Proceedings of the Eighth IASTED International Conference
SIGNAL AND IMAGE PROCESSING
August 14–16, 2006, Honolulu, Hawaii, USA
ISBN Hardcopy: 0-88986-583-3 / CD: 0-88986-585-X

PROCESSING AND ANALYSIS OF VOICE ANOMALIES IN COURSE OF PARKINSON'S DISEASES

Izworski A., Augustyniak P., Orzechowski T.

izwa@agh.edu.pl, august@agh.edu.pl, cybertomas@agh.edu.pl

Department of Automatics, AGH University of Science and Technology, Krakow, Poland

ABSTRACT

The results presented here cover beginning of tests concentrated on analysis of voice anomalies induced by Parkinson's Diseases. This work contains an outline of theoretical basics of voice physiology and pathological changes in speech caused by PD and dysarthria. The selection of linguistic material was characterized according to the place and manner of articulation in the phonetic system of Polish. Another section of this work contains the description of the applied methods of voice recording and preliminary voice analysis e.g. continuous sound analysis and changes in the realization of sounds. The phenomena heard in the subjective examination of speech pathologist or neurologist have been borne out with precise objective examination. The parameters obtained allow to parameterize the research results, making complex classification feasible.

KEY WORDS: speech processing, voice analysis, Acoustic Signal Processing, Parkinson disease, dysarthria.

1. INTRODUCTION

The study presented in this publication is the first from the planned complex, interdisciplinary studies. The examination was carried out on patients of CM-UJ clinic in Krakow who suffered from neurodegenerative disease with the damage of the extrapiramidal system with dysarthria-type changes in speech. Control examinations of healthy persons have also been carried out. The elements whose realization was tested had been chosen based on the linguistic knowledge in the scope of phonetics as well as on experience resulting from long-term practice as a speech pathologist. The linguistic material was selected in such a way as to pinpoint voice changes characteristic for patients with PD and dysarthria. During the examination, phrases based on Polish idioms were also recorded for further analyses.

2. VOICE PHYSIOLOGY

Voice and speech production requires close cooperation of numerous organs which from the phoniatric point of view may be divided into organs:
- lungs, bronchi, trachea: (producing expiration air stream necessary for phonation);
- larynx: (amplifying the initial tone);
- root of the tongue, throat, nasal cavity, oral cavity: (forming tone quality and forming speech sounds).

3. VOICE PATHOLOGY

Apart from typical changes caused by neurodegenarative disease (e.g. shivering of the body, limbs, muscle stiffness) changes in the voice may also be observed. The research shown in [4] indicate the serious problem of speech pathology occurrence with as much as 75% of patients. Thus it may be concluded that voice constitutes one of the more crucial components of neurological diagnosis.

Patients suffering from neurodegenerative diseases (and such patients were examined by the authors) show dysarthria-type speech alternations. Dysarthria is a group phonation and articulation disorder which result from damage to the movement control systems of the central or peripheral nervous system also responsible for the speech apparatus. The disorders occur although the speech plan is preserved [3]. Other definitions characterize dysarthria as handicapped production of articulated speech sounds resulting from disturbances to nervous mechanisms of voice production, modulation, intensity, timbre and resonance [2]. Nowadays dysarthria is described as a group of motor speech impairment result from a disruption of muscular control due to lesions of either the central or peripheral, or both, nervous systems (Communication Independence for the Neurologically Impaired CINI – 1994).

Due to the dominating symptom of disorder 6 types of dysarthria have been specified [3]. In our study, patients suffered from hypokinetic and hyperkinetic types. Parkinson disease and Parkinson syndrome (damage to the extrapiramidal system; speech impairment related to slowness) are accompanied by hypokinetic dysarthria-type changes in speech. Its most important characteristics in relation to isolated sounds are: distortions, loudness limitations. Distorted articulation is caused by quick and limited tongue and lips movements, sounds reduced down to slurring. Impairment in the speech process consist in sudden pauses in phonation. The voice is monotonous, quiet, weak and vanishing. The other type of dysarthria occurring in neurodegenerative diseases of the extrapyramidal system is hyperkinetic dysarthria. Phonation is distorted, sudden pauses in speech may occur. Moreover, incorrect articulation occurs as well as irregular breaks in articulation, sound elongation, repetition of sounds caused by abnormal muscular tension. Hypernasality may also occur, and the loss of air caused by throat and palate impairment result in the shortening of phrases. There are variations of speech loudness, the voice is trembling, tense and stifled, weak, with breaks.

4. CHARACTERIZATION AND CLASSIFICATION OF SOUNDS USED IN THE EXAMINATION

During the examination both consonants and vowels were used. Patients were asked to pronounce the sounds in isolation.
The vowel group consisted of [a], [e] and [i]. This particular choice was related to the difference in the elevation of the tongue as well as to the gap between the lips. In our examination we used front vowels.
The consonant group consisted of fricatives [s], [x] and plosives [p], [k], [g].

5. EXAMINATION METHOD

The examinated group consisted of 18 patients between the ages of 20 and 80 and a comparative group of healthy persons with similar age range. Patients suffered from hypokinetic and hyperkinetic types of movement disorders. The voice of the examined patients was recorded with high-quality digital equipment in a soundproof room in order to eliminate any undesirable factors which could negatively affect the results. First, particular sounds were isolated from the recorded voice and then they were processed (filtration and spectrum analysis). Spectrum analysis contains numerous details, thus parameterization was necessary for automatic classification.

5.1. Continuous sound analysis

The patients were asked to pronounce the tested sounds [a], [e], [i], [s] or [x] on one breath. The sound emitted for a long period of time allowed for power analysis.

The signal of each sound was splited up into (40ms overlapping) frames (160ms length).
Power value of each frame is calculated by summing up the values of the signal energy within the respective x(a) and x(a+m) limits of the t frame.

$$P(t) = \sum_{k=a}^{a+m} \left| x(k)^2 \right|, \quad m - frame\, length \qquad (1)$$

The polynomial p(x) of degree 4 that fits the data, p(x(t)) to P(t), in a least squares sense is calculated to represent the average values of P(t).

$$p(x) = p_1 x^n + p_2 x^{n-1} + \ldots + p_n x + p_{n+1}. \qquad (2)$$

The sum of differences between vector P(t) and p(x) can be used as a voice stability parameter.

$$Stab = \sum_{k=1}^{t} \left| P(k) - p(k) \right|. \qquad (3)$$

The values received were compared with the values obtained in the control group. The values of *Stab* parameter were much greater for patients (Stab>0.6) than for persons from the control group.

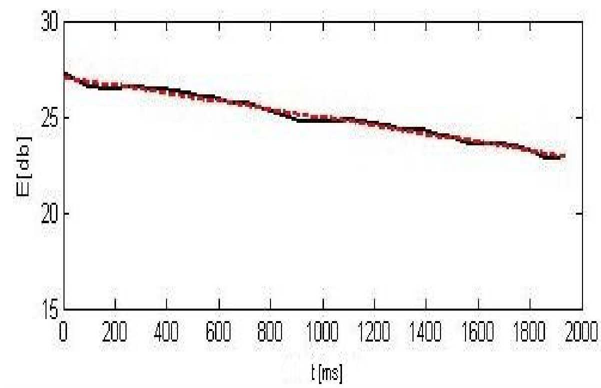

Fig.1a. Example of voice signal [a] from control group, Stab=0.14

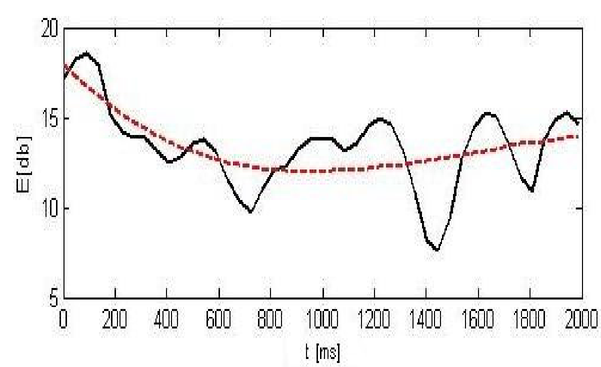

Fig.1b. Example of voice signal [a] from patients group, Stab=1.43

With many patients, distinct and varying breaks in phonation were observed. With healthy persons, gradual quietening took occurred, whereas the patients ended the emission abruptly.

5.2. Changes in sounds realization

Voice signals consist of several waves with different frequencies and amplitudes. The inner ear of humans decomposes the incoming acoustical waves into separate frequencies. Thus, it is appropriate to transform the voice signal into the frequency domain before analyzing it further. This can be achieved using Fourier Transformations.

$$H(f) = \int_{-\infty}^{\infty} h(t) e^{-j2\Pi ft} dx \qquad (4)$$

The audio signal is sampled at a fixed sampling rate 48kHz. Consider a series x(k) with N samples of the form $x_0, x_1, \ldots, x_{N-1}$

$$X(n) = \sum_{k=0}^{N-1} x(k) e^{-jk2\Pi n/N} \; for \;\; n = 0..N-1 \qquad (5)$$

The power spectrum matrix P(*n; t*), where *n* is the index for the frequency and *t* for the time frame:

$$P(n,t) = \left| X_t(n) \right|^2 \frac{1}{N} \qquad (6)$$

The index *n* ranges from 1 to *N/*2 + 1.

Bark Frequency Scale was used instead of Hz. The Bark scale ranges from 1 to 24 Barks, corresponding to the first 24 critical bands of hearing [Hz].
A critical-band value is calculated by summing up the values of the power spectrum within the respective $f_{low}(i)$ and $f_{high}(i)$ frequency limits of the *i* critical-band.

$$CB = \sum_{n \in I(i)} P(n,t)$$
$$I(i) = \{ n : \mathrm{f_{low}}(\mathrm{i}) < \mathrm{f(n)} \le \mathrm{f_{high}}(\mathrm{i}) \} \qquad (7)$$

where *i*, *t*, *n* are indexes, CB is a matrix containing the power within the *i*-th critical band at a specific time interval *t*.

With the patients, changes in sounds articulation are visible (precisely, transition into another sound during realization). It is both audible and detectable through spectrum comparison. These changes were particularly observable for the following consonants, for which the occurring change has also been indicated:

- [k] ➔ [a] / [k] ➔ [y]
- [g] ➔ [y] / [g] ➔ [e]
- [s] ➔ [y]
- [x] ➔ [a]

Each piece of voice is represented by CB matrix. Firstly, the information represented each group of people was combined using median method. The median proved to be the simplest approach with a comparable quality to other more complex methods. Classification was done using simple distance comparation between CB matrixes.

Transition into another sound during realization can be detectable through modulation frequencies analysis. Voice loudness rises and falls in a critical-band several times per second. The sound pressure level in dB and our hearing sensation measured in sone is not linear. The modulation amplitude of the *i*-th critical-band is calculated from the loudness sensation matrix using Fourier Transformation.

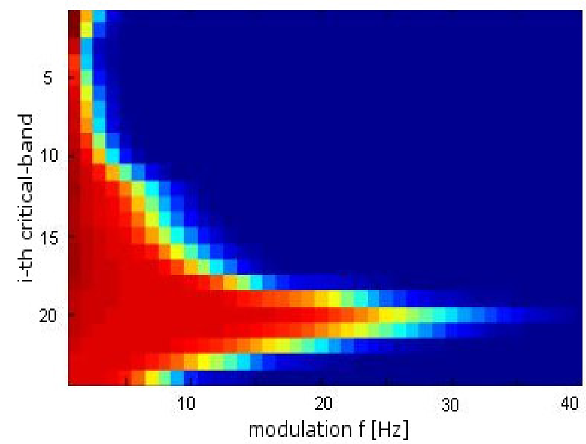

Fig.2a. Modulation amplitude,
example of voice signal [s] without changes

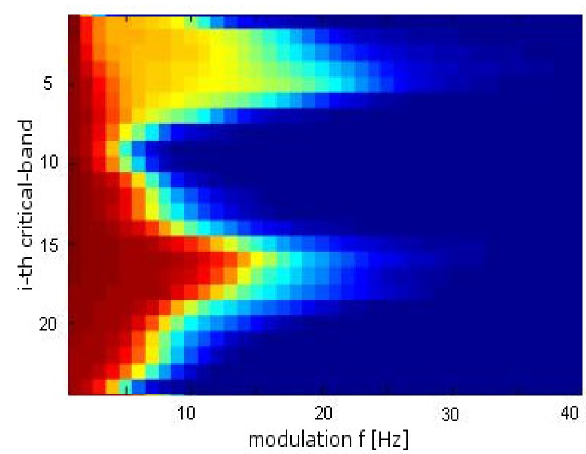

Fig.2b. Modulation amplitude,
example of voice signal [s] with changes [s] ➔ [y]

The result of this distortion is caused by the weakening of the elasticity of the larynx muscles that is why a consonant is followed by a vowel, which does not require as much tension.

6. CONCLUSION

The results presented here constitute the beginning of tests concentrated on automatic voice classification. The authors referred both to the question of duration parameterization as well as voice spectrum parameterization. The main goal set for the future studies is such defining of descriptors, which together with particular search algorithms will enable proper interpretation of a patient’s voice changes. The proposition of recording, processing and analysis of speech as a digital signal is also presented.
Further analysis of the isolated sounds is planned, compared in realization between the patients and healthy persons, taking sex, age and phrase analysis into consideration. With patients the dynamics of disease progression is also registered.

Moreover, some linguistic material on the level of phrases was recorded and a technical analysis is being prepared. In this study, the prosodic elements of speech – rhythm, pace, intonation, accent and melody will be analyzed. The elements mentioned above are available in subjective diagnostic (defining the type of dysarthria). The authors wish to examine the characterization of changes in speech unavailable in subjective examination as well as to create a complex model of automatic classification.

ACKNOWLEDGEMENT

Scientific work supported by the Polish State Committee for Scientific Research resources in years 2004-2007 as a research project No. 3 T11E 00127.

REFERENCES

[1] P. Duus, *Neurologisch-Topische Diagnostik* (Stuttgart, 1983)
[2] D. F. Johns, *Clinical management of neurogenic communicative disorders*, (Boston, College Hill, 1985)
[3] F. Darley, A. Aronson, J. Bron: Cluster of deviant speech dimension in the dysartrias, *Journal of Speech and Hearing Research*, 12, pp. 462-496, 1969.
[4] L. Ramig, S. Sapir, S. Countryman: Intensive voice treatment (LSVT®) for patients with Parkinson's disease: a 2 year follow up, *J Neurol Neurosurg Psychiatry 71*, pp. 439 – 498, 2001.
[5] Tadeusiewicz Ryszard, *Speech signal*, WKiL, Warszawa 1988 (in Polish).
[6] A. Pruszewicz, *Foniatria kliniczna*, PZWL, Warszawa, 1992 (in Polish).
[7] A. Izworski, R. Tadeusiewicz, “Artificial Intelligence Methods in Diagnostics of the Pathological Speech Signals”, *Speech and Language Technology*, vol. 6, PPA, Poznań, pp. 183 – 197, 2002.
[8] A. Ralston, *A first course in numerical analysis*, New York, 1965.
[9] T. Körner, *Fourier Analysis*, England, Cambridge University Press, 1988.
[10] D. C. Hanselman, B. L. Littlefield, *Mastering MATLAB 7*, 2004.