Janusz DULAS

Politechnika Opolska, Instytut Elektrowni i Systemów Pomiarowych

The new method of the inter-phonemes transitions finding

Streszczenie. *Artykuł przedstawia nową metodę lokalizacji przejść międzyfonemowych opartą o analizę obrazów. Automatyczne określenie miejsc przejść międzyfonemowych jest równoznaczne z określeniem liczby fonemów występujących w danym wyrazie. Jest to ważny parametr wykorzystywany w systemach automatycznej identyfikacji sygnałów mowy. (**Nowa metoda lokalizacji przejść międzyfonemowych**).*

Abstract*. This article describes the new method of the inter-phonemes transition finding based on the image recognition. Automatic borders between phonemes finding is the same as the number of phonemes finding. This is an important factor used in Automatic Speech Recognition systems.*

Słowa kluczowe: Automatyczna segmentacja mowy, analiza czasowa, sterowanie za pomocą mowy, automatyczne rozpoznawanie mowy
Keywords: Automatic speech segmentation, time domain analysis, speech controlling, automatic speech recognition.

Introduction

Phonemes – the smallest parts of speech are the elements which build words and phrases. For many languages, the number of basic phonemes is found [1,2,3]. For Polish, for example, 37 phonemes can build 95% of all words [4]. For English, 41 but as was written in many science reports this number is not constant and changes according the language changes [5,6]. Nowadays' Automatic Speech Recogniton systems use mostly spectrum analyses and Hidden Markov Models (HMM) or Artificial Neural Network for recognition. Many spectrum and cepstrum parameters are counted there and the probability metods are used for fitting unknown words to the words placed in data base. In this article another approach is presented. Inter-phonemes zones were analysed and some typical images of these places were found. This method was tested for 100 words spoken in Polish by speakers of different age and sex, including 350 inter-phonemes zones, which enabled recognition of 450 phonemes.

Inter-phonemes zones' characteristic

Inter-phonemes zones are the places where one phoneme ends and another starts. As author's research shows the signal in these places could change in different ways. In figure 1, the time characteristic of the word "zero" is shown, also the phonemes and inter-phonemes zones are matched.

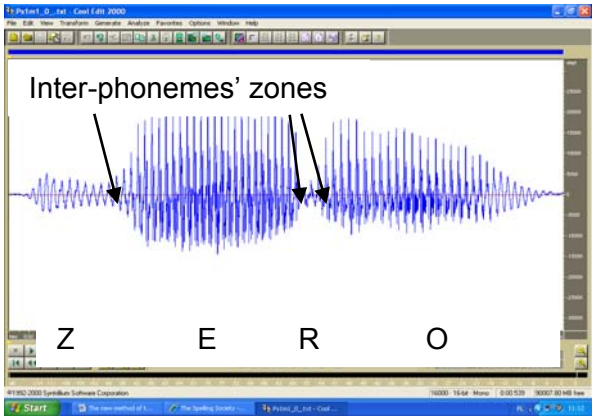

Fig.1. The word "zero" time characteristic with matched phonemes and inter-phonemes zones

As is easy to observe, the first inter-phoneme zone (between "Z" and "E") is connected with big signal's amplitude increase. Similar situation occurs between "R" and "O" phonemes. Another phenomena take place between "E" and "R" phonemes. Here a big decrease of the signal's amplitude is observed. Let's see another example. In figure 2 the word "JEDEN" (Eng. "one") time characteristic is placed.

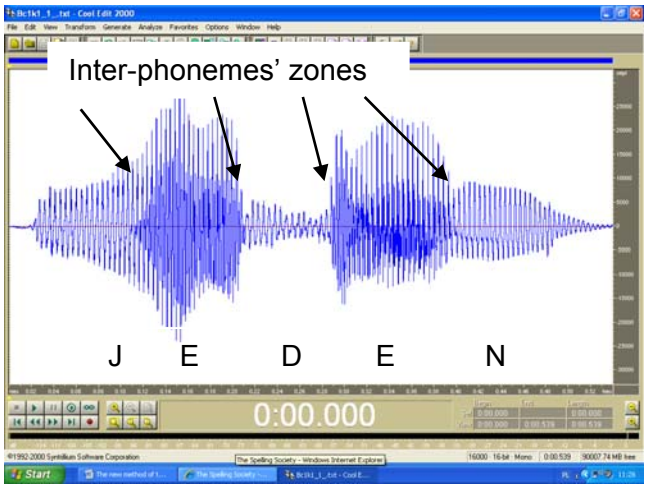

Fig.2. The word "jeden" time characteristic with matched phonemes and inter-phonemes zones

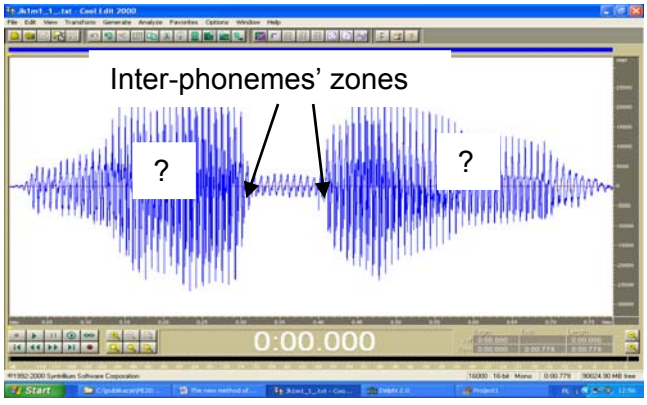

Fig.3. The word with hidden inter-phonemes zones

Here also exists a strong decrease and increase of the signal amplitude (phonemes E-D-E), but between "J" and "E" phonemes, a slow increase of the signal amplitude is observed. Also the shape of the signal changes which is observed as a "darker" zone on the time characteristic. Similar phenomena occurs between phonemes "E" and "N" where the amplitude is decreasing slowly and the signal becomes "lighter". Sometimes inter-phoneme zones are difficult to find. This situation is shown in figure 3. Here only 2 zones are easy to observe. In order to find the other 2, more advanced tools must be used. For this purpose, the

grid method is used [8,9]. First, the local minima are found and then the distance among them is calculated. If the distance between neighboring local minima is bigger than 20ms or there are exist some grids with strongly varying duration it means that this place could be the inter-phonemes zone. For signal from figure 3, the local minima are shown in figure 4. In figure 5 there are the grid durations in ms counted from data from figure 4.

37.75 ; 45.63 ; 53.38 ; 61.63 ; 69 ; 77.25 ; 85.38 ; 93.38 ; 101.6 ; 110.1 ; 118.5 ; 126.9 ; 134.9 ; 143.3 ; 151.6 ; 159.9 ; 168.1 ; 181 ; 192.6 ; 205.3 ; 216.9 ; 225.3 ; 233.6 ; 242 ; 250.4 ; 258.8 ; 267.1 ; 275.5 ; 288.1 ; 309.9 ; 319.3 ; 327.9 ; 336.9 ; 345.5 ; 353.6 ; 362.1 ; 370.6 ; 379.1 ; 387.6 ; 396.5 ; 404.8 ; 412.5 ; 420.5 ; 428.8 ; 436.8 ; 445.1 ; 453.4 ; 461.5 ; 469.8 ; 478 ; 486.3 ; 494.6 ; 502.9 ; 511.1 ; 519.5 ; 527.8 ; 536 ; 548.9 ; 560.9 ; 569.1 ; 577.3 ; 585.8 ; 594.5 ; 602.6 ; 610.9 ; 619.1 ; 627.4 ; 635.5 ; 643.9 ; 652.1 ; 660.3 ; 668.5 ; 676.5 ; 684.9 ; 692.6 ; 700.9 ; 709.4 ; 717.8 ; 727.1 ; 735.5 ; 743.8 ; 752.4

Fig.4. Local minima localization

7.88 7.75 8.25 7.38 8.25 8.13 8 8.25 8.5 8.38 8.38 8 8.38 8.38 8.25 8.25 12.9 11.6 12.6 11.6 8.38 8.38 8.38 8.38 8.38 8.38 8.38 12.6 21.8 9.38 8.63 9 8.63 8.13 8.5 8.5 8.5 8.5 8.88 8.25 7.75 8 8.25 8 8.38 8.25 8.13 8.25 8.25 8.25 8.38 8.25 8.25 8.38 8.25 8.25 12.9 12 8.25 8.13 8.5 8.75 8.13 8.25 8.25 8.25 8.13 8.38 8.25 8.13 8.25 8 8.38 7.75 8.25 8.5 8.38 9.38 8.38 8.25 8.63

Fig.5. Grid durations calculated from data from figure 4

The places where the grid durations are bigger than neighboring ones were matched in grey. As is easy to observe, three such places were found. The first and the last are our hidden zones. The places between them show the other two inter-phonemes zones.

The new set of parameters enabling inter-phonemes' automatic localization

As was mentioned above, in places where inter-phonemes transitions occur, different phenomena could be observed. They were divided into 10 parameters.

1.**Parameter 1** – big grid's length
Usually the grid's length equals or is near to the pitch period. The length bigger than 20,76ms shows the place where one phoneme ends and another starts. This case is shown in figure 6 and the place on the time characteristic is shown in figure 7.

16.9 14.4 7.63 7.88 8.13 8 8.13 8.13 8.13 8.13 8.25 8.25 8.25 8.25 8.38 8.38 8.5 8.38 8.63 8.63 8.63 8.75 22.1 8.63 14.8 8.88 8.88 8.75 9.25 9.25 14.4 9.63 9.75 14.1 9.5 14.5 9.13 15.1 9.25 9.25 10.8 9.13 8.63

Fig.6. The big grid's duration showing the inter-phoneme transition

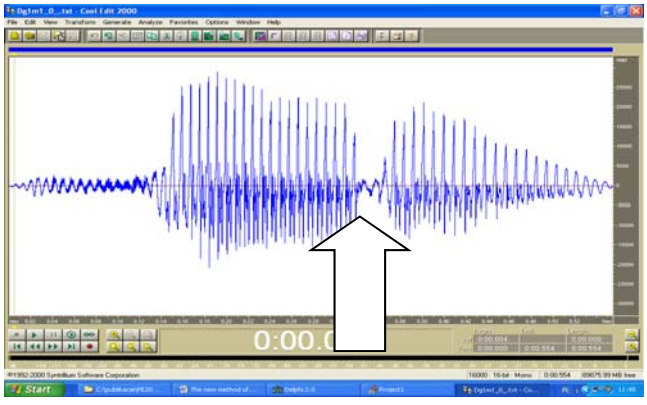

Fig.7. The inter-phoneme transition found after the grid's length analysis

2. **Parameter 2** – minimum phoneme's duration
This parameter could vary from 50-70ms depending on the place of the phoneme in the word. As authors research showed, typical voiced phonemes aren't shorter than 50 ms. This feature is matched in figure 8.

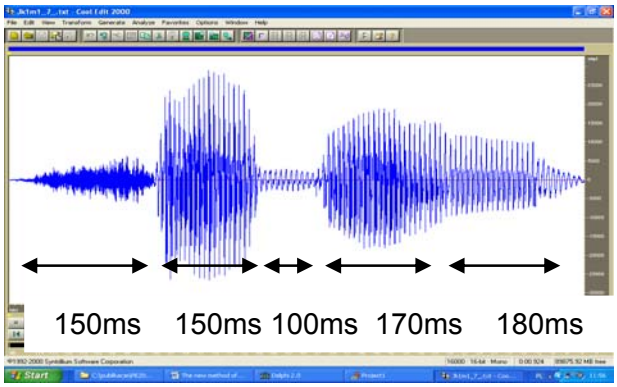

Fig.8. Typical phonemes' durations

3. **Parameter 3** – Minimum amplitude value
This parameter is necessary in order to check if the part of the time characteristic is a phoneme or a brake. If signal has the amplitude of more than 5% of the maximum amplitude existing in the whole word it could be considered as a phoneme. This situation is shown in figure 9.

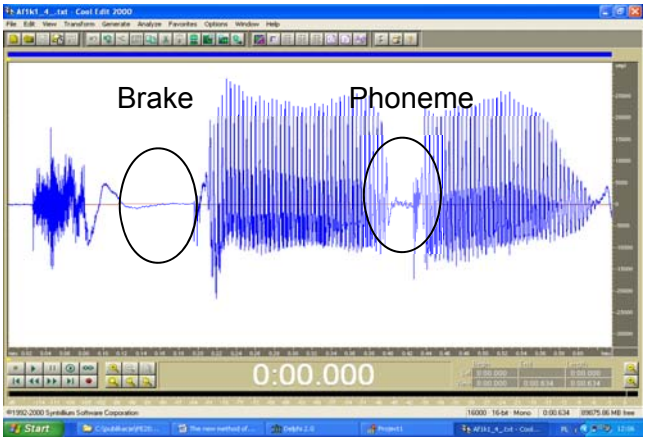

Fig.9. Brakes and phonemes inside the word

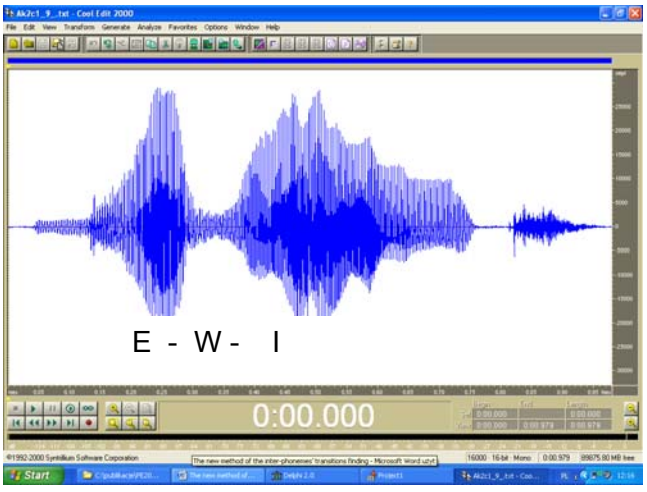

Fig.10. The part of signal with small amplitude existing between the parts with big amplitude

4. **Parameter 4** – The part of signal with small amplitude existing between the parts with big amplitude

This situation usually takes place for consonants placed between vowels. The example of phoneme “w” placed between “e” and “i” is shown in figure 10.

5. **Parameter 5** – Big duration differences among neighboring grids.
Sometimes the signal amplitude is almost unchanged but in inter-phonemes zone big grid’s duration variety is observed. This case is shown in figure 11, and matched places in figure 12.

15.8 12 12.6 7.5 7.63 7.5 7.88 7.75 12.9 10.3 7.75 7.75 7.75 7.25 8.25 7.88 7.88 7.5 7.88 12.9 10.8 8.25 8.5 8.5 8 8.13 8 8.13 8 8.13 8 8.13 8 8.13 8.13 8.38 8.5 8.75 44.4 17.6 9.38 8.75 8.13 8.5 14 17.8 8.88 7 7.75 7.63 16.1 6.88 26.6

Fig.11. Local big grids’ duration differences

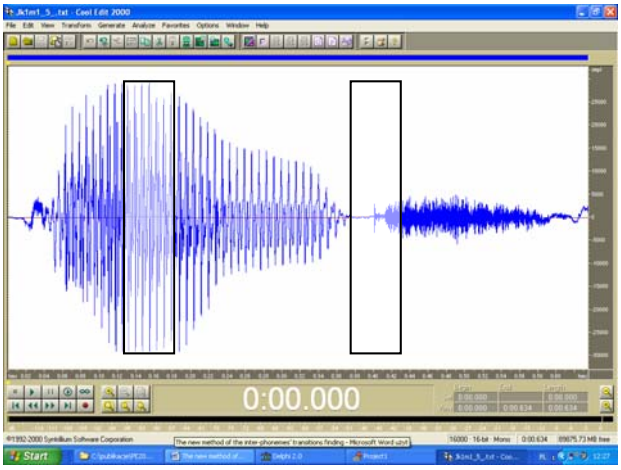

Fig.12. Inter-phonemes zones found after grids’ duration analyses.

6. **Parameter 6** – Sudden amplitude increase.
As author’s research showed in many cases one phoneme changes to another together with strong amplitude decrease. Usually the next phoneme’s amplitude is bigger than 1.6 of the previous phoneme amplitude. This feature is shown in figure 13.

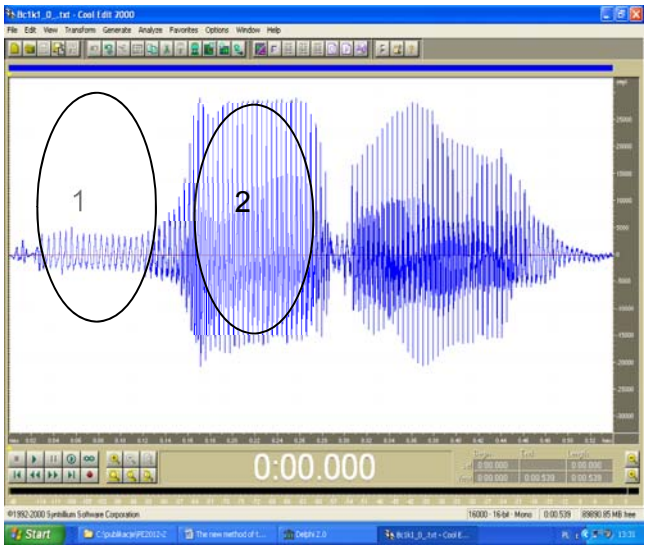

Fig.13. Two neighboring phonemes with different amplitude

7. **Parameter 7** - The amplitude decrease connected with long grid occurence.
This feature usually take place between two voiced phonemes. The example is shown in figure 14 and 15.

5.25 5.38 5.63 6 5.63 5.88 5.88 5.88 6 5.88 5.88 6 6 6.13 6 6.13 6 6.13 6.13 6 5.75 6.25 6 5.75 6 5.88 6.13 5.63 5.88 5.88 5.75 5.5 5.88 5.75 5.5 6 5.75 5.5 5.63 5.63 5.13 5.63 11.5 5.88 5.75 5.88 5.75 5.75 5.75 5.75 5.75 5.63 6 7.88 11.3 5.5 5.88 64 16.3 5.75 5.5 5.38 5.63 5.38 5.63 5.75 5.75 5.63 5.75 5.75 5.75 5.75 5.5 5.75 5.75 7.25 10.3 5.5 5.75 6.75 5.25 5.75 5.88 5.63 5.75 5.75 5.75 5.63 5.75 5.75 5.63 5.75 5.75 5.88 5.75 5.75 5.75 6.75 5.75 5.38 5.38 5.63 5.5 5.38 5.13 5.75

Fig.14. Long grid in inter-phoneme zone

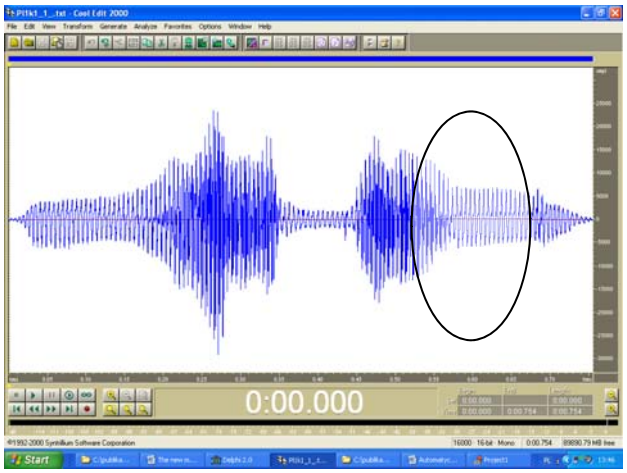

Fig.15. Inter-phone zone with long grid and amplitude decrase

8. **Parameter 8** – Sudden amplitude decrease.
This phenomenon usually takes place in moments where a vowel ends and consonant starts. The example is shown in figure 16.

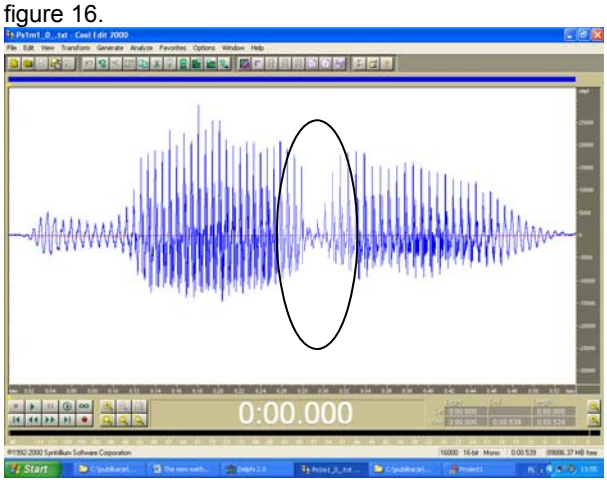

Fig.16. Sudden amplitude decrease in inter-phoneme zone

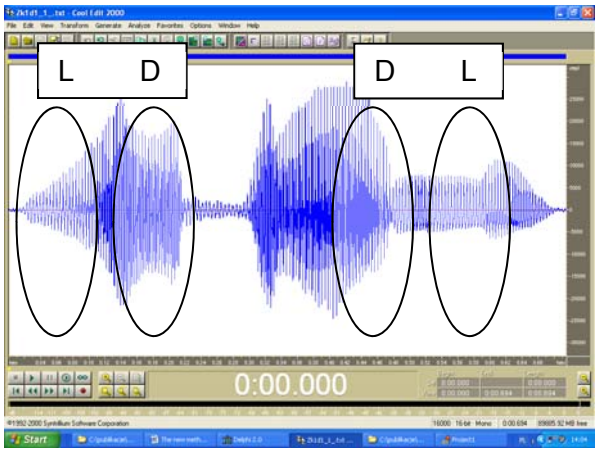

Fig.17. Neighboring “lighter” and “darker” signals.

9. **Parameter 9** – Sudden zero crossing points increase. This feature is observed on the time characteristic as a "darker" signal (D) occuring after "lighter" signal (L). The example is shown in figure 17.

10. **Parameter 10** - Sudden zero crossing points decrease. This feature is observed on the time characteristic as a "darker" signal (D) occuring before "lighter" signal (L). The example is shown in figure 17.

Results of research

The new method of inter-phonemes zones finding is based on 10 parameters' analyses (described above). The zone exists if one of the parameter number 1 or 5 or 6 or 7 or 8 or 9 or 10 is true and the parameter number 2 is true. If also parameters 3 and 4 are true it means that the short phoneme was found and two inter-phonemes zones exist one after another. This method was tested on the 100-record set obtained from the Corpora database containing men's, women's and children's voices. 350 inter-phoneme zones occuring in those records were found properly. In the future further research are planned and an application for the number of phonemes finding will be done.

REFERENCES

[1] International Phonetic Association, Phonetic description and the IPA chart, *Cambbridge University Press, 1999*

[2] Burquest D., Payne L., Phonological analysis: A functional approach, *Summer institute of Linguistics, Dallas, 1993*

[3] Dulas J., Speech recognition based on the grid method and image similarity, *Speech technologies, INTECH 2011, 321- 340*

[4] Basztura Cz., Rozmawiać z komputerem, *Wydawnictwo Format, Wrocław 1992*

[5] Bett S., Can we pin down the number of phonemes In English?, *Simple Spelling Newsletter*, 3/1999, p.7

[6] Brown A.,The number of phonemes in English: not a simple answer to a simple question, JSSS 27/2000, 11-13.

[7] Chakraborty R., Sangupta D., Sinha S.Pitch tracking of acoustic signals based on average square mean difference function, *Signal image and video processing, Springer London vol.3, number 4, 2008.*

[8] Dulas J., Speech recognition based on the grid method and image similarity, *Speech technologies, INTECH 2011, 321- 340*

[9] Dulas J., Automatyczna identyfikacja cyfr dla mówców polskojęzycznych, *PE 5/2010, 15-18*

[10] Dulas J., Szybka metoda identyfikacji fonemów szumowych występujących w cyfrach wypowiadanych w języku polskim, *PE 2/2011, 242-245*

[11] Wydra S. Recognition quality improvement In automatic speech recognition system for Polish, *EUROCON 2007,Warszawa, 218-223*

[12] Dulas J., Automatyczna segmentacja sygnałów mowy w oparciu o metodę siatek o zmiennych parametrach, *PE 1/2010, 229-232*

[13] Dulas J., Automatic words' recognition algorithm used for digits clessification, *PE 11/2011, 230-233.*

[14] Dulas J., Rozpoznawanie jednostek fonetycznych zawierających okresy podstawowe tonu krtaniowego, *Konferencja Podstawowe Problemy Metrologii, Sucha Beskidzka 2008*

[15] Dulas J., Analiza obwiedni jako parametr wspomagający automatyczną identyfikację wyrażeń, *PAK 5/2009, 308-309*

[16] Dulas J., Wspomaganie rozpoznawania wyrazów za pomocą opisu ich obwiedni, *Konferencja Podstawowe Problemy Metrologii, Sucha Beskidzka 2009, s.152-156*

[17] Dulas J., Automatyczne rozpoznawanie cyfr w języku polskim – identyfikacja fonemów szumowych, *PE 1/2011*

[18] Kłosowski P. Usprawnienie procesu rozpoznawania mowy w oparciu o fonetykę i fonologię języka polskiego, *Rozprawa Doktorska, Politechnika Śląska 2000*

[19] Nishida M., Horiuchi Y.,Ichikawa A.,Automatic speech recognition based on adaptation and clustering using temporal-difference learning, *INTERSPEECH 2005,* Lisbon, Portugal, 285-288

***Autor**: dr inż. Janusz Dulas, Politechnika Opolska, Instytut Elektrowni i Systemów Pomiarowych, ul. Prószkowska 76, 45-758 Opole, e-mail: j.dulas@po.opole.pl*