Janusz DULAS

Politechnika Opolska, Instytut Elektrowni i Systemów Pomiarowych

Pitch period's properties and the new method used for finding them

Streszczenie. *Artykuł przedstawia interesujące właściwości okresów podstawowych tonu krtaniowego występującego we wszystkich samogłoskach i spółgłoskach dźwięcznych oraz nową metodę ich odnajdywania i wyznaczania ich długości. Poprawne odnajdywanie okresów podstawowych i wyznaczanie czasu ich trwania jest ważnym elementem algorytmu automatycznej identyfikacji słów opracowanego przez autora. (**Wyszukiwanie okresów podstawowych w tonach kataniowych i wyznaczanie czasu ich trwania**)*

Abstract*. This article describes the pitch's periods interesting properties. These periods are included in each vowel and voiced consonant. It also describes the new method of pitch period finding and their duration counting. These parameters are very important elements of the automatic speech recognition algorithm worked out by the author.*

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

Pitch periods included in human speech are the effect of the air flowing from lungs through the vocal folds to the lips. The vibrating vocal folds produce a sound which frequency depends on the vocal fold length and vocal tract size. As research [1, 2] showed the male vocal folds are between 17 and 25 mm in length while the female ones are between 12,5 and 17,5 mm in length. Also the male's vocal tract size is larger than the female's so the male voice is usually lower-sounding. There are known methods of finding pitch period duration but in this article the new, simpler method is described. As results of research show it works properly for speakers different sex and age.

Pitch period's properties

Pitch periods are included in all vowels and voiced consonants. They are well visible on the time characteristics as a repeatable parts of the signal with similar shape. Figure 1 shows these periods for male's(a), female's(b) and child's(c) voice.

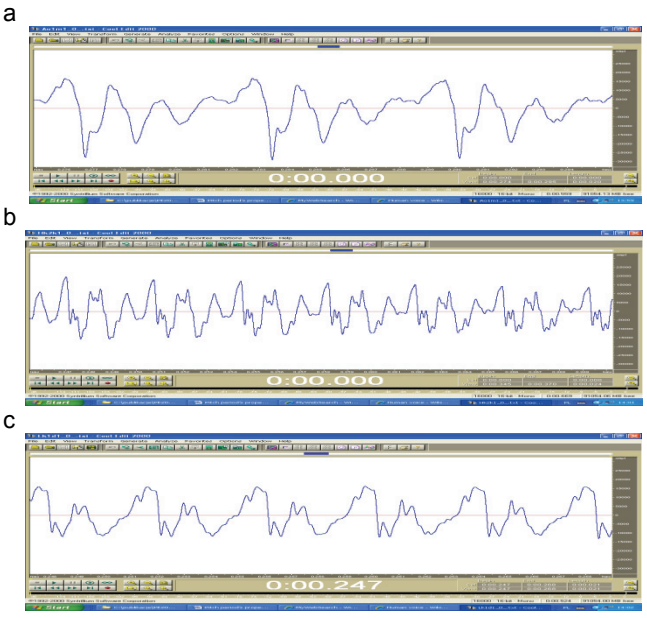

Fig.1. The pitch periods included in male's(a), female's(b) and child's(c) voice

As is easy to observe the male's pitch periods have bigger duration than female's and child's. As author's research showned for women and children the duration is between 2 and 5 ms, whereas for men is between 5 and 10 ms. Another interesting pitch periods' property is their duration's changeability. Although this parameter depends on the fold's length and vocal tract's size, it also could be changed by the speaker. This phenomenon is shown in figure 2.

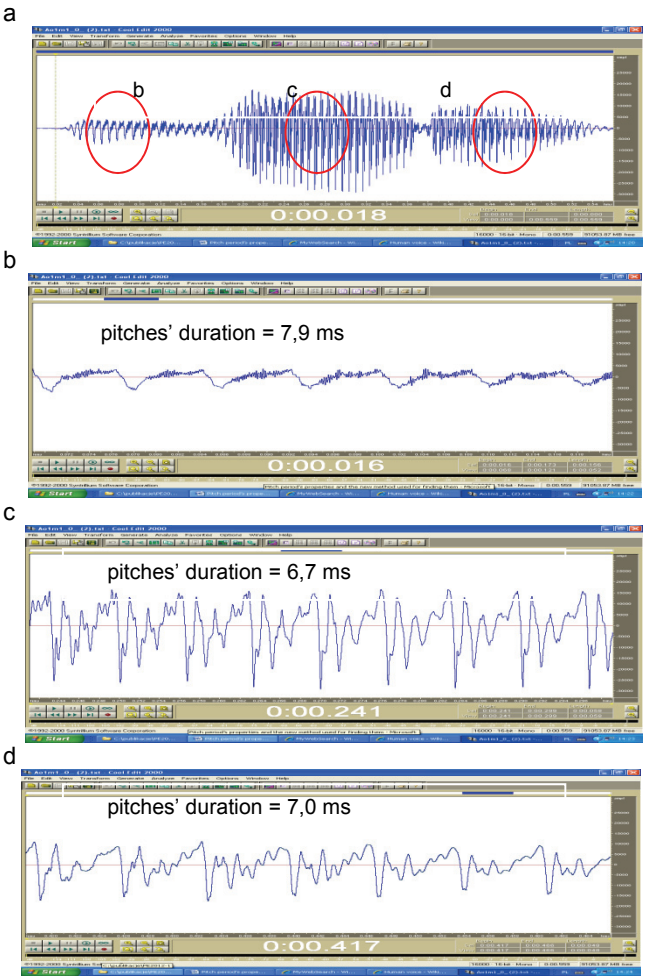

Fig.2. Different pitch periods' durations within one word (a-the whole word, b- the first part, c- the second part, d- the third part)

Here, the whole word (which time characteristic is shown in “a”) was divided on 3 parts. First part (Fig.2b) includes pitch periods with duration 7,9ms, second part (Fig.2c) with duration 6,7ms and the third part (Fig.2d) with duration 7,0ms. So the difference between the biggest and the smallest duration equals 7,9 – 6,7 = 1,2 ms. This is about 16% of the average pitch period’s value. It shows that there is no one constant duration’s value but the interval of values.

Existing methods of the pitch period’s duration finding

The methods which are used for pitch period duration finding could be divided into time and frequency domain[4,5,6]. In time domain the most popular is method called “zero crossing” where the time between neighboring crossing signal and “zero level” points is measured. This method gives good results for signals without high harmonics. For speech signal where these harmonics exist additional low pass filter must be applied. Another methods used in time domain are autocorrelation methods. In this case segments of the signal are compared with other segments offset by a trial period to find a match. These are more sophisticated methods and sometimes they have problems with noisy signals and false detection. In frequency domain first the FFT transformation must be done. Then the cepstrum (reverse FFT transform) is counted. The FFT transformation must be done in “windows” which length is fitted to the basic period duration and usually equals 10ms. A number of problems appeared here because the pitch period changes its duration value during measurement.

The new method of the pitch period finding

During the authors research connected with automatic speech recognition the new method of pitch period finding was worked out. This is a new approach based on the image of the time characteristic analyses. The first step is a zero level signal finding. It is counted as a average value of the 100 samples before recognition word. Next step is finding a local signal’s minimums. It is made according to the equation:

$$x(n) = \min \text{ if: } \begin{array}{l} x(n-z) > x(n) \ \ and \\ x(n+v) \geq x(n) \ and \\ x(n) < \text{"0"} level \end{array} \quad (1)$$

where:
x(n)- the sample’s “n” value
z=1..15
v=1..10

This condition is tested for 15 samples before and 10 after a sample which could be a local minimum. This is done because this way only main local minima are found. This case is shown in figure 3.

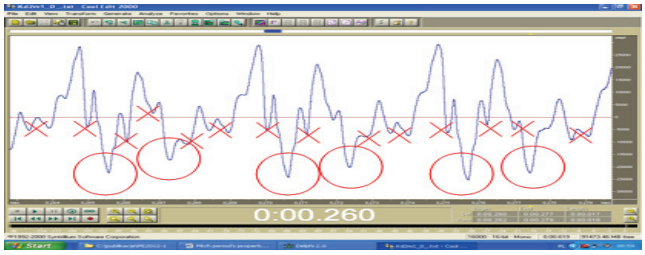

Fig.3 Local minimums finding

As was shown above the local minimums which are not a main minimums are omitted (they are matched by a crosses). For further analyses only main minimums are taken.

Next step is withdrawing a minimums which values are bigger than the previous minimums’ values about more than 5% of the signal’s amplitude.

Next the time between local minimums is tested. If it is lower than 2 ms the following minimum is withdrawn. This is made because as an author’s research shown the minimum pitch period duration is bigger than 2 ms. Figure 4 shows the result of this operation.

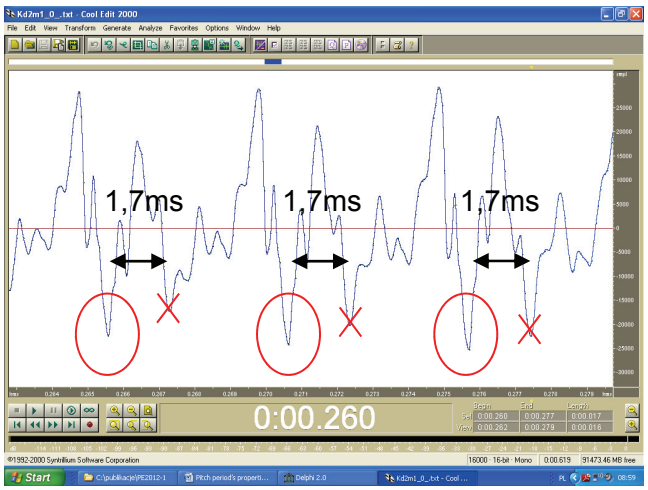

Fig.4 Too near placed local minimums’ withdrawing

Another step is checking if the pitch period’s duration is long (bigger than 8 ms). For this goal the number of long periods is counted. If this number is higher than 25% of all periods, than periods with less than 7 ms duration are withdrawn. This operation is necessary because sometimes for long periods local minimums were treated as a main minimums. This case is shown in figure 5.

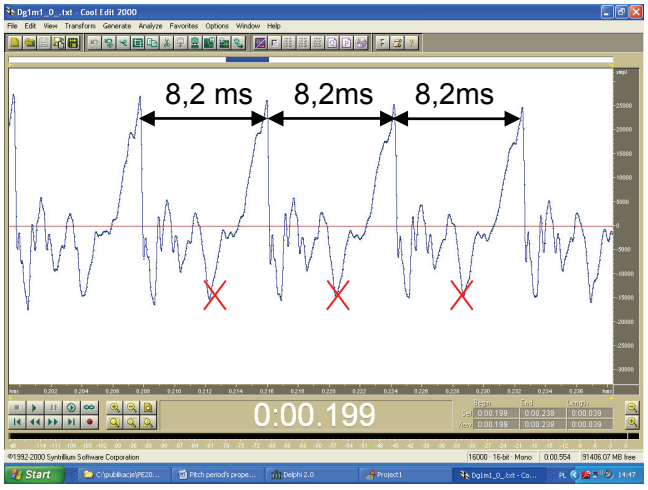

Fig.5 Long periods finding

Next step involves counting an average value of period’s duration from all periods which remain after previous steps. In order to eliminate big errors the duration values are sorted and 20% the lowest and 20% the biggest results are withdrawn. Then the average value is counted. In the next step all periods with duration less than 80% of the average value are withdrawn. This operation is repeated once again. After this the right pitch period’s duration is obtained. The algorithm which makes above operation is shown in figure 6.

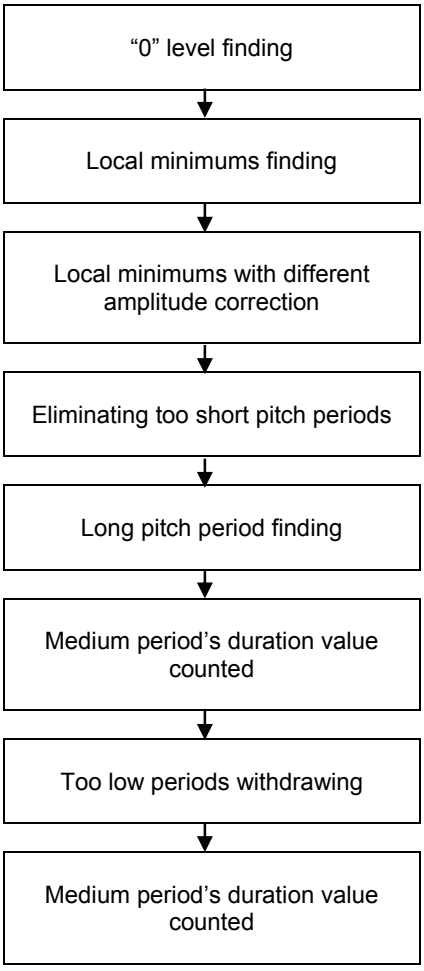

Fig.6 Automatic pitch period's finding algorithm

As was shown above the algorithm is simple and after eight steps the right result is obtained.

The result of research

The new method of the pitch period finding was tested on the set of 50 files with recorded one word made by speakers different sex and age. The results are shown in table 1.

Table 1. Results of research

Speaker number	File name	Pitch period duration red from the time characteristic			The result obtained from the algorithm	Error [%]
		min.	max.	mean		
1	Af1k1_0_	4,5	4,7	4,6	4,5	2,2
2	Ak1c1_0_	2,3	3,1	2,7	3,1	-14,8
3	Ak2c1_0_	3,0	4,3	3,7	3,2	12,3
4	Ao1m1_0_	6,7	7,5	7,1	7,0	1,4
5	Bc1k1_0_	4,2	5,0	4,6	4,4	4,3
6	Bc1m1_0_	7,5	9,0	8,3	7,7	6,7
7	Bw1k1_0_	4,5	5,1	4,8	4,8	0,0
8	Ck1c1_0_	6,2	7,0	6,6	6,7	-1,5
9	Dg1m1_0_	8,5	9,5	9,0	8,9	1,1
10	Hk1k1_0_	3,7	4,5	4,1	3,6	12,2
11	Hk2k1_0_	3,6	3,8	3,7	3,8	-2,7
12	Is1k1_0_	4,8	5,2	5,0	4,6	8,0
13	Jc1m1_0_	7,0	8,0	7,5	7,1	5,3
14	Jd1m1_0_	8,0	8,6	8,3	8,3	0,0
15	Jd2m1_0_	7,5	7,7	7,6	8,3	-9,2
16	Jd3m1_0_	7,2	8,0	7,6	7,2	5,3
17	Jd4m1_0_	7,7	8,0	7,9	8,0	-1,9
18	Jd5m1_0_	8,5	9,0	8,8	8,8	-0,6
19	Jd6m1_0_	8,5	9,0	8,8	8,7	0,6
20	Jk1m1_0_	8,2	8,5	8,4	8,3	0,6
21	Jo1m1_0_	8,8	9,0	8,9	8,9	0,0
22	Jp1m1_0_	7,7	8,0	7,9	8,5	-8,3
23	Jp2m1_0_	7,5	7,7	7,6	7,6	0,0
24	Js1m1_0_	7,5	8,2	7,9	8,0	-1,9
25	Kd1m1_0_	6,2	7,5	6,9	7,2	-5,1
26	Kd2m1_0_	4,5	5,0	4,8	5,2	-9,5
27	Ld1k1_0_	4,5	4,7	4,6	4,5	2,2
28	Ld2k1_0_	4,0	4,3	4,2	4,2	-1,2
29	Lk1d1_0_	3,9	4,2	4,1	4,2	-3,7
30	Mr1m1_0_	8,8	9,2	9,0	9,2	-2,2
31	Ms1m1_0_	8,5	9,7	9,1	8,9	2,2
32	Oj1k1_0_	4,2	4,5	4,4	4,5	-3,4
33	Pb1k1_0_	4,7	5,0	4,9	4,9	-1,0
34	Pl1k1_0_	5,8	5,9	5,9	6,2	-6,0
35	Pl1m1_0_	6,7	9,8	8,3	7,6	7,9
36	Ps1m1_0_	6,8	7,5	7,2	7,4	-3,5
37	Pw1m1_0_	7,2	9,0	8,1	8,2	-1,2
38	Rg1m1_0_	7,8	8,5	8,2	9,2	-12,9
39	Sg1m1_0_	8,5	8,8	8,7	9,3	-7,5
40	Sg2m1_0_	9,0	9,1	9,1	9,3	-2,8
41	Sg3m1_0_	9,0	9,2	9,1	9,2	-1,1
42	Sp1m1_0_	8,8	10,0	9,4	9,9	-5,3
43	Sw1m1_0_	8,0	10,0	9,0	9,3	-3,3
44	Ts1m1_0_	7,2	9,5	8,4	7,3	12,6
45	Tz1m1_0_	7,6	8,8	8,2	8,3	-1,2
46	Wb1m1_0_	6,5	6,7	6,6	7,0	-6,1
47	Wm1m1_0_	9,0	10,5	9,8	9,2	5,6
48	Zb1m1_0_	8,1	10,0	9,1	8,8	2,8
49	Zk1d1_0_	3,8	3,9	3,9	3,9	-1,3
50	Zk1m1_0_	6,3	6,4	6,4	6,5	-2,4

As is easy to observe the greatest errors value equals -14,8%. This not a big error as we consider that in typical word pitch's period could vary about 16%. The mean errors value equals -0,6% what is a very good result and shows that this algorithm works properly.

Summary

This article described a new method of the pitch period duration finding. As results of research shown this method works properly. This is a great progress and simplifying comparing to well known methods used up to now. This algorithm was used for automatic speech recognition system described in [12].

REFERENCES

[1] Flanagan J.L., Speech analysis, synthesis and perception, , *Speech technologies, Springer Verlag Berlin Heidelberg, 1970*

[2] Kacprowski J., Phisical models of the larynx source, Archives of Acoustics, 1977, vol.12, no 1, pp-47-70

[3] Gerhard D., Pitch extraction and fundamental frequency – history and the current techniques, technical report, University of Regina, 2003

[4] McLeod P., Wyvill G., A smarter way to find pitch, International computer music conference ICMC'2005.

[5] Massaoud A., Bouzid A., Ellouze N., A new method of Speech estimation and voicing decision based on spectral multi- scale product analisys, *Publié dans Signal Processing: An International Journal, Vol. 3(5), September 2009.*

[6] 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.*

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

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

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

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

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

[12] Dulas J., Automatic words' recognition algorithm used for digits clessification, *PE 11/2011, 230-233.*
[13] Dulas J., Rozpoznawanie jednostek fonetycznych zawierających okresy podstawowe tonu krtaniowego, *Konferencja Podstawowe Problemy Metrologii, Sucha Beskidzka 2008*
[14] Dulas J., Analiza obwiedni jako parametr wspomagający automatyczną identyfikację wyrażeń, *PAK 5/2009, 308-309*
[15] Dulas J., Wspomaganie rozpoznawania wyrazów za pomocą opisu ich obwiedni, *Konferencja Podstawowe Problemy Metrologii, Sucha Beskidzka 2009, s.152-156*
[16] Dulas J., Automatyczne rozpoznawanie cyfr w języku polskim – identyfikacja fonemów szumowych, *PE 1/2011*
[17] Basztura Cz., Rozmawiać z komputerem, *Wydawnictwo Format, Wrocław 1992*
[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
[20] Liu D., Kiecza D., Srivastava A., Kubala F., Online speaker adaptation and tracking for real-time speech recognition, *INTERSPEECH 2005,* Lisbon, Portugal, 281-284
[21] Xiang B., Nguyen L., Guo X. Fu D., The BBN Mandarin Broadcast News Transcription System, *INTERSPEECH 2005,* Lisbon, Portugal,1649-1652
[22] Lamel L., Adda G., Bilinski E., Gauvain J.L., Transcribing lectures and seminars, *INTERSPEECH 2005,* Lisbon, Portugal,1657-1660
[23] Trancoso I., Nunes R., Neves L.,Recognition of classroom lectures in european Portuguese *INTERSPEECH 2006,* Pittsburgh, USA,281-284
[24] Chang-wen H., Lin-shan L., Extended powered cepstral normalization (P-CN) with range equalization for robust teatures in speech recognition, *INTERSPEECH 2007,* Antwerp, Belgium, 1106-1109
[25] Weifeng L., Herve B., Non-linear spectral contrast stretching for in-car speech recognition, *INTERSPEECH 2007,* Antwerp, Belgium, 1122-1125
[26] Seymour R., Stewart D., Ming J.Audio-visual integration for robust speech recognition using maximum weighted stream posteriors, *INTERSPEECH 2007,* Antwerp, Belgium, 654-657
[27] Zhu B., Hazen J., Glass R., Multimodal speech recognition with ultrasonic sensors, *INTERSPEECH 2007,* Antwerp, Belgium, 662-665
[33] Neiberg D., Ananthakrishnan G., Gołaś A.Blomberg M., On Acquiring Speech Production Knowledge from Articulatory Measurements for Phoneme Recognition, *INTERSPEECH 2009, Brighton, United Kingdom, 1387-1390*

***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*