

2.23 — akustyka mowy

5/1984

Piotr Domagała

AUTOMATYZACJA PROCESU SEGMENTACJI
SYGNAŁU MOWY
W UKŁADZIE ANALOGOWO-CYFROWYM

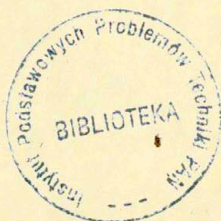
5/1984

P. 209



WARSZAWA 1984

Praca wpłynęła do Redakcji dnia 24 października 1983 r.



56953



N a p r a w a c h r ę k o p i s u

Instytut Podstawowych Problemów Techniki PAN
Nakład 140 egz. Ark.wyd. 1. Ark. druk. 1,5.
Oddano do drukarni w lutym 1984 r.
Nr zamówienia 111/84

Warszawska Drukarnia Naukowa, Warszawa,
ul. Śniadeckich 8

Piotr Domagała
Pracownia Fonetyki Akustycznej
IPPT PAN

AUTOMATYZACJA PROCESU SEGMENTACJI SYGNAŁU MOWY
W UKŁADZIE ANALOGOWO-CYFROWYM^{1/}

Streszczenie

Zbiór 18 wyrazów polskich wymówionych w studio bezpogłosowym przez 12 głosów (7 męskich i 5 kobiecych) zapisanych na taśmie magnetofonowej i zanalizowanych w układzie hybrydowym dokonującym cyfrowej analizy widma dynamicznego poddano automatycznej segmentacji implementując algorytm składający się z prostych zdań logicznych na minikomputerze MERA 303. W porównaniu z segmentacją wizualną przeprowadzoną na komputerowych obrazach spektrograficznych uzyskano poprawność segmentacji w około 94 % przypadków.

1. Wstęp.

Badania dotyczące automatycznej segmentacji sygnału mowy stanowią zazwyczaj wstępny etap złożonego procesu automatycznego rozpoznawania fonemów i w związku z tym mają istotne znaczenie w pracach badawczych nad automatycznym rozpoznawaniem mowy. Niniejsza praca jest rozwinięciem, modyfikacją i weryfikacją na liczniejszym zbiorze mówców metody przedstawionej w pracy [7]. Jako cel pracy przyjęto uzyskanie skutecznej metody segmentacji izolowanych wyrazów. Wypowiedziany wyraz jest dzielony na segmenty, które powinny możliwie najdokładniej odpowiadać jednostkom fonetycznym. Jako jednostkę fonetyczną przyjęto głoskę monosegmentalną lub jeden segment głoski polisegmentalnej [5,6]. Granice między segmentami uzyskanymi w sposób automatyczny

^{1/} Praca wykonana w ramach problemu węzłowego 06.9.

nazwano cezurami. W przypadku automatycznej segmentacji sygnału mowy pojęcie segmentu nabiera znaczenia interwału czasowego wyrazu (frazy), który może pokrywać się z pojedynczą jednostką fonetyczną, z kilkoma lub nie pokrywać się z żadną jednostką fonetyczną.

Należy wyodrębnić następujące przypadki :

- segmentacja właściwa,
- segmentacja dodatkowa (oversegmentation),
- brak segmentacji (undersegmentation).

Jako segmentację właściwą przyjmuje się pełną zgodność jednostki fonetycznej z automatycznie wyodrębnionym segmentem.

Segmentacja dodatkowa wystąpi w przypadku pojawienia się dwóch lub większej liczby segmentów w ramach pojedynczej jednostki fonetycznej.

Przypadek, w którym jeden segment obejmuje dwie lub więcej kolejnych jednostek fonetycznych, definiowany jest jako brak segmentacji.

2. Realizacja automatycznej segmentacji sygnału mowy (ASM).

Opracowana metoda charakteryzuje się prostotą i wysokim stopniem skuteczności, nie gorszym od wyników uzyskiwanych bardziej skomplikowanymi metodami (por. [1,2,3,4,9,10]).

Materiałem wyjściowym do ASM jest uśredniony spektrogram cyfrowy zapamiętany w pamięci operacyjnej minikomputera jako tablice o współrzędnych czasu i częstotliwości. Istota opracowanego algorytmu polega na tym, że dla każdej pary chwil czasowych $k-1$ i k oraz dla każdego kanału obliczono różnicę między wartościami poziomu sygnału. Jej znak odnotowywano w przypadku gdy bezwzględna wartość różnicy przekraczała pewien założony próg P . Na powyższej podstawie obliczono i zapamiętano dla każdej chwili k liczbę zmian monotoniczności, penetrując kolejno kanały od częstotliwości najwyższej do najniższej. Odnotowano również znak różnicy w kanale analizowanym jako ostatni.

Przykład :

Niech założony próg $P=4$

Niech spektrogram cyfrowy przyjmie postać przedstawioną poniżej :

Nr kanału			
4	15	+	20
3	14		12
2	10	+	15
1	20	-	10
	k- 1		k
			czas

Obliczona liczba jest równa 2 i znak różnicy w kanale analizowanym jako ostatni (nr 1) jest ujemny - przebieg malejący. Oznaczono jako $c(k)$ liczbę równą mocy zbioru ciągów binarnych o jednakowych wyrazach, będących wynikami analizy monotoniczności dla wszystkich kanałów w k -tej chwili czasu. Jako $z(k)$ przyjęto wartość wyrazów ostatniego ciągu według następującej zasady :

$$\begin{aligned} z/k/ &= 0 \quad \text{dla przebiegu rosnącego} \\ z/k/ &= 1 \quad \text{dla przebiegu malejącego} \end{aligned}$$

Przyjęto, że cezura zostanie położona w chwili k , w następujących przypadkach :

A/

$$c/k/ = 1 \wedge [c/k+1/ \neq 1 \vee [c/k+1/ = 1 \wedge z/k/ \neq z/k+1/]] \quad (1)$$

B/

$$c/k/ = c/k+i/ = 1 \wedge z/k/ = z/k+i/ \wedge [c/k+n/ \neq 1 \vee [c/k+n/ = 1 \wedge z/k/ \neq z/k+n/]] \quad (2)$$

gdzie : $i = 1, 2, \dots, n-1$

D/

$$c/k/ = 2 \wedge c/k-1/ \neq 1 \wedge c/k+1/ \neq 1 \wedge c/k+1/ \neq 2 \quad (3)$$

F/

$$c/k/ = 2 \wedge c/k-1/ \neq 1 \wedge c/k+1/ \neq 1 \wedge c/k+i/ = 2 \wedge z/k/ = z/k+i/ \wedge z/k/ \neq z/k+n/ \wedge c/k+n/ = 2 \quad (4)$$

gdzie : $i = 0, 1, 2, \dots, n-1$ oraz $n \gg 1$

G/

$$c/k/ = 2 \wedge c/k-1/ \neq 1 \wedge c/k+1/ \neq 2 \wedge c/k+i/ = 2 \wedge z/k/ = z/k+i/ \wedge c/k+n/ \neq 1 \wedge c/k+n/ \neq 2 \quad (5)$$

gdzie : $i = 0, 1, 2, \dots, n-1$ oraz $n \gg 2$

Powyższe zależności stały się formalną podstawą do komputerowej realizacji ASM. W porównaniu z pracą [7] zmieniono definicję cezury B oraz zrezygnowano z cezury C.

Konfiguracja sprzętowa została przedstawiona na rys. 1. Z przyczyn technicznych zakres analizowanych widm ograniczono do około 7 kHz wykorzystując 60 kanałów. Okres próbkowania wynosił 23 ms przy 5-milisekundowym czasie obiegu wszystkich kanałów. Po hardware'owej logarytmizacji przebiegu i konwersji a/c spektrogram został zapamiętany w pamięci operacyjnej, w relacji 1 bit \approx 0,6 dB. W następnym kroku dokonywano uśrednienia spektrogramu według następującej zasady /8/ :

$$A_{\text{sr}_{ik}} = \frac{A_{ik-2} + 3A_{ik-1} + 4A_{ik} + 3A_{ik+1} + A_{ik+2}}{8} \quad (6)$$

gdzie A_{ik} jest wartością próbki w i-tej chwili czasu i w k-tym kanale $i/k \geq 3/$.

Dla kanałów o numerach 1 i 2 utworzono lustrzane odbicie względem osi czasu i wzór (6) zmodyfikowano do następujących postaci :

$$A_{\text{sr}_{12}} = \frac{4A_{11} + 4A_{12} + 3A_{13} + A_{14}}{8} \quad (7)$$

$$A_{\text{sr}_{11}} = \frac{4A_{12} + 7A_{11} + A_{13}}{8} \quad (8)$$

Taki spektrogram może być już wyprowadzony przy pomocy wykonanej w Pracowni Fonetyki Akustycznej przystawki komputerowej zwanej memoskopem na ekran typowego odbiornika TV rys. 2 . Wielkość poziomu oznaczono na ekranie TV szerokością plamki stosując skalę 5-stopniową.

Na uzyskany w ten sposób spektrogramie przeprowadza się segmentację zgodnie z klasycznymi zasadami (5,6) celem uzyskania wzorca dla porównania i oceny automatycznej segmentacji, którą wykonujemy w następnym kroku.

W trakcie realizacji pracy okazało się, że istotne znaczenie dla jakości segmentacji ma wysokość dobranego progu P, o którym wspomniano poprzednio. Jeśli wartość progu jest za niska rzędu kilku bitów , to segmentacja jest nieskuteczna. Występuje duża liczba cezur dodatkowych oraz cezur brakujących. Widać to dobrze

na rys. 3 i rys. 4 przedstawiających słowo ŁAJANIE (głos JI). Zwiększanie wartości progu prowadzi do poprawy segmentacji (rys. 5,6,7), aż do uzyskania pewnej optymalnej postaci (rys. 8). Dalsze zwiększanie tego progu powoduje stopniowy zanik cezur (wzrost liczby brakujących cezur). Sytuację taką przedstawia rys. 9. W oparciu o liczne próby dobrano optymalną wielkość progu i zastosowano do całości eksperymentu. Uzyskane wyniki przedstawiono w tabelach.

3. Wyniki eksperymentu.

W tablicach 1 ÷ 6 przedstawiono wyniki automatycznej segmentacji. Kolumna 1 jest zestawieniem wypowiedzianych wyrazów, a kolumnę 2 tworzą ich transkrypcje fonetyczne. Znak "," oznacza tutaj zwarcie, a występujący po nim fonem /t,k,p/ fakt wystąpienia ploszji. W kolumnie 3 jest podana teoretyczna liczba cezur, nie uwzględniająca jednak występowania pewnych elementów fonetycznych, np. aspiracji, afrykacji itp. W kolumnie 4 przedstawiono wyniki analizy wizualnej spektrogramu otrzymanego na ekranie TV. Kolumna 5 podaje liczbę uzyskanych prawidłowych cezur. Cezury, których nie można było ustalić na ekranie TV, zostały wyszczególnione w kolumnie 6. Kolumna 7 jest zestawieniem braków segmentacji wraz z podaniem ich liczby, natomiast w kolumnie 8 wypisano segmentacje dodatkowe, podając również ich liczbę, np. : 3 ɸa,ɨ oznacza, że występują trzy nadsegmentacje, jedna między fonemami ɸ i i `a ` oraz dwie nadsegmentacje w ramach fonemu `ɨ`. W kolumnie 9 dokonano redukcji nadsegmentacji, która polegała na odrzuceniu tych dodatkowych cezur, jakie pojawiały się w następujących uzasadnionych przypadkach :

- w bezpośrednim sąsiedztwie cezury właściwej,
- w samogłosce nosowej / ẽ, õ /,
- w realizacji głoski " r ".

Liczbę nadsegmentacji w ramach jednego fonemu oznaczono odpowiednią liczbą kresek nad symbolem tego fonemu.

4. Analiza uzyskanych danych i wnioski końcowe.

4.1. Przedstawiona i zastosowana metoda ASM charakteryzuje się wysokim stopniem skuteczności segmentacji, nie gorszym od wyników uzyskiwanych za granicą [1,2,3,9,10]. Dodatkowo wyróżnia ją prostota i minimalne zapotrzebowanie na moc obliczeniową.

Ocena jakości segmentacji została przedstawiona na rys. 10.

4.2. Wyodrębnienie plozji w głoskach zwartych oraz zwarto-trących ma charakter przypadkowy. Związane jest to z faktem, że czas trwania plozji (ok. 10 ms) jest krótszy od okresu próbkowania.

4.3. W zależności od sposobu i tempa wypowiedzi pojawiały się podwójne granice między segmentami, np. AA, AB, BA. Wynika to z występujących kolejno zaników i wzrostów energii w pasmach.

4.4. Mówcy wypowiadali poszczególne wyrazy w sposób naturalny (dla siebie), a więc nie zawsze zgodnie z przyjętymi zasadami poprawnej wymowy, np. wyraz "gęsty" bywał wymawiany zarówno jako gęst¹, jak i gęst².

4.5. Uzyskano wysoki procent poprawnie rozsegmentowanych dyftongów. Na ogólną ich liczbę wynoszącą 96 rozdzielono poprawnie 80, co stanowi 83 %.

4.6. Do celów bieżących wprowadzono miarę skuteczności segmentacji

$$s = \frac{n_5}{2n_4 - n_5 + 0,3n_9} \quad (9)$$

gdzie n_i oznacza wartość pobraną z i-tej kolumny. Ponieważ do celów rozpoznawania fonemów nadsegmentacja dodatkowa jest mniej "szkodliwa" niż brak segmentacji, wprowadzono współczynnik korygujący 0,3. Najniższą wartość s uzyskał głos AM w wyrazie "chrząszcz" (0,41). Korygując wartość progu (podwyższenie o 10 bitów) uzyskano następujące rozsegmentowanie (por. Tab. 4, głos AM) :

	3	4	5	6	7	8	9
chrząszcz	6	5	4	1 tʃ	1xʃ	1ʃc	-

Widać, że skuteczność segmentacji wzrosła, a jej miara wynosi obecnie 0,67. Z powyższego wynika, że dobierając do każdego głosu nagrania indywidualny próg, można podwyższyć skuteczność segmentacji.

4.7. Nie zauważono różnic w jakości segmentacji między

głosami żeńskimi i męskimi (rys. 10)).

4.8. Dalsze prace badawcze i techniczne zmierzające do podwyższenia efektywności ASM powinny zmierzać w następujących kierunkach :

- redukcji parametrów fonetyczno-akustycznych,
- zwiększenia częstotliwości próbkowania,
- penetracji przebiegu sygnału mowy w pewnych wydzielonych pasmach częstotliwościowych.

Tablica 1.

	PU												JI				
	3	4	5	6	7	8	9	4	5	6	7	8	9				
wstrętny	10	10	10	-	-	-	-	-	9	8	1tr	1en	8	9			
wiedźma	7	6	6	1dz	-	1zm	-	5	5	2dz	-	-	-	-			
piaszczyna	10	9	8	1ts	1s	2s, a	2	10	10	-	-	-	2s, a	2			
chciałbym	8	7	7	1tp	-	2a, s	2	8	7	-	1tp	1pa	7	-			
właściwy	9	8	8	1tp	-	2p, v	1	9	7	-	2p	1w	1	1			
pastrągi	8	8	7	-	1, 1, 3	3r, a	1	7	6	1ps	1, 0, 3	-	-	-			
sklepy	7	6	6	1, 1, p	-	2s, e	2	6	5	1, 1, p	1e, 1	-	-	-			
sprawca	9	8	8	1, 1, p	-	2a, s	2	8	8	1, 1, p	-	1a	1	1			
barszczyk	9	8	8	1ts	-	3a, r	1	8	8	1ts	-	1a	1	1			
kształty	8	7	7	1, 1, t	-	3s, a, s	3	8	8	-	-	1ks	-	-			
listwy	6	6	6	-	-	2s	2	6	6	-	-	211, f	-	-			
płatkarz	8	7	7	1, 1, k	-	-	-	8	8	-	-	1a	-	-			
chrząszcz	6	5	5	1ts	-	2x, s, 0	1	6	6	-	-	4x, 0	3	3			
gęsty	5	5	5	-	-	3e	1	5	5	-	-	3e, e, s	1	1			
część	8	7	7	1ts	-	3e, e, p, 0	1	7	7	1ts	-	1e	1	1			
sąsiedzi	9	8	8	1ts	-	3o, 0, p	1	8	8	1ts	-	2o, pe	-	-			
pluskwa	7	6	6	1pl	-	3us, kf, fa	-	5	5	2, kf	-	-	-	-			
błądny	7	6	6	1, 1, d	-	1s	1	6	5	1, 1, d	1n	1s	2	2			
	441	427	455	47	2	37	21	429	422	42	7	23	42	42			

Tablica 2.

	WJ												LR				
	3	4	5	6	7	8	9	4	5	6	7	8	9				
1		2															
wstrętny	10	10	9	7	1tn 2 br	fs br	-	-	9	8	1, n	1tr	-				
wiedźma	7	6	5	1dz	1ma	2 v, a		2	6	6	1dz	1vj	-				
piaszczyczna	10	9	9	1tf	-	4 pwi, t, a		2	9	7	1tf	4pw, n	3				
chciałbym	8	8	8	-	-	1pa		-	8	8	-	1x	1				
właściwy	9	8	7	1, t	1iv	-		-	8	7	1tp	2v	2				
pstrągi	8	8	7	-	1tr	2 o, 3		1	7	7	1, t	-	-				
sklepy	7	7	7	-	-	3s, kl, k		2	7	7	-	4kl, i, e	2				
sprawca	9	7	7	2, 3	-	6s, sa, a		5	7	6	2, 3	1ra	2s, sa				
barszczyk	9	9	9	-	-	2a, r		1	8	8	1, t	-	3a, r, s				
kształty	8	7	7	1, t	-	1f		1	6	6	2, 3	-	2				
listwy	6	5	5	1, t	-	2i, f, k		1	6	5	-	1l	1				
plotkarz	8	8	7	-	1pw	2ka, a, f		-	8	7	-	3pw, a	2				
chrząszcz	6	5	5	1, t	-	3x, j, o, s		-	5	5	1, t	-	-				
gęsty	5	5	5	-	-	3g, es, t, k		1	5	4	-	3e, es	1				
częściej	8	6	5	2, 3	1ej	3e, p		2	6	5	2, 3	2 e, o	1				
sąsiedzki	9	8	8	1, t	-	4s, o, p		3	8	8	1, t	-	10 // o				
pluszka	7	7	7	-	-	4u, fa, a		3	6	6	1, k	-	1u				
biędny	7	7	6	-	1n, s	4e, dn, k		3	7	7	-	1bw	-				
	44/129/12/12				8	46		27	126	117	15	9	33				

Tablica 3.

	HK												PD				
	3	4	5	6	7	8	9	4	5	6	7	8	9				
1	2																
wstrętny	10	9	9	1st	-	-	-	-	9	1fs	-	3tn, n ^h , ^h	1				
wiedźma	7	6	5	1dz	1je	3ma, ^h	2	6	5	1dz	1je	3v, ^h , a	3				
piaszczyna	10	9	7	1ts	2 ^{wa} 2 ^{bx}	1zn	-	10	8	-	2ts, wa	1na	-				
chciałbym	8	7	7	1tp	-	2x, b ^h	1	7	7	1tp	-	5x, ^h , a, aw ^m	2				
właściwy	9	8	8	1tp	-	-	-	9	8	-	1vw	1v ^h	-				
pstrągi	8	7	5	1ps	2 ^{tr} 2 ^{of}	4s, r ^o , ^o , i	2	8	7	-	1tr	1 ^o j	-				
sklepy	7	6	6	1kl	-	1p ^h	-	6	6	1p	-	3kl, ^h	2				
sprawca	9	7	7	2 ^{eb}	-	2f, s ^h , a	1	8	8	1ts	-	4s, af, f, ^{aa}	2				
barszczyk	9	9	9	-	-	-	-	8	8	1ts	-	2ar, r ^h	-				
kształty	8	6	6	2 ^{kl} 2 ^{sl}	-	1 ^h	1	7	7	1ks	-	2s, ^h	2				
listwy	6	6	6	-	-	4ll, s ^h , f	3	6	5	-	1tf	1f ^h	-				
piotkarz	8	8	8	-	-	5o, a ^h , ^h	5	8	8	-	-	1k ^h , a, f	1				
chrząszcz	6	5	4	1ts	1xf	3 ^o , ^o , ^h	1	5	5	1ts	-	2s, ^o	1				
gęsty	5	5	5	-	-	3g, e ^h	1	5	5	-	-	3ge, ^h	2				
częściej	8	6	5	2 ^{ts} 2 ^{ep}	1pe	4ep, e ^h	3	6	5	2 ^{ts} 2 ^{ep}	1ej	6 ^o , ^o , ^o	1				
sąsiedzki	9	8	8	1ts	-	7 ^o , ^o , ^h , ki	3	8	8	1ts	-	1pe	-				
pluszka	7	6	6	1pl	-	3l, s, kf	2	7	6	-	1pl	-	-				
błądny	7	7	6	-	1, d	3e ^h , ^h	2	6	6	1, d	-	4bw, e, ni	3				
	44/125/117	16	8			46	27	129	121	16	8	43	10				

Tablica 4.

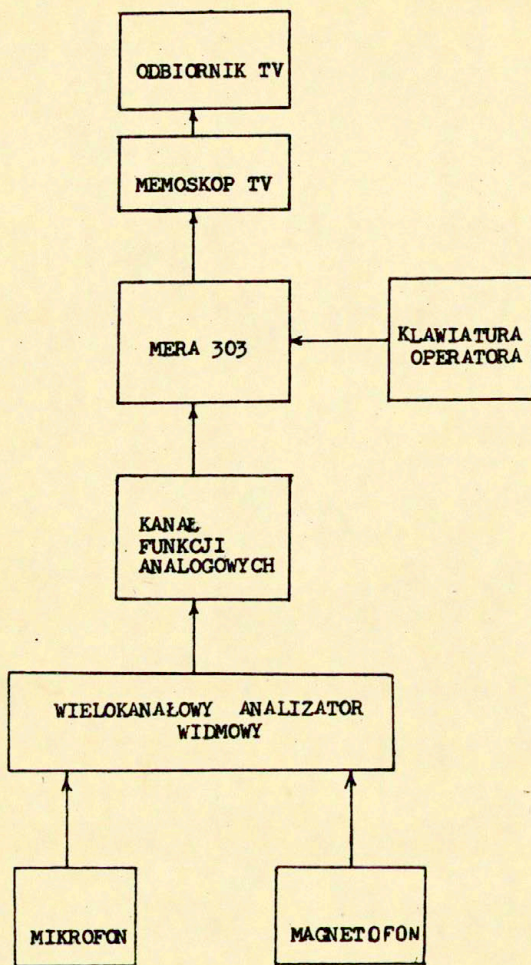
	AM															GD				
	1	2	3	4	5	6	7	8	9	4	5	6	7	8	9					
wstrętny		fstrenstni	10	8	8	2 ^{fs} by	-	1tn		-	10	8	-	1re	3r, tn ^z	1				
wiedźma		vjedzma	7	6	6	1dz	-	2e, z ^r		2	7	6	-	1je	1a	1				
piaszczyna		pwajszczyna	10	10	9	-	1wa	3af, sz, a		1	10	8	-	2 ^{wo} ll	1pw	-				
chciałbym		xstgambim	8	7	7	1tp	-	3x, pa		1	8	8	-	-	1pa	-				
właściwy		wnawstgure	9	8	7	1tp	1uw	4ap, st, i		1	7	7	2 st ll	-	2a, pl	1				
pstrągi		psstrągzi	8	6	6	2 ^{ps} tr	-	5, d, r, ō, i		3	8	7	-	1tr	2s, rō	1				
sklepy		sklepy	7	6	6	1>p	-	2z		2	6	6	1>p	-	-	-				
sprawca		soprawtsa	9	8	8	1>p	-	2r, f, b		-	8	7	1,p	1ra	2sf	2				
barszczyk		barszyczk	9	8	7	1tj	1yt	2ar		1	8	8	1tj	-	5r ^{ll}	2				
kształty		kjstawsté	8	6	6	2 ^{ta} te	-	3aw, z		2	8	7	-	1ta	2a, z	2				
listwy		listfē	6	6	6	-	-	2li, z		1	6	6	-	-	-	-				
plotkarz		pwostokas	8	7	7	1ka	-	2as, f		1	8	7	-	wo 1	3a, as	2				
chrząszcz		xšofotf	6	5	3	1tj	2 ^x z, t	2s, š		1	6	6	-	-	1š	-				
gesty		gēstē	5	5	5	-	-	5ē		1	5	4	-	1tž	2gē, es	-				
częście		fšoptgej	8	5	4	2 ^{tf} p	1ej	4j, ē		2	8	7	-	1tq	-	-				
sąsiedzki		sšeretszki	9	8	8	1tj	-	3op, p, p		1	8	8	1ts	-	3š, šp, p	1				
pluskwa		plusskfa	7	6	6	1kf	-	3lu, fa		2	7	6	-	1kf	2u a	2				
błądny		bwenodni	7	6	6	1yd	-	1d, e		2	7	5	-	2en, nō	-	-				
			141	122	115	20	6	49		24	135	122	6	13	30	15				

Tablica 5.

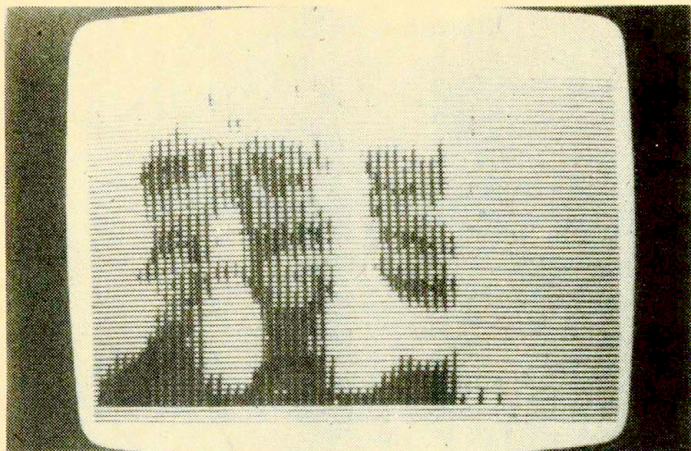
	ES										BS					
	3	4	5	6	7	8	9	4	5	6	7	8	9			
wstrętny	10	9	9	1fs	-	2r, tn	-	9	4	5	6	7	8	9		
wiedźma	7	6	5	1dz	1je	2vj, zm	-	5	5	5	2 _{ic} ^{vj}	-	3j, za	3		
piaszczyna	10	10	10	-	-	5 ^{pw, s, z, zn}	3	3	9	8	1tj	1±z	4 ^{pw, s, t, d}	2		
chciałbym	8	7	6	1tq	1wb	2p _a , a	1	1	7	6	1tq	1wb	2ca ⁿ	-		
wiśniowy	9	8	8	1tq	-	3 _{av} , v±	2	2	8	8	1tq	-	2v, ±	2		
pstrągi	8	7	7	1tr	-	3 ⁿ , j1	-	-	7	7	1tr	-	4 ⁿ , n, 1	2		
sklepy	7	7	7	-	-	4s, k, kl	3	3	7	6	-	1p±	4sk, s, ±	3		
sprawca	9	7	7	2 _{ts} ^{pv}	-	1sa	-	-	8	7	1ts	1pr	2se, a	1		
barszczyk	9	8	8	1tj	-	2t _s , ±	-	-	8	8	1tj	-	4 ⁿ , r, j, j	3		
kszały	8	8	7	-	1, t	2t±, ±-	1	1	7	6	1k _s	1ta	3(a, ±)	3		
listwy	6	6	6	-	-	2l1	-	-	6	6	-	-	5 _{ts, s, t, t, t}	2		
plotkarz	8	8	8	-	-	5 ^{pw, a, j}	3	3	8	8	-	-	2pw, ka	-		
chrząszcz	6	5	5	1tj	-	5x _s , o, o ⁿ	2	2	5	5	1tj	-	3 _s , s, o ⁿ	1		
gęsty	5	5	5	-	-	4es, t±	-	-	5	5	-	-	3ge, e, es	-		
częściec	8	6	6	2 _{ts} ^{ts}	-	4 _{ts} , e _{ts} , e _{ts}	1	1	6	6	2 _{ts} ^{ts}	-	5 ⁿ , e, j, j	3		
sąsiedzki	9	9	9	-	-	4 _{ts} , o _{ts} , e, ki	-	-	9	9	-	-	7 ⁿ , i, o _{ts} , o _{ts}	5		
pluskwa	7	6	6	1kf	-	4u, fa, a	2	2	7	7	-	-	5 _{ts} , u, s, t, i	2		
biędný	7	7	6	-	1n±	1dn	-	-	6	5	1bd	1en	-	-		
	141	129	125	12	4	55	18	18	127	121	14	6	60	32		

Tablica 6.

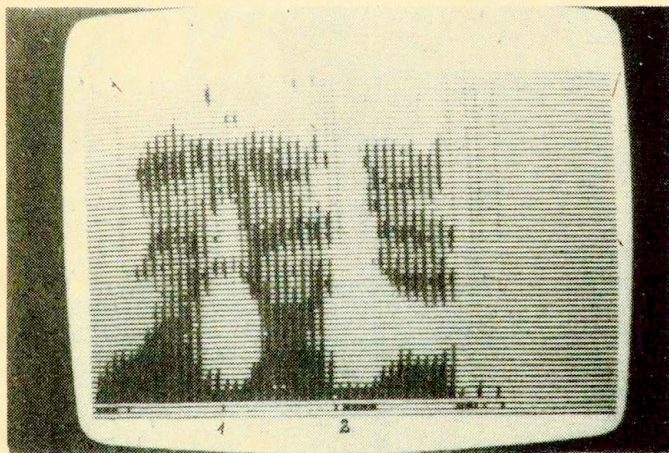
	DG										RC									
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
wstrętny		1strętnie	10	10	9	-	1n	2tn, z		1	10	10	-	-	5tr, n, tn	1				
wiedza		wiedza	7	6	5	1dz	1je	4vj, ma, z		2	6	5	1dz	1je	1v	1				
plaszczyna		plaszczyna	10	8	8	2 ^{wa} 2 ^z	-	3a, z, zn		2	9	8	1t ^s	1z	2e ^s , zn	-				
chciałbym		chciałbym	8	7	7	1t ^p	-	2x, w		2	7	6	1t ^p	1wb	3x, pa, a	2				
właściwy		właściwy	9	8	8	1t ^p	-	2v, a ^p		1	8	7	1t ^p	1vw	1 ^p	1				
ostrągi		ostrągi	8	7	5	1ps 2 ^{tr} 2 ^z	2s, i			2	7	5	1tr 2 ^z	2r, i		1				
sklepy		sklepy	7	7	7	-	1s			1	6	6	1p ^z	1le	-	-				
sprawca		sprawca	9	8	7	1ts	1pr	1s, a,		-	8	7	1ts	1pr	2re, af	-				
barszcyk		barszcyk	9	8	8	1t ^s	-	2r, k, w ^p		-	8	8	1t ^s	-	5d ^z k	3				
kształty		kształty	8	7	6	1k ^s 1 ^s	4 ^s , a, t ^z			3	5	5	3 ^z 3 ^z	-	-	-				
listwy		listwy	6	6	6	-	2l, li			1	6	6	-	-	3li, i, t ^f	1				
piotkarz		piotkarz	8	8	7	-	1ka	1pw		-	8	8	-	-	3pw, w, e ^s	1				
chrząszcz		chrząszcz	6	6	6	-	2 ^s , 2 ^z			1	4	4	2 ^z	-	3 ^z , o ^s , s	1				
gesty		gesty	5	5	5	-	-	4 ^z , e ^s , t ^z		1	4	4	1t ^z	-	-	-				
częściej		częściej	8	7	6	1t ^s	1ej	2 ^z , ^z		1	6	5	2 ^z	1ej	4 ^z , e ^z , e ^z	1				
sąsiedzki		sąsiedzki	9	8	8	1ts	-	6s ^z , e ^s , k ⁱ		-	8	8	1ts	-	4 ^z , e ^s , s	1				
pluszka		pluszka	7	7	7	-	-	7 ^z , u ^s , 2 ^z , f		3	7	7	-	-	5 ^z , u ^s , f	1				
błądny		błądny	7	7	7	-	-	2bw, z		1	7	7	-	-	3b, bw, w	1				
			44	130	122	11	8	49		22	124	116	17	8	47	16				



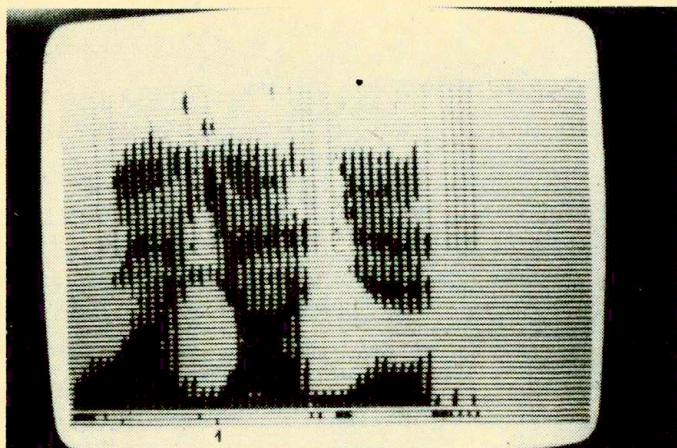
Rys.1. Schemat blokowy układu do ASM.



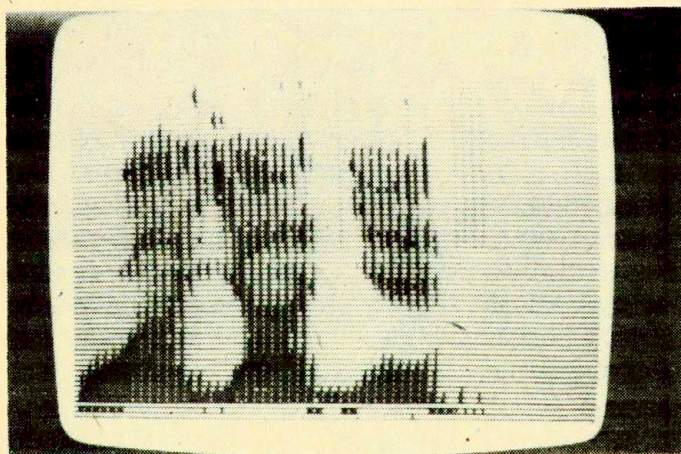
Rys.2. Uśredniony spektrogram cyfrowy wyrazu "łajanie" /wajane/. Głos JI.



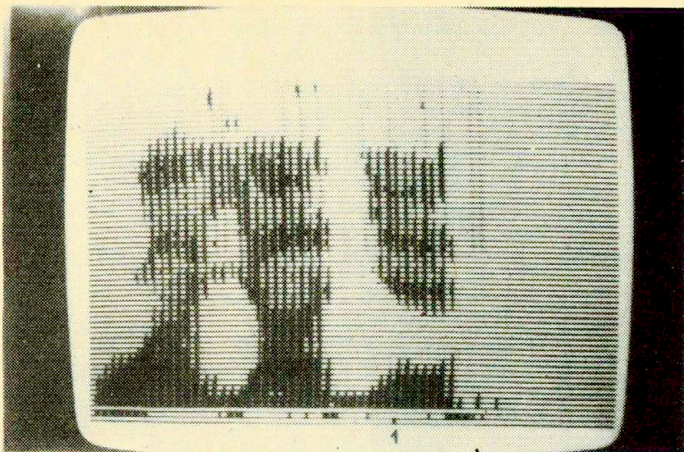
Rys.3. Automatyczna segmentacja wyrazu "łajanie" /wajane/. Głos JI. Próg $p = 0$ bitów.
1 - cezura typu A, 2 - cezura typu B.



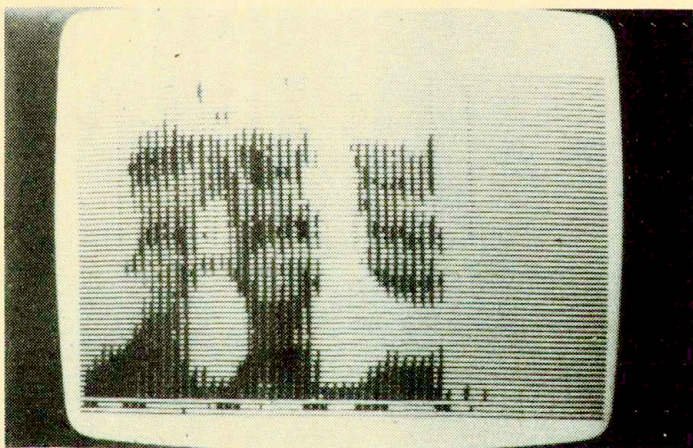
Rys.4. Automatyczna segmentacja wyrazu "łajanie"
/wajane/. Głos JI. Próg $p = 3$ bity.
1 - cezura typ D.



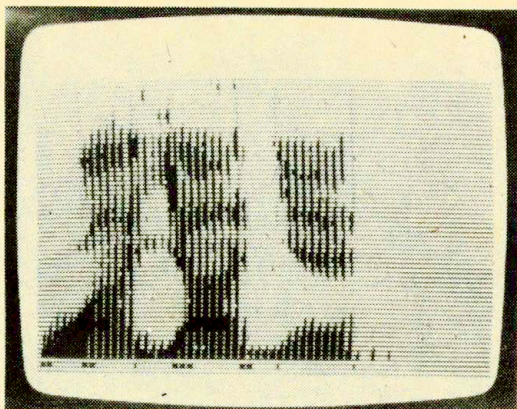
Rys.5. Automatyczna segmentacja wyrazu "łajanie"
/wajane/. Głos JI. Próg $p = 14$ bitów.



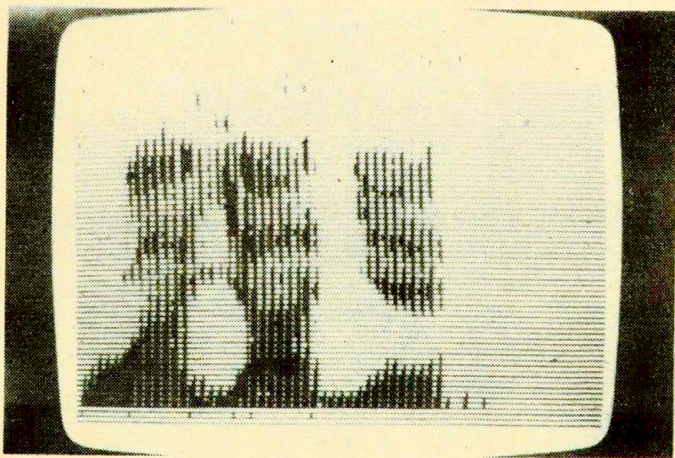
Rys.6. Automatyčna segmentacja wyrazu "łajanie"
/wajane/. Głos JI. Próg $p = 31$ bitów. 1-cezura F



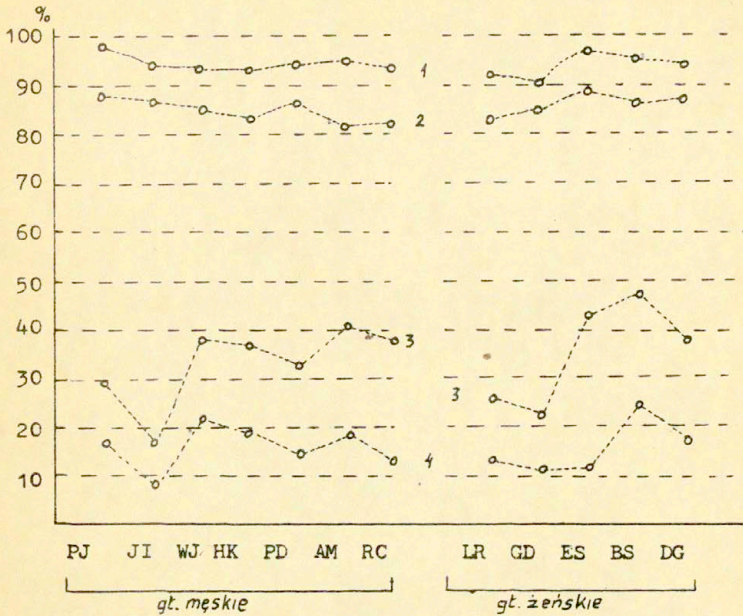
Rys.7. Automatyčna segmentacja wyrazu "łajanie"
/wajane/. Głos JI. Próg $p = 63$ bity.



Rys.8. Automatyczna segmentacja wyrazu "łajanie"
/wajane/. Głos JI. Próg $p = 95$ bitów.



Rys.9. Automatyczna segmentacja wyrazu "łajanie"
/wajane/. Głos JI. Próg $p = 111$ bitów.



- 1 - $\frac{\text{liczba automatycznie uzyskanych cezur}}{\text{teoretyczna liczba cezur wg TV}} \times 100\%$ wartość średnia = 94 %
- 2 - $\frac{\text{liczba automatycznie uzyskanych cezur}}{\text{teoretyczna liczba cezur}} \times 100\%$ wartość średnia = 85 %
- 3 - $\frac{\text{całkowita liczba segmentów dodatkowych}}{\text{teoretyczna liczba cezur wg TV}} \times 100\%$ wartość średnia = 34 %
- 4 - $\frac{\text{zredukowana liczba segmentów dodatkowych}}{\text{teoretyczna liczba cezur wg TV}} \times 100\%$ wartość średnia = 15 %

Rys.10. Podsumowanie wyników ASM.

BIBLIOGRAFIA

- [1] BUISSON, L., MERCIER, G., GRESSER, J., QUERRE, J., VIVES, R. :
Phonetic decoding for automatic recognition of speech, w :
Proceedings of the speech Communication Seminar,
Stockholm, 1974.
- [2] FRERICH, R., BROUWER, P. : Spectral analysis and segmenta-
tion in speech recognition new developments and
experiences, w : Proceedings of the speech Communica-
tion Seminar, Stockholm, 1974.
- [3] HATON, J.P. : Current work in segmentation and analitical
recognition of speech, w : Proceedings of the speech
Communication Seminar, Stockholm, 1974.
- [4] IZWORSKI, A., TADEUSIEWICZ, R. : Metody analizy skupień
w segmentacji i rozpoznawaniu prostych elementów
mowy polskiej, Prace XXX Otwartego Seminarium z
Akustyki, Gdańsk, 1983.
- [5] JASSEM, W. : Podstawy fonetyki akustycznej, PWN,
Warszawa, 1973.
- [6] JASSEM, W. : Fonetyczno-akustyczne założenia automatycznego
rozpoznawania fonemów, Prace IPPT 14/1970, Warszawa,
1970.
- [7] JASSEM, W., KUBZDELA, H., DOMAGAŁA, P. : Segmentacja sygnału
mowy na podstawie zmian rozkładu energii w widmie,
Prace IPPT 13/1983, Warszawa, 1983.
- [8] KUBZDELA, H. : Metoda wyznaczania położenia formantów
w widmie sygnału mowy, Prace XXX Otwartego Seminarium
z Akustyki, Gdańsk, 1983.
- [9] LIENARD, J.S., MLOUKA, M., MARIANI, J.J., SAPALY, J. : Real
time segmentation of speech, w : Proceedings of the
speech Communication Seminar, Stockholm, 1974.
- [10] RUSKE, G., SCHOTOLA, T. : The Efficiency of Demisyllable
Segmentation in the Recognition of Spoken Word, w :
Automatic Speech Analysis and Recognition, red.
J.P.Haton, D.Reidel Publishing Company, Dordrecht
Holland , 1981.