Annales UMCS Informatica AI 1 (2003) 133-140

Annales UMCS
Informatica
Lublin-Polonia
Sectio AI

http://www.annales.umcs.lublin.pl/

Prolongation detection with application of fuzzy logic

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]*Department of Computer Science, Management Department, Technical University of Lublin, Nadbystrzycka 38, 20–618 Lublin, Poland*

Abstract

The article presents the method elaborated by the authors for automatic prolongation detection in utterances by the people who stutter. Fricative and nasals consonants were focused upon, as they are the most frequently prolonged ones. Fuzziness was applied in the scales of time, frequency and level of distinctive features of a given manner of articulation. The presented method was verified in continuous speech. It characterizes with almost 90% effectiveness of recognition and high precision of duration measurements of non-fluent episodes.

1. Introduction

The application of fuzzy logic in speech recognition results directly from the processes of its production and perception. Characteristics of particular utterance components frequently depend on a great number of individual features of phonic string creation related to the construction and functioning of speech organs as well as language structure, precision, expressiveness and articulation rate which are characteristic of a particular speaker. What also adds to the fuzziness of the features is a vast variety of language forms which create segment compositions appearing with varying probability. Many systems of automatic recognition treat the process of natural perception as their basis and they attempt to mirror it closely. Perception is a complex mechanism related to brain activity, only partly studied, thus in that case automatic recognition is linked with a widening range of uncertainty. In the case of speech pathology, and especially stuttering, other unpredictable features are added, due to which there is an ambiguity in the disturbance assessments by particular listeners [1-3].

* Corresponding author: *e-mail address:* waldemar.suszynski@umcs.lublin.pl.

The authors of the article have elaborated a method of automatic prolongation detection in the speech by people who stutter in which fuzzy logic is applied [4-6]. Sound prolongations appear frequently in the speech of people with a very advanced form of stuttering. Their detection is significant in diagnosing this fluency disturbance. As the results of the research show, the most frequently prolonged sounds are fricative and nasal, less frequently vowels, glides and liquids. In some stutterers there has observed a distinct tendency for this type of disturbances for sounds of a particular articulation manner.

Determination of the fluency disturbance level is very significant for diagnosing, forecasting and therapy, and the detection and duration measurements of stuttering episodes are of great importance in a logopaedist's work. However, there are not many studies aiming at automation of speech assessment of people who stutter. The studies are carried out by Howell and colleagues [7-8] and by our team. Howell applies neural networks in his research, and the authors of this article employ fuzzy logic.

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. The created wave files were subjected to the FFT analysis with the Hamming window in the time intervals of 20 ms and the level values were obtained in the logarithmic scale in 1/3 octave frequency band with additional correction using of A filter.

3. Feature extraction

The features useful for prolongation recognition of sounds of a certain articulation manner were isolated on the basis of systematic analyses of the spectrograms of non-fluent words and their fluent counterparts. The features were: 1) the band where the sound peak location was, 2) the range of width alteration of the band containing the peak, 3) the range of alteration of the average sound level and 4) time, in which the peak was located in a given band and changed within a certain range. To illustrate this, in Figure 1 non-fluent articulation is presented of *zzzjego* with prolonged *z* fricative sound at the beginning while in Figure 2 the articulation of *jjjesteś* with *jjj* prolonged is shown.

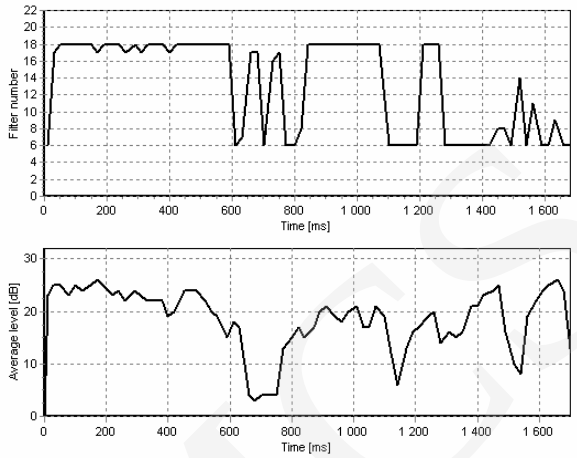

Fig. 1. Spectrum peak location and alteration of average sound level in the articulation of “z ɲego” with the prolonged “z” sound at the beginning

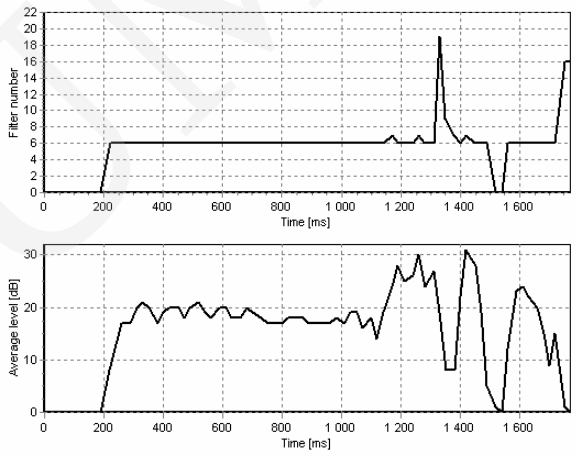

Fig. 2. Spectrum peak location and alteration of average sound level in the articulation of “ɲesteti” with the prolonged “ɲ” sound at the beginning

4. Fuzzy sets

In the elaborated detection method it was assumed that the areas of fuzziness may concern all the fields of speech signal alteration: frequency, level and duration. In the first one the area to be considered was the range between 100 and 10,000 Hz, determined by twenty-one 1/3 octave filters (1, 21).

The area was divided into two parts, described as *high* and *low* (Fig. 3a). The membership functions to proper fuzzy sets were described as follows:

$$m_{LOW}(i)=\begin{cases}1 & for\ i\le 6\\ \dfrac{16-i}{10} & for\ \ 6<i\le 16,\\ 0 & for\ i>16\end{cases} \tag{1}$$

$$m_{HIGH}(i)=\begin{cases}0 & for\ i\le 6\\ \dfrac{i-6}{10} & for\ \ 6<i\le 16,\\ 1 & for\ i>16\end{cases} \tag{2}$$

where *i* is the filter number in which the peak was located.

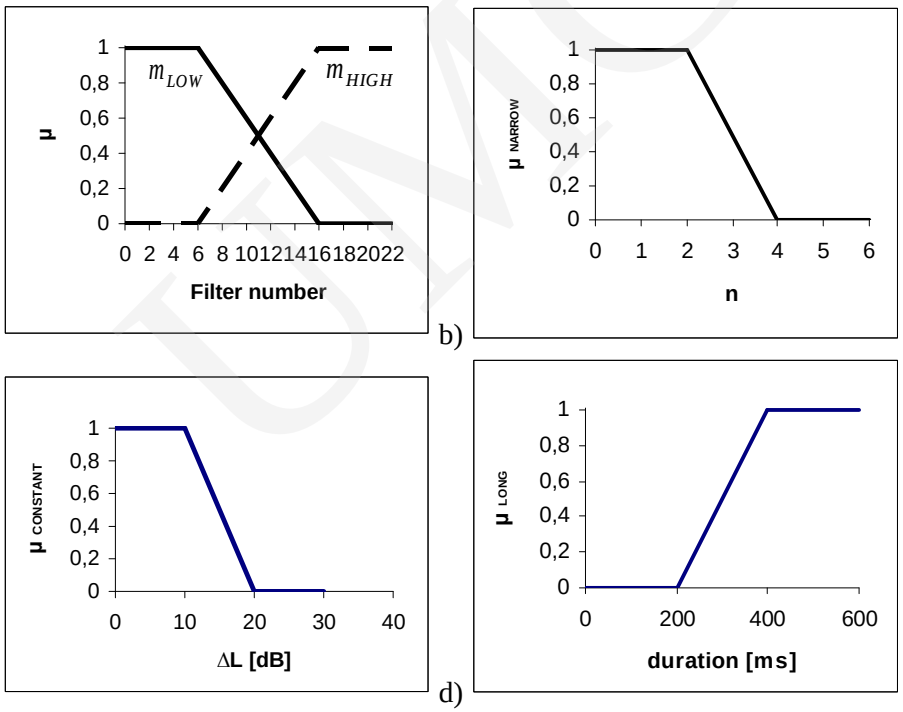

Fig. 3. Fuzzy set membership functions: a) "low" (μ_{low}) and "high" (μ_{high}) frequencies, b) "narrow", c) "constant", d) "long"

In the case of fricative consonants the peak location should fall into the range of "high" frequencies [11-12] while in the case of nasals – in the range of "low" frequencies [13-14]. The conditions, which could testify to the fuzzy set membership is the maximum (peak) staying for a long enough time in a "narrow" range of frequencies or remaining of a "constant" average level.

The "narrow" membership function (Fig. 3b) has been defined as:

$$m_{NARROW}(n) = \begin{cases} 1 & \text{for } n \le 2 \\ \dfrac{4-n}{2} & \text{for } 2 < i \le 4 \\ 0 & \text{for } i > 4 \end{cases} \tag{3}$$

where n is the number of frequency bands, where the peak level is located.
The "constant" membership function (Fig. 3c) has been defined as:

$$m_{CONSTANT}(\Delta L) = \begin{cases} 1 & \text{for } \Delta L \le 10dB \\ \dfrac{20-\Delta L}{10} & \text{for } 10dB < \Delta L \le 20dB \\ 0 & \text{for } \Delta L > 20dB \end{cases} \tag{4}$$

where *DL* is the range change of average sound level.

A significant feature is duration, which determines the prolongation set membership. From the research carried out for fricatives it appears that within the range of 200 to 400 ms these consonants are perceived by one listener as non-fluent while the other judges them as fluent. Long enough (over 400 ms) consonants were assessed by listeners as prolonged. For this reason, the "long" fuzzy set membership function (Fig. 3d) has been defined as:

$$m_{LONG}(t) = \begin{cases} 0 & \text{for } t \le 200ms \\ \dfrac{t-200}{200} & \text{for } 200ms < t \le 400ms \\ 1 & \text{for } t > 400ms \end{cases} \tag{5}$$

where t is duration.

5. Fuzzy inference

The basic "fricative consonant prolongation" set membership criterion was as follows:

If peak was located in the *high* frequency and duration is long then *long fricative*. Thus the fuzzy set was a product of two sets: *high* and *long*.

$$\text{LONG FRICATIVE} = \text{HIGH} \wedge \text{LONG}\ , \tag{6}$$

$$m_{LONG\ FRICATIVE}(i,t) = \min\left(m_{HIGH}(i), m_{LONG}(i)\right). \tag{7}$$

In some utterances, when the high fuzzy set membership function was below 1 or the peak was located in a band lower than the 16th, and the additional condition was introduced:

If the peak was located in the *narrow* frequency range or level is *constant* then the same *fricative*. Thus defined *fricative* fuzzy set was a sum of *constant* and *narrow* sets:

$$FRICATIVE = CONSTANT \vee NARROW\ , \tag{8}$$

$$m_{FRICATIVE}(\Delta L, n) = \max\left(m_{CONSTANT}(\Delta L), m_{NARROW}(n)\right). \tag{9}$$

As it was an additional condition, eventually the fricative prolongation set was defined as follows, with the assumption that $m_{\mathrm{HIGH}} > 0.5$:

$$FRICATIVE\ PROLONGATION = FRICATIVE \oplus LONG\ FRICATIVE\ , \tag{10}$$

$$\begin{aligned} m_{FRICATIVE\ PROLONGATION}(\Delta L, n, i, t) = \\ = \min\left(1, m_{FRICATIVE}(\Delta L, n) + m_{LONG\ FRICATIVE}(i, t)\right). \end{aligned} \tag{11}$$

The conditions of nasal consonant prolongation set membership were as follows:

If peak location was in the *narrow* frequency range and the mean level was *constant* and was in the *low* frequency then *nasal*. If *nasal* and *long* then *nasal prolongation*.

$$NASAL = CONSTANT \wedge NARROW \wedge LOW\ , \tag{12}$$

$$m_{NASAL}(\Delta L, n, i) = \min\left(m_{CONSTANT}(\Delta L), m_{NARROW}(n), m_{LOW}(i)\right), \tag{13}$$

$$NASAL\ \ PROLONGATION = NASAL \wedge LONG\ , \tag{14}$$

$$m_{NASAL\ PROLONGATION}(i, \Delta L, n, t) = \min\left(m_{NASAL}(i, \Delta L, n), m_{LONG}(t)\right). \tag{15}$$

5. Effectiveness of automatic prolongation detection method

The sound files of 10 stuttering people's utterances (reading aloud and describing) were analyzed with the presented program. In total, they contained approximately 3,000 words. In the utterances fricative prolongations prevailed, nasals were prolonged in two people's utterances. Figs 4 and 5 present the appropriate function courses for fricative and nasal prolongation sets in the utterances "z ɲego" and "ɲesteti", whose acoustical characteristics are presented in Figs 1 and 2.

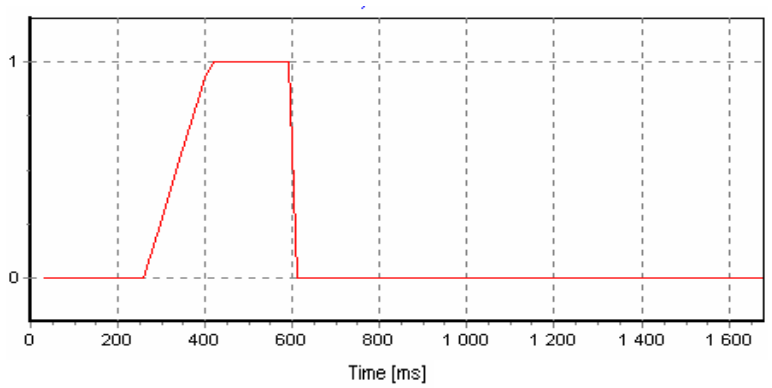

Fig. 4. Fricative prolongation set membership function in the utterance "z ɲego" with a prolonged "z" sound at the beginning

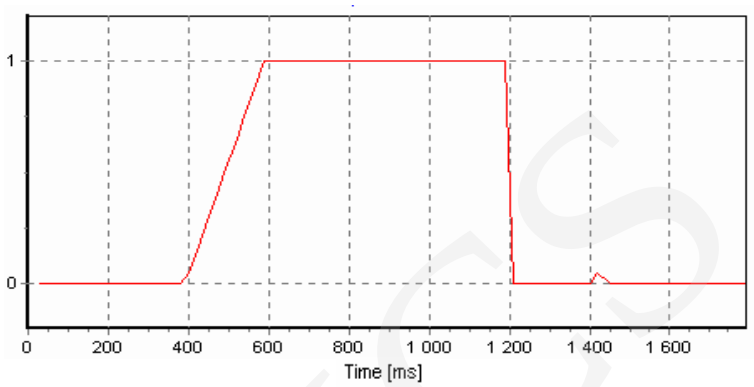

Fig. 5. Nasal prolongation set membership function in the utterance "ɲesteti" with a prolonged "ɲ" sound at the beginning

The experimenters compared the results obtained from the spectrographic pictures with those obtained with the automatic method. The computer program recognized 91% of all the words which had been qualified by the judges as prolongations. Their location in the sound file corresponded to that automatically found.

Conclusion

The basis of the elaborated method of automatic recognition and duration measurement of prolongation is the generalization of acoustic features within the range of the same manner of articulation. Such an approach is related to the reasons for the occurrence of a particular type of disfluency which are similar in respect of articulation. The features characterizing the prolongations have been described with the use of fuzzy sets. In the present article the authors have limited themselves to fricatives and nasals only, because they were dominant in the utterances which were used as the verification model and they made up sets numerous enough. The authors work on another disfluency type recognition, namely consonant and syllable repetitions, with the use of the methodology described above.

Acknowledgements

The research was supported by Grant No. 4 T11E 035 22 from the State Committee for Scientific Research in Poland.

The authors wish to thank Natalia Fedan for translation of the paper into English.

References

[1] Cordes AK., *The reliability of observational data: I. Theories and methods for speech-language pathology*, Journal of Speech and Hearing Research, 37 (1994) 264.
[2] Cordes AK. & Ingham RJ., *The reliability of observational data: II. Issues in the identification and measurement of stuttering events*, Journal of Speech and Hearing Research, 37 (1994) 279.
[3] Ingham RJ., Cordes AK., Gow L., *Time-interval measurement of stuttering: Modifying interjudge agreement*, Journal of Speech and Hearing Research, 36 (1993) 503.
[4] Łachwa A., *Fuzzy world of sets, numbers, relationships, facts, rules and decisions*, Academic Publishing Company EXIT, Warsaw, (2001), (in Polish).
[5] Rutkowska D., Piliński M., Rutkowski L., *Neural Networks, genetics algorithms, and fuzzy systems*, Polish Scientific Publishers, PWN, Warszawa-Łódź, (1999), (in Polish).
[6] Zadeh L.A., Fu K.S., Tanaka K., Shimura M., *Fuzzy sets and their applications to cognitive and decision processes*, New York: Academic Press, (1975).
[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: II ANN recognition of repetitions and prolongations with supplied word segment markers*, Journal of speech, Hearing and Language Research, 40 (1997) 1085.
[8] 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 (1997) 1073.
[9] 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), (1997).
[10] Howell P., Staveley A., Sackin S., Rustin L., *Method of interval selection, presence of noise and their effects on detectability of repetitions and prolongations*, The Journal of the Acoustical Society of America, 104 (1998) 3558.
[11] Jassem W., *Basis of Acoustical Phonetic*, Polish Scientific Publishers PWN, Warsaw, (1973), (in Polish).
[12] Jongman A., Wayland R., Wong S., *Acoustic characteristics of English fricatives*, The Journal of the Acoustical Society of America, 108 (1993) 1252.
[13] Fujimura O., *Analysis of nasal consonants*, The Journal of the Acoustical Society of America, 34 (1962) 1865.
[14] Jassem W., Łobacz P., *Analysis of Polish nasal consonants*, Speech and Language Technology, 3 (1993) 217.