speech therapy, stutter, logopaedics, signal analysis,

Elżbieta SZABELSKA, Anna KRUCZYŃSKA[*]

COMPUTER-BASED SPEECH ANALYSIS IN STUTTER

Abstract

The paper presents computer aided approach to speech disorder analysis. The three levels definition of stutter is presented for describing this particular problem. The effective speech analysis algorithm used for segmentation of signal and detecting stutter is discussed and samples are presented.

1. INTRODUCTION

Speech therapy is a science incorporated in many other fields of science. Its interdisciplinary character allows it to reach for modern analysis and description methods. At present, speech therapy diagnosis can be supported by computer software used for both normal and pathological speech analysis. Owing to cutting edge technologies, the diagnosis becomes wider and more accurate, while the therapy is faster and more efficient than with the application of traditional linguistic and mechanical methods. Such techniques find application in virtually all types of speech disorder, especially in stutter, autistic patients or children suffering from alalia (speech delay). Modern computer technology allows programming therapy for persons unable of using active speech through adjusting one of the nonverbal communication channels. Additionally, computer programmes are widely used in biomedical applications [2]. They are extremely useful in diagnosis and classification of voice pathology, allow voice quality monitoring during rehabilitation (e.g. following larynx cancer removal), research on the influence of hearing injuries on speech quality as well as support treatment of speech organs injuries.

Modern electronic multimedia systems can be useful in speech signal analysis and processing in development norms. Specialised computer software allows identifying characteristic speech parameters and accurate voice synthesis [2].

[*] PhD students, Maria Curie-Skłodowska University, Zakład Logopedii i Językoznawstwa Stosowanego, ul. Sowińskiego 17, 20-040 Lublin, Poland, e-mail: ela.szabelska@wp.pl

This is achieved thanks to mathematical description of speech signal (as any other acoustic wave).

This article presents the issue of computer-based speech analysis of stuttering patients, as the support of computer technologies frequently allows full recovery.

2. STUTTER

Stuttering has earned a number of definitions, which present a picture of speech fluency, pace and rhythm disorder caused by discoordination of respiratory, phonetic and articulatory muscles (with possible excessive tension), which may be accompanied by anxiety [6]. This disorder can be extremely stigmatising and may cause difficulties in interpersonal communication as well as functioning in a society. This is caused by the fact that stuttering definition is threefold: linguistic (lack of speech fluency), psychological (logophobia – fear of speaking) and neurophysiological (spasticity) [7].

Nevertheless, the disorder varies from patient to patient in terms of the set of symptoms as well as their intensity [3]. Stutter is a persistent and complicated disorder with continuously unspecified etiology. Neuromuscular disorders seem to be most frequently named as the reason of stutter among the abundance of hypotheses concerning its causes and mechanisms [7]. The fundamental elements of neuromuscular disorders are minimal delays and pauses in coordination of articulatory movements. This, in turn, may cause prolongations and automatic repetitions of parts of words. Additionally, it is commonly believed that stuttering may result from a combination of multiple factors of biological, psychological or even social nature [3].

According to the three levels of stutter description, its definition comprises three different types of symptoms [7]. First of all, it is manifested on the linguistic level as a disrupted flow of speech, as irregular rhythm, frequent sound prolongation, hesitation or pausing before speech as well as pronouncing additional sounds. The psychological sphere encompasses frustration and fear, in particular fear related to the act of speaking, reluctance to speak, the sense of guilt and low self-esteem. The third sphere – neurophysiological – encompasses spastic muscle movements as well as other physiological reactions (including sinkinesis).

To recapitulate, a pause in a standard rhythm and pace of speech, during which the person knows exactly what they intend to say, but, at that moment is entirely unable of articulating their thoughts due to involuntary repetition, elongation or pause of a sound is referred to as stutter [3].

3. STUTTER ANALYSIS METHODS

Speech analysis, both correct and pathological, involves a precise acoustic description, which can be achieved thanks to computer software devised especially for this purpose.

Voice is an individual feature, typical of a specific person. It can be characterised by a specific height of the basic pitch and variable frequency of subsequent formants. The frequencies of particular formants are closely related to timbre, which is also determined by the structure and arrangement of resonators (larynx, pharynx, oral and nasal cavity, paranasal sinuses, thorax) [1].

To date, speech analysis based on merely subjective assessment of speech quality. Rapid technological development has enabled recording speech and subjecting it to objective assessment based on the analysis of physical parameters of particular sounds [1].

In the case of stuttering patients articulatory changes, resulting from muscle spasms, are noticeable. These may be observed in spectral and cepstral analysis results. The speech disorder analysis is conducted by speech signal segmentation and parameterisation of obtained speech sample sequences. The parameters which are subjected to analysis are: the frequency of laryngeal tone (fundamental frequency) and formants frequencies and amplitudes [3].

In speech assessment the disordered speech is characterised by the fluctuation apparent in such format parameters as amplitude and frequency. In normal speech such fluctuations are considerably lower. Loss of formants or emergence of additional formants in frequencies unrelated with the uttered vowel may be observed in disordered speech. The aforementioned stem from uncontrollable voice box muscles spasms [3].

The development of modern methods of stutter detection provides higher precision in recording particular occurrences of disorder, to date based on methods of subjective assessment of speech quality. Regardless of the type of stutter (single sound, syllable or word repetition), special algorithms employing correlation analysis are applied in case of repetition detection (as in Fig. 1). This method of speech analysis automatically detects the place where the repetition occurred by indicating maximum value of the autocorrelation function. The procedure begins with the parameterisation of the stuttering patient's speech by the segmentation of the signal and calculation of the energy of each of the marked segments according to the formula:

$$E = \frac{1}{l}\sum_{i=1}^{l} p_i^2 \tag{1}$$

where:

E – segment energy,

l – the number of samples in a given segment,

p_i – normalised value of the subsequent sample.

The subsequent stages in repetition detection comprise: a) correlation analysis of particular parameters and b) control of exceeding the threshold value of the autocorrelation function. Research conducted at Gdańsk University of Technology show that the effectiveness of automatic detection of repetition amounts to 77% [3].

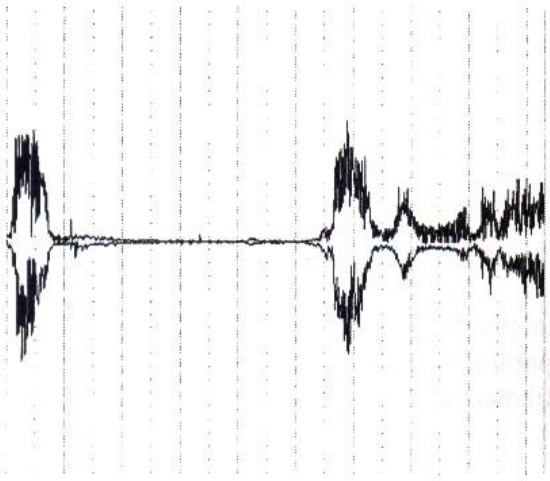

Fig. 1. 'Ad_admiral' sequence with the repetition of the anlaut syllable [3]

Acoustic analysis of speech of stutterers has developed methods of phonation break detection. The detection method of these speech phenomena in stutterers includes: speech signal segmentation, equalising energy levels of higher-frequency components and calculating average signal energy (according to the formula, as in repetition). The programme then calculates upper and lower value of the energy state for each component. These values allow determining the trailing edge preceding phonation break. The trailing edge describes energy decrease rate. It has emerged that phenomena of such type are characteristic of patients with speech disfluency and occur as a consequence of speech muscles spasms, stopping articulation of subsequent sounds of a given speech signal. In stuttering patients, the decrease rate of recorded trailing edge of the energy function is significant, whereas in non-stutterers the envelope of energy function preceding phonation break decreases gently [3].

Protracted articulation, particularly connected with vowels, constitutes another feature distinctly characterising the speech of stutterers. The analysis of elongation of vowels is considerably uncomplicated as the time segment when particular parameters change is extensive and, therefore, enabling high precision measurement.

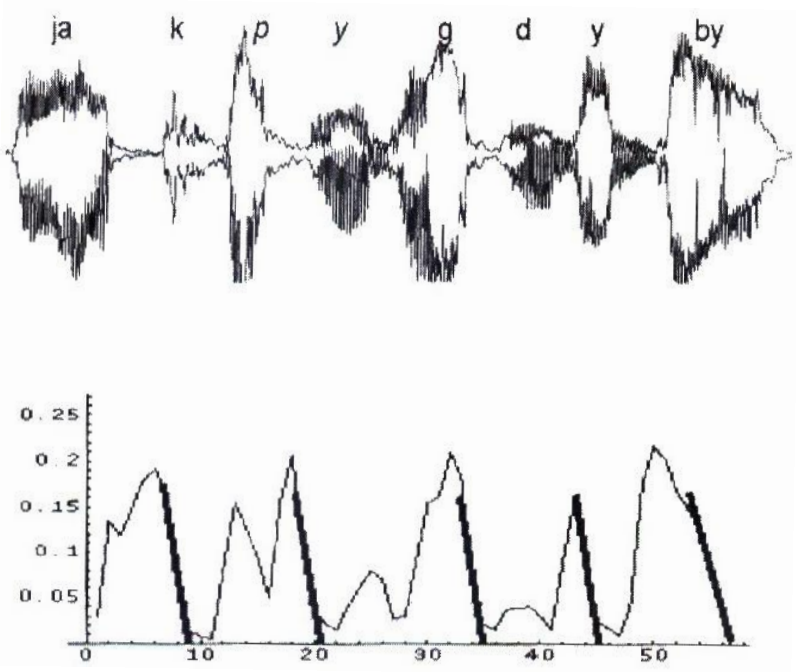

Fig. 2. The rate of decrease of the trailing edge preceding phonation breaks [3]

From the viewpoint of acoustic analysis, these are speech disorders involving the repetition of syllables or vowels which prove more demanding. The prevailing reason for the inconvenience of such measurements is that the articulation time of voiced phonemes is considerably shorter, which results in a significantly smaller amount of information obtained from the analysis of a single recording of the disorder [3].

4. WAVESURFER – SPEECH SOUNDS ANALYSIS SOFTWARE

WaveSurfer is one of solutions for conducting research on speech samples. It is an Open Source tool (such as MySQL or OpenOffice) which enables sound visualisation (mainly speech analysis) as well as its further manipulation and processing. Owing to uncomplicated interface, both beginners and experienced

users will easily find and apply desired functions. WaveSurfer can be used as a stand-alone tool or become an element of advanced sound processing platform, which can be achieved by additional installation of custom plugins or by embedding its components in other applications. The most recent version of the programme, i.e. WaveSurfer 1.8.8, can be downloaded free of charge from WaveSurfer home page: http://sourceforge.net/projects/wavesurfer

WaveSurfer 1.8.8 features:
– type: Freeware
– size: 1.58 MB
– OS: w98 WNT w2000 wxp
– software language: English

WaveSurfer is equipped in numerous features for sound processing, including the following:
– Waveform – a graph representing the shape and form of an analysed sound wave in time.
– Spectrogram – is a graph representing the amplitude spectrum in time. The horizontal axis represents time and vertical – frequency (default in kHz).
– Intensity (the amount of energy in spectrum) is represented by the degree of saturation in given frequencies.
– Formants – lines representing the course of particular formants. Each formant is marked with a different colour.
– Time – a time bar which allows keeping track of sound time.
– WaveBar – a bar for scrolling across the recording – it facilitates sampling.

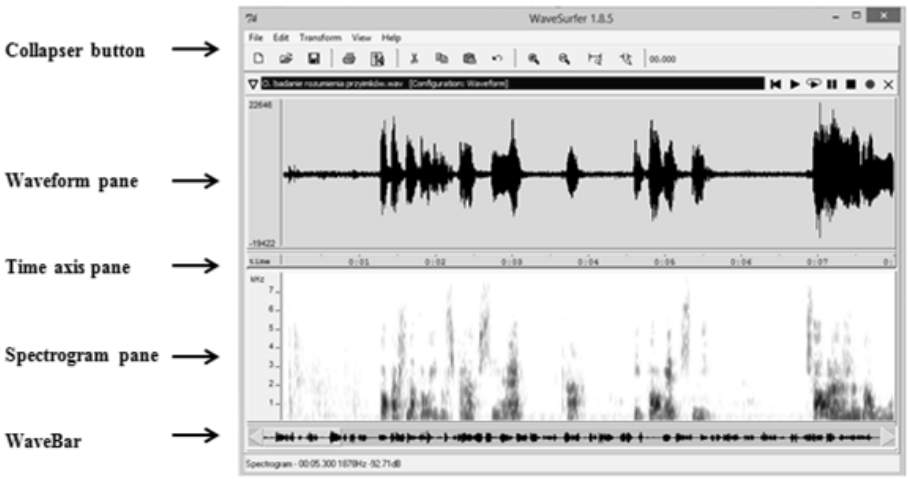

Fig. 3. WaveSurfer user interface [source: own study]

One of numerous features of WaveForms tool is creation of spectrograms, using which a speech therapist, or any other researcher dealing with speech analysis, is presented with a precise record of particular sounds in articulation and can therefore evaluate it and compare with model patterns. The image may prove valuable not only for the therapist but for the patient as well, for whom it can provide the source of feedback, considerably helpful in auto-correction (a system employing auditory feedback in stutter therapy).

The spectrogram record is a two-dimensional representation of the 3D reality. The horizontal axis comprises values referring to time (in milliseconds) while the vertical axis – spectrum elements values connected with frequency analysis (in Hz). The third dimension is visualised by the degree of saturation of image in black or colour. This dimension refers to the intensity of persistence of an analysed element in a given frequency [4].

To present differences in spectrograms in the case of stutter, two utterances are compared – fluent and non-fluent.

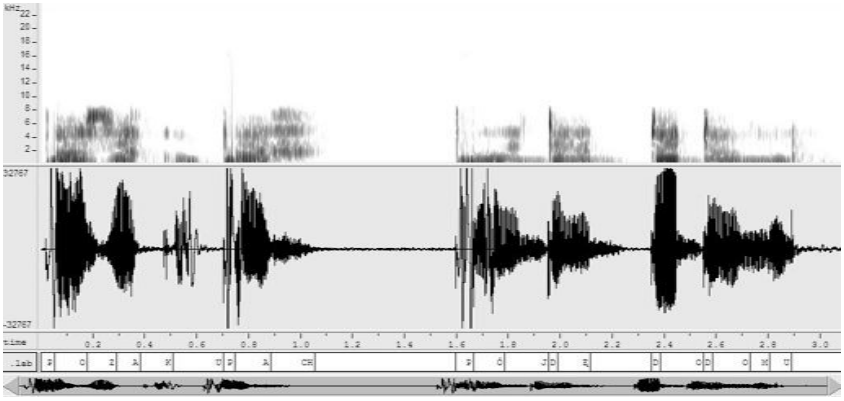

Fig. 4. Fluent utterance: "Po zakupach pójdę do domu" (I'll go home after shopping) [source: own study]

In non-fluent utterance the wave and its spectrogram distinctly show numerous repetitions of 'p' sound as well as 'po' and 'do' syllables, both typical in repetition of the anlaut sound or syllable stutter. Moreover, by increased saturation the spectrogram shows the amount of energy present in repetitions.

What is more, WaveSufrfer software allows the user to cut and manipulate the material as required. The prepared samples can be subsequently compared. In the case in question, the comparative analysis could, for instance, delve into the differences between the correct and non-fluent utterances of 'po' syllable. Another method of examination could consist in comparing the syllables in the course of repetition of non-fluency – whether their times, volume or formants are identical or different, even though the samples have been obtained from the same patient.

Both utterances differ also in total time of utterance as well. Fluently uttered "Po zakupach pójdę do domu" lasted 2.9 s, whereas in the case of stutter, the time was naturally longer, and equal to 5.9 s. The direct comparison of the preceding figures fails to represent the time span of both utterances in an ideal way (different time scale applied), which leads to the necessity to visualise full sentences.

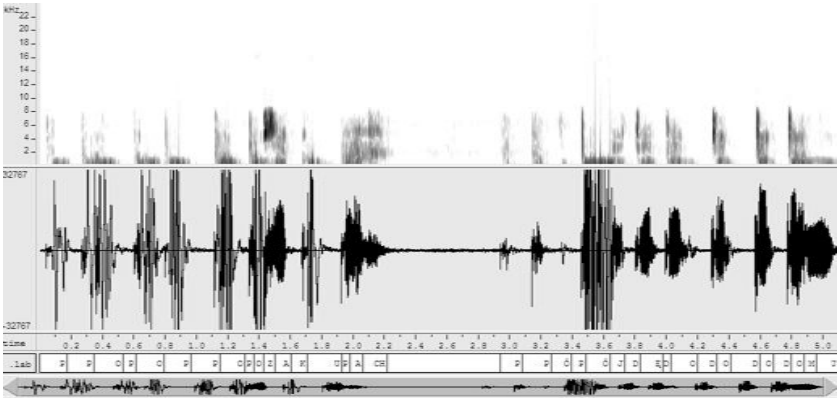

Fig. 5. Non-fluent utterance: "Po zakupach pójdę do domu." (I'll go home after shopping) [source: own study]

As it was presented, WaveSurfer is a programme combining the simplicity of operation with a number of useful and extended functions. What it provides is genuinely in-depth analysis of data, i.e. speech sample, in the area of a researcher's interest. What must not be forgotten is the fact that the programme is freeware, which adds to its unquestionable advantages.

5. CONCLUSION

Computer-aided speech analysis plays an important role in modern-day logopaedics. The application of cutting edge methods of speech analysis enables the therapist to provide the patient with more precise diagnosis, and it, moreover, facilitates speech therapy process. Modern software produces repeatable results which can be represented by numerical or graphical data, suitable for comparative studies and, therefore, is applicable in different fields of science. Computer acoustic speech analysis of stutterers enables notable and objective evaluation of changes in the speech signal structure, and increasingly accurate tests provide the insight into mechanisms accompanying the stutter phenomenon.

REFERENCES

[1] BUDREWICZ S.: *Akustyczna analiza zaburzeń mowy w chorobie Parkinsona z uwzględnieniem wpływu stymulacji dopaminergicznej na wybrane parametry mowy*. Uniwersytet Medyczny im. Piastów Śląskich, Wrocław, 2012.

[2] CIOTA Z.: *Metody przetwarzania sygnałów akustycznych w komputerowej analizie mowy*. Akademicka Oficyna Wydawnicza EXIT, Warszawa, 2010.

[3] CZYŻEWSKI A., KOSTEK B., SKARŻYŃSKI H.: *Technika komputerowa w audiologii, foniatrii i logopedii*. Akademicka Oficyna Wydawnicza EXIT, Warszawa, 2002.

[4] GONET W.: *Próba określenia normy wymowy polskich samogłosek ustnych*. Opuscula Logopedica, Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej, Lublin, 1993.

[5] KUNISZYK-JÓŹKOWIAK W.: *Akustyczna analiza i stymulacja płynności mówienia*. Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej, Lublin, 1996.

[6] TARKOWSKI Z.: *Jąkanie*. Wydawnictwo Naukowe PWN, Warszawa, 2002.

[7] WOŹNIAK T.: *Pojęcie jąkania a metodyka postępowania logopedycznego*. Opuscula Logopedica, Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej, Lublin, 1993.