MECHANICS
vol. 27 No. 2 2008

Wiesław WSZOŁEK*

TECHNIQUES OF NEURAL NETWORKS FOR SIGNAL PARAMETERIZATION USED IN MONITORING SYSTEMS

SUMMARY

The paper presents selected aspects of research concerning a new concept in application of computer technology to the analysis of acoustic signal. This new concept assumes, that during the analysis of signal the study is not focused on determining some or other signal parameters, neither it is focused on the signal classification, but it is supposed to lead to an automated understanding of the origins of the deformation, which can be revealed in analyzed signal.

Keywords: *pattern recognition, neural networks, signal analysis, automatic diagnostics, speech recognition*

WYKORZYSTANIE SIECI NEURONOWYCH DO PARAMETRYZACJI SYGNAŁU W UKŁADACH MONITORUJĄCYCH

W pracy przedstawiono wybrane aspekty badań dotyczących nowej koncepcji zastosowania techniki komputerowej do analizy sygnału akustycznego. Ta nowa koncepcja zakłada, że podczas analizy sygnału nie dążymy do ustalenia parametrów sygnału ani nie usiłujemy dokonać jego klasyfikacji, lecz zmierzamy do automatycznego zrozumienia przyczyn deformacji, jakie dają się zaobserwować w rozważanym sygnale.

Słowa kluczowe: *rozpoznawanie wzorców, sieci neuronowe, analiza sygnałów, diagnostyka automatyczna, rozpoznawanie mowy*

1. INTRODUCTION

The characteristic feature of tasks related to the object analysis and recognition based on acoustic signals is the fact that selection of optimal analysis rule or recognition algorithm is very difficult. The methods of signal analysis and methods of its recognition should be strictly adjusted to the specific features of the particular task. The methods of low- and high-pass filtration, compression techniques, or the methods of spectral transformation of the acoustic signal do not essentially depend on the nature of the signal itself or the purpose of its registration and processing. An impressive progress has been achieved recently in the field of acoustic signal processing. In tasks related to analysis and recognition of acoustic signals the progress has been much slower, and the unification of methods and standardisation of algorithms encounters considerable problems. The main origin of those problems is the fact that in every single task of signal analysis the objectives comprise the extraction of different set of characteristics and parameters. In most problems of technical or medical diagnosis it is necessary to evaluate the signal generated by the objects under consideration. The evaluation might probably concern the measure of signal deformation level. In such a case the task comprises the quantitative (preferably scalar) expression of the measure of examined signal deviation from an abstract signal, which can be considered as a reference signal.

In present paper author describes the premises for a general approach for the task of design and construction of such a diagnostic system, which could be able to cope with automated evaluation of objects. The present study deals with analysis of objectively registered signals, generated by the objects subject to diagnostic evaluation.

A frequent requirement imposed on the diagnostic systems is the diagnostic efficiency better than the one resulting from human actions. At present the performance of many methods of automated signal classification is much worse than the one offered by an experienced human expert. It is probably true for all the tasks related to object diagnosis based on the measured signals.

The most important and also the most difficult element of the research, preceding the practical use of signal (e.g. the speech signal in medicine) as a source of useful diagnostic and prognostic information, is the identification and description of the signal parameters, which are highly independent of the object features that are irrelevant for the diagnosis. In addition the requested signal features have to be highly sensitive to even small signal deformations in the layer that is directly related to essential parameters of the evaluated object. During the research particular attention has been focused on the analysis and structure description for the space of features describing the evaluated object. The detailed knowledge of the feature space topology (which is not simple for a direct description because of its multidimensional nature) allows in further stages an effective application of proper methods of automated recognition (Wszolek and Wszolek 2004; Wszolek and Klaczyński 2005).

* AGH University of Science and Technology, Krakow, Poland

2. THE BASIC FUNCTIONAL ELEMENTS OF THE SYSTEM

A typical model of the monitoring system, assuming unidirectional information flow, is presented in Figure 1.

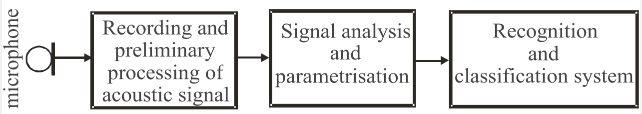

Fig. 1. Structure of the monitoring system with unidirectional information flow

The functional description of the registration and preliminary signal processing stage comprises the problems related to the dynamics of the analogue part (microphone, preamplifier) defined as

$$D \overset{def}{=} 20 \cdot \log \frac{X_{\max}}{X_0} \text{ [dB]} \tag{1}$$

where:

$X_{\max}$ – $\max_{T_W}\{x(t)\}$ (2)

X_0 – reference level,

$x(t)$ – time course of the measured signal (e.g. acoustic pressure)

$$T_W \overset{def}{=} [0, t_W] \subset \Re \tag{3}$$

T_W – the set of the considered time values,

t_W – signal duration time [s].

as well as the preliminary signal formation call preemphasis[1].

The acoustic signal, when it is received, registered and analysed by an arbitrary technical system, is represented as a time course

$$x \in A^{T_W} \tag{4}$$

where:

$A \subseteq \Re_+$ – the set of momentary values of the acoustic pressure,

A^{T_W} – set of functions of a scalar argument t_W, taking values from the ***A*** set.

In the diagnostic systems various physical phenomena are measured and registered, which are later used as a basis for a human (expert) elaborated decision, or used by the considered technical system, elaborating diagnostic decisions in an automated way. However in order to make the decision making process functional and effective – it is necessary to subject the "raw" signals to purposeful and often very sophisticated processing, directly connected with the extraction of the requested features. As a result of that signal processing a selective condensation takes place for that part of information, which can be useful for the diagnostic purposes, instead of attempted elimination (in most cases incomplete) of those signal components, which are related to object features that are irrelevant for solution of a given task. Such a selective measurement, acquiring only information that is important and useful, seems in practice impossible for realisation. It should be stressed, that in all cases of diagnostic studies known to the authors the signals coming from the evaluated objects (technological or biological) have been registered in all its versatile complexity, and only later they were analysed and classified – merely on the basis of certain parameters, calculated in the second stage of the analysis. In practice it however denotes that in real diagnostic systems a drastic information reduction is decided before the proper decision making stage, because the parametric signal description is usually much more concise, than the description resulting from the original signal registration.

If we assume that[2] $\#A = m$ and $\#T_W = n$ (m – depends on the reproduction accuracy of the signal amplitude scale, n – depends on the frequency of signal sampling), then the informational volume of the signal $I(x)$ can be determined from the well-known formula:

$$I(x) = n \log_2 m \tag{5}$$

One of the most important functional subsystems of the monitoring system is the part transforming the "raw" signal to the space of distinctive features (the "Signal analysis and parametrisation" block in Fig. 1). Such a feature space can be constructed on the basis of purely technical premises (in such a way many indicators have been created and adjusted, which describe the spectral or temporal features of vibro-accoustic signals and which are used up to date), but the practical experience of the author and many published studies suggest, that it is often favourable to search for perfectly new, diverse signal features. For example the primary pitch in the speech signal can be strongly distorted (e.g. in the frequency domain or in its time structure – see Figs 2 and 3) in reference to the normal speech signal, and still it can be fully acceptable for the listening person.

An opposite situation is also possible – a signal with seemingly very minor distortions will be received as incomprehensible or so unpleasant for listening, that it will present a serious obstacle in interpersonal relations. Similar problem is encountered in evaluation of the voice quality after surgical treatments. Usually the last stage in monitoring aided by artificial intelligence methods is the classification of the monitored phenomena.

[1] Preemphasis – a preliminary signal formation, which can be carried out both in time or frequency domain (Tadeusiewicz 1988).
[2] The # symbol denotes the set's cardinal number.

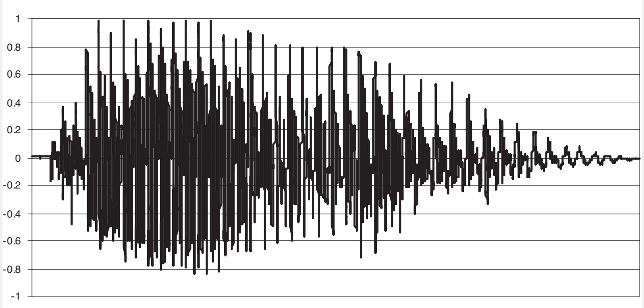

Fig. 2. Primary pitch – distorted case

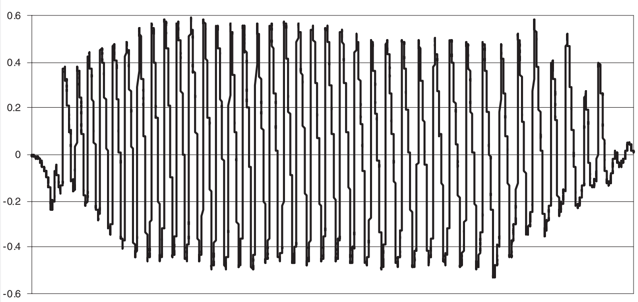

Fig. 3. Primary pitch – normal case

3. PARAMETERIZATION AND ANALYSIS OF THE SIGNAL

Procedures of the processing and analysis of acoustic signal usually lead to creation of the so-called **acoustic pattern** of the object, mapping the signal into an ordered set of real numbers. Therefore in the signal processing one has to take into account, which characteristics of the determined feature vector will be used as a basis for the new feature space. In most cases the recognition methods of the monitored deformations of the acoustic signal can be based on the spectral, temporal and spectro-temporal characteristics. The general form of such mapping is presented below

$$F : A^{T_W} \rightarrow H^{T_W \times f} \tag{6}$$

where H – the set of considered signal frequencies.

The usefulness of the above mapping lies in its two aspects:

1) it enhances the signal features, which are particularly interesting and valuable for the signal recognition and interpretation;
2) it allows the reduction of the information volume $I(w)$ acquired from the signal.

As a result of mapping (6) it can be assumed that $\#T_W = n_1 << n$ as well as $\#H = m_1 << m$ and then $I(w) << I(x)$, because:

$$I(x) = n_1 \log_2 m_1 \tag{7}$$

In order to reduce the information volume of the spectro-temporal signal and a considerable dimensionality of the created feature space further processing of the analysed signal is carried out, depending on the monitored signal, what leads to specific feature vectors ***Xi***.

An exemplary selection of feature vectors is shown for monitoring of the acoustic signal of pathological speech, where the following groups of feature vectors have been used:

– **Parameters ignoring the time factor**

$$\langle f_1, f_2, f_3, \ldots, f_N \rangle = X_1 \tag{8}$$

where:

f_i – averaged amplitude of the i-th band in the time and frequency domain spectrum,
N – number of considered spectrum bands.

For example in Figure 4 averaged spectra of the "a" vowel with prolonged phonation have been shown (normal and pathologically deformed pronunciation).

– **Parameters describing the spectrum shape**

$$\langle F_1, F_2, F_3, M_0, M_1, M_2 \rangle = X_2 \tag{9}$$

where:

F_1, F_2, F_3 – formant frequencies,
M_0, M_1, M_2 – spectral moments.

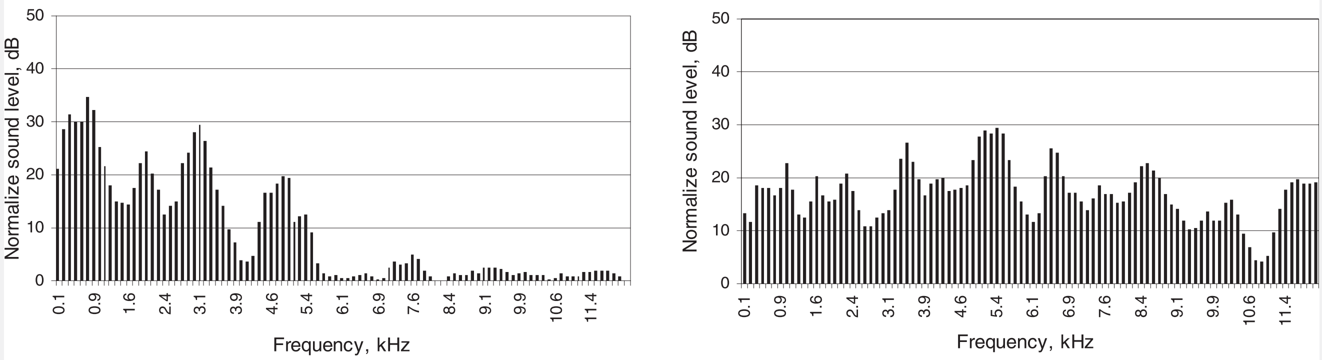

Fig. 4. Spectrum of a normal and deformed vowel

$$F_k \equiv f_k \Leftrightarrow \left(\left. \frac{\partial G(x(t))}{\partial f} \right|_{f_k} = 0 \right) \wedge \left(\left. \frac{\partial^2 G(x(t))}{\partial f^2} \right|_{f_k} < 0 \right)$$

$$\wedge_{j=1}^{k=1} \left\{ \left[\left[\exists f_j \left\langle f_{j+1} \div \left(\left. \frac{\partial G(x(t))}{\partial f} \right|_{f_k} = 0 \right) \wedge \left(\left. \frac{\partial^2 G(x(t))}{\partial f^2} \right|_{f_k} \right\rangle 0 \right) \right] \right]$$

$$\left. \wedge \; (F_{kd} \le F_{kg}) \right\} \tag{10}$$

$$M_m = \sum_{k=0}^{\infty} \left| G(k) \right| \cdot [f_k]^m \tag{11}$$

where:

- f_k – central frequency of the k-th band, selected in frequency analysis,
- $G(k)$ – signal spectrum.

– **Parameters describing spectrum shape and power distribution of the signal in selected bands**

$$\left\langle M_0, M_1, M_2, C_{FbP}, C_{Fb_1}, C_{Fb_2}, C_{Fb_3} \right\rangle = X_3 \tag{12}$$

where:

- C_{Fbk} – power coefficients for every k-th frequency band, defined as:

$$C_{Fbi} = \frac{\sum_{j=1}^{m} \sum_{FL}^{Fu} w_j^k}{\sum_{j=1}^{m} \sum_{k=1}^{n} w_j^k} \tag{13}$$

- w_j^k – signal value in the k-th frequency band in the j-th time instance,
- FL – lower frequency of the selected band,
- Fu – upper frequency of the selected band
- FbP – ed. frequency band of the reference phoneme.

– **Parameters directed towards detection of speech generation irregularities**

$$\left\langle M_0, M_1, M_2, C_{FbP}, C_{FbR}, J, S \right\rangle = X_4 \tag{14}$$

where:

- C_{FbP} – power coefficient for the reference phoneme band,
- C_{FbR} – power coefficient for the remaining frequency band,
- J – jitter,
- S – shimmer, defined as

$$J = \frac{\sqrt{\frac{1}{N-1} \sum_{i-1}^{N-1} (F_i - F_{i+1})^2}}{\frac{1}{N} \sum_{i-1}^{N} F_i} \tag{15a}$$

$$S = \frac{\sqrt{\frac{1}{2N-1} \sum_{i-1}^{2N-1} (A_i - A_{i+1})^2}}{\frac{1}{2N} \sum_{i-1}^{N} A_i} \tag{15b}$$

On the basis of exemplary configurations of feature vectors, depending on the objects examined by the monitoring systems, various vectors of distinctive features can be created. For example for evaluation of post-operative laryngeal damages the following feature vector has been applied

$$\left\langle M_0, M_1, M_2, F_1, F_2, F_3, F_4, AF_1, AF_2, AF_3, AF_4, C_{Fb_1}, C_{Fb_2}, C_{Fb_3}, C_1, C_2, C_3, C_4, C_5, J, S \right\rangle = X_E \tag{16}$$

where:

- AF_i – amplitude value of the i-th formant,
- C_n – Mel-Frequency Cepstral Coefficients, defined by relation (Grocholewski 2001)

$$c_n = \sqrt{\frac{2}{3}} \sum_{i=1}^{N} ws_i \cos\left(\frac{\Pi n}{N} (i - 0.5) \right) \tag{17}$$

- ws_i – i-th coefficient, obtained from signal processing by the triangular filters,
- N – number of filters in the system.

4. SELECTED RESULTS OF THE STUDY

The last element of the “intelligent” monitoring system can be the part in which recognition of the registered signals is carried out. As an example of application of the artificial intelligence methods in monitoring systems results have been shown for a group of patients with abnormal changes in the larynx area. The goal of the system was to classify the patients on the basis of registered acoustic samples of the speech signal.

The final acoustic material has been collected from 140 persons divided into two groups:

1) The reference group (standard group), 25 persons with correct pronunciation.
2) The group of patients (115 persons) treated by the following surgery methods:
   – removing of inflammation polyps of the larynx (35 persons),
   – partial surgery of the larynx (80 persons).

In this study the classificator was an artificial neural network, learned by the error back-propagation method. The input data for the network were the parameters of the feature vector (21). The hidden layers have been selected in the experiment (7, 15, 30 or 50 neurons). The output was presented by three neurons (representing three categories: reference group, patients with polyps, and patients with cancer). Figures 5 and 6 show the results of correct recognitions for selected network configurations. The best results achieved were at the level of 90%, with 7 neurons in the hidden layer. On the other hand the results with 50 neurons in the hidden layer were the worst (Fig. 7). The level of correct recognitions was below 70%.

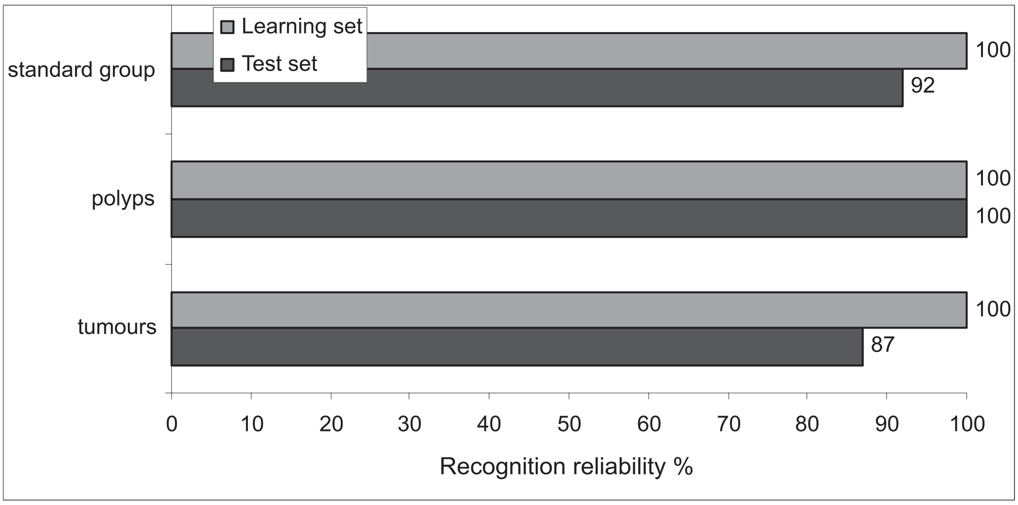

Fig. 5. Results of learning and recognition of pathological speech signal (network with 7 neurons in the hidden layer)

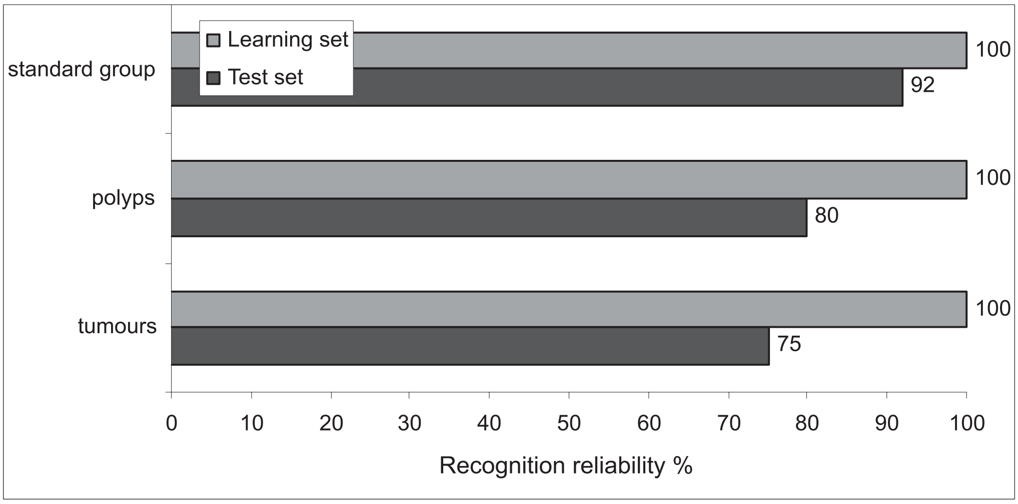

Fig. 6. Results of learning and recognition of the pathological speech signal (networks with 15 neurons in the hidden layer)

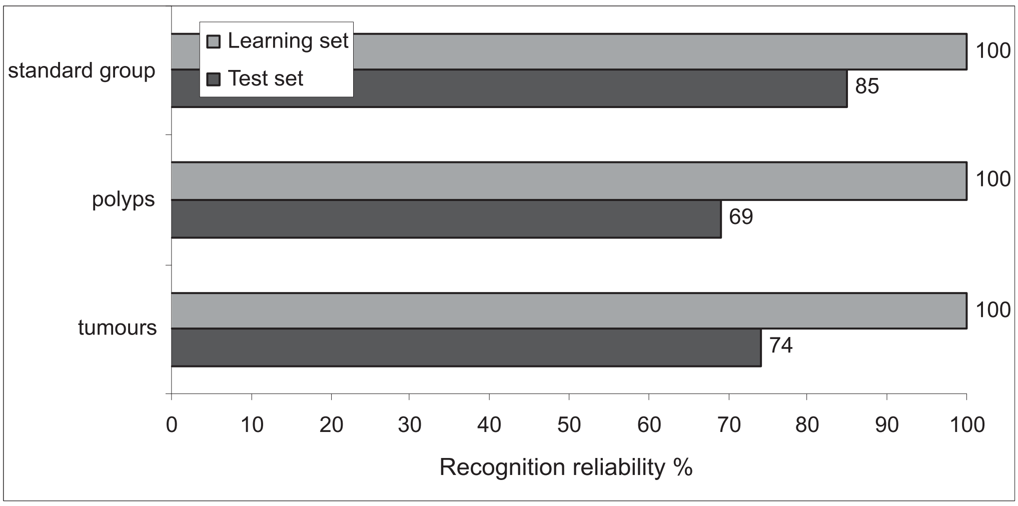

Fig. 7. Results of learning and recognition of the pathological speech signal (networks with 50 neurons in the hidden layer)

5. SUMMARY

The problems considered in the present work are related to the possibility of extending the diagnostic abilities of the acoustic monitoring systems by application of the artificial intelligence methods, aiding the regular diagnostic processes. In the paper a general functional structure is presented, for the system in which one of the last elements is a block (with unidirectional information flow) elaborating a diagnostic decision. Such a structure has been presented for the case of deformed speech signal evaluation. Also for that signal it has been confirmed by the research, that it is possible to create a system for classification of pathological speech, based on the artificial intelligence methods. Such a system was intended to be able (using the speech signal) to classify patients with respect to the type of preliminary medical diagnosis of the pathology in the larynx area. In the works (Wszołek 2006; Tadeusiewicz *et al*. 1999; Wszołek *et al*. 2001) results have been presented for a study concerning the application of this technique to aircraft recognition with positive effects.

The completed research has confirmed, that in many monitoring systems dealing with acoustic signals their functionality should be aided by the systems based on the artificial intelligence methods.

REFERENCES

Adamczyk J.A., Wszołek W. 1990: *The analysis and evaluation of machine emitted vibroacoustics signal patterns*. Zeszyty Naukowe AGH, Mechanika, t. 9, z. 4, pp. 5–15.

Grocholewski S. 2001: *Statystyczne podstawy systemu ARM dla języka polskiego*. Rozprawy, nr 362, Politechnika Poznańska, Poznań.

Rutkowski L. 2005: *Metody i techniki sztucznej inteligencji*. WN PWN, Warszawa.

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

Tadeusiewicz R. 1993: *Sieci neuronowe*. Akademicka Oficyna Wydawnicza RM, Warszawa.

Tadeusiewicz R., Wszołek W., Izworski A. 1999: *Classification of acoustic and vibration signals by the artificial intelligence methods*. Sixth International Congress on Sound and Vibration, 5–6 July 1999, Copenhagen, Denmark, pp. 3181–3188.

Wszołek W. 2006: *Metody sztucznej inteligencji w systemach monitoringu hałasu lotniczego – Artificial intelligence methods in air-noise monitoring systems*. Przegląd Mechaniczny, nr 3, pp. 15–21.

Wszolek W., Klaczynski M. 2005: *Acoustic methods of voice estimation after surgical treatment of the vocal tract*. Archives of Acoustics, vol. 30, No. 4, pp. 193–197.

Wszołek W., Tadeusiewicz R., Chyla A. 2001: *Recognition of selected helicopter types based on the generated acoustic signal with application of artificial intelligence methods*. Proceedings of Inter-Noise, 27–30 August 2001, The Haque, The Netherlands.

Wszolek W., Wszolek T. 2004: *Automatic Understanding of Speech Signal*. Chapter in book: Kłopotek A., Wierzchoń S., Trojanowski K. (eds), 2004: Intelligent Information Processing and Web Mining, Springer-Verlag, Berlin – Heidelberg – New York, pp. 609–620.