TOWARDS COMPUTER DIAGNOSIS OF LARYNGOPATHIES BASED ON SPEECH SPECTRUM ANALYSIS
A Preliminary Approach

Jan Warchoł
Department of Biophysics, Medical University of Lublin, Poland

Jarosław Szkoła, Krzysztof Pancerz
Institute of Biomedical Informatics, University of Information Technology and Management in Rzeszów, Poland

Keywords: diagnosis support system, laryngopathies, Reinke's edema, laryngeal polyp, spectrum analysis.

Abstract: The main goal of this paper is to give the outline of a preliminary approach to creating a computer tool being a diagnosis support system for laryngopathies. This approach is based on speech spectrum analysis. A simple parameter based on a statistical approach is calculated. Two diseases are considered: Reinke's edema and laryngeal polyp. The paper presents a medical background, basic problems, a proposed procedure for the computer tool, and experiments carried out using this tool.

1 INTRODUCTION - MEDICAL BACKGROUND

A model of speech generation is based on the "source - filter" combination. The source is larynx stimulation, i.e., passive vibration of the vocal folds as a result of an increased subglottis pressure. Such a phenomenon of making speech sonorous in the glottis space is called phonation. The filter is the remaining articulators of the speech canal creating resonance spaces. A signal of larynx stimulation is shaped and modulated in these spaces. A final product of this process is called speech.

Pathological changes appearing in the glottis space entail a bigger or smaller impairment of the phonation functions of the larynx. The subject matter of presented research concerns diseases, which appear on the vocal folds, i.e., they have a direct influence on phonation (Lalvani, 2008).

We are interested in two diseases: Reinke's edema (*Oedema Reinke*) and laryngeal polyp (*Polypus laryngis*).

1.1 Reinke's Edema

Reinke's edema appears often bilaterally and usually asymmetrically on the vocal folds. It is created by transudation in a slotted epithelial space of folds devoid of lymphatic vessels and glands, called the Reinke's space. In the pathogenesis of disease, a big role is played by irritation of the laryngeal mucosa by different factors like smoking, excessive vocal effort, inhalatory toxins or allergens. The main symptoms are the following: hoarseness resulting from disturbance of vocal fold vibration or, in the case of large edemas, inspiratory dyspnea. In the case of Reinke's edemas, conservative therapy is not applied. They are microsurgically removed by decortication with saving the vocal muscle.

1.2 Laryngeal Polyp

Laryngeal polyp is a benign tumor arising as a result of gentle hyperplasia of fibrous tissue in mucous membrane of the vocal folds. In the pathogenesis, a big role is played by factors causing chronic larynx inflammation and irritation of the mucous membranes of the vocal folds: smoking, excessive vocal effort, reflux, etc. The main symptoms are the following: hoarseness, aphonia, cough, tickling in the larynx. In the case of very big polyps, dyspnea may appear. However, not big polyps may be confused with vocal tumors especially when there is a factor of the load of the patient voice. The polyp may be pedunculated or may be placed on the wide base. If it is necessary, polyps are microsurgically removed with saving a free edge of vocal fold and vocal muscle.

Warchoł J., Szkoła J. and Pancerz K. (2010).
TOWARDS COMPUTER DIAGNOSIS OF LARYNGOPATHIES BASED ON SPEECH SPECTRUM ANALYSIS - A Preliminary Approach.
In *Proceedings of the Third International Conference on Bio-inspired Systems and Signal Processing*, pages 464-467
DOI: 10.5220/0002757204640467

2 BASIC PROBLEMS

Research proves that a subjective assessment of voice is always reflected in the basic acoustic parameters of a speech signal. Sound parameters correlating with the anatomical structure and functional features of the voice organ are a subject of interest for researchers. However, the diversity of anatomical forms, inborn phonation habits, and the diversity of an exploratory material cause that researches are performed on different grounds.

A voice generation is conditioned by a lot of factors, which give that voice an individual, peculiar character. However, analysis of individual features of a speech signal in an appropriate group of people, suitably numerous, shows some convergence to values of tested parameters. This enables differentiation of changes of characteristics of the source (larynx stimulation) caused by different pathologies.

Since a colloquial speech is a stochastic process, an exploratory material is made up often by vowels uttered separately with extended articulation. Together with a lack of intonation, it enables eliminating phonation habits.

We can distinguish two types of the acoustic measurement methods: objective and subjective. Both of them belong to indirect exploratory methods. Comparing them to direct methods (e.g., computer roentgenography, stroboscopy, bioelectrical systems) shows that they have several advantages. They are convenient for the patient because a measurement instrument (in this case, a microphone) is located outside the voice organ. This enables free articulation. The advantage of acoustic methods is the possibility of automating measurements by using a computer technique. It is also possible to visualize individual parameters of a speech signal. Subjective auscultatory methods are used, among others, in laryngology and phoniatrics in the case of both correct or pathological voice emission.

Objective methods base on physical features of the voice. They become especially popular, when a computer technique reaches a high extent of specialization. They enable the objective assessment of voice and deliver information in the case of pathology and rehabilitation of the voice organ. Examined parameters aid the doctor assessment of the patient's health state.

In the literature we may notice that parameters of the source (larynx stimulation) are often examined (e.g. (Orlikoff et al., 1997)). However, it is possible to modify an exploratory method so that it encompasses wider range of a material analyzed. A crucial role is played by further mathematical processing of basic acoustic parameters. In this way, we can take into consideration and examine dynamic changes during the phonation process resulting from functions of the speech apparatus as well as from additional acoustic effects occurring in the whole voice organ.

3 PROCEDURE

The proposed approach bases on analysis of distribution of harmonics in the speech spectrum. Clinical experience shows that harmonics in the speech spectrum of a healthy patient are distributed approximately steadily. However larynx diseases may disturb this distribution (cf. (Warchoł, 2006)). Therefore, analysis of a degree of disturbance can support diagnosis of larynx diseases. Disturbance of distribution is expressed by basic parameter *SDA* based on standard deviation. The presented further approach is the first step towards creating a computer diagnosis support system for laryngopathies. The quality of the proposed method is unsatisfactory, but it shows the direction of further research. In our approach we use a basic statistical parameter, which can be replaced by calculations based on computational intelligence methods (Rutkowski, 2008). An important role is played by the quality of speech recording. Quality of results is also dependent on chosen preprocessing methods (e.g. filtration) and signal processing methods (e.g. Fourier transformation). Especially, extraction of the correct signal is a difficult task.

In our approach, we analyze the speech spectrum of a patient. Input data are tuples (f,a), where f is a frequency of the component whereas a is a magnitude of the frequency component. We are interested in peaks and their distribution. An example of the speech spectrum is shown in Figure 1.

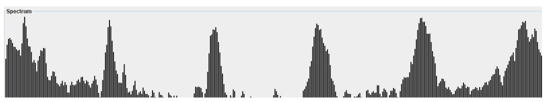

Figure 1: An example of the speech spectrum.

We can describe an algorithm for calculating the *SDA* factor using the following steps:

- **Step 1**: Sort a set $\mathbb{T}$ of tuples in non-decreasing order according to the magnitude. The first tuple has the greatest magnitude.
- **Step 2**: Calculate the average $\overline{a}$ of magnitudes of all frequency components and next remove from the set of tuples $\mathbb{T}$ each tuple with the magnitude less than $1.5 \cdot \overline{a}$. This step is called magnitude filtration.

- **Step 3**: Create an empty array $\mathbb{SDA}$.
- **Step 4**: For each tuple (f,a) from $\mathbb{T}$:
  - if $f < \frac{f'}{3}$, where f' is a frequency of some tuple from $\mathbb{SDA}$ closest to f, then ignore (f,a),
  - otherwise, add (f,a) to $\mathbb{SDA}$.
- **Step 5**: Remove from $\mathbb{SDA}$ tuples with $f < 100$ Hz.
- **Step 6**: Group tuples from $\mathbb{SDA}$ into clusters in the following way. Tuples (f_i,a_i) and (f_j,a_j) belong to the same cluster if a distance between f_i and f_j is less than 3.1 Hz.
- **Step 7**: For each cluster, choose a tuple with the greatest magnitude. After this step we obtain a sequence of frequencies $f_0, f_1, f_2, \ldots, f_k$ of tuples with the greatest magnitudes in clusters, where $f_0 < f_1 < \cdots < f_k$, f_0 is a basic frequency whereas $f_1, \ldots, f_k$ are harmonics of speech signal.
- **Step 8**: Calculate the *SDA* parameter:

$$SDA = \sqrt{\frac{\sum_{i=1}^{k}(F_i - \overline{F})^2}{k(k-1)}},$$

where:

$$F_i = \frac{f_i - f_{i-1}}{f_0},$$

$$\overline{F} = \frac{\sum_{i=1}^{k} f_i}{k}.$$

Division by f_0 enables us to eliminate personal features.

The algorithm given above has been implemented in the computer program prepared by authors of this paper. It is used for speech signal analysis to detect certain anomalies, which suggest a state of sickness of a speech engine. The input file contains information about signal magnitudes of speech for selected frequencies. Based on the correlation between the selected signal frequency spectrum we can determine the status of the patient. The program allows the user to visualize the samples, allowing a review of normalization, magnitude distribution, harmonic analysis.

The construction of the input file is as follows (fragment):

```
Spectrum (power) No.1 [aggreg.degree = 4]
-------------------------------------------
f[Hz] Lev[dB] f[Hz] Lev[dB] f[Hz] Lev[dB]
-------------------------------------------
TOTAL 76.6    1870.6 36.9    3745.6 26.7
19.0  46.7    1894.0 31.1    3769.0 34.9
```

Since the beginning of the file to the word TOTAL, all characters are ignored. From the word TOTAL, all data are entered in columns containing three pairs of frequency - magnitude. The exception is the first row of data as the first parameter TOTAL means the total magnitude of the signal spectrum. The amount of data is not limited.

4 EXPERIMENTS

In experiments, sound samples were analyzed. Experiments were carried out on two groups of patients. The first group included 40 patients (20 men and 20 women) without disturbances of phonation. Most of them was confirmed by phoniatrist' opinion. Patients came from different social groups (students, laborers, office workers). They were classified into four age groups (20 to 30, 31 to 40, 41 to 50, 51 to 60), ten patients in each group. All patients were non-smoking, so they did not have contact with toxic substances which can have an influence on the physiological state of vocal folds. The second group included patient of Otolaryngology Clinic of the Medical University of Lublin in Poland. They had clinically confirmed dysphonia as a result of Reinke's edema (Oedema Reinke) or laryngeal polyp (Polypus laryngis). Information about diseases was received from patient documentations. A group of patients with Reinke's edema included 16 women (at the age of from 41 to 56) and 6 men (at the age of from 43 to 65). A group of patients with laryngeal polyp included 12 women (at the age of from 38 to 72) and 6 men (at the age of from 23 to 62).

Experiments were carried out by a course of breathing exercises with instruction about a way of articulation. A task of all examined patients was to utter separately polish vowels: "A", "I", and "U" with extended articulation as long as possible, without intonation, and each on separate expiration.

Microphone ECM-MS907 (Sony) was used during recording. This is an electret condenser microphone with the directional characteristics. Each sound sample was recorded on MiniDisc MZ-R55 (Sony). In MiniDisc, an analog signal is converted into digital signal according to the CD (Compact Disc) standard (16 bits, 44.1 kHz), and next it is transformed by using the ATRAC (Adaptive Transform Acoustic Coding for MiniDisc) system. It is the compression system proposed by Sony. A data size is reduced in the ratio of 5 to 1. Psychoacoustic effects like audibility threshold and masking quiet sounds by strong neighboring sounds are the basis for compression. Compression system is based on separating harmonics on which a human is most sensitive. Such harmonics are encoded with high precision. However, the less significant harmonics are encoded with the higher compression ratio.

In the case of acoustic experiments, the most precise mapping of speech signal is important. In the ATRAC system the majority of compression is performed for sounds over 5.5 kHz. The voice examination encompasses sounds below 4 kHz. Therefore, the MiniDisc can be used successfully. Effectiveness of such analysis was confirmed by Winholtz and Titze in 1998 (Winholtz and Titze, 1998). They compared recording of speech by using the MiniDisc and recording of speech without compression by using DAT (Digital Audio Type) taking into consideration perturbations and the shape of the acoustic wave. The analysis did not reveal any significant differences.

From the MiniDisc the speech signal was sent to SVAN 912AE analyser (Svantek). Using this tool the fast Fourier transformation (FFT) was performed with the following parameters:

- sampling frequency: 48 kHz,
- 16 quantization bits per sample,
- Hanning Window.

The spectral range was restricted to 5.66 kHz. The frequency resolution was 2,94 Hz.

Selected results are collected in Tables 1, 2, and 3. One can notice that in the case of diseases a value of the SDA parameter is a bit higher.

Table 1: Values of parameter *SDA* for the control group - women.

ID	*SDA*	*ID*	*SDA*
W_1	0.311	W_6	0.205
W_2	0.159	W_7	0.118
W_3	0.167	W_8	0.127
W_4	0.012	W_9	0.008
W_5	0.139	W_{10}	0.129

Table 2: Values of parameter *SDA* for women with laryngeal polyp.

ID	*SDA*	*ID*	*SDA*
W_1^{LP}	0.147	W_6^{LP}	0.219
W_2^{LP}	0.84	W_7^{LP}	0.191
W_3^{LP}	0.2	W_8^{LP}	0.501
W_4^{LP}	0.333	W_9^{LP}	0.138
W_5^{LP}	0.169	W_{10}^{LP}	0.084

5 CONCLUSIONS AND FURTHER WORK

In the paper, a basic approach to diagnosis of larynx diseases has been showed. This approach bases on analysis of distribution of harmonics in the speech spectrum. The quality of the proposed method is unsatisfactory, but it shows the direction of further research. In the performed experiments, a spectrum was calculated in an external device (SVAN 912AE analyser). Generally, this device is used in engineering measurements. We are going to implement own signal processing methods. Moreover, a basic statistical parameter (based on standard deviation) has been calculated. In many cases it is not enough to use only this parameter. Further work will be concentrated on implementation effective preprocessing methods, adequate signal processing methods, and efficient decision support methods based on computational intelligence methodologies (cf. (Rutkowski, 2008)).

Table 3: Values of parameter *SDA* for women with Reinke's edema.

ID	*SDA*	*ID*	*SDA*
W_1^{RE}	0.139	W_6^{RE}	0.216
W_2^{RE}	0.124	W_7^{RE}	0.368
W_3^{RE}	0.187	W_8^{RE}	0.205
W_4^{RE}	0.142	W_9^{RE}	0.175
W_5^{RE}	0.101	W_{10}^{RE}	0.227

REFERENCES

Lalvani, A. (2008). *Current Diagnosis and Treatment in Otolaryngology - Head and Neck Surgery*. McGraw-Hill.

Orlikoff, R., Baken, R., and Kraus, D. (1997). Acoustic and physiologic characteristics of inspiratory phonation. *Journal of the Acoustical Society of America*, 102(3):1838–1845.

Rutkowski, L. (2008). *Computational Intelligence - Methods and Techniques*. Springer-Verlag.

Warchoł, J. (2006). *Speech Examination with Correct and Pathological Phonation Using the SVAN 912AE Analyser (in Polish)*. PhD thesis, Medical University of Lublin.

Winholtz, W. and Titze, I. (1998). Suitability of minidisc (MD) recordings for voice perturbation analysis. *Journal of Voice*, 12(2):138–142.