Automatic recognition of non-fluent stops[+]

Waldemar Suszyński[a*], Wiesława Kuniszyk-Jóźkowiak[a],
Elżbieta Smołka[a], Mariusz Dzieńkowski[b]

[a]*Institute of Physics, Maria Curie-Skłodowska University,*
Pl. Marii Curie-Skłodowskiej 1, 20-031 Lublin, Poland
[b]*Management Department, Lublin University of Technology,*
Nadbystrzycka 38, 20-618 Lublin, Poland

Abstract

The presented article constitutes a fragment of research which aims at automatic recognition of disfluencies in the speech of stuttering people. In the utterances of such people difficulties frequently appear in words beginning with stop consonants. On the basis of acoustical analyses of the episodes, procedures for their automatic detection have been elaborated. The procedures apply fuzzy logic.

1. Introduction

In logopaedic practice there is great need for computer systems supporting both therapy and diagnosis. At present, in practice the only diagnostic tool is the logopaedist's ear. Very seldom in logopaedic clinic current recordings of therapy subjects' utterances are made and on their basis corrective measures are undertaken. This results from lack of simple tools for the analysis of such records. Stuttering is the most complex of speech disorders. The judgement of this disorder severity as well as therapy requires determining the frequency with which particular types of disfluencies occur (the disfluency types being: sound prolongations, sound and syllable repetitions, blocks, interjections) and measurement of their duration times. What is significant in the diagnosis is also determination of acoustic and linguistic conditions of disfluency occurrences in particular people. It appears that, for many people, especially adults, who struggle with the disorder, there exist certain words in which difficulties in uttering quite often occur. There is also a tendency to more frequent non-fluent realisation of words beginning with certain sounds. It concerns fricative, nasal

---

[+] After second revision.
[*] Corresponding author: *e-mail address*: waldemar.suszynski@umcs.lublin.pl

and stop sounds. Disfluencies when uttering these sounds often result from uncontrolled, long-lasting contractions of articulatory muscles – so called blockades and they constitute the most acute form of stuttering. In the case of fricatives and nasals the result of such a contraction is prolongation of the sound. The procedures for automatic detection of these episodes have been described earlier by the authors of the present article [1-3]. During the articulation of stops the disfluencies are multiple short impulse articulations, divided with, often very long-lasting, periods of so-called tenses pause or interpolated sound. These features repeat in utterances of various people and they may constitute the acoustical basis for recognition of the type of disorder. In the case of such a complex disorder, however, there can be no set clear-cut borders between measurable parameters, and that is why the authors have applied fuzzy logic [4,5]. To give an example, one of the distinctive features of the type of disfluency is the duration time of the speech fragment, separated with pauses. If it is "short", there is a probability of a non-fluent stop. However, no clear-cut border can be determined which would allow us to classify a given fragment into the "short" set. The situation is similar as far as other parameters are concerned.

Attempts at automatic recognition of speech disfluencies in recent years have been only made in two centres in the world. They concerned application of neural networks [6-8] and rough sets [9] in the analysis of non-fluent speech.

2. Speech signal processing

Microphone speech signals of the examined people were transformed into digital ones with the use of a Sound Blaster card and they were recorded directly on a computer disk. 20050 Hz sampling frequency and 16-bit amplitude quantization were used. Sound files were analysed with the use of 21 digital 1/3-octave filters in the range between 100 and 10000 Hz and A-weighting filter. A logarithmic scale of the amplitude was used. This processing was used as approximation to characteristic of the human hearing. The auditory system behaves like bank of a band filters. The bandwidths of the auditory filters were different estimated [10]. A change of the 1/3-octave scale in the described procedure was a result of comparison of the Bark scale [10], 1/3-octave scale and measurements obtained by Moor et. el. [12] at low centre frequencies. The subjective loudness we attribute to a sound varies not only with intensity level but also with frequency. The A-weighting filter approximates the contribution of sounds of different frequencies so that the response of the average human ear is simulated.

The mean sound level was calculated as the average value in the all filters for each time moment.

3. Fuzzyfication

On the basis of the average level value the following parameters were calculated: speech fragment times (S), pause times before speech fragments (B) and pause times after speech fragment (A).

These values were measured from spectrograms in non-fluent and fluent realisations of stop consonants in utterances of 12 stuttering people. Fig. 1 presents the histogram of duration times of non-fluent realisations of a stop consonant.
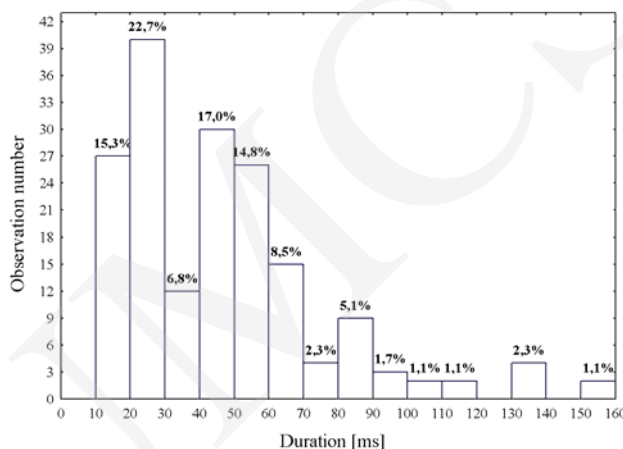

Fig. 1. Histogram of non-fluent stop times

Fig. 2 presents the histogram of duration times of pauses after a non-fluent realisation of a stop consonant.
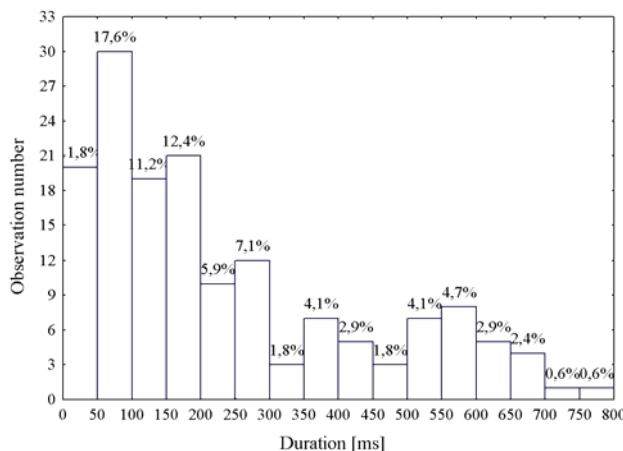

Fig. 2. Histogram of pause times after non-fluent stops

For comparison, Fig. 3 shows the histogram of pause times after fluent realisation of a stop consonant.
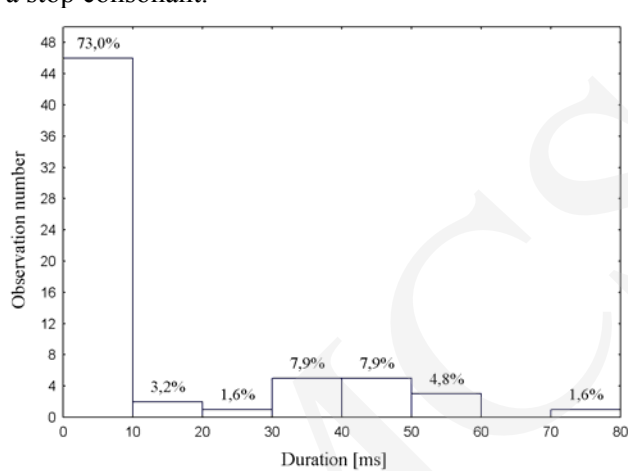

Fig. 3. Histogram of pause times after fluent stops

The membership functions were determined in the "short", "medium" and "long" sets (Figs. 4 and 5) as well as principles allowing one to ascribe a given speech fragment to the following sets: non-fluent stop, another non-fluent sound and fluent sound. What is also important apart from the three enumerated values is the vicinity, i.e. the membership of the preceding or following fragment in the set of non-fluent stops.
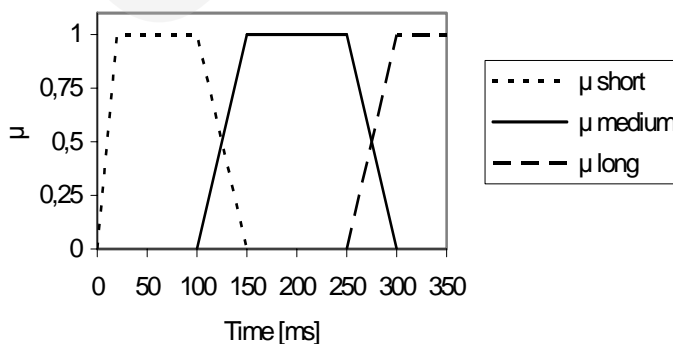

Fig. 4. The membership functions in fuzzy sets: "short", "medium"
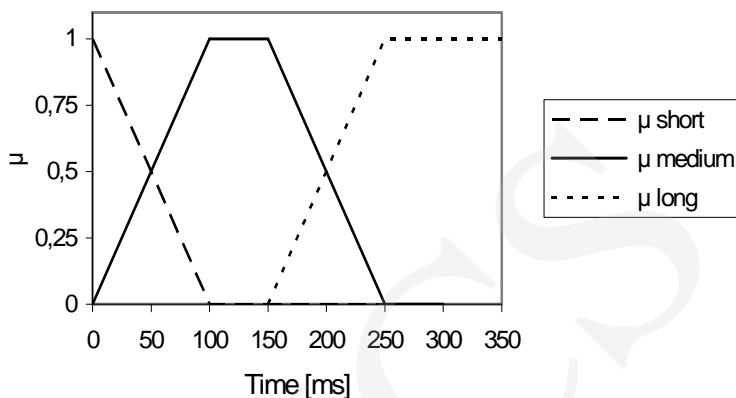and "long" speech fragment times (S)

Fig. 5. The membership functions in fuzzy sets: "short", "medium" and "long"
pause times before (B) and after (A) speech fragments

Fig. 6 presents an example spectrogram of the utterance "życzył b b b b bym sobie" of a stuttering person (a), the course of sound level averaged in 1/3-octave bands (b) and the non-fluent realisations of a stop consonant found by the computer (c).

Fig. 6. An example spectrogram of the utterance "życzył b b b b bym sobie" of a stuttering person
(a), the course of sound level averaged in 1/3-octave bands (b) and the non-fluent realisations
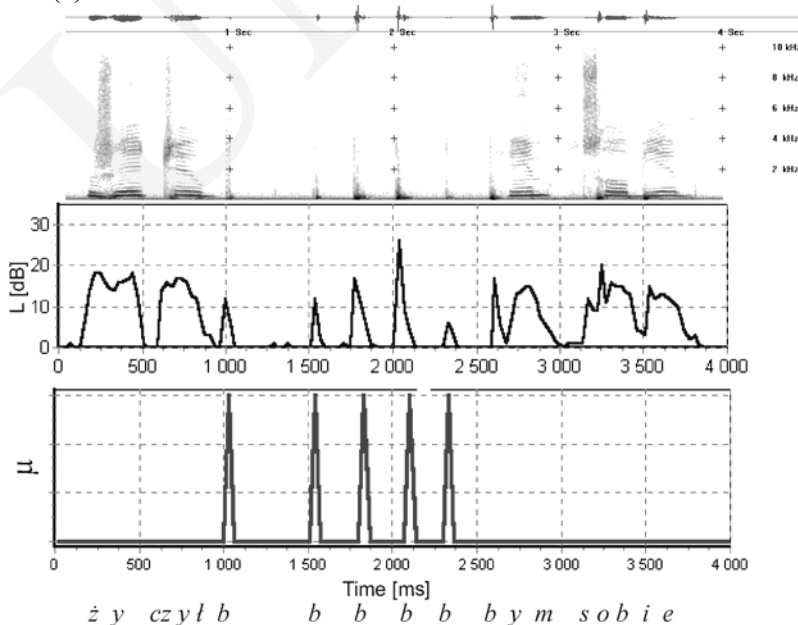of a stop consonant found by the computer (c)

4. Fuzzy rules

The speech fragments were classified as nonfluent stops, nonfluent other sounds and fluent on the basis of the following rules:

1. If *S short* and *A long* and *B long* then *nonfluent stop*.
2. If *S short* and *A medium* and *B medium* and *vicinity nonfluent* then *nonfluent stop*.
3. If *S short* and *A long* and *B medium* then nonfluent stop.
4. If *S short* and *A medium* and *B long* then nonfluent stop.
5. If *S medium* and *A long* and *B long* then *nonfluent sound*.
6. If *S medium* and *A medium* and *B long* and *vicinity nonfluent* then *nonfluent sound*.
7. If *S medium* and *A long* and *B medium* and *vicinity nonfluent* then nonfluent sond.
8. If *S medium* and *A medium* and *B medium* and *vicinity nonfluent* then nonfluent sond.
9. *If S long then fluent speech or prolongation.*
10. If S *short and A short or B short then fluent sound*

5. Verification

The procedure has been verified on the basis of 150 utterances lasting 4 sec each by 10 stuttering people, half of which contained non-fluent stops and half were fluent. The identification exactness of non-fluent fragments amounts to approximately 95%. In the fluent utterances there were no error disfluency identifications.

6. Conclusions

The described computer procedure for automatic recognition of disfluencies related to the articulation of stop consonants constitutes, as mentioned at the beginning, a fragment of the stuttering episode detection programme which is being elaborated by the authors. The authors' aim is to be able to recognise disfluencies independently of the speaker and type of utterance. Variety of disfluencies in particular people, individual and context features make fuzzy logic the most adequate description of the acoustical parameters of the episodes.

The authors also plan to incorporate the autocorrelation function with adaptive and fuzzy windows to the recognition of repetitions and other fluency distortions described in the paper.

Acknowledgements

References

[1] Suszyński W., Kuniszyk-Jóźkowiak W., Smołka E., Dzieńkowski M., *Prolongation detection with aplication of fuzzy logic,* Annales Informatica UMCS, Lublin, 1 (2003).

[2] Suszyński W., Kuniszyk-Jóźkowiak W., Smołka E., Dzieńkowski M., *Automatic recognition of Nasals Prolongations in the Speech of person Who Stutter*, Structures-Waves – Human Health, Kraków, XII(2) (2003).

[3] Suszyński W., *Automatyczne rozpoznawanie niepłynności mowy*, 50 Otwarte Seminarium z Akustyki, Szczyrk-Gliwice, (2003), in Polish.

[4] Zadeh LA., Fu KS., Tanaka K., Shimura M, *Fuzzy sets and their applications to cognitive and decision processes,* New York: Academic Press, (1975).

[5] Łachwa A., *Fuzzy world of sets, numbers, relationships, facts, rules and decisions,* Academic Publishing Company EXIT, Warsaw, (2001), in Polish.

[6] Howell P., Sackin S., Glenn K., *Development of two stage procedure for the automatic recognition of dysfluencies in the speech of children who stutter: II ANN recognition of repetitions and prolongations with supplied word segment markers,* Journal of speech, Hearing and Language Research, 40 (1997a).

[7] Howell P., Sackin S., Glenn K., *Development of two-stage procedure for the automatic recognition of dysfluencies in the speech of children who stutter: I Psychometric procedures appropriate for selection of training material for lexical dysfluency classifiers,* Journal of Speech, Hearing and Language Research, 40 (1997b).

[8] Howell P., Sackin S., Glenn K., Au-Yeung, J., *Automatic stuttering frequency counts,* In Speech Motor Production and Fluency Disorders, edited by H. Peters, W. Hulstijn, and P. van Lieshout, Elsevier, Amsterdam, (1997b).

[9] Czyżewski A., Kaczmarek, A., Kostek, B., *Intelligent processing of stuttered speech*. *Journal of Intelligent Information Systems*, 21(2) (2003).

[10] Gold B., Morgan N., *Speech and audio signal processing*. John Wiley & Sons, Inc. New York, (1999).

[11] Zwicker E., *Subdivision of the audible frequency range into critical bands (Frequenzgruppen). The Journal of the Acoustical Society of America*, 33 (1961).

[12] Moore B.C.J., Peters, R.W., Glasberg, B.R., *Auditory filter shapes at low center frequencies*. The Journal of the Acoustical Society of America, 88(1) (1990).