Models and Analysis of Vocal
Emissions for Biomedical Applications
(MAVEBA 1999)
Firenze, Italy, September 1-3, 1999

METHODS OF DEFORMED SPEECH ANALYSIS

Ryszard Tadeusiewicz[1], Wies aw Wszo ek[2], Andrzej Izworski[3], Tadeusz Wszo ek[4].
University of Mining and Metallurgy, Kraków, Al. Mickiewicza 30, Poland
[1]e-mail: rtad@biocyb.uci.agh.edu.pl, [2]wwszolek@uci.agh.edu.pl, [3]izwz@biocyb.uci.agh.edu.pl, [4]twszolek@uci.agh.edu.pl

Abstract

In the present paper a set of artificial intelligence methods are presented, with special focus on the pattern recognition algorithms and neural networks techniques, applied to the evaluation of deformed speech signal. The principal idea of the paper is the statement that in the presented problems the standard methods of signal processing and classification, widely applied for analysis and recognition of normal speech, are totally ineffective. In the present work particular attention has been focused on the evaluation of structure for the feature space describing the pathological speech signal. The main original result presented in the paper is the choice of proper vectors of acoustic features adapted for description of those properties of the speech signal, which turned out to be useful for the medical diagnosis, as the ultimate goal of the study is a construction of a diagnostic system for a wide variety of pathological speech signals.

Introduction

In many problems of medical diagnosis, as well as planning and monitoring of therapy and rehabilitation of speech organs and other elements related to the speech abilities (e.g. dental prosthetics), the evaluation of deformed speech signal quality is necessary. In the present paper excerpts from the authors' long-time research concerning the application of modernized methods of acoustic signal processing in selected problems of medical diagnosis. In the course of preliminary studies a few years ago it has been shown that in problems connected with analysis, evaluation and classification of pathological speech signals the standard methods of signal processing and classification, used in semantic speech recognition (comprehension of the statement's content) or voice recognition (speaker's identification), totally fail [1]. It has been also shown [3] that the standard techniques of the speech signal parametrization, like linear prediction coefficients or cepstral coefficients, cannot satisfactorily describe the pathological speech signal, because of its different phonetic and acoustic structure when compared to the regular speech signal, and also because the goal of the recognition process is totally different in that case [2]. More exhaustive discussion of this problem is contained in the papers referenced above. It has been found that the most important (and most difficult !) element of the research work, preceding the practical application of speech as a source of medically useful diagnostic and prognostic information, is the recognition and description of the signal parameters which can be extremely independent both from the semantic context and the personal characteristics of the investigated voice. Additionally the required signal characteristics have to be highly sensitive to even small deformations in the layer related to the structure and functioning of the speech signal generating organs (larynx and the constrictions being the source of the speech's noise components) and the structure of the vocal tract, used in the articulation process. 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.

The methods of the research material's collection

The search for the proper structure of the space of distinctive features in the task of analysis and recognition of pathological speech signal required the collection of appropriate number of signal samples related to various forms of pathologies being considered. 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. It is worth mentioning that because of the specific nature of the study (frequently carried out on small children handicapped by congenital defects or patients after serious operations in the speech organs area) it was our concern to minimize the tediousness of the samples registration for the examined patient. Therefore the words in the test have been comprehensible, easy to repeat and they contained only one or two syllables. The block diagram of the measurement setup has been presented in Fig.1.

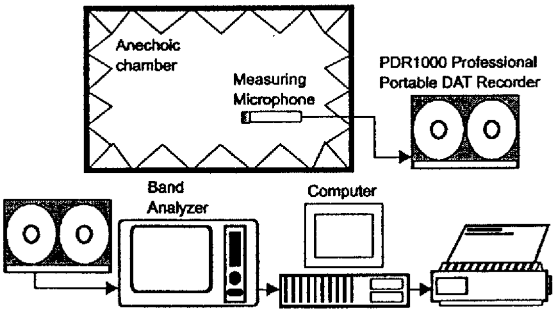


Fig.1 The measurement setup

After preliminary processing of the registered signal the result is a multispectrum, digitized in time, frequency and amplitude by the acoustic analyzer. In order to standardize the measurement procedure and ensure comparability of the results the same signal processing scheme has been used, with amplitude resolution Δs=1dB, signal frequency digitization in a band of f_d = 125Hz, f_g = 12kHz with frequency step Δf = 125Hz. It is worth noticing that linear, not logarithmic scale has been used in the frequency quantization , what has been motivated by the fact that subtle spectra deformations, distinguishing pathological from normal speech, are often localized in the disregarded high frequency bands, which are (correctly) regarded as useless in speech recognition with respect to semantic or personal aspects. This is one of many singularities distinguishing the problems of pathological speech processing considered here from the typical problems encountered in analysis and recognition of normal speech. The multispectra have been obtained using digital analyzer and quantized in time (sampled) with a time interval Δt = 9ms. The implemented professional registration system ensured the transmission band from 20Hz to 20kHz with a dynamic range not less than 80dB.

Signal parametrization

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 studied pathology forms, but because of redundancy and considerable dimension of such a feature space it has been usually transformed to seven-dimensional feature vectors X of the following form:

$$< M_0, M_1, M_2, WS_S, ,WS_1, ,WS_2, ,WS_3 >= X \quad (1)$$

where:

M_0, M_1, M_2 - spectral moments defined and described in previous papers,

WS_s - relative power coefficient, describing the ratio of the signal power in the standard phoneme band to the signal power in the whole band used by the pathological speech signal,

WS_i - relative power coefficient, describing the ratios of the signal power in the i-th band (i=1,2,3) to the signal power in the whole band (the choice of upper and lower limits for the selected bands was one of main research problems solved in the course of the present study).

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 listed set of components was not the only one used in the present study. Additionally the sets of features described below by the formulas 2, 3 and 4 have been also studied

$$< f_1, f_2, \dots, f_{96} > = X \qquad (2)$$

$$< F, F_2, F_3, M_0, M_1, M_3 > = X_2 \qquad (3)$$

$$< M_0, M_1, M_3, C_w, C_p, J, S > = X_3 \qquad (4)$$

where:

f_i - averaged amplitude in the i-th frequency band,

F_1, F_2, F_3 - formants frequencies,

M_0, M_1, M_2 - spectral moments defined earlier,

J, S - special parameters introduced for evaluation of the stage of the speech pathology, called *Jitter* and *Shimmer*

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

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

Quality tests of the parameters proposed for evaluation of the deformed speech signal

Unique and accurate quality evaluation for the proposed set of distinctive parameters determined for the samples of pathological speech is very difficult, because the phonetic data collected from the examined persons differ also in aspects being outside the scope of our analysis (e.g. the articulation rate is varied), similarly as different are the statements of persons in good health and correct (standard) articulation, regarded as a reference data in the present study. Already at the preliminary stage of the research work it has been found that the quality evaluations of the selected parameters are difficult for quantitative estimation because of the multitude of measured acoustic parameters and also because of the great variety of registered acoustic phenomena. It is also very difficult to construct for various feature vectors a quality measure general enough to be used as a criterial function in the search described here. Exactly this (i.e. the inaccessibility of formal definition of a mathematical quality measure) was the reason that made impossible the application of any automated or computer aided optimalization methods (e.g. gradient search method, but also Monte Carlo methods or genetic algorithms techniques) in the considered task. Therefore in the study described here the evaluation has been done by the author's themselves and the structural analysis technique has been used instead of quantitative measures. Specifically the cluster analysis has been used as a preliminary method of data analysis, allowing at least a rough estimation of the distribution of objects representing the particular speech samples in the multidimensional space of the analyzed acoustic features.

Problems considered

The research oriented towards determination of the utility of specific methods of untypical speech signal parametrization in the pathological speech diagnosis have been carried out using the pathological speech samples (and the correct speech samples, forming a reference set constructed separately for each method) in many specific tasks, belonging to the general problem of pathological speech analysis. Altogether several tens of such studies have been carried out, but in the present work the attention will be focused on the following fields:

Dental prosthetics

It seems obvious that every change in the teeth layout affects destructively the speech abilities. Evident irregularities, concerning particularly the dental phones, are observed for the edentulous patients. These disturbances cannot be corrected without using dental prostheses. The prosthesis changes the geometry of the vocal tract and during the initial period it considerably hinders the articulation. The aim of the authors' research in this field was the determination of the submaxilla teeth setting, optimal for the speech abilities in the edentulous patients. Specifically in a series of studies (published separately) we searched for submaxilla teeth setting in total dentures which was optimal for the "s" consonant pronunciation. The task faced by the speech pathology evaluation system in this case was the assistance in the choice of the denture's optimal shape based on the evaluation of speech signal deformation's level. Typical phoneme spectrograms obtained in this examination of correct and pathological speech samples (with confidence intervals marked) for various shapes of the considered dentures are shown in Fig.2.

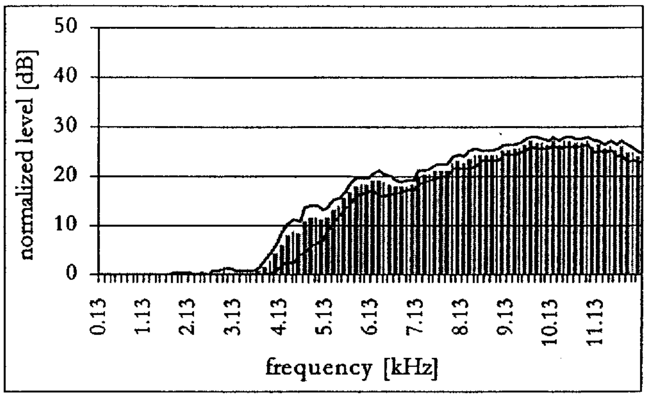


Fig.2a Typical spectrogram of the averaged "S" phoneme - the reference group

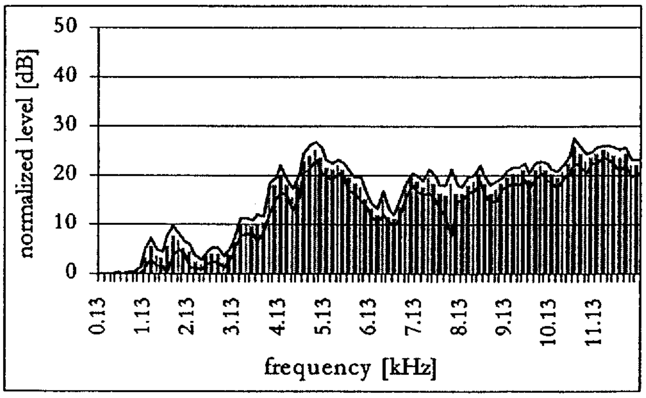


Fig.2b Averaged spectrogram of the "S" phoneme - without the denture.

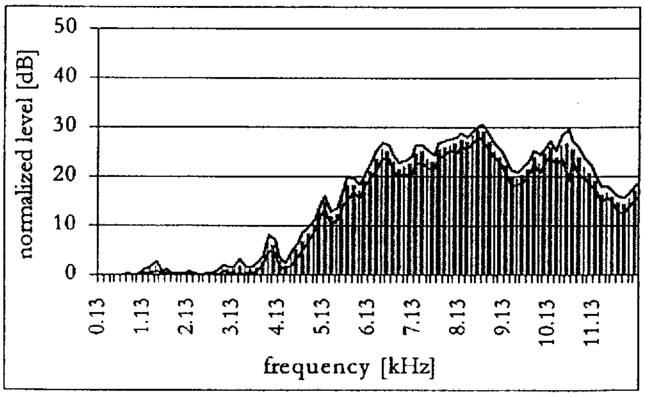


Fig.2c Averaged spectrogram of the "S" phoneme - the denture's structural parameter 1mm

Maxillofacial surgery

Both the postoperative losses in the palate or jaw area and the losses resulting from the cleft palate induce severe aesthetic and functional impairment of the patient. The betterment in physical appearance and considerable reduction of the impairment of speech functionality, chewing, swallowing and breathing can be achieved by means of surgical reconstruction and prosthetic rehabilitation. But even many surgical treatments and application of the best prosthetic reconstruction methods does not always lead to complete restitution of the lost anatomic structures in the jaw and oral cavity area. It shows by the fact that the patient's speech after the surgical treatment and denturing still differs considerably from standards of correct speech. Thus there are many indications for application of the prosthetic rehabilitation. Therefore also in this field the implementation of objective (acoustic) evaluation of the speech deformation level after various types of treatment (often equivalent from the medical point of view, i.e. resulting in identical therapeutic effects) can assist the surgeon's decision making process (during the consecutive future operations of course). The evaluations of the speech deformation level (precisely speaking - the evaluation of improvement for the deformed speech observed after the operation) are also an important factor in the choice and optimization of the denture. Typical deformed phoneme runs with marked confidence intervals for articulation before the surgery, after the surgery but without the denture and after the surgery and denturing are shown in Fig.3.

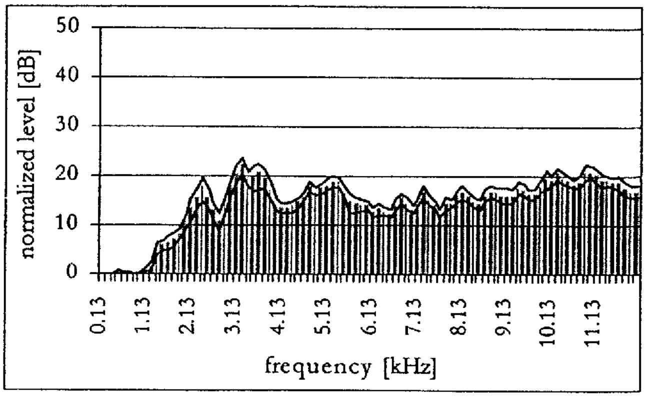


Fig.3a The averaged spectrogram of the "S" phoneme - before the surgery

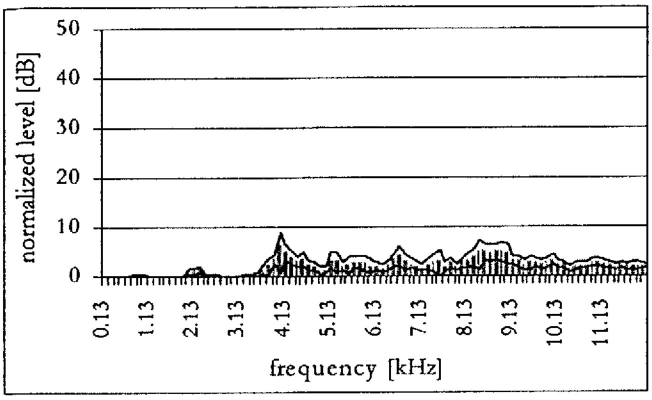


Fig.3b Average spectrogram of the "S" phoneme - after the surgery without the denture

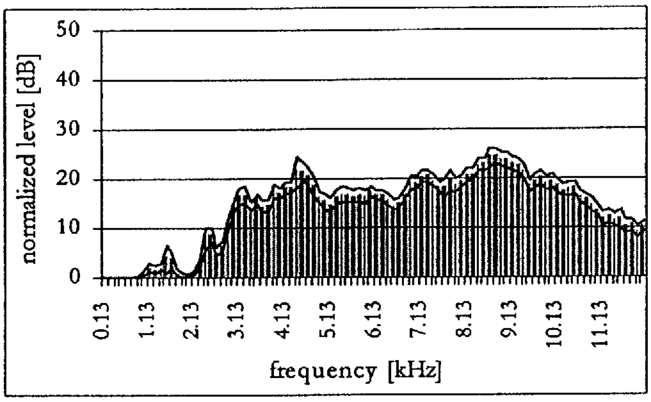


Fig.3c Average spectrogram of the "S" phoneme - after the surgery and denturing

Cleft palate in children

The techniques of objective processing and automated recognition of the pathological speech play a particularly important role in the case, when the evaluation required for the diagnosis and therapy has to be done for patients who are not able to cooperate effectively with the physician - e.g. small children. Directing the investigations towards tasks of that type extended research has been undertaken in years 1992-1999 concerning the choice (based on computer analysis of the articulation of several words in a purposely selected sentence) of the optimal type of surgical treatment eliminating a serious congenital anomaly, namely the cleft palate in children. The study included sixty children (of age between 4 and 7 years) divided into three groups with respect to the defect type. The aim of the research in this case was the evaluation of the surgical treatment results for children with primary type cleft palate operated by the Skog method and the secondary type cleft palate operated by the Wardill-Killner method

Orthodontics

The speech deformation by the orthodontic apparatus depends on its construction and the patient's adaptive abilities. The adjustable apparatus changes the parameters of the vocal tract to a various degree. For an objective determination, which of the apparatuses applicable in the therapy induced the lowest speech disturbances, the studies have been carried out for 11 patients with teeth anomalies.

Laryngology

In practice there is often a possibility to choose one of several methods of surgical larynx treatment and therefore evaluation of the speech deformation level expected after the surgery is often required. Before starting with the problem experimental data have been collected about the forms and levels of speech signal deformation after various types of surgical treatments in order to complete the so called learning set for the later application of artificial intelligence techniques which require the learning process (among others the neural networks technique). The studies of the speech articulation have been carried out for male patients after various types of larynx cancer operations. The exemplary differences in articulation between the reference group and the group of patients are shown by the spectrograms if Fig. 4.

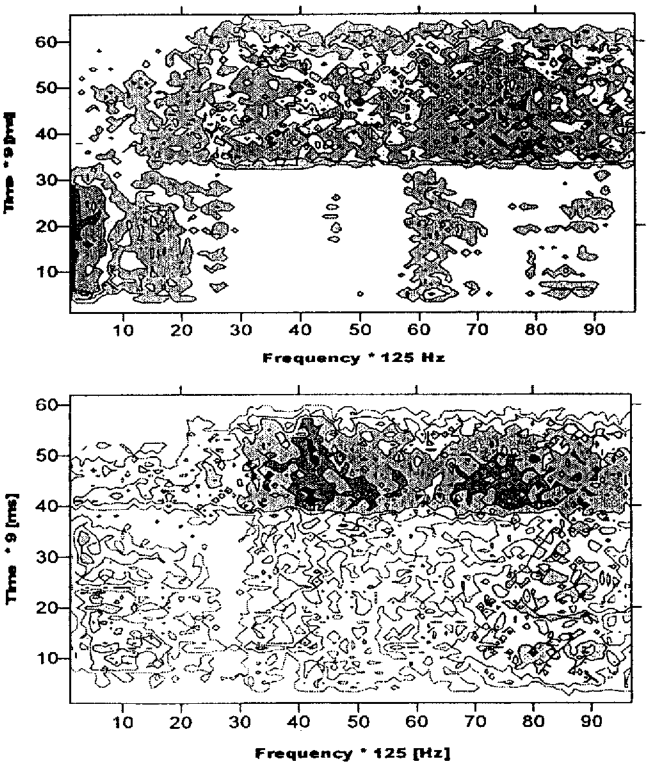


Fig. 4 Spectrogram of the "AS" statement for the reference group and the group of patients

Next in Fig. 5 the typical spectra of a single vowel "A" are presented for the reference group and the group of patients after selected operation types

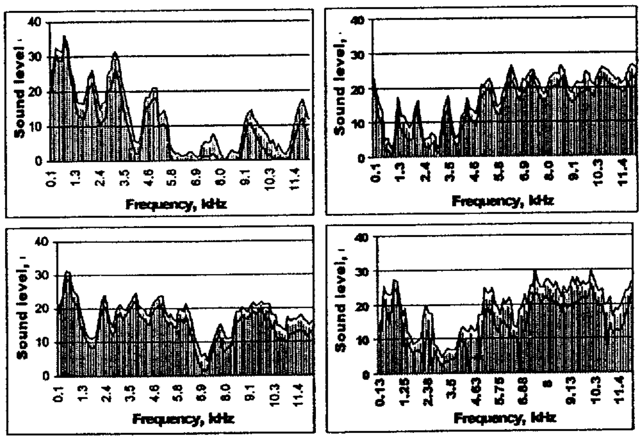


Fig. 5 Spectra of the "A" vowel, correct and deformed (after chordectomy, hemilaryngectomy and fronto-lateral laryngectomy)

Analysis of the feature space structure

For analyzing the correctness of the feature space structure central agglomeration procedure. In connection with the above mentioned method selected (during previous studies) space metrics have been used.

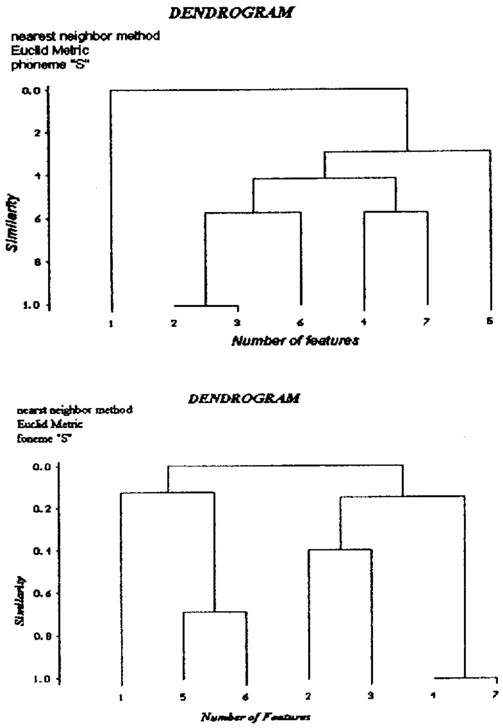


Fig.6 Dendrograms

The algorithms of cluster analysis executing the connections of close objects (for a specific metric and specific feature space) creating clusters, which in further considerations can be replaced by single representatives. The results obtained from the clustering algorithms have been presented in the form of dendrograms, showing which objects and with what similarity level have been connected or separated in consecutive algorithm steps. The similarity level has been calculated from the following formula:

$$d = \frac{1}{1+\rho} \qquad (5)$$

where: d - similarity
ρ - distance (for the selected metric)

Exemplary dendrograms have been presented in Fig.6.

Methods of analysis and classification of pathological speech signals

The above mentioned results of applications of the automated clustering methods and cluster analysis algorithms for various problems (discussed above) concerning the pathological speech analysis provide a general orientation in the topology of specific feature spaces listed above. Performing such analysis for various sets of signal parameters constituting the basis in the space for which the analysis is being performed or in which the recognition is to be done leads to a general concept concerning this space. In particular by confrontation of the automated object clustering results for a given feature space with the known (from the medical documentation) classification of analyzed cases associated with specific signal samples, containing the studied forms of speech pathology, one can evaluate the utility and adequacy of the selected parameters and investigate (in a synthetic way) the data topology in the selected feature space. The ultimate measure of utility of the specific data and parameters is always the confrontation with their application to solution of the final task which, for the problem considered here, is the automated classification and recognition of the registered samples of pathological speech.

In the discussion of methods and recognition results mentioned above the following factor should be taken into account: in the tasks considered here the process of deformation recognition for the acoustic signal of pathological speech is equivalent to attribution of *acoustic patterns* obtained in during the study to one of the classes, in general **not known** in advance. Thus the recognition in the case discussed here is combined with the detection and specification of the essential features qualifying the individual patterns into respective groups, the number and features of which are not known in advance. The results of classification and recognition presented below have been obtained by using feature vectors describing the pathological speech discussed in the present work. In order to increase the reliability the testing of classification and recognition have been carried out using two methods: by the pattern recognition algorithms derived from artificial intelligence techniques and by the neural network methods, based on the achievements of biocybernetics.

Pattern recognition

The following groups of automated pattern recognition methods have been selected for the study:

- ❑ nearest neighbors
- ❑ spherical neighborhood
- ❑ approximation

The summary of the obtained results is presented in Table 1

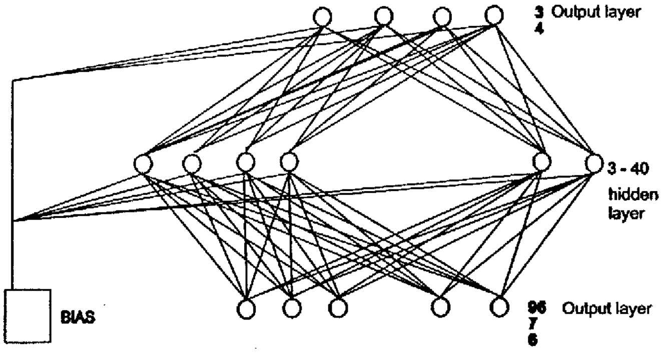
Table 1 Results obtained by various pattern recognition methods, Euclid Metric

Recognition method	Recognition reliability[%]		
	Dental prosthet.	Maxillofacial surgery	Orthod ontics
NN algorithm	84	90	79
α -NN algorithm	76	90	75
NM algorithm	84	85	79
Optimal spherical neighborhoods	-	79	-
Quadratic approximation	67	82	-

The presented results definitively confirm the opinion that the selected parametrs properly describe the analyzed phenomena connected with the forms of speech signal degradation in the context of selected pathologies

Neural networks

On the basis of general recommendation it has been decided to use in the study three layer networks of the *feedforward* organization type, built of elements with sigmoidal characteristics of the nonlinear transfer functions, connected according to the rule of full network connection between the input layer and the next layer of the network. The structure of such a network is shown in Fig. 7.



Topology of the neural network

Fig. 7 Network structure

The network's task was the transformation of the parametrized speech signal samples to the values which can be interpreted as the evaluation of the speech signal deformation level. As a results of the study it has been found that the number of neurons in the hidden layer is of critical importance for the obtained results. The factor describing the sum square error in an epoch has been used as an auxiliary quantity, helpful in the evaluation of influence of hidden layer's neuron number on the learning process. The criteria for the termination of the learning process have been also determined during the study. This element should be particularly stressed, as it is an original proposition, not found in the other authors' works. The criterion for the learning process termination has been connected with the changes (monitored during the learning process) of a purposely constructed coefficient:

$$\mathbf{DELTA = 1 - pos(Y) - neg(Y)} \qquad (6)$$

where:

$$\mathrm{neg}(Y) = \max_{1 \le i \le n} (1 - z_i) y_i \qquad \mathrm{neg}(Y) = \max_{1 \le i \le n} (1 - z_i) y_i$$

This coefficient is a measure of domination of the recognition accepted by the network over the other competing recognitions. The description of application methodology for this original criterion can be found in the paper [4]. In Table 2 the study's results have been shown in the form of recognition reliability (expressed in %) for neural networks applied for three groups of the research material.

Table 2

Type of neural network	Recognition reliability [%]		
	Dental prosthettcs	Maxillofacial surgery	Orthodontics
Triple layer type I	-	95	91
Triple layer type II	90	93	89
Triple layer type III	88	89	85

On the other hand the results of selected simulations presented in Table 3 are based on the "Laryngology" research material group.

Table 3

No of neurons in the input layer	Group1 [%]	Group2 [%]	Group3 [%]	Group4 [%]
96	73	72	70	67
6	79	77	75	74
7	86	86	84	83

The presented results also can be considered as satisfactory, what proves, that the selected parameters in general correctly describe the analyzed phenomena of pathological speech signal deformation.

Summary

The utility of sets of features (parameters of the signal description) dedicated for pathological speech evaluation tasks revealed as a result of the present study confirms the opinion that dedicated features should be used in the considered task and not the widely used parameters and signal description elements, applied in the analysis and recognition of the normal speech signal. The revealed advantage of the neural networks technique in the solution of problems formulated here is an obvious consequence of the generally known utility of this tool in the tasks characterized by great shape complexity for the areas of the particular considered classes in the feature space. This result indicates a possibility of application of the neural network technique as a favorable alternative for the other techniques (like pattern recognition or statistical methods), however this was not the primary objective of the study. The crucial role of the proper preparation of speech signal parameters is worth stressing once more, because in the case when the problem cannot be parametrized in a way ensuring a proper resolution of areas belonging to the particular classes in the feature space, the results obtained by any methods (particularly the pattern recognition methods) are highly unsatisfactory.

It is worth stressing that in spite of the basic nature of the study, directed towards the analysis of properties of various elements in the feature space, our study is very helpful in practical applications like selection of dentures and orthodontic apparatuses or qualification of patients for specific types of surgical treatments in order to minimize the patient's voice deformation after the treatment.

References

1. Tadeusiewicz R., Izworski A., Wszo ek W., (1997), *Pathological Speech Evaluation Using the Artificial Intelligence Methods*, w materia ach konferencji: “World Congress on Medical Physics and Biomedical Egineering”, September 14-19, 1997, Nice, France
2. Tadeusiewicz R., Wszo ek W., Izworski A., (1998), *Application of Neural Networks in Diagnosis of Pathological Speech*, w materia ach konferencji NC’98, “International ICSC/IFAC Symposium on Neural Computation”, Vienna, Austria, 1998, September 23-25
3. Tadeusiewicz R., Wszo ek W., Modrzejewski M., (1998), *The Evaluation of Speech Deformation Treated for Larynx Cancer Using the Neural Network and Pattern Recognition Methods*, w materia ach konferencji “IV International Conference on Engineering Applications of Neural Networks”, Gibraltar, 1998, June 10-12, str. 613-616
4. Mikrut Z., Tadeusiewicz R.: Metodyka eksperymentów z sieciami neuronowymi rozpoznaj cymi obrazy, Zeszyty Naukowe AGH, "Automatyka", nr 66, 1993, ss. 7-30
5. Engel Z., Tadeusiewicz R., Wszo ek W.:Estimation of Applicability of the Neural Networks for Automatic Evaluation of Pathological Speech. 15th International Congres on Acoustic, Trondheim, Norway , June, 1995