

Janusz DULAS

POLITECHNIKA OPOLSKA, INSTYTUT ELEKTROWNI I SYSTEMÓW POMIAROWYCH

Analiza obwiedni jako parametr wspomagający automatyczną identyfikację wyrażzeń

Dr inż. Janusz DULAS

W 1995r ukończył Studia na Politechnice Opolskiej, na kierunku Elektrotechnika. W 2002r. obronił rozprawę doktorską pt. "Metoda siatek o zmiennych parametrach w zastosowaniu do rozpoznawania fonemów mowy polskiej". Obecnie adiunkt na Politechnice Opolskiej, na wydziale Elektrotechniki, Automatyki i Informatyki. Zajmuje się badaniami nad automatycznym rozpoznawaniem i sterowaniem za pomocą sygnałów mowy.



e-mail: j.dulas@po.opole.pl

Streszczenie

W badaniach nad automatycznym rozpoznawaniem sygnałów mowy notuje się stały postęp, choć różnorodność języków utrudnia wprowadzenie jednakowych rozwiązań. Przykładem rozwoju i upowszechnienia metod identyfikacji mowy może być system operacyjny Windows XP, w którym zamieszczono narzędzia do sterowania aplikacjami za pomocą sygnałów głosowych. Brak jednak nadal rozwiązań dla języka polskiego, co sprawia że potrzebne są badania zmierzające do opracowania niezawodnych algorytmów identyfikujących i sterujących. W artykule przedstawiono wyniki badań obwiedni sygnałów mowy, będących cyframi z zakresu 0-9, uzyskanych dla grupy 50-ciu osób różnych płci i w różnym wieku. Celem przeprowadzonych badań było uzyskanie odpowiedzi na pytanie, czy analiza obwiedni może stanowić parametr w procesie automatycznego rozpoznawania sygnałów mowy i czy jest możliwe stworzenie modeli obwiedni dla każdej z cyfr, które byłyby wspólne dla wszystkich (50) mówców.

Słowa kluczowe: Automatyczne rozpoznawanie sygnałów mowy, analiza obwiedni.

The envelope analysis as a useful parameter in automatic phrase identification

Abstract

In scientific research on the speech signal recognition there can be noted great development, although differences between languages make it difficult to work out the same algorithms for all of them. A good example of the big progress in this field can be Windows XP, an operating system which enables controlling some applications by voice (but not in Polish). There is still lack of good working programs controlled by Polish. In this paper the results of investigations on the voice signal envelope are described. There were tested digital recordings, from the range 0 – 9, obtained for 50 persons of different age and sex. The main goal was to find out if the envelope analysis could be helpful in automatic speech recognition. During the investigations basing on the analysis of the digit time characteristic, each digit was divided into parts (from 2 to 5) having the similar envelope. Also the minimum duration and the amplitude range were found for each part. The results are given in Table 1. Table 2 contains the results of fitting the envelope to each digit. It is shown that the envelope patterns are common for all the speakers and digits. Although the envelope analysis is not sufficient alone for automatic speech recognition (some digit patterns fit to the others), it can be used as one of the parameters employed for this purpose.

Keywords: automatic speech recognition, envelope analysis.

1. Wstęp

Problem automatycznego rozpoznawania sygnałów mowy nie jest nowy. W Polsce już w latach pięćdziesiątych prowadzono prace zmierzające do określenia parametrów użytecznych w procesie rozpoznawania [1], lecz ówczesny stan techniki uniemożliwił szybki postęp w tej dziedzinie. Dopiero po wprowadzeniu komputerów do laboratoriów, co nastąpiło w końcu lat osiemdzie-

siątych XX w. na nowo zostały podjęte badania w tej dziedzinie [2, 3, 4]. Obecnie problematyka ta obecna jest na konferencjach i warsztatach międzynarodowych np. Interspeech, gdzie wymiennie są doświadczenia związane z wytwarzaniem, rozpoznawaniem i kodowaniem sygnałów mowy [5, 6, 7]. W polskojęzycznych opracowaniach temat ten pojawia się rzadko [8, 9] stąd postęp w tej dziedzinie nie jest duży. Niniejszy artykuł przedstawia wyniki badań nad możliwościami wykorzystania analizy obwiedni do celów identyfikacji wyrazów w języku polskim.

2. Sposób przeprowadzania badań

Badania zostały przeprowadzone z wykorzystaniem bazy nagrań „Corpora” opracowanej przez Politechnikę Poznańską oraz nagrań własnych. W sumie baza stanowiła 500 nagrań cyfr z zakresu 0-9, dla 50 osób. W przeprowadzonych testach postawiono pytania:

- 1) Czy jest możliwe opracowanie jednakowych wzorców obwiedni dla każdej z cyfr dla wszystkich mówców?
- 2) Czy opracowane wzorce (o ile istnieją) mogą być pomocne w procesie rozpoznawania sygnałów mowy?

Analizując kolejne nagrania, początkowo badano poziom sygnału (amplitudę) w czasie trwania całego nagrania danej cyfry i maksymalną jego wartość przyjmowano jako 100%. Przyjęcie procentowej skali do porównania poziomów sygnałów niezależnie od wyniku rozpoznawania od natężenia dźwięku z jakim dany mówca wypowiada dany wyraz. Wszystkie pozostałe poziomy sygnałów określano w % względem tej wartości. Czasy trwania poszczególnych fragmentów wypowiedzi, dla których określono poziom sygnału są czasami minimalnymi, gdyż każdy z mówców może niemal dowolnie wydłużać pewne fragmenty wypowiedzi (szczególnie będące samogłoskami). W ten sposób przebadano 500 nagrań, a wyniki pokazały, iż możliwe jest opracowanie uniwersalnych wzorców dla każdej z cyfr.

3. Wyniki badań

Poniżej przedstawiono uzyskane wzorce dla cyfr 0-9, które są spełnione dla przebadanej grupy 50-ciu mówców.

Tab. 1. Wzorce obwiedni dla cyfr 0-9
Tab. 1. Envelope patterns for digits 0-9

Cyfra	Fragmenty wypowiedzi [minimalny czas trwania/zakres amplitud]				
	1	2	3	4	5
0	80ms/ A<50%	30ms/ A>53%	10ms/ A>5-64%	50ms/ A>31%	-
1	60ms/ A<62%	70ms/ A>44%	50ms/ 1<A<26%	60ms/ A>25%	100ms/ A<60%
2	50ms/ A<30%	20ms/ A>29%	50ms/ A>43%	-	-
3	130ms/ 6<A<72%	40ms/ A>68%	-	-	-
4	30ms/ A>1%	40ms/ A<4%	100ms/ A>40%	10ms/ 3<A<35%	60ms/ A>13%
5	140ms/ A>29%	20ms/ A<23%	110ms/ >0%	-	-
6	80ms/ 2<A<53%	100ms/ A>26%	100ms/ 1<A<56%	70ms/ A>1%	-
7	100ms/ 3<A<65%	50ms/ A>36%	30ms/ 2<A<28%	110ms/ 6<A<59%	-
8	70ms/ A>46%	90ms/ 6<A<61%	100ms/ 16<A<92%	90ms/ 5<A<43%	-
9	40ms/ 1<A<45%	70ms/ A>46%	150ms/ A>9%	80ms/ A<50%	-

Gdzie: A-zakres amplitud, przy założeniu, że pierwszy fragment wypowiedzi rozpoczyna się nie później niż „x[ms]” od początku wypowiedzi, gdzie „x” dla poszczególnych cyfr wynosi 0-40ms, 1-40ms, 2-70ms, 3-0ms, 4-80ms, 5-80ms, 6-80ms, 7-10ms, 8-120ms, 9-150ms.

Jak widać w tabeli 1, wypowiedź każdej z cyfr została podzielona na fragmenty, w których został określony przedział amplitud występującego sygnału. Liczba fragmentów jest zależna od cyfry i może się wahać od 2 do 5 (dokładną liczbę fragmentów określono na podstawie analizy charakterystyki czasowej dla każdej z cyfr). Pozostaje jednak pytanie, czy wyniki te mogą być użyteczne w procesie rozpoznawania sygnałów mowy. Zobrazowanie ich użyteczności przedstawia tabela 2.

Tab. 2. Wyniki dopasowania obwiedni do wypowiedzi cyfr 0-9 dla 50-ciu osób
Tab. 2. Envelope fitting results for digits 0-9 for 50 persons

Wypowiedź	Dopasowanie wzorca obwiedni do poszczególnych cyfr [%]									
	0	1	2	3	4	5	6	7	8	9
0	100	2	98	70	4	40	8	6	0	78
1	80	100	78	56	4	78	6	58	22	92
2	48	0	100	56	0	2	6	6	0	26
3	34	0	70	100	0	8	6	0	0	14
4	88	12	78	36	100	55	4	12	0	76
5	14	4	34	4	4	100	10	0	12	58
6	42	6	96	82	0	70	100	2	0	74
7	94	90	98	84	0	52	20	100	2	90
8	2	10	4	18	0	44	2	0	100	34
9	86	62	98	52	16	84	58	46	8	100

Jak wynika z tabeli 2, różne są zgodności wzorców obwiedni poszczególnych cyfr z rzeczywistymi wypowiedziami. Dla każdej z cyfr jej zgodność ze wzorcem jest równa 100% (wyniki na głównej przekątnej w tabeli), lecz jest również częściowo zgodna z innymi wzorcami. Jak łatwo zauważyć występują jednak wykluczające się przypadki:

- z wypowiedzią cyfry „0” nie jest zgodny żaden wzorec z cyfry 8,
- z wypowiedzią cyfry „2” nie jest zgodny żaden wzorec z cyfr 1, 4, 8,
- z wypowiedzią cyfry „3” nie jest zgodny żaden wzorec z cyfr 1, 4, 7, 8,
- z wypowiedzią cyfry „4” nie jest zgodny żaden wzorec z cyfr 8,
- z wypowiedzią cyfry „5” nie jest zgodny żaden wzorec z cyfry 7,
- z wypowiedzią cyfry „6” nie jest zgodny żaden wzorec z cyfr 4, 8,
- z wypowiedzią cyfry „7” nie jest zgodny żaden wzorec z cyfr 4,

- z wypowiedzią cyfry „8” nie jest zgodny żaden wzorec z cyfr 4, 7.

Na podstawie wyżej poczynionych spostrzeżeń można stwierdzić, iż nie jest możliwe całkowite rozróżnienie wszystkich cyfr na podstawie analizy ich obwiedni, jednak daje ona cenne wsparcie przy testowaniu wypowiedzi, np. jeżeli wyniki zgodności obwiedni z cyfrą „3” są równe 100%, oznacza to iż na pewno nie jest to cyfra 1, 4, 7, 8, choć z coraz mniejszym prawdopodobieństwem może to być cyfra 2, 0, 9, 5, 6.

4. Podsumowanie

Przedstawione wyniki badań stanowią odpowiedź na postawione na wstępie pytania. Jak pokazują – jest możliwe określenie wzorców obwiedni dla cyfr 0-9 jednakowych dla wszystkich (50) mówców. Otrzymane rezultaty pozwalają również stwierdzić, iż analiza obwiedni może być wykorzystana jako jeden z parametrów do identyfikacji wypowiedzi.

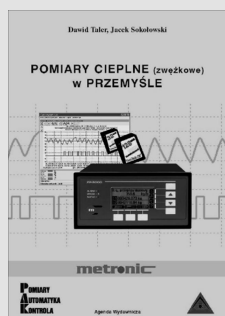
5. Literatura

- [1] M. Steffen: „Częstość występowania głosek polskich”, biuletyn PTJ XVI, Warszawa 1957r.
- [2] R. Tadeusiewicz: „Sygnał mowy”, WKŁ, Warszawa 1988r.
- [3] Z. Brysiak: „Lektor 4.0 - Mów do mnie jeszcze”, Magazyn Komputerowy CHIP, nr 10/1998.
- [4] Cz. Basztura: „Źródła, sygnały i obrazy akustyczne”, WKŁ, Warszawa 1988r.
- [5] A. Waibel, K. Bernardin, M. Wolfel: Computer-Supported Human-Human Multilingual Communication, Interspeech conference August 27-31, Antwerp, Belgium, p. 14-21, 2007.
- [6] J. Li, C.Li: Soft Margin Feature Extraction for Automatic Speech Recognition, Interspeech conference August 27-31, Antwerp, Belgium, p.30-33, 2007.
- [7] P. Lucey, G. Potamianos, S. Sridharan: A Unified Approach to Multi-Pose Audio-Visual ASR, Interspeech conference August 27-31, Antwerp, Belgium, p. 650-653, 2007.
- [8] A. Gołaś: Problemy sterowania dźwiękiem, 51st OSA, Gdańsk 2004.
- [9] A. Wojciechowska: Kompresja a zrozumiałość sygnału mowy, 51st OSA, Gdańsk 2004.
- [10] J. Dulas: Rozpoznawanie jednostek fonetycznych zawierających okresy podstawowe tonu kraniowego, PPM 2008, Sucha Beskidzka.

Artykuł recenzowany

INFORMACJE

Książka Wydawnictwa PAK



Książka „Pomiary cieplne (zwęzkowe) w przemyśle” przedstawia problematykę pomiarów strumienia masy i ciepła płynów przepływających w przewodach przy użyciu zwęzek pomiarowych. Książka przeznaczona jest dla inżynierów i techników zajmujących się zagadnieniami ciepło-przepływowymi w przemyśle, energetyce i ogrzewnictwie. W książce omówiono przyrządy i układy do pomiarów zwęzkowych strumienia ciepła, produkowane przez firmę Metronic.

Zamówienia prosimy składać na adresy PAK:

Wydawnictwo PAK
00-050 Warszawa, ul. Świętokrzyska 14A,
tel./fax: 022 827 25 40

Redakcja PAK
44-100 Gliwice, ul. Akademicka 10, p. 30b,
tel./fax: 032 237 19 45
e-mail: wydawnictwo@pak.info.pl