzaman K., Rapiejko P. 2006, *Laryngeal cancer risk factors*. Polski Merkuriusz Lekarski, 21(121), pp. 94–8.

Koike Y. 1971, *Application of some acousitc measures for the aevaluation of laryngeal dysfunction*. Studia Phonologica 7, pp. 45–50.

Marasek K. 1997, *Electroglottography description of voice quality*. Phonetic AIMS, Univesitat Stuttgard.

Modrzejewski M., Olszewski E., Wszolek W., Reroń E., Strek P. 1999, *Acoustic assessment of voice signal deformation after partial surgery of the larynx*. Auris Nasus Larynx, International Journal of ORL & NNS, 26, Japan, pp. 183–190.

Rabiner L.R. 1993, *Fundamentals of speech recognition*. Prentice-Hall, Inc.

Reroń E., Tadeusiewicz R., Modrzejewski M., Wszolek W. 1998, *Application of Neural Networks and Pattern Recognition Methods to the Evaluation of Speech DeformationDegree for Patients Surgically Treated for Larynx Cancer*. Neuroendocrinology Letters, vol. 19, No. 3, Mattes-Heidelberg-Germany, pp. 147–157.

Szaleniec J., Modrzejewski M., Wszolek W. 2003, *Application of Acoustic Analysis of Speech Signal for Evaluation of Intubation-Related Damages of the Speech Organ*. 3rd International Workshop MAVEBA Proceedings 2003, pp. 269–272.

Szaleniec J., Modrzejewski M., Wszolek W. 2005, Research *on the Influence of Endotracheal Intubation on the Speech Signal*. Speech Analysis, Synthesis and Recognition in Technology, Linguistics and Medicine. Materiały konferencji naukowej Szczyrk 23–26.09.2003, 127–133.

Tadeusiewicz R. 1988, *Sygnał mowy*. WKiŁ, Warszawa.

Titze I.R. 1994, *Principles of Voice Production*, Englewood Cliffs, Prentice Hall.

Wszołek W. 2006, *Selected methods of pathological speech signal analysis*. Archives of Acoustics, vol. 31, no. 4, pp. 413–430.