amcs

ACOUSTIC ANALYSIS ASSESSMENT IN SPEECH PATHOLOGY DETECTION

Daria PANEK [a,*], Andrzej SKALSKI [a], Janusz GAJDA [a], Ryszard TADEUSIEWICZ [b]

[a] Department of Measurement and Electronics
AGH University of Science and Technology, Al. Mickiewicza 30, 30-059 Kraków, Poland
e-mail: {dpanek,skalski,jgajda}@agh.edu.pl

[b] Department of Automatics and Biomedical Engineering
AGH University of Science and Technology, Al. Mickiewicza 30, 30-059 Kraków, Poland
e-mail: rtad@agh.edu.pl

Automatic detection of voice pathologies enables non-invasive, low cost and objective assessments of the presence of disorders, as well as accelerating and improving the process of diagnosis and clinical treatment given to patients. In this work, a vector made up of 28 acoustic parameters is evaluated using principal component analysis (PCA), kernel principal component analysis (kPCA) and an auto-associative neural network (NLPCA) in four kinds of pathology detection (hyperfunctional dysphonia, functional dysphonia, laryngitis, vocal cord paralysis) using the a, i and u vowels, spoken at a high, low and normal pitch. The results indicate that the kPCA and NLPCA methods can be considered a step towards pathology detection of the vocal folds. The results show that such an approach provides acceptable results for this purpose, with the best efficiency levels of around 100%. The study brings the most commonly used approaches to speech signal processing together and leads to a comparison of the machine learning methods determining the health status of the patient.

Keywords: linear PCA, non-linear PCA, auto-associative neural network, validation, voice pathology detection.

1. Introduction

Despite the highly developed digital technology used in acoustic analysis, there is still the problem of extracting useful information about the state of a patient's health. Voice is dynamic and complex, whilst speech can be conducted in different languages, intonations and with different emotions. In the last few years, voice pathology detection has been studied intensively in the signal processing research community using various digital signal processing methods (Arroyave *et al.*, 2012; Fong *et al.*, 2013). Because of this, the reliability of the developed software which is currently being used to automatically detect pathology in the vocal folds and the underlying variability of speech has increased (Fraile *et al.*, 2009). It depends on the language analysed and the kind of pathologies and techniques used in the analysis. The researchers are still working on the development of different tools for automatic evaluation of pathological voices. Due to this fact, the calculations in this work are provided separately for women and men.

In speech clinical practice, the patient's voice quality is assessed using sustained vowel phonations and/or conversational speech. Sustained vowels are particularly useful because they circumvent linguistic artefacts and are thus considered sufficient for many voice assessment applications (Tsanas, 2013). During speech examination, the patient is asked to maintain the phonation of sustained vowels for as long as possible and as steady as possible (in terms of pitch and amplitude).

In order to make quantitative voice evaluation, acoustic, aerodynamic, endoscopic, perceptual or patient-self assessment examination may be needed. If the diagnosis is not ambiguous, the methods can be combined (Godino-Llorente *et al.*, 2006b). There are different acoustic features being analysed, with a different focus. Closer attention can be given to measuring instabilities in the voice signal or the noise content and general articulatory problems. The most commonly used measures are fundamental frequency (Manfredi *et al.*, 2000), jitter (frequency perturbation), shimmer (amplitude perturbation) (Farrus *et al.*, 2007; Lieberman, 1963; Horii, 1980; Steinecke

---
*Corresponding author

and Herzel, 1995), harmonic-to-noise ratio (Yumoto *et al.*, 1982) and mel-frequency coefficients (Rabiner and Juang, 1993; Godino-Llorente *et al.*, 2006a; Godino-Llorente and Gomez-Vilda, 2004; Steinecke and Herzel, 1995; Jothilakshmi, 2014; Saldanha *et al.*, 2014).

The detection of voice impairments in vocal fold pathology reviews the excitation of the signal (Wong *et al.*, 1979). Non-parametric approaches use a magnitude spectrum of speech, where short-term mel-frequency cepstral coefficients (MFCCs) are used for representing the magnitude spectrum. It is still a challenge to attain high accuracy and an automatic, robust and inexpensive approach in voice pathology detection and classification. Automatic detection of voice pathologies enables non-invasive, low cost and objective assessment of the presence disorders and speed-up, and improves the process of diagnosis and clinical treatment given to patients.

In the past few years, researchers have paid a lot of attention to automatic systems assessing voice disorders due to their non-invasiveness and low cost implementation compared with traditional diagnostic approaches. The accuracy detection levels of voice and speech disorders has increased during the last few years (Arroyave *et al.*, 2012). Unfortunately, no universal solution has yet been found to detect the condition of the voice. Most of the features and algorithms are trained using limited databases, including few types of disorders. One hundred percent of accuracy in voice pathology detection was achieved by Eadie and Doyle (2005) using long-term average spectral measures, glottal noise measures, and measures based on linear prediction modelling, which in turn formed the inputs into conditional logistic regression analysis. This research included only 24 patients.

Hadjitodorov and Mitev (2002) also report the level of 100% accuracy in the detection for well-manifested voice pathologies, and 96,1% accuracy of weakly manifested pathologies were achieved for the K-nearest neighbours using turbulent noise in voice signals (turbulent noise index, TNI) and for breathy voice characterization (normalized first harmonic energy, NFHE). The database contained 744 patients, of whom 638 suffered from various functional and organic larynx disorders. Usage of a few parameters (mel-frequency cepstral coefficients, harmonics to noise ratio, normalised noise energy and glottal to noise excitation) and principal component analysis ensured an 89.3% accuracy level in pathology detection of only one disease—asthenia (Saenz-Lechon *et al.*, 2006). The accuracy of 93,4% was achieved with the use of an adaptive time-frequency transform algorithm and several features, such as the octave max, octave mean, energy ratio, length ratio and frequency ratio, which were classified using linear discriminant analysis (Umapathy *et al.*, 2005). An identification accuracy of 96% was achieved

using a neural network: a multilayer perceptron and learning vector quantization fed using short-term vectors calculated according to mel-frequency coefficients (Godino-Llorente *et al.*, 2006b).

In the literature, there are different ways of classifying vocal disorders for clinical applications. One of the most commonly used is principal component analysis (Saenz-Lechon *et al.*, 2006), which helps in extracting the most useful information regarding the signal and reduces the number of parameters. Reduction in parameters may cause a smoothed distribution of voices with an inferior classification (Matassini *et al.*, 2000). Goddard *et al.* (2009) present reduction in dimensionality for visualization of normal and pathological speech data using PCA and kernel PCA methodology. Another method used for reducing the dimensionality is non-linear PCA based on a neural network. It has been demonstrated that NLPCA, which minimizes the mean square reconstruction error from a reduced dimensionality space, can be very effective for representing the data which lie in a curved subspace, and may be very effective in voice pathology detection (Hu and Zahorian, 2008).

Moreover, there are studies combining acoustical signal and imaging of the vocal folds. There are already techniques available such as kymography and high-speed digital videoendoscopy (HSV) that enable the clinician to record the oscillating vocal folds in real time during the phonation. What is more, HSV enables one to visualise and quantify pathologies that affect the dynamic behaviour of the vocal folds (Skalski *et al.*, 2008). Acoustic analysis and a system visualizing vocal fold vibrations enable direct correlations to be established between acoustic parameters and measurements of glottal closure and vibratory symmetry (Mehta *et al.*, 2010).

In this paper we tackle the pathology detection challenges, by creating a feature vector made up of 28 different parameters derived from voice signal analysis. The parameters are used for analysis, because they are increasingly being adopted and understood by physicians due to the adjunctive programs used in the diagnosis making process. Other parameters, such as mel-frequency cepstrum coefficients, formants or zeroth-, first-, second-, and third-order moments, are used due to their wide functionality (Rabiner and Juang, 1993; Imai, 1983). Based on the vector made up of 28 parameters, we compare pathology detection accuracy, precision and specificity using machine learning methods such as principal component analysis, kernel principal component analysis and an auto-associative neural network. This provides a conceptual basis for the development of the current state of biomedical speech signal processing and mathematical methodology, highlighting the most clinically valuable information from the recorded speech signals which may not be adequately quantified using the currently available algorithmic tools contained in the

Table 1. Number of pathological patients taken under examination.

name of the pathology	women	men
hyperfunctional dysphonia	165	45
functional dysphonia	75	36
laryngitis	56	82
vocal cord paralysis	138	74

literature.

This paper is organised as follows. Section 2 contains the description of the material used in this work, Section 3 presents the description of the pathologies that are analysed, Section 4 showcases the methodologies used in the examination of all results, whose validation is presented in Section 5. Section 6 presents the final results and Section 7 contains a conclusion of all the work featured in this paper.

2. Material

Experiments have been performed on the Saarbrucken Voice Database, which was published online by the Institute of Phonetics of the University of the Saarland.[1] This database holds a collection of voice signals from more than 2000 healthy and pathological German speakers. Each of the recording sessions contains recordings of the sustained phonation of the a, i, u vowels spoken at four different pitches: high, normal, low and high-low-normal. The voice recordings were sampled with a frequency of 50 kHz and 16-bit resolution. The length of the recordings with sustained vowels lasts from 1 to 4 seconds. For the purpose of this work, all the vowels were used at a high, normal and low pitch. Recordings that were incomplete or damaged were excluded from the database. In our work we use the female and male recordings of four pathologies: hyperfunctional dysphonia, functional dysphonia, laryngitis and vocal cord paralysis. The number of pathological patients in the database who suffered from exact pathology is presented in Table 1. The number of healthy patients selected for examination was the same as for pathology cases.

3. Pathologies

Functional dysphonia (FD) is an enigmatic and controversial voice disorder that is frequently encountered in multidisciplinary voice clinics. Functional dysphonia occurs predominantly in women, commonly follows upper respiratory infection symptoms, can be transient and varies in its response to treatment (Morrison *et al.*, 1986; Roy, 2003). FD is often regarded as a disorder on a continuum of severity and it implies

---

[1] www.stimmdatenbank.coli.uni-saarland.de.

phonation is preserved but disordered in terms of quality, pitch, or loudness. There are two types of functional dysphonia, namely, disorders of hypofunction, with inadequate apposition of the vocal cords, and hyperfunctional dysphonia, in which accessory laryngeal muscles are used in voicing.

Hyperfunctional dysphonia is the most frequently occurring kind of dysphonia, which is a result of the voice organ failure with an excessive muscle tensioning inside and outside the larynx during the phonation (the so-called hard attitude voice). It is usually observed in people using their voice at work, and comes across as hoarse with limited resonance. Hyperfunctional dysphonia is characterised by strained vocal folds, congested due to the hyperkinetic mechanism of phonation (Belafsky *et al.*, 2002; Jiang *et al.*, 1998). In the advanced stage of hyperfunctional dysphonia, the vocal folds are hidden under the atrial folds, which take over the function of the vocal folds in the pathological phonation, or the so-called vestibular phonation.

Vocal fold paralysis causes a loss of vocal cord abduction and adduction. It may also influence phonation, respiration and deglutition, hence food and fluids may be easily aspirated into the trachea, resulting in pain in the throat when speaking, and a reduction in vocal volume. In unilateral paralysis the voice can be hoarse and breathy. In bilateral paralysis the voice is in a good condition, but with limited intensity. The airway is inadequate, which results in stridor and dyspnea with moderate exertion as each cord is drawn to the mid-line glottis by an inspiratory Bernoulli effect (Jiang *et al.*, 1998; Sulica and Blitzer, 2006; Maran, 1983).

Laryngitis is an inflammation of the mucous membrane lining the larynx and mostly results from viral infection (viral laryngitis). Laryngitis is usually associated with hoarseness or a temporary loss of speech and is often accompanied by an upper respiratory infection. In laryngitis, the tissues below the level of the epiglottis are swollen and inflamed. This causes swelling around the area of the vocal cords, which hinders their normal vibration (Mathieson *et al.*, 2009; Nicolosi *et al.*, 2004).

4. Methods

This study was carried out in order to assess the accuracy of classification methodologies in the detection of pathological speech. Each method was used with a reference to correct and pathological speech samples for both women and men separately. In the first step of this examination, the focus was on preliminary transformations of speech waveforms into a set of 28 parameters. Those parameters represented a dataset for a diagnosis of the patient's disease.

Preliminary signal processing and registration by

itself does not make it completely useful for the process of identifying and assessing changes in the deformation and pathology. Thus one needs to develop and explain phonetic recordings using set parameters and an accurate classification method. In this paper, the analysis of the speech signal is performed with the use of 28 parameters, namely, fundamental frequency, jitter and shimmer coefficients, energy, zeroth-, first-, second-, third-order moment, kurtosis, power factor, 1-, 2- and 3-formant amplitude, 1-, 2- and 3-formant frequency, maximum and minimum values of the signal and 10 mel-frequency ceptrum coefficients (MFCCs). The latter are commonly used in speech recognition because they simulate the subjective human perception of voice pitch by enhancing audible frequency and are less sensitive to noise (Rabiner and Juang, 1993). Cepstral analysis was used to estimate the fundamental frequency and was performed using a Fourier transformation of the sound spectrum.

In order to facilitate the extraction of the main components of the spectrum, before the transformation from the time domain to the frequency domain, the signal was subjected to the windowing operation using a Hamming window. In this paper we use a window with a width of $w = 384$ samples. Hence, knowing the position of the maximum frequency means, we can determine the basic frequency of the analysed sound using the following formula:

$$f_0 = \frac{l_s z}{wm}, \tag{1}$$

where l_s is the number of elements of the spectrum, z is the audio sample rate, w is the width of the analysed window (number of samples) and m represents the position of maximum cepstrum.

The sounds of laryngeal stimulation should have a relatively constant amplitude and frequency of the basic tone. When there are pathological changes to the larynx, the level of signal and its fundamental frequency change. Jitter (relative) is the average absolute difference between consecutive periods, divided by the average period, and is expressed as a percentage. The jitter coefficient (J) is calculated using

$$J = \frac{\frac{1}{N-1} \sum\limits_{i=1}^{N-1} |T_i - T_{i+1}|}{\frac{1}{N} \sum\limits_{i=1}^{N-1} T_i} 100\%, \tag{2}$$

where T_i are extracted F_0 period lengths and N is the number of extracted F_0 periods. The amplitude perturbation known as shimmer (relative) S is defined as the average absolute difference between the amplitudes of consecutive periods, divided by the average amplitude,

expressed as

$$S = \frac{\frac{1}{N-1} \sum\limits_{i=1}^{N-1} |A_i - A_{i+1}|}{\frac{1}{N} \sum\limits_{i=1}^{N} A_i} 100\%, \tag{3}$$

where A_i is the extracted peak-to-peak amplitude data and N represents the number of extracted fundamental frequency periods. A measure of the energy carried by the signal x is achieved by integrating the square of the signal E_x:

$$E_x = \int_{t_2}^{t_1} x^2(t)\, \mathrm{d}t \tag{4}$$

$$E_x = \sum_{n=n_1}^{n_2} x^2(n), \tag{5}$$

where n_1 and n_2 are the boundaries of intervals of time, $x(n)$ means the signal value of the n-th moment of time, i.e.,

$$x(n) = x(n\Delta t). \tag{6}$$

Having defined the signal time-frequency domain $G(t, f)$, the shape of the signal spectrum can be described as the zeroth-order spectral moment (Fujinaga, 1996):

$$M_0(t) = \sum_{i=0}^{\infty} G(t, f_i), \tag{7}$$

where $G(t, f)$ represents the time-frequency spectrum considering the centre frequency of the i-th band highlighted in frequency analysis.

The first-order moment is the centre of gravity of the spectrum (frequency-weighted average),

$$M_1(t) = \frac{\sum\limits_{i=0}^{\infty} G(t, f_i) f_i}{M_0(t)} \tag{8}$$

The second-order moment is defined as the square of the spectrum width,

$$M_2(t) = \frac{\sum\limits_{i=0}^{\infty} G(t, f_i)[f_i - M_1(t)]^2}{M_0(t)}. \tag{9}$$

The third-order moment is described as the asymmetry of the spectrum and means skewness,

$$M_3(t) = \frac{\sum\limits_{i=0}^{\infty} G(t, f_i)[f_i - M_1(t)]^3}{M_0(t)}. \tag{10}$$

Due to correlation, the standardised higher-order spectral moments are less suitable. Further calculations were made to obtain kurtosis, which is interpreted as

flattening the spectrum measurement (Joanes and Gill, 1998):

$$\text{kurtosis} = \frac{M_4(t)}{M_2(t)^2}. \tag{11}$$

The next parameter calculated was the so-called coefficient of relative power, which means the ratio of the power of the signal in the desired frequency wide-band $f \in [f_d, f_g]$ to signal power across the bandwidth $[f_0, f_\infty]$ (Engel *et al.*, 2007),

$$W_m(t) = \frac{\sum\limits_{t=t_b}^{t_g} \sum\limits_{f=f_{d2}}^{f_{g2}} G(t, f_i)}{\sum\limits_{t=t_b}^{t_g} \sum\limits_{f=f_{d1}}^{f_{g1}} G(t, f_i)}, \tag{12}$$

where f_{g1}, f_{d1} represent the lower and upper frequencies of the power wide-band, f_{g2}, f_{d2} are the upper and lower frequency ranges of the selected frequency wide-band and t_b, t_g are the beginning and the end of the recorded voice sample. The power factor was calculated for the first power coefficient describing the ratio of signal power in the first band to the overall signal power.

Other computed parameters were the formants. Formant frequencies reflect the size and shape of the supraglottal filter system, whereas the formant's amplitudes provide information pertaining to vocal intensity levels as well as coupling between the supraglottal cavities and subglottal spaces (Rachida and Amar, 2009). In order to determine the formant parameters, we needed spectral envelope, which was computed using linear predictive coding (LPC). It is a speech signal analysis technique which consists in the presentation of the speech signal as a response to the all-pole filter on the signal pitch (Atal and Hanauer, 1971). Linear prediction treats the signal as if it were generated by a signal of minimum energy, which was passed through a purely recursive IIR filter. An important thing when calculating formant frequencies from the filter is to find the locations of the resonances that make up the filter. To do this, we adopted the filter coefficients as a polynomial and solved the roots of the polynomial (Makhoul, 1975).

The formant amplitudes are influenced by the intensity of the signal source, frequency and damping within the vocal folds. Formants can be defined from the envelope which can be drawn to smoothly embrace the harmonics within the spectral maximum, i.e., the sound pressure level in dB of the envelope peak. The formant's frequency is related to the frequency position of the envelope maximum ($F1$, $F2$, $F3$, respectively). The formants calculated for five women, of whom four suffered from four voice pathologies and one was a healthy individual are presented in Fig. 1. Many authors show interest in cepstral factors in diagnostic evaluation of pathological changes in the glottis (Huber *et al.*, 1999;
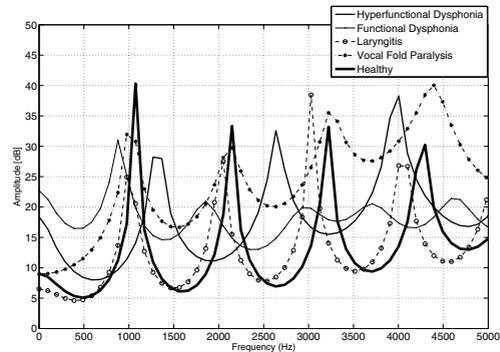

Fig. 1. Formant representation of five women: one suffering from functional dysphonia (age 41), one suffering from hyperfunctional dysphonia (age 48), one suffering from laryngitis (age 59), one who had vocal fold paralysis (age 54) and a healthy individual (age 21).

Fong *et al.*, 2013; Brinca *et al.*, 2014). The cepstrum is defined as the power spectrum of the logarithm of the power spectrum. It has a strong peak corresponding to the pitch period of the voiced-speech segment (Noll, 1967). The cepstral coefficients in this project are extended to the so-called Melow filtration, which consists of an additional non-linear frequency scale signal spectrum transformation. As a result, we achieved 10 MFCCs.

Since the feature vector was composed of a large number of parameters, we needed to use a method that would extract the most significant information. In order to perform reduction in the number of essential parameters we needed to organise their features according to the discriminant ability and in the final analysis obtain stable and consistent results, which better reflected the overall system performance.

The first method used for preserving the most relevant information from the data was the mathematical technique called principal component analysis (PCA). The significance of PCA has been discussed by Oja (2002) and Bishop (2006). It is a powerful tool for feature extraction based on minimizing the mean square error between original data vectors and other data vectors that could be estimated from the reduced dimensionality. PCA has two distinct versions: linear and non-linear. The objective of linear PCA is to find t orthogonal vectors (principal components) from the set of data in a lower dimensional data space that exhibit the largest variance. PCA has been proved to be the optimal linear transformation of the dataset and provides for efficient reduction of the initial dataset.

Theoretically, non-linear mathematical techniques have the potential to be more "efficient" than linear methods as they provide better representations with fewer dimensions. Non-linear transformation methods presented in this paper are non-linear PCA (NLPCA)

(Scholz *et al.*, 2008) and kernel PCA (kPCA) (Schölkopf *et al.*, 1999). The aim of NLPCA is to minimize the mean square error between the features estimated from the reduced dimensionality featured and the original ones. In other words, NLPCA determines a non-linear mapping function that projects N-dimensional features onto their corresponding highly informative M-dimensional representation (Tadeusiewicz *et al.*, 2013; Makki *et al.*, 2010; Zahorian and Hu, 2011).

One significant potential parametrization of NLPCA is based on auto-associative multilayer neural networks. Generally, an auto-associative neural network is a cascade subnet responsible for compression and decompression of data, as shown in Fig. 2. The values of output data from the auto-associative neural network algorithm are the same as the values of input data. In the middle of the network there is a layer that works as a bottleneck in which the dimension of the data is reduced. This bottleneck-layer provides the component values. This method is patterned after PCA. For the purpose of this work we chose hierarchical non-linear PCA. It forces the non-linear components to have the same hierarchical order as the linear components of the standard PCA, which yields a more significant meaning for individual components (Scholz *et al.*, 2008). The algorithm used in this work was proposed by Scholz and Vigário (2002), and it ensures optimal non-linear subspace spanned by components and restricts the non-linear components to have the same hierarchical order as the linear components in the standard PCA. Hierarchical NLPCA removes complex non-linear relationships between components. The auto-associative mapping is based on a 28-m-m-m-28 network, where m is the number of features that was calculated using PCA. PCA pre-processing and linear weight initialisation are used to start NLPCA with the linear solution of PCA. Figure 3 shows that the data set has non-linear characteristics.

The second method is referred to as kernel PCA. The goal of the non-linear transformation is to map the input space into a feature space via a non-linear map and then compute the principal components in that feature space. kPCA may compute the principal components in higher-dimensional feature space, which is non-linearly related to the original input dataset (Bishop, 2006). For the calculations of kPCA, the Gauss kernel function was used. The kPCA method was implemented using the algorithm created by Wang (2012). One of the most important factors for kernel PCA is parameter selection. In the case of a Gaussian kernel which is used in this work, it is σ in a kernel function described by

$$K(\vec{x}) = \exp\left(-\frac{(\vec{x} - \widetilde{\vec{x}})^2}{2\sigma^2}\right), \qquad (13)$$

where $\vec{x}$ is the vector data containing points, and $\widetilde{\vec{x}}$ equals
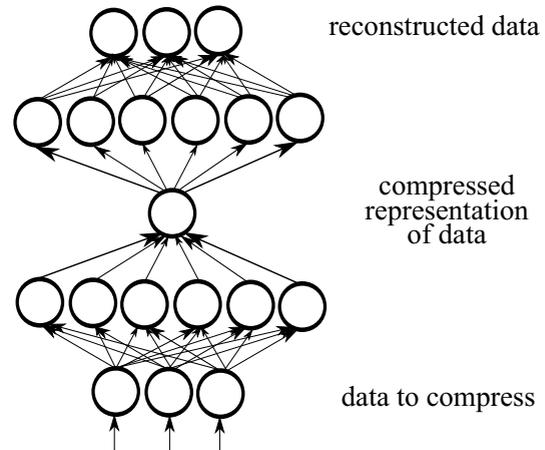

Fig. 2. Diagram of an auto-associative neural network performing non-linear PCA.

the mean value of $\vec{x}$.

In general, if we want to separate different classes in the new feature space, the parameter σ should be smaller than the inter-class distances, and larger than inner-class distances in the parameters domain (Wang, 2012). In this work, σ is selected in an experimental manner separately for female and male data. As a selection criteria the mean accuracy was from a function depending on the number of features and σ (Table 2), for the i vowel for all analysed pathologies. The vowel i was chosen due to the best results in relation to the other vowels. The results for vowel i for women and men are presented in Figs. 4 and 5. The calculation was made based on the data which did not enter the database used to validate the algorithms. For male signals, the maximum value of σ was achieved for 28 features and the $\sigma = 16.75$. The maximum value for women was also achieved for 28 features and $\sigma = 20.25$. These values were selected and used in the subsequent experiments.

5. Validation

To validate the calculated feature vector, a quality estimation of the classification operator was carried out using cross-validation. The classification operator consisted of results from the analysis for each patient. 10-fold cross-validation was used, where the data set was divided into $k = 10$ subsets. In this way, one of the subsets containing 10% of the data was used as a test set, with the 90% representing a training set. The analysis was carried out k times, where the voice samples belonged to random groups. The analysis was completed individually for each vowel at different intonations, separately for men and women for each pathology and each vowel at a different pitch. The confusion matrix, which is the classification standard tool, was used to summarise the

Table 2. σ value which characterised the highest accuracy result for each vowel.

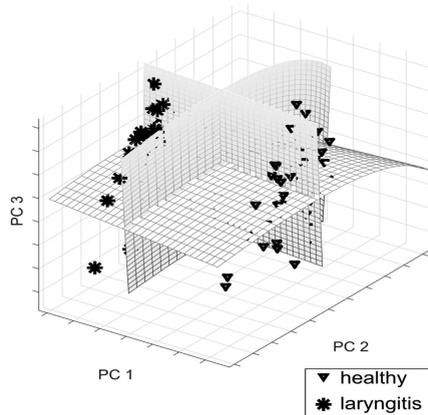| vowel	a		i		u	
gender	women	men	women	men	women	men
number of features	28	28	28	28	28	28
σ	20,5	17,5	20,25	16,75	21,25	18,00

Fig. 3. Hierarchical non-linear PCA applied to a 28-feature vector for women for the vowel a at a low pitch suffering from laryngitis showing non-linear behaviour. The first three non-linear components are visualised in the space of the first three PCA components, where the grids represent the new coordinate system of the component space.

Fig. 4. Graph presenting the mean σ distribution in relation to the number of parameters and detection accuracy for vowel i for women after PCA.

Fig. 5. Graph presenting the mean σ distribution in relation to the number of parameters and detection accuracy for vowel i for men after PCA.

results of the work. To achieve a quality assessment, parameters such as accuracy, precision and specificity for healthy and pathological detection were calculated. Sensitivity relates to the test's ability to detect positive results, whereas specificity relates to the identification of negative results. Definitions of the parameters placed in the proposed confusion matrix are shown in Table 3 and Eqns. 14–16.

An accuracy (ACC) is defined as the ratio of the parameters from a test set that were properly classified to all the parameters from a test set, as presented in Eqn. (14). The accuracy of the results achieved using the proposed confused matrix analysing each of the mentioned voice pathologies is presented in Table 6. Calculations were made separately for men and women:

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{TN} + \text{FP}}. \tag{14}$$

In pattern recognition and data mining, precision is the fraction of relevantly classified cases. It is defined as the number of all the cases that were classified into a particular class, and how many of them actually belonged to that class. In other words, it is defined as the ratio of cases that have been c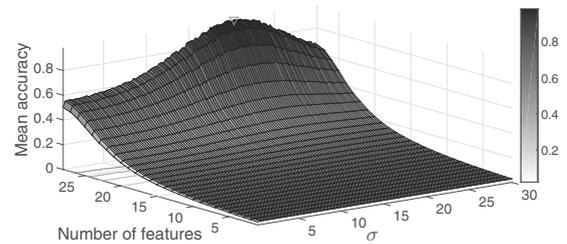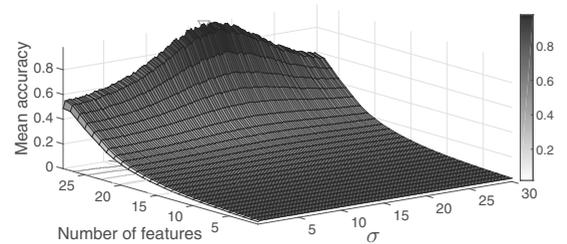lassified by the system correctly to the corresponding class to all results taken into the classification. Precision was calculated separately for health

$$\text{HD} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{15}$$

where TP and FP respectively mean true and false positive, and pathology detection

$$\text{PD} = \frac{\text{TN}}{\text{FN} + \text{TN}}, \tag{16}$$

where TN and FN respectively mean true and false negative.

6. Experimental results

Recordings for 1342 patients from the Saarbruecken Voice Database were used to evaluate the pathology detection and classification models. After signal processing analysis, the feature vector made up of 28 parameters

Table 3. Confusion matrix used in the analysis

		Results from classification	
		healthy	**pathology**
Diagnosed	**healthy**	true positive (TP)	false positive (FP)
	pathology	false negative (FN)	true negative (TN)
		Sensitivity =TP/(TP+FN)	**Specificity** =TN/(TN+FP)

was created, (cf. Section 4). These parameters were calculated automatically, by characterizing the clinically useful properties of the speech signals. The objective of our experiments is to compare the performance of 3 machine learning methods on 4 kinds of pathologies and relate them to the origin feature vector. Performance evaluation covers 5 main aspects: accuracy, healthy and pathology precision and sensitivity in a healthy and pathology state. The accuracy results of each method are presented in Table 6.

In the overall analysis, PCA enabled dimension reduction of up to 14 parameters from a 28-dimensional space, depending on the vowel and its pitch, as presented in Tables 4 and 5. The required number of parameters was chosen to cover at least 90% of the information contained in the signal, in order to make calculations faster (a lower dimensional space). This method showed mostly the same result as the initial 28-parameter vector, and in many cases even improved the effectiveness of the detection.

In general, the highest pathology detection accuracy rate is shown using the kPCA method and NLPCA, which means that the analysed data are non-linear and both methods deal with detection giving promising results, as shown in Table 6. What is more, NLPCA ensured dimension reduction to the same number of parameters (Tables 4 and 5). Going further, usage of kernel principal component analysis yielded a visibly higher level of accuracy than the initial 28-parameter vector and linear principal component analysis. There was no result that was worse than that provided by principal component analysis. The non-linear PCA based on an auto-associative neural network also showed the same trend as kernel PCA, and it enabled detection of voice impairments at the same level as kPCA. The main difference between kPCA and NLPCA is calculation time, as kPCA was four times shorter than NLPCA.

After analysing the mean accuracy of women for all the methods for all the vowels and their pitches, the best result was achieved by kernel PCA, whereas for men it was kernel PCA and non-linear PCA. While analysing

which vowel and pitch could detect with the highest accuracy all the pathologies presented in this paper, it was noted that there are vowel i at a normal pitch, u at a low and normal pitch using the kernel PCA method, vowel i at a low pitch and u at a normal pitch using kPCA, and vowel u at a low pitch using NLPCA. The poorest accuracy for all the analysed methods showed vowel a at a low pitch. Keeping all the vowels and their pitches in the analysis of laryngitis showed that the highest accuracy of voice pathology detection was by the 28-parameter feature vector, the PCA method, kPCA for women and by the 28-parameter feature vector, kPCA and NLPCA. In fact, the analysis for which voice pathology showed the highest level of accuracy is very difficult, because the differences between the mean accuracies were up to 1.5%. The results of precision for women (HD) for the initial 28 parameters vector, kPCA and NLPCA were between 94 and 100% and for PCA between 87 and 100%. For men, all of the methods showed HD of between 90 and 100%. The precision of pathological cases (PD) for women was kept between 94 and 100% for the 28-parameter vector, kPCA and NLPCA and for PCA from between 86 and 100%, whilst for men it was between 88 and 100%. The sensitivity for women ranged from 89 and 100%, where the lowest value belonged to PCA. The sensitivity calculated for men was located between 87 and 100% with the same trend for PCA. The specificity for women was between 90 and 100% and for men between 90 and 100%.

7. Conclusion

Traditional classification algorithms require the input of initial data to be formatted in a data matrix, where the columns represent the parameters describing the signal, and the rows represent the observations. We can find in the literature many types of proposed features describing the human voice, which are analysed further using different processing techniques that concentrate on covering more information (Bishop, 2006). Those analysis techniques provide a lot of advantages in transforming voice data. Extracting information about the state of a patient's health might be challenging while gathering significant information from many features.

The goal of this paper was to form a multi-dimensional feature vector with a proper classification that could discriminate between healthy and pathological voices. We created a vector that consisted of 28 features and used it for different machine learning methods such as principal component analysis, kernel principal component analysis and non-linear PCA based on an auto-associative neural network. The data set consists of 1342 patients of which half were healthy individuals. The database included patients suffering from 4 voice pathologies: functional dysphonia, hyperfunctional dysphonia, laryngitis and vocal fold

Table 4. Number of principal components covering up to 90% variance for women taken under examination.

number of principal components	a_h	a_l	a_n	i_h	i_l	i_n	u_h	u_l	u_n
functional dysphonia	16	16	16	14	14	14	15	14	15
hyperfunctional dysphonia	16	16	16	15	15	14	15	14	14
laryngitis	16	16	16	15	14	14	15	14	14
vocal cord paralysis	16	16	16	15	14	14	15	14	14

Table 5. Number of principal components covering up to 90% variance for men taken under examination.

number of principal components	a_h	a_l	a_n	i_h	i_l	i_n	u_h	u_l	u_n
functional dysphonia	16	16	16	14	14	14	15	14	15
hyperfunctional dysphonia	16	16	16	15	15	14	15	14	14
laryngitis	16	16	16	15	14	14	15	14	14
vocal cord paralysis	16	16	16	15	14	14	15	14	14

paralysis. The analysis was completed separately for recordings of 3 vowels: a, i, u spoken at 3 different pitches: high, low and normal, by both women and men. The results from different techniques were subjected to 10-fold cross-validation. Results from all the analysis were put in the confusion matrix, which enabled us to summarise the detection process.

Principal component analysis was used to extract relevant information from a huge dataset, by excluding a correlation between the parameters and as a result obtain reduced dimensions of data. In the end, classification based on a reduced number of principal components showed the very similar level of accuracy to the classification based on the 28 feature vector. The experimental results show that the PCA methodology helped to reduce the data in an efficient manner and we still had 90% of the variance data. The results did not lose significant signal information. The proof for this is that PCA achieved an optimal feature subset with a similar classification accuracy for most analysed vowels at all the vocal pitches for both female and male recordings having a reduced number of principal components in comparison with the 28 parameters vector.

By adding the "kernel trick" with non-linearity and proper sigma to PCA, we scaled the non-linear uncorrelated components to the unit variance. By this process we obtained a complex non-linear whitening (sphering transformation), which was a useful step for further classification. The results of kPCA were slightly better than of PCA analysis and the classification completed for a 28-parameter feature vector. The disadvantage of adding the "kernel trick" is still having an initial number of parameters (28), which required a much longer calculation time compared with the PCA method.

In addition, as illustrated through the last experiment, the hierarchical NLPCA method detected pathology in voice analysis with a similar accuracy, precision and sensitivity to kernel PCA, and in a few cases it was even higher. It enabled us to decompose the data non-linearly to distinguish application of dimensionality. The idea behind this approach is that the NLPCA method locates and eliminates linear and non-linear relationships in the data and ensures non-linear components have the same hierarchical order as linear components in the standard PCA (Wang, 2012).

The contribution presented in this paper shows that the chosen classifying methodology is relevant to the pathology detection process. Acoustic analysis using proposed parameters can be a useful, objective tool for confirming the pathological changes of the glottis in the analysed four types of voice pathology: functional dysphonia, hyperfunctional dysphonia, laryngitis and vocal fold paralysis.

Acknowledgment

References

Arroyave, J.R.O., Bonilla, J.F.V. and Trejos, E.D. (2012). Acoustic analysis and non linear dynamics applied to voice pathology detection: A review, *Recent Patents on Signal Processing* **2**(2): 1–11.

Atal, B.S. and Hanauer, S.L. (1971). Speech analysis and synthesis by linear prediction of the speech wave, *The Journal of the Acoustical Society of America* **50**(2B): 637–655.

Belafsky, P.C., Postma, G.N., Reulbach, T.R., Holland, B.W. and Koufman, J.A. (2002). Muscle tension dysphonia as a sign of underlying glottal insufficiency, *Otolaryngology—Head and Neck Surgery* **127**(5): 448–451.

Bishop, C.M. (2006). *Pattern Recognition and Machine Learning*, Vol. 1, Springer, New York, NY.

Table 6. Detection accuracy of voice pathology for four diseases (functional dysphonia, hyperfunctional dysphonia (hyperfunc. dysphonia), laryngitis and vocal fold paralysis) for three vowels spoken at three pitches for women and men, based on the initial feature vector (28 param.) and three machine learning methods expressed as a [%]. The results are compared separately for each vowel at a different pitch, where a_h represents the vowel a at a high pitch, a_l is the vowel a at a low pitch, a_n is the vowel a at a normal pitch, i_h is the vowel i at a high pitch, i_l is the vowel i at a low pitch, i_n is the vowel i at a normal pitch, u_h is the vowel u at a high pitch, u_l is the vowel u at a low pitch, u_n is the vowel u at a normal pitch.

| | | WOMEN | | | | MEN | | | |
		functional dysphonia	hyperfunc. dysphonia	laryngitis	vocal fold paralysis	functional dysphonia	hyperfunc. dysphonia	laryngitis	vocal fold paralysis
a_h	28 param	93	93	99	95	94	95	96	98
	PCA	93	92	97	92	91	93	96	98
	kPCA	93	93	100	95	92	94	97	98
	NLPCA	93	100	100	94	95	93	96	98
a_l	28 param	92	93	93	94	88	90	90	90
	PCA	91	89	90	94	87	87	88	91
	kPCA	92	99	93	95	88	87	89	91
	NLPCA	93	100	93	100	93	93	90	95
a_n	28 param	94	94	94	93	95	92	93	93
	PCA	92	92	95	90	95	89	90	92
	kPCA	94	97	98	95	98	90	93	94
	NLPCA	94	94	100	100	96	95	97	94
i_h	28 param	94	96	93	94	98	97	93	92
	PCA	92	96	93	96	99	97	90	90
	kPCA	94	100	97	99	100	98	91	92
	NLPCA	93	93	96	96	96	94	96	96
i_l	28 param	97	93	95	95	97	98	98	97
	PCA	99	92	95	95	97	96	98	95
	kPCA	100	93	97	96	99	98	99	96
	NLPCA	93	93	96	96	96	94	97	95
i_n	28 param	95	95	99	97	96	95	96	94
	PCA	94	95	99	96	95	94	96	92
	kPCA	97	99	100	98	96	94	97	94
	NLPCA	99	93	95	96	97	95	97	94
u_h	28 param	97	92	99	96	96	91	95	94
	PCA	98	93	97	96	97	89	96	93
	kPCA	100	95	98	98	97	90	98	95
	NLPCA	99	95	96	97	97	94	97	94
u_l	28 param	99	97	97	97	96	98	94	97
	PCA	98	98	99	98	93	98	92	97
	kPCA	100	100	100	100	94	100	94	99
	NLPCA	99	98	97	97	97	98	97	98
u_n	28 param	97	99	96	98	95	98	99	99
	PCA	99	99	95	98	94	97	98	97
	kPCA	100	100	98	97	95	97	98	97
	NLPCA	99	99	97	97	94	98	98	97

Brinca, L.F., Batista, A.P.F., Tavares, A.I., Goncalves, I.C. and Moreno, M.L. (2014). Use of cepstral analyses for differentiating normal from dysphonic voices: A comparative study of connected speech versus sustained vowel in European Portuguese female speakers, *Journal of Voice* **28**(3): 282–286.

Eadie, T.L. and Doyle, P.C. (2005). Classification of dysphonic voice: Acoustic and auditory-perceptual measures, *Journal of Voice* **19**(1): 1–14.

Engel, Z.W., Klaczynski, M. and Wszolek, W. (2007). A vibroacoustic model of selected human larynx diseases, *International Journal of Occupational Safety and Ergonomics* **13**(4): 367.

Farrus, M., Hernando, J. and Ejarque, P. (2007). Jitter and shimmer measurements for speaker recognition, *Annual Conference of the International Speech Communication Association (Interspeech 2007), Antwerp, Belgium*, pp. 778–781.

Fong, S., Lan, K. and Wong, R. (2013). Classifying human voices by using hybrid SFX time-series preprocessing and ensemble feature selection, *BioMed Research International* **2013**:1–27, DOI: 10.1155/2013/720834.

Fraile, R., Saenz-Lechon, N., Godino-Llorente, J., Osma-Ruiz, V. and Fredouille, C. (2009). Automatic detection of laryngeal pathologies in records of sustained vowels by means of mel-frequency cepstral coefficient parameters and differentiation of patients by sex, *Folia phoniatrica et logopaedica* **61**(3): 146–152.

Fujinaga, I. (1996). *Adaptive Optical Music Recognition*, Ph.D. thesis, McGill University, Montreal.

Goddard, J., Schlotthauer, G., Torres, M. and Rufiner, H. (2009). Dimensionality reduction for visualization of normal and pathological speech data, *Biomedical Signal Processing and Control* **4**(3): 194–201.

Godino-Llorente, J.I. and Gomez-Vilda, P. (2004). Automatic detection of voice impairments by means of short-term cepstral parameters and neural network based detectors, *IEEE Transactions on Biomedical Engineering* **51**(2): 380–384.

Godino-Llorente, J.I., Gomez-Vilda, P. and Blanco-Velasco, M. (2006a). Dimensionality reduction of a pathological voice quality assessment system based on Gaussian mixture models and short-term cepstral parameters, *IEEE Transactions on Biomedical Engineering* **53**(10): 1943–1953.

Godino-Llorente, J.I., Sáenz-Lechón, N., Osma-Ruiz, V., Aguilera-Navarro, S. and Gómez-Vilda, P. (2006b). An integrated tool for the diagnosis of voice disorders, *Medical Engineering & Physics* **28**(3): 276–289.

Hadjitodorov, S. and Mitev, P. (2002). A computer system for acoustic analysis of pathological voices and laryngeal diseases screening, *Medical Engineering & Physics* **24**(6): 419–429.

Horii, Y. (1980). Vocal shimmer in sustained phonation, *Journal of Speech, Language, and Hearing Research* **23**(1): 202–209.

Hu, H. and Zahorian, S.A. (2008). A neural network based nonlinear feature transformation for speech recognition, *9th Annual Conference of the International Speech Communication Association (Interspeech 2008), Brisbane, Australia,* pp. 1533–1536.

Huber, J.E., Stathopoulos, E.T., Curione, G.M., Ash, T.A. and Johnson, K. (1999). Formants of children, women, and men: The effects of vocal intensity variation, *The Journal of the Acoustical Society of America* **106**(3): 1532–1542.

Imai, S. (1983). Cepstral analysis synthesis on the mel frequency scale, *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP'83, Boston, MA, USA*, Vol. 8, pp. 93–96.

Jiang, J.J., Diaz, C.E. and Hanson, D.G. (1998). Finite element modeling of vocal fold vibration in normal phonation and hyperfunctional dysphonia: Implications for the pathogenesis of vocal nodules, *Annals of Otology, Rhinology and Laryngology* **107**(7): 603–610.

Joanes, D. and Gill, C. (1998). Comparing measures of sample skewness and kurtosis, *Journal of the Royal Statistical Society: Series D (The Statistician)* **47**(1): 183–189.

Jothilakshmi, S. (2014). Automatic system to detect the type of voice pathology, *Applied Soft Computing* **21**: 244–249.

Lieberman, P. (1963). Some acoustic measures of the fundamental periodicity of normal and pathologic larynges, *The Journal of the Acoustical Society of America* **35**(3): 344–353.

Makhoul, J. (1975). Linear prediction: A tutorial review, *Proceedings of the IEEE* **63**(4): 561–580.

Makki, B., Hosseini, M.N. and Seyyedsalehi, S.A. (2010). An evolving neural network to perform dynamic principal component analysis, *Neural Computing and Applications* **19**(3): 459–463.

Manfredi, C., D'Aniello, M., Bruscaglioni, P. and Ismaelli, A. (2000). A comparative analysis of fundamental frequency estimation methods with application to pathological voices, *Medical Engineering & Physics* **22**(2): 135–147.

Maran, A. (1983). Description of specific diseases of the larynx, *in* R. Harden and A. Marcus (Eds.), *Otorhinolaryngology*, Vol. 4, Springer, Dordrecht, pp. 99–104.

Matassini, L., Hegger, R., Kantz, H. and Manfredi, C. (2000). Analysis of vocal disorders in a feature space, *Medical Engineering & Physics* **22**(6): 413–418.

Mathieson, L., Hirani, S., Epstein, R., Baken, R., Wood, G. and Rubin, J. (2009). Laryngeal manual therapy: A preliminary study to examine its treatment effects in the management of muscle tension dysphonia, *Journal of Voice* **23**(3): 353–366.

Mehta, D.D., Deliyski, D.D., Zeitels, S.M., Quatieri, T.F. and Hillman, R.E. (2010). Voice production mechanisms following phonosurgical treatment of early glottic cancer, *The Annals of Otology, Rhinology, and Laryngology* **119**(1): 1.

Morrison, M.D., Nichol, H. and Rammage, L.A. (1986). Diagnostic criteria in functional dysphonia, *The Laryngoscope* **96**(1): 1–8.

Nicolosi, L., Harryman, E. and Kresheck, J. (2004). *Terminology of Communication Disorders: Speech–Language–Hearing*, Lippincott Williams & Wilkins, Philadelphia, PA.

Noll, A.M. (1967). Cepstrum pitch determination, *The Journal of the Acoustical Society of America* **41**(2): 293–309.

Oja, E. (2002). Unsupervised learning in neural computation, *Theoretical Computer Science* **287**(1): 187–207.

Rabiner, L.R. and Juang, B.-H. (1993). *Fundamentals of Speech Recognition*, Vol. 14, PTR Prentice Hall, Englewood Cliffs, NJ.

Rachida, D. and Amar, D. (2009). Effects of acoustic interaction between the subglottic and supraglottic cavities of the human phonatory system, *Canadian Acoustics* **37**(2): 37–43.

Roy, N. (2003). Functional dysphonia, *Current Opinion in Otolaryngology & Head and Neck Surgery* **11**(3): 144–148.

Saenz-Lechon, N., Godino-Llorente, J.I., Osma-Ruiz, V., Blanco-Velasco, M. and Cruz-Roldan, F. (2006). Automatic assessment of voice quality according to the GRBAS scale, *28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS'06, New York, NY, USA*, pp. 2478–2481.

Saldanha, J.C., Ananthakrishna, T. and Pinto, R. (2014). Vocal fold pathology assessment using mel-frequency cepstral coefficients and linear predictive cepstral coefficients features, *Journal of Medical Imaging and Health Informatics* **4**(2): 168–173.

Schölkopf, B., Smola, A. and Müller, K.-R. (1999). Kernel principal component analysis, *in* B. Schölkopf, C.J.C. Burges and A.J. Smola (Eds.), *Advances in Kernel Methods—Support Vector Learning*, MIT Press, Cambridge, MA.

Scholz, M., Fraunholz, M. and Selbig, J. (2008). Nonlinear principal component analysis: Neural network models and applications, *in* A.N. Gorban *et al.* (Eds.), *Principal Manifolds for Data Visualization and Dimension Reduction*, Springer, Berlin/Heidelberg, pp. 44–67.

Scholz, M. and Vigário, R. (2002). Nonlinear PCA: A new hierarchical approach, *10th European Symposium on Artificial Neural Networks (ESANN), Bruges, Belgium*, pp. 439–444.

Skalski, A., Zielinski, T. and Deliyski, D. (2008). Analysis of vocal folds movement in high speed videoendoscopy based on level set segmentation and image registration, *International Conference on Signals and Electronic Systems, ICSES'08, Kraków, Poland*, pp. 223–226.

Steinecke, I. and Herzel, H. (1995). Bifurcations in an asymmetric vocal-fold model, *The Journal of the Acoustical Society of America* **97**(3): 1874–1884.

Sulica, L. and Blitzer, A. (Eds.) (2006). *Vocal Fold Paralysis*, Springer, Berlin/Heidelberg.

Tadeusiewicz, R., Korbicz, J., Rutkowski, L. and Duch, W. (Eds.) (2013). *Neural Networks in Biomedical Engineering*, Inżynieria biomedyczna. Podstawy i zastosowania, Vol. 9, Akademicka Oficyna Wydawnicza EXIT, Warsaw, (in Polish).

Tsanas, A. (2013). Acoustic analysis toolkit for biomedical speech signal processing: Concepts and algorithms, *Models and Analysis of Vocal Emissions for Biomedical Applications* **2**: 37–40.

Umapathy, K., Krishnan, S., Parsa, V. and Jamieson, D.G. (2005). Discrimination of pathological voices using a time-frequency approach, *IEEE Transactions on Biomedical Engineering* **52**(3): 421–430.

Wang, Q. (2012). Kernel principal component analysis and its applications in face recognition and active shape models, *ARXIV* **1207.3538**.

Wong, D., Markel, J. and Gray Jr, A. (1979). Least squares glottal inverse filtering from the acoustic speech waveform, *IEEE Transactions on Acoustics, Speech and Signal Processing* **27**(4): 350–355.

Yumoto, E., Gould, W.J. and Baer, T. (1982). Harmonics-to-noise ratio as an index of the degree of hoarseness, *The Journal of the Acoustical Society of America* **71**(6): 1544–1550.

Zahorian, S. and Hu, H. (2011). *Nonlinear Dimensionality Reduction Methods for Use with Automatic Speech Recognition*, Vol. 06, Speech Technologies Source: InTech, Rijeka.

Daria Panek received her diploma in biomedical engineering from the AGH University of Science and Technology in Cracow in 2013. At present, she is a Ph.D. student of biocybernetics and biomedical engineering at AGH. Her research includes signal processing, pattern recognition, biomedical imaging, medicine and biocybernetics.

Andrzej Skalski is an assistant professor at the AGH University of Science and Technology in Cracow, Poland. His principal research interests are in the field of image processing and analysis, in particular segmentation, registration and tracking algorithms. He has participated in a number of research projects in the biomedical signal and imaging area. He has published about 50 research papers.

Janusz Gajda received his M.Sc., Ph.D. and D.Sc. degrees in electrical engineering at the AGH University of Science and Technology, Cracow (Poland), in 1978, 1985 and 1993, respectively. He is currently a full professor with the Department of Measurement and Electronics at AGH and the head of this department. Since 2008 he has been a member of the International Society of Weigh in Motion. Currently he is a topic editor of the journal *Metrology and Measurement Systems*. Professor Gajda's fields of expertise are electrical metrology, object identification, signal processing, biomedical measurements, measurements of road traffic parameters.

Ryszard Tadeusiewicz studied at the Electrical Engineering Department of the AGH University of Science and Technology in Cracow, from which he graduated (with honours) in 1971. Additionally, after receiving his degree in automatic control engineering, he studied at the Faculty of Medicine of the Medical Academy in Cracow, and undertook studies in the field of mathematical and computer methods in economics. Since 1971 he has been conducting research in the areas of bio-cybernetics, automatic control engineering, and computer science. In 1975 he was awarded the Ph.D. degree, and in 1981 the D.Sc. (habilitation) degree. In 1986 he became a professor and in 1991 a full professor at AGH. His research interests cover neural networks, computer vision, biomedical engineering, and distance learning. He has written and published over 950 scientific papers, which have appeared

in prestigious Polish and foreign scientific journals, as well as numerous conference presentations, both national and international. Professor Tadeusiewicz has also authored over 90 scientific monographs and books, including highly popular, frequently reprinted textbooks. The total number of Ph.D. graduate students supervised: 58, the total number of graduate students he has advised as the principal advisor: 228. Elected three times as the rector of the AGH University, and several times as a member of the Governmental Central Committee for Scientific Degrees and Titles as well as the Ministry Council for Scientific Research. Awarded honorary doctorates by twelve Polish and foreign universities.