Analysis of Voice Modifications for Persons After Tonsillectomy

Ł. Potępa[a,*], J. Szaleniec[b], W. Wszołek[a], A. Steczko[b] and J. Składzień[b]

[a]AGH University of Science and Technology, Al. A. Mickiewicza 30, 30-059 Krakow, Poland

[b]Jagiellonian University Collegium Medicum, Chair and Clinic of Otolaryngology

Śniadeckich 2, 31-531 Kraków, Poland

The goal of the research described in the present paper was the determination of modification range for voice acoustic parameters resulting from tonsillectomy. Within the scope of the described research program, an attempt has been made to determine the changes of selected voice parameters for persons after such a treatment and also to elaborate work out some premises for prediction of potential voice modifications for persons who have not yet decided to undergo such a treatment. In order to achieve the goal, analyses have been carried out for voice utterances of persons before the tonsillectomy surgery and after the treatment. The first voice recordings took place between one and three days before the surgery. The post-treatment recordings have been carried out about 6 weeks after the surgery, as a procedure accompanying the follow-up examinations. In the present paper, an analysis has been carried out concerning phonemes /a/, /e/, /i/, and /u/ with prolonged phonation. The completed research shows that for evaluation of voice modification in the aspect of changes resulting from tonsillectomy, the most useful parameters are some of the mel-cepstral coefficients, the formant frequencies, and also the relative power coefficients.

1. Introduction

Tonsillectomy, or excision of palatine tonsils, is a surgery treatment executed on modulation structures of the vocal tract. Morphological changes resulting from such a treatment are not very extensive if the size is considered, however they may lead to various changes in voice quality. Potential deterioration of the voice quality may be of great importance for some patients (e.g. performing actors or singers) and it may significantly affect their eventual decision to undergo the palatine tonsils excision. Therefore it seems purposeful to determine objectively the changes in voice quality resulting from tonsillectomy treatment. The determination of extensiveness and characteristics of the voice changes in patients after the surgery may provide an indication for prediction of potential voice changes for patients who have not decided yet to undergo such a treatment and before the decision they consider all the pro and contra arguments regarding the operation. The research undertaken by the authors of the present work are focused exactly on that topic and may provide some scientific premises for rational elaboration of proper decisions for taking proper decision rationally.

From the analysis of collected recordings, values of selected speech signal parameters have been determined before and after the surgery with special attention focused on objectively detectable differences. The study included only the effects of changes in the morphological structures of the voice channel. There is therefore expected to changes in the fundamental tone generation process. This is where changes in the fundamental tone generation process are expected.

In spite of the fact that in the opinion of the majority of patients their voice has not changed after the treatment, acoustic time-frequency analysis of their voice structure and statistical assessment of the obtained results showed that there are certain voice parameters that significantly distinguish the utterances before and after the tonsillectomy. Such a result indicates that considering the pro and contra arguments in the selection of that type of treatment, one has to take into account also the certain secondary effects like the changes in the voice sound of the patient that can be of essential importance in some specific situations. In addition to the above-mentioned special situations of persons for whom the voice is the tool that makes their living, the changes in voice sound may create real problems if the biometric identification methods and personal identity verification methods, based on the voice evaluation of the person that is being identified e.g. in a bank system or is supposed to get access to confidential data, will be introduced into common practice. Voice changes in the case of persons subjected to the above-mentioned biometric identification methods may result in effects similar to forgetting their PIN code or losing their chip-card. Therefore the subject of the study should be considered as important and needful, however taking into account the small number of samples contained in the research material analyzed in the present study it should be stated that the presented results are of preliminary nature. Still they provide a basis for indi-

---

*corresponding author; e-mail: lpotepa@agh.edu.pl

cation of phonetic and acoustic parameters of the utterances that are being modified as a result of tonsillectomy – and this is the main objective of the present paper.

2. Hypertrophy of palatine tonsils and recommendations for their excision

The palatine tonsils of the regular size take the shape of an ellipsoid with the long and short axis equal to 2 and 1 cm, respectively. Their name has been taken after the shape similar to natural almonds. They are located in the pharynx grooves on both lateral walls of the pharynx, between the palato-pharyngeal arch and the palato-glossal arch. Together with the pharyngeal tonsil they form the Waldeyer's pharyngeal ring located in the pharyngeal narrowing [1].

Location of tonsils on the crossing of respiratory and digestive tract allows the contact with pathogenic factors both in the inhaled air and the consumed food (e.g. bacteria, viruses, fungi). In the tissues of tonsils, the first exposure of the pathogenic factors to the human immunological system takes place allowing early reaction to the threat posed by presence of pathogens.

Fig. 1. Hypertrophied tonsils (source: authors' own archives).

The origins of the tonsils hypertrophy (Fig. 1) have not been uniquely determined so far. Specific cases of tonsils hypertrophy do not also lack any commonly accepted etiology. Hypertrophied tonsils may be related to both physiological hypertrophy (because of the antibodies production) as well as with recurring acute pharynx infections or other diseases. It should be pointed out that the size of tonsils should not be treated as the primary indication for the surgery treatment. The decision regarding excision of tonsils should be taken in the context of all the clinical symptoms. Because of the essential role played by tonsils in the physiology of the immunological system, their extraction (especially in children) may result in some impairment of the immunological functions of the organism. Taking into account all the negative aspects of the treatment it should be admitted that not in all cases the tonsils excision is beneficial for the patient. For each specific case, an individual evaluation is required to maintain balance of potential benefits and threats re-

sulting from the treatment. The present work provides additional (new) arguments located in the threats area.

3. The research procedures

Recordings of the patient's voices have been executed out in a properly insulated room dedicated to audiometric measurements in the Otolaryngology Clinic of the UJ Collegium Medicum. The utterances of persons recommended for the tonsillectomy treatment have been recorded twice. The first recording has been carried out during the waiting period, namely one to three days before the treatment, while the second recording has been carried out about six weeks after the treatment, as a procedure accompanying the check-up examinations. In such a mode, 20 persons recommended for tonsillectomy (12 males and 8 females) have been examined. From the collected recordings, signal time samples that contained utterances of selected phonemes i.e. /a/, /e/, /i/, and /u/ have been selected for further analysis. During the recording session, patients were asked to read out a pre-selected text. The registration was executed three times in order to reduce the probability of the getting recordings deformed by accidental interferences that would make them unsuitable for further analysis and to control to ensure repeatability of the acquired data. For all the recordings, the text was the same. The text consisted of the following set of utterances:

a, e, i, u,
ala, as, ula, ela, igła,
Dziś jest ładna pogoda.
(English: The weather is good today.)
aaa, eee, iii, uuu[†]

The signal recording has been carried out using the measurement setup presented as a block diagram in Fig. 2. The sampling frequency for the signal was 48 kHz. The collected data have been registered in the WAVE format.
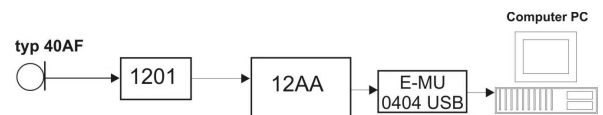

Fig. 2. A schematic diagram of the measurement setup where: 40AF — measurement microphone by G.R.A.S.; 1201 — preamplifier by Norsonic; 12AA — amplifier by G.R.A.S, EMU 0404 USB — 24 bit soundcard.

4. Signal parametrization

In many scientific papers, including the papers by the present authors [5, 6] it has been found that the most important (and most difficult) element of the research work preceding the practical use of the speech signal as a source of medically useful diagnostic and prognostic

---

[†]The listed vowels read with prolonged phonation.

information was the detection and description of those signal parameters that were as indifferent as possible to the context and the personal features of the voice being studied.

The factors that are important in the description of pathological speech are the factors that can disturb the natural articulation process. These may be factors deforming the functions of the laryngeal generator, resulting from various diseases or from surgical treatments performed within the voice tract. The extraction of these factors from the acoustic signal requires a series of transformations executed on the registered "raw" acoustic signal.

In the present paper, for the purpose of determination of the variation range of the speech signal resulting from the tonsillectomy treatment, several phonetic and acoustic parameters of the patient's voice have been distinguished. The parameters have been calculated from utterances of the above-mentioned vowel phonemes of prolonged phonation. In consecutive parts of the present work, the main parameters characterizing the differences in utterances before and after the treatment have been discussed.

The signal parameters that have been considered among others were formants. As it is commonly known, these parameters of the speech signal can be determined as local maxima of its envelope curve in the amplitude vs. frequency characteristic. Amplitudes of these maxima in general do not contain interesting information while the frequencies at which the maxima are observed are of great importance in the speech analysis and recognition. They are called the formant frequencies.
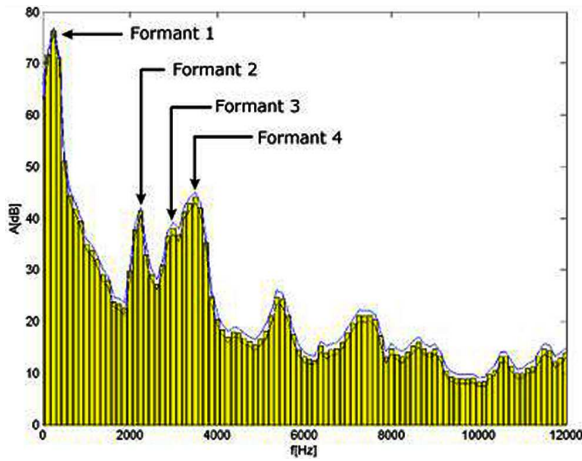

Fig. 3. Acoustic spectrum with its envelope curve for the utterance of /i/ phoneme by a male voice. Four consecutive formants have are marked.

Fig. 3 presents a typical frequency characteristic of the speech signal in its quasi-stationary part related to the prolonged phonation of the vowel together with its envelope curve and the marked sequence of formants. Clearly distinguishable formants in the Fourier spectrum are vis-

ible for the vowel utterances. In spite of the fact that the formant frequencies are individual features of the speaker, their relative values are approximately constant for individual phonemes.

Up to date, several analytical methods have been elaborated for formant determination [2–4, 13]. If by $G(t_i, f_i)$ we denote the result of Short-Time Fourier Transform (STFT) [14] of the speech signal, then i-th formant F_i can be defined as [15]

$$F_i(j) \equiv f_i \Leftrightarrow$$

$$\left(\frac{\partial G(t_j, f_i)}{\partial f}\bigg|_{f_i} = 0\right) \wedge \left(\frac{\partial^2 G(t_j, f_i)}{\partial f^2}\bigg|_{f_i}\langle 0\rangle\right)$$

$$\wedge_{j=1}^{i=1}\left\{\left[\exists f_j \langle f_{j+1} : \left(\frac{\partial G(t_j, f_i)}{\partial f}\bigg|_{f_i} = 0\right)\right.\right.$$

$$\left.\left.\wedge \left(\frac{\partial^2 G(t_j, f_i)}{\partial f^2}\bigg|_{f_i}\langle 0\rangle\right)\right] \wedge \left(F_{il} \leq f_i \leq F_{ih}\right)\right\}, \qquad (1)$$

where: F_{il} and F_{ih} are values limiting the frequency band sought of formants, t_j — j-th time interval for which the short term spectrum is calculated, f_i — center frequency of the ith band.

Changes within the set of formant parameters do not determine the changes in the shape of the spectrum envelope function outside the close vicinity of formant frequencies. Therefore it is necessary to determine also such parameters that will be appropriate for the changes in the whole structure of the spectrum (the shape of the spectrum). Such parameters that describe the shape of the spectrum are the spectral moments.

The spectral moments are used for evaluation of the contribution from the high frequency energy components in the total signal energy. In general, such moments are described by

$$M(m) = \sum_{k=1}^{K} |X(k)| [f_k]^m, \qquad (2)$$

where $|X(k)|$ is the value of the power spectrum for the k-th analyzed frequency band, f_k is the center frequency for the k-th band, and K is the number of analyzed frequency bands.

The parameter m determines the sensitivity of the spectral moment to changes of the signal energy in higher frequency bands. The higher the rank of the moment, the higher sensitivity one gets to such energy distribution changes.

Comparison of the utterance recordings of various signal energy values requires the normalization of the signal. In the present study, in most cases the normalized spectral moments are applied. The normalized spectral moment of the mth rank is given by

$$MN_m = \frac{M(m)}{M(0)}. \qquad (3)$$

The zero rank moment that is most frequently applied to normalization of the other moments is defined as

$$M(0) = \sum_{k=1}^{K} X(k). \qquad (4)$$

One of the observations that can be made while keeping track of the consequences of pathology in the speech signal articulation field is the statement that in the pathological speech signal, the energy is distributed differently (in comparison to the correct speech pattern) within the individual frequency bands of the signal. That feature is additionally emphasized by the relative power coefficients W_m, proposed by the authors as additional components of the feature vector.

In the present study, the above-mentioned relative power coefficients describing the ratio of the signal energy in a given (selected) frequency band to the signal energy in the band being analyzed have been applied for evaluation of changes in the frequency range close to the formant frequencies . The frequency band for the power coefficients is usually determined by experimentally found boundary frequencies that limit the occurrence of formant frequencies for individual phonemes [6]. The formula for the power coefficient in the range of the m-th formant has the following form:

$$W_m = \frac{\sum_{f=f_d}^{f_g} |X(f)|}{\sum_{f=f_p}^{f_k} |X(f)|}. \qquad (5)$$

Frequencies f_d and f_g are the boundary frequencies limiting the occurrence of formant frequencies in individual phonemes, while the f_p and f_k determine the starting (low) and ending (high) frequencies, respectively, for the analyzed frequency band. The frequency ranges used for individual formants can be found e.g. in [7]. Yet another methods for determination of of further speech signal parameters are the cepstral techniques.

The concept of cepstrum is particularly useful for speech analysis and synthesis. On the basis of the so-called perceptual approach it is generally considered that the most effective recognition results are obtained by mimicking the recognition mechanisms that are used by humans. One of such methods is the transformation of the frequency scale in such a way that it is naturally fit to the subjective frequency reception of the human hearing organ. The perceptual scales applied in the speech recognition are characterized by nonlinear frequency mapping and the nonlinearity is particularly noticeable for higher frequencies (above 1 kHz).

The idea of making use of cepstral parameters calculated in the mel scale follows from the nonlinear dependence between the frequency scale expressed in Hz and the subjective reception of the signal level by the human ear. With increasing frequency values, the minimum frequency differences for two sinusoidal signals resolved by the human ear are also higher. The dependence is observed for two simultaneously received signals as well as for two non-simultaneous signals. In order to compensate for this feature in the sound signal analysis, a special scale (mel scale) has been applied reflecting the ability of subjective frequency perception by the human hearing organ [15]. The mel-cepstral analysis applied in the present study makes use of the results obtained from signal spectrum filtering by a set of filters with triangular characteristics and the same width in the mel scale [9]. The use of such scale leads to enhancement of low frequency components in relation to high frequency components in the analyzed signal.

After signal filtering by using consecutive filters of the set, integrals of the filtered results are calculated. Such a vector, obtained from the integration results, is then subjected to the cepstral analysis. The result of the operation is a set of mel-cepstral parameters the number of which is determined by the number of filters used.

The cepstral coefficients are calculated from

$$C_n = \sqrt{\frac{2}{n}} \sum_{i=1}^{N} \log(s_i) \cos\left(\frac{\pi n}{n}\left(i - \frac{1}{2}\right)\right), \qquad (6)$$

where C_n is the n-th cepstral parameter, s_i is the i-th parameter obtained by integration of the spectrum filtered by the i-th filter in the set, and N is the number of filters. For the present study, 20 triangular filters have been used, each with the 300 mel width. A schematic diagram of the MFCC determination method is shown in Fig. 4.

Fig. 4. A schematic diagram of the classical algorithm for determination of MFCC coefficients.

The values of cepstral coefficients with lower indices reflect the shape of the spectral envelope function, while the other coefficients carry information about the excitation signal (which for the voiced consonants can be identified with the laryngeal tone). The presentation of cepstral coefficients calculated for the spectrum reduced by the mel filtration often allows allows to obtain results that are more clear in the interpretation in the aspect of the changes that have taken place.

5. Research results obtained so far

Based on ANOVA analysis, the set of 41 considered parameters is reduced to a set of 15 parameters which have significant (in statistical sense) effect on appraisal and classification of patients' condition before and after tonsillectomy.

The parameters are different, depending on the specific phonemes and and gender of the person uttering the phonemes. Tables I and II present the set of features that significantly differentiate between the patients' utterances before and after the tonsillectomy treatment.

In the Tables and Figs. 5 and 6, the following symbols have been used:
F_1, F_2, F_3, F_4 — consecutive formant frequencies (obtained from the LPC algorithm);
$C_1, \ldots, C_{20}$ — consecutive mel-cepstral parameters according to formula (6);

TABLE I

Listing of average values for parameters that significantly differentiate between the phoneme read out by a male before and after the tonsillectomy.

Name of feature	Before		After		Before		After	
	tonsillectomy				tonsillectomy			
	mean	std.	mean	std.	mean	std.	mean	std.
F_1	877.92	121.99	893.83	116.08	460.00	99.18	418.92	74.52
A_1	24.92	6.20	22.33	10.50	28.42	4.80	26.92	7.55
A_2	20.92	5.88	21.17	7.32	15.00	5.53	14.83	4.86
F_4	4925.58	825.73	5013.00	885.78	5588.08	615.50	5391.50	611.07
C_1	8.69	2.45	9.62	2.18	−2.96	4.65	−3.87	2.35
C_5	−1.98	1.42	−1.81	1.61	−0.39	1.27	0.45	1.14
C_{11}	0.48	0.85	0.60	1.02	−0.84	0.71	−0.87	0.99
C_{12}	0.26	0.87	−0.03	0.64	−0.80	0.74	−1.44	0.96
C_{14}	−0.85	0.77	−0.72	0.86	−0.04	0.58	0.43	0.49
C_{15}	−1.34	0.65	−1.55	0.58	0.39	0.60	−0.37	0.80
C_{19}	0.08	0.55	0.17	0.40	−0.75	0.59	−0.21	0.59
MN_2	7.27×10^6	2.76×10^6	6.51×10^6	2.67×10^6	5.33×10^6	2.89×10^6	5.74×10^6	2.81×10^6
MN_3	7.12×10^{10}	3.20×10^{10}	5.77×10^{10}	2.80×10^{10}	6.33×10^{10}	2.78×10^{10}	7.28×10^{10}	3.04×10^{10}
W_1	0.22	0.05	0.23	0.06	0.19	0.14	0.27	0.11
W_2	0.13	0.05	0.17	0.07	0.10	0.04	0.11	0.03

TABLE II

Listing of average values for parameters that significantly differentiate between the phoneme read out by a male before and after the tonsillectomy.

Name of feature	Before		After		Before		After	
	tonsillectomy				tonsillectomy			
	mean	std.	mean	std.	mean	std.	mean	std.
F_1	784.75	234.97	1018.63	164.89	440.88	56.48	480.38	67.91
A_1	19.38	5.21	28.13	9.92	27.38	7.33	28.75	5.82
A_2	17.75	6.25	20.63	12.47	13.50	11.05	17.38	9.47
F_4	4702.75	476.78	5182.00	551.85	5068.50	792.56	5431.50	672.10
C_1	9.86	2.48	9.69	2.46	−5.43	3.63	−3.51	3.22
C_5	−3.02	0.81	−3.49	2.26	−4.07	2.18	−4.48	2.49
C_{11}	−0.02	1.36	−1.10	0.78	−1.23	0.69	−1.51	0.73
C_{12}	−1.08	1.50	−1.09	1.70	−0.68	1.30	−0.47	1.32
C_{14}	−0.67	0.62	−1.00	0.71	−0.31	0.56	−0.57	1.05
C_{15}	−0.46	0.83	−0.56	0.74	0.00	0.99	−0.32	0.72
C_{19}	−0.31	0.50	−0.39	0.56	−0.48	0.46	−0.49	0.62
MN_2	8.82×10^6	2.03×10^6	1.13×10^6	2.62×10^6	6.65×10^6	2.07×10^6	7.30×10^6	2.18×10^6
MN_3	8.47×10^{10}	2.15×10^{10}	1.17×10^{10}	3.66×10^{10}	8.80×10^{10}	2.23×10^{10}	9.59×10^{10}	2.84×10^{10}
W_1	0.22	0.06	0.23	0.06	0.15	0.15	0.17	0.16
W_2	0.19	0.08	0.14	0.07	0.11	0.08	0.08	0.03
W_3	0.03	0.01	0.04	0.02	0.01	0.00	0.02	0.02

$C_1, \ldots, C_4$ — power coefficients for consecutive formants. according to formula (5);
MN_1, MN_2, MN_3 — normalized spectral moments of rank 1, 2, and 3 according to formula (3).

The criteria used for selection of parameters (components) of the feature vector that were sufficient to distinguish between utterances before and after the treatment were formed by the values of ANOVA based on Snedecor's F distribution calculated for particular acoustic parameters of vowels read out with prolonged phonation. Additionally, from the distinguishing set of parameters, some have been excluded by the condition that they did not failed to fulfill the necessary conditions regarding the distribution of values and homogeneity of variance required by the above-mentioned statistical tests. The average values of the parameters distinguishing between the utterances (vowel /a/) are presented as radar plots separately for female voices (Fig. 5) and male voices (Fig. 6). To improve the readability of the figures values
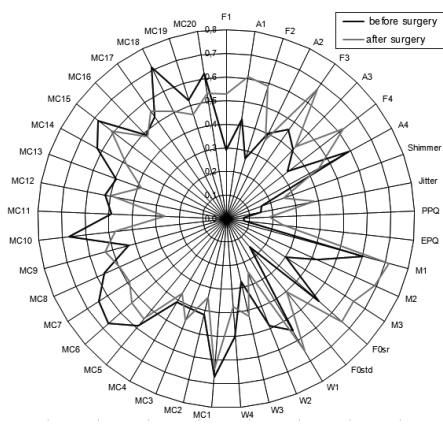
Fig. 5. A representation of normalized average values for parameters from the set differentiating between female voices before and after the treatment.
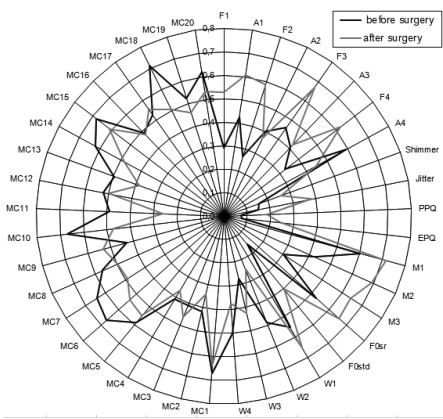

Fig. 6. A representation of normalized average values for parameters from the set differentiating between male voices before and after the treatment.

of parameters are standarized. From such a presentation of the results (values of individual components of the feature vector) one can determine which are the ones that change their values considerably compared to the values after the treatment to the respective values calculated before the treatment. These parameters can be used for evaluation of changes in the speech after the surgery in comparison with the speech before the surgery (e.g. by applying a properly chosen metric) [6].

When analyzing the obtained results (presented in Figs. 5 and 6) one should emphasize the great usefulness of the mel-cepstral parameters, spectral moments, and power coefficients in evaluation of voice changes resulting from palatine tonsils excision for both male and female voices. For the female voices, a statistically significant change has been also observed for such parameters as format frequencies for the forth formant of the /a/ vowel (the average value before the treatment was 4.7 kHz and after the treatment 5.2 kHz) and second and third spec-

Mapping multi-variant data onto a plane is not a real reflection of the original data set. However, it retains

tral moments (values for /a/ vowel are 8.82×10^6 and 8.47×10^{10} before the treatment and 1.13×10^6 and 1.17×10^{10} respectively after the treatment). A change in those parameters can be also observed in male voices but it is less pronounced).

The feature vectors represent the speech standard and deformed signals in a 15-dimensional space. This multidimensional structure is difficult to interpret by visual inspection. Therefore, it is important to develop an understandable and perceptible graphical illustration of such set of data. In certain studies [17], solving this problem by applying the Sammon's representation [16] is proposed.. The Sammon's representation allows to present multi-variant data on a plane. The relevant transformation is based upon selection of two-dimensional vectors created a the way which minimises the error function of the distance between the considered objects.
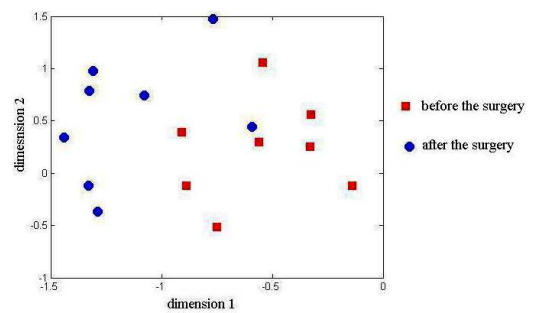

Fig. 7. A visualization of 15-dimensional feature vectors by Sammon's representation – female speech, vowel /a/.
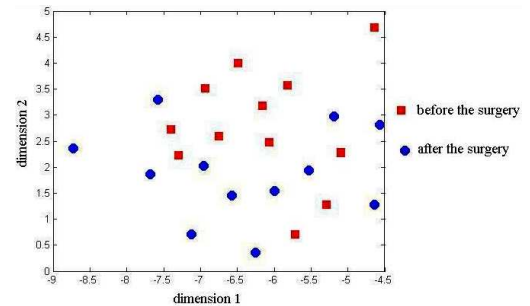

Fig. 8. A visualization of 15-dimensional feature vectors by Sammon's representation — male speech, vowel /u/.

Example results of the Sammon's representation for the examined group of voices are presented in Figs. 7 and 8.

Summaries of Sebestyen measures for distractions and separability of classes before and after application of the mapping are shown in Tables III and IV.

such of the features as recognition of the inter-and intraclass in quity good agreement with the original vectors.

Summary of Sebestyen measures for the data before and after the Sammon's representation (Fig. 7). TABLE III

	Measure of dispersion between classes	Measure of dispersion inside classes	Quantitative assessment of the separability of classes $\log_2(M_1/M_2)$
Original data	0.99	0.75	0.39
Sammon's representation data	1.10	0.66	0.73

Summary of Sebestyen measures for the data before and after the Sammon's representation (Fig. 8). TABLE IV

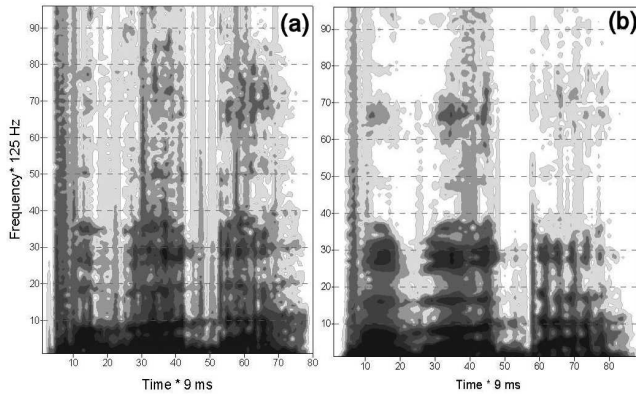|	Measure of dispersion between classes	Measure of dispersion inside classes	Quantitative assessment of the separability of classes $\log_2(M_1/M_2)$
Original data	0.51	0.38	0.53
Sammon's representation data	0.58	0.36	0.68

Fig. 9. Spectrogram comparison for /*pogoda*/ word utterance by female voice: (a) before the treatment and (b) after the treatment.

For male speech signal, differences are smaller than in the observed female speech signals.

In the plots, one can distinguish separate regions occupied by signal parameters obtained for recordings acquired before and after the treatment, respectively. These regions are not fully disjunctive which gives a strong argument that one can not distinguish between a person's voice before and after the treatment using only these selected parameters. Comparing the average values of individual parameters one can often indicate the most probable direction that the voice changes will take after the surgery. The obtained results of voice analysis show that only for the /a/ phoneme uttered by the female voices, an increase occurs of the forth formant frequency and of the third formants resulting from the tonsillectomy. Additional information regarding the changes in formant frequencies can be obtained by analyzing the spectrum of the speech signal before and after surgery (Fig. 9).

Comparison of the spectrogram plots examples of which can be seen in Fig. 9, may provide the information concerning the direction of voice changes in the region of formant frequencies. The observed differences between the utterances before and after the treatment for the collected recordings of the /a/ phoneme provide a confirmation of the results presented in papers [10–12]. From the

differences in the spectrogram plots of these results one can conclude that the formant structure of the patient's utterances after the treatment changes much more than it might result from a natural variation.

6. Conclusions

Results obtained from analysis of selected parameters extracted from the utterances indicate that the voice changes related to palatine tonsils excision mainly regard the changes in values of parameters determined from the Fourier spectrum of the speech signal. These changes mainly follow from the changes in shape of the organs modulating the sound generated by the vibrating vocal folds. The results of the present analysis partially confirm the results of the papers by other authors (in particular the results regarding the increase of the F_3 and F_4 formant frequency for the /a/ phoneme uttered by female voices).

As a consequence of the present study it has been also confirmed that the selected mel-cepstral parameters, spectral moments, and power coefficients distinguish the persons' utterances before and after the tonsils excision. The fact indicates the possibility of using such an analysis for evaluation of voice changes related to the surgical treatment in the area of modulation tract.

Results of the present study allowed also to identify the changes concerning the values of formant frequencies F_3 and F_4 determined for the /a/ phoneme utterances by female voices and the changes related to the speech signal energy distribution in the region of the second and third formant frequency (power coefficients W_2 and W_3), and respectively in the region of the second formant frequency (power coefficient W_2) from male voices. The results show also that there are considerable changes in the structure of the speech signal spectrum caused by the surgical treatment.

Concluding the completed research one can state that differentiation between the utterances before and after the treatment is based on wide application of selected mel-cepstral parameters, power coefficients, and formant frequencies.

These parameters can be used as input to computer systems supporting the work of doctors involved in the diagnosis and treatment of voice channel. It seems however necessary to carry out further studies in order to

verify usefulness of the selected mel-cepstral parameters for distinguishing of utterances for bigger groups of patients. Further studies are also necessary for verification of changes noticed in the frequency areas close to the first and second formant of the utterance, because such changes have not been described before in the widely circulated publications, despite the fact they are related to research dealing with the format structure of the utterances.

In spite of the fact that the presented results should be verified for a larger group of patients, the results described here suggest that in further perspective one can expect they will provide a basis for defining a new, extended set of parameters distinguishing the utterances of persons before and after the tonsillectomy treatment.

References

[1] A. Bochenek, M. Reicher, *Human Anatomy, vol. II*, Wydawnictwo Lekarskie PZWL, Warszawa 2006, p. 112, (in Polish).

[2] R. Snell, *IEEE Trans. Speech Audio Process.* **1**, 129 (1993).

[3] P. Zolfaghari, T. Robinson, *Fourth Int. Conf. Spoken Language ICSLP 96*, vol. 2, 1996.

[4] A. Acero, *EUROSPEECH* **99**, (1999).

[5] R. Tadeusiewicz, *Speech Signals*, WKiŁ, Warszawa 1988, (in Polish).

[6] W. Wszołek, *Archiv. Acoust.* **31**, 413 (2006).

[7] W. Jassem, *Fundamentals of acoustic phonetics*, PWN, Warszawa 1973, (in Polish).

[8] M. Konior, M.Kłaczyński, W.Wszołek, *Acta Phys. Pol. A* **119**, 1000 (2011).

[9] V. Makinen, *Front-end Feature Extraction with Mel-scaled Cepstral Coefficients*, Helsinki University of Technology, 2000.

[10] P. Lin, W.J. Gould, T. Fukazawa, et al., *J. Voice* **3**, 81 (1989).

[11] Y. Hori, Y. Koike, G. Ohyama, *Acta Otolaryngol (Stockh.) Suppl.* **523**, 258 (1996).

[12] H. Saida, H. Hirose, *Acta Otolar. (Stockh.) Suppl.* **523**, 239 (1993).

[13] W. Wszołek, M. Kłaczynski, *Acta Phys. Pol. A* **118**, 190 (2010).

[14] T.P. Zieliński, *Digital signal processing*, WKiŁ, Warszawa 2005, (in Polish).

[15] W. Wszołek, *Methods of cognitive categorization for analysis and classification of selected cases of pathological speech*, Rozprawy Monografie vol. 232, Wydawnictwa AGH, Kraków 2011, (in Polish).

[16] J. Sammon, *IEEE Trans. Comp.* **18**, (1969).

[17] W. Sobczak, W. Malina, *Methods of Selection and Reduction of Information*, WNT, Warszawa 1985, (in Polish).