

2.23 — akustyka

**Henryk Kubzdela**

**WSTĘPNA OCENA REPREZENTACJI  
SYGNAŁU MOWY W FORMIE CIĄGÓW  
WEKTORÓW DWUBAJTOWYCH  
WYZNACZONYCH NA PODSTAWIE  
FUNKCJI WAGI WIDMA**

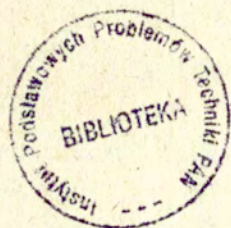
17/1989

P.269



**WARSZAWA 1989**

Praca wpłynęła do Redakcji dnia 29 listopada 1988 r.



56766



Na prawach rękopisu

---

Instytut Podstawowych Problemów Techniki PAN

Nakład 140 egz. Ark.wyd. 1 Ark.druk. 1,25

Oddano do drukarni w lutym 1989 r.

Nr zamówienia 245/89

---

Warszawska Drukarnia Naukowa, Warszawa,  
ul. Śniadeckich 8



Wstępna ocena reprezentacji sygnału mowy  
w formie ciągów wektorów dwubajtowych  
wyznaczonych na podstawie funkcji wagi widma

S t r e s z c z e n i e

W ramach poszukiwań prostych form reprezentacji wypowiedzi wyrazów zbadano wstępnie trzy koncepcje. Łączy je wspólna cecha, którą jest dynamiczna redukcja widma oparta o zmodyfikowane pojęcie wagi widma. Przekształcenie redukcyjne polega na przemartościowaniu składowych widma wygładzonego zgodnie z przebiegiem pewnego zmiennego progu, określonego jako próg maskowania. Pierwsze dwie koncepcje charakteryzują się tym, że reprezentację wypowiedzi wyrazu stanowi obraz binarny o szerokości szesnastu punktów dwuwartościowych. Według pierwszej koncepcji obraz ten powstaje na podstawie dwóch kryteriów - pierwszego zwanego kryterium wypukłości obwiedni widma i drugiego określonego jako kryterium progu maskowania. Według drugiej koncepcji podstawę przy tworzeniu obrazu binarnego stanowi jedynie drugie kryterium. Badane obrazy wyrażone były przez ciąg wektorów dwubajtowych. Trzecia koncepcja zakłada jako reprezentację wypowiedzi wyrazu ciągu liczb wyrażających punkty założonego podziału widma wygładzonego, zredukowanego. Wstępna weryfikację pierwszych dwóch koncepcji przeprowadzono w drodze pomiarów odległości obrazów binarnych dziesięciu wypowiedzi dwoma głosami dwóch podobnych i popularnych w automatycznym rozpoznawaniu wyrazów jeden i siedem. W 24 seriach pomiarowych zbadano odległości obrazów dla różnych kombinacji głosowych i dla różnych wartości współczynnika wyznaczającego zmienny próg maskowania. Na podstawie uzyskanych wyników stwierdza się wyższość reprezentacji tworzonej za pomocą jedynie kryterium progu maskowania. Obraz wypowiedzi według tego kryterium uzyskuje się łatwiej niż obrazy binarne stosowane w automatycznym rozpoznawaniu wyrazów za pomocą modelu ROWBIR [6]. Koncepcja reprezentacji wypowiedzi wyrazu w formie pary ciągów liczb wyrażających punkty założonego podziału widma nie spełnia oczekiwań. Okazało się, że ciągi takie nie wyrażają jednoznacznie wypowiedzi wyrazu i nie mogą pełnić roli jej reprezentacji w automatycznym

## rozpoznawaniu wyrazów.

### 1. WSTĘP

Praca obecna stanowi kontynuację wcześniejszych badań autora nad przydatnością określonego typu reprezentacji sygnału mowy w rozpoznawaniu wypowiedzi wyrazów [3],[4],[5],[7]. Rozpatrywane w nich reprezentacje sygnału mowy charakteryzowały się zawsze tym, że wyrażały pewne cechy widmowe w formie obrazu o strukturze binarnej. Badane wersje tego typu parametryzacji różniły się kryteriami tworzenia obrazu binarnego z widm dyskretnych 63 punktowych, wygładzonych przy użyciu specjalnej funkcji wygładzającej. W rezultacie szerszych rozważań przyjęto kryterium, które uzależniało wartość punktu tworzonego obrazu binarnego od charakteru obwiedni widma w odnośnym zakresie częstotliwości. Później wprowadzono umowne pojęcie funkcji wagi widma i jej granicy określonej przez założoną wartość tej funkcji. Funkcję wagi widma wyznaczano z widma uprzednio wygładzonego a ustalenia wartości granicznej dokonano na podstawie efektywności rozpoznawania wyrazów. Obraz binarny tworzono jedynie w odniesieniu do tych punktów widma, które przypadały poniżej wspomnianej granicy. Na podstawie wyników testów automatycznego rozpoznawania wypowiedzi wyrazów uznano wyższość reprezentacji sygnału mowy utworzonej przy wykorzystaniu umownego pojęcia funkcji wagi widma. Dla obrazu binarnego nie ograniczonego za pomocą funkcji wagi widma wprowadzono wcześniej redukcję szerokości obrazu i zbadano wpływ jej na efektywność rozpoznawania wyrazów. Okazało się, że zastosowanie odpowiedniego przekształcenia pierwotnego obrazu binarnego o szerokości 63 punktów w obraz czterokrotnie węższy nie wpłynęło negatywnie na efektywność rozpoznawania. Nie użyto dotychczas analogicznej redukcji szerokości w stosunku do obrazu ograniczonego za pomocą funkcji wagi widma. Równocześnie z zamiarem zbadania wpływu jej użycia na efektywność rozpoznawania postanowiono zastosować nową zasadę tworzenia obrazu binarnego, znacznie prostszą od dotychczasowej. Poszukiwanie nowych zasad tworzenia reprezentacji sygnału mowy w formie obrazu binarnego inspirowane było dążeniem do uproszczenia i przez to przyspieszenia procesu rozpoznawania wyrazów. Wśród różnych koncepcji wychodzących naprzeciw takiemu



dążeniu znalazła się koncepcja zbadania wartości dyskryminacyjnej parametrów określających punkty założonego podziału widma ważonego.

## 2. CEL PRACY I ZAŁOŻENIA

Głównym celem pracy było uzyskanie wstępnej odpowiedzi na pytanie, czy dla potrzeb automatycznego rozpoznawania wyrazów reprezentantem wypowiedzi może być obraz utworzony przy użyciu funkcji wagi widma i wyrażony przez ciąg wektorów dwubajtowych. Założono zbadanie dwóch typów wektorów. Wektor każdego z tych typów wyznaczano przy użyciu zmodyfikowanej funkcji wagi widma. Funkcja ta określa poziom maskowania poszczególnych elementów wygładzonego widma amplitudowego. O wysokości tego poziomu w danym punkcie widma decyduje suma składowych po lewej stronie tego punktu. Zasadę tę wyraża wzór:

$$pm_i = k \sum_{j=1}^{i-1} s_j \quad (1)$$

w którym  $s_j$  oznacza składową widma z przedziału  $(1, \dots, i-1)$ ,  $i$  jest numerem składowej, której dotyczy poziom maskowania  $pm_i$ ,  $k$  jest założonym współczynnikiem maskowania.

Założono, że wektor pierwszego typu będzie miał dwa parametry 1-bajtowe, co odpowiada szesnastu parametrom binarym uzyskanym w wyniku przekształcenia ciągu 63 liczb binarnych. Rozpatrywano dwa rodzaje binaryzacji prowadzące do uzyskania takiego ciągu. Pierwszy polegał na binarnej klasyfikacji poszczególnych punktów widma dyskretnego na podstawie dwóch kryteriów; pierwszego, uzależniającego przynależność punktu do danej klasy od istnienia lub braku odpowiedniej wypukłości obwiedni widma wygładzonego w tym punkcie i drugiego, uzależniającego tę przynależność od położenia obwiedni widma względem poziomu maskowania. Drugi rodzaj binaryzacji opierał się jedynie na drugim z wymienionych kryteriów. Redukcja długości ciągu z 63 liczb binarnych do 16 następowała według zasady podanej w pracach [5] i [6]. Postanowiono przeprowadzić

badania dla trzech wartości współczynnika maskowania  $k$ . Podstawą ich przyjęcia były informacje zaczerpnięte z pracy [7]. Przedmiotem oceny było zatem 6 wariantów reprezentacji każdej wypowiedzi wyrazu. Każdy wariant wynikał z jednego z dwóch przyjętych rodzajów binaryzacji oraz jednej z trzech wartości współczynnika  $k$ .

Założono, że parametrami wektora drugiego typu będą numery punktów założonego podziału części widma wygładzonego, przypadającej powyżej poziomu maskowania. Obliczenie takiego parametru wymaga więc wpiery wyznaczenia poziomu maskowania według wzoru (1), wyzerowania rzędnych widma w punktach, w których nie sięgają one poziomu maskowania i zmniejszenia pozostałych rzędnych odpowiednio o wartość tego poziomu. Następnie obliczona zostaje suma  $S_p$  rzędnych tak zredukowanego widma. Poszukiwanym parametrem  $n$  jest numer punktu widma, od którego wzwyż spełniona jest nierówność:

$$S_p \cdot n \geq S_c, \quad (2)$$

gdzie  $S_c$  oznacza sumę częściową rzędnych widma w punktach od -1 do  $n$ .

Wielość badanych reprezentacji wypowiedzi wyrazu oraz ukierunkowanie zainteresowania na podobieństwo obrazów różnych wypowiedzi tego samego wyrazu oraz wypowiedzi wyrazów podobnych zadecydowały o wyborze do badań małej liczby wyrazów i głosów. Założono, że przedmiotem badań powinny być reprezentacje wypowiedzi wyrazów podobnie brzmiących i jednocześnie popularnych w automatycznym rozpoznawaniu. Ten warunek spełniają wyrazy jeden i siedem. Postanowiono, iż przedmiotem badań będą obrazy wypowiedzi tych wyrazów jednym głosem męskim i jednym żeńskim.



### 3. SPOSÓB PRZEPROWADZENIA BADAŃ

Badania przeprowadzono za pomocą układu analogowo - cyfrowego. Część analogową oraz fragment interface'u tworzyły urządzenia przedstawione szerzej w pracach :[1],[2] i [6]. Część cyfrową stanowił mikrokomputer spectrum + wyposażony w pamięć dyskową, monitor ekranowy oraz drukarkę. Między innymi dla celów tej pracy skonstruowano przystawkę rozszerzającą istniejący interface wcześniejszej konfiguracji analogowo-cyfrowej o elementy umożliwiające dołączenie mikrokomputera spectrum +. W języku asemblera mikroprocesora Z80 ułożono programy następujących operacji:

- wpisu danych z analogowego analizatora widma z równoczesnym wyglądaniem widm,
- wyznaczania obrazów binarnych według dwóch wyżej podanych zasad,
- wizualizacji obrazów wypowiedzi na ekranie monitora,
- porównywania obrazów binarnych,
- wyznaczania punktów założonego podziału zredukowanego widma.

Programy tych operacji były procedurami asemblerowymi, do których odwoływały się 3 różne programy główne napisane w basicu. Każdy z nich zarządzał pewnymi spośród następujących operacji:

- wpis danych z analizatora widma,
- wyznaczenie obrazu binarnego według podanej zasady,
- wizualizacja kontrolna obrazu wypowiedzi,
- przesłanie spektrogramu cyfrowego i zbioru obrazów binarnych do/z pamięci dyskowej z/do komputera.
- porównanie wskazanych obrazów binarnych i wizualizacja na ekranie lub wydruk wyniku porównania,
- znalezienie punktów zadanego podziału widma i wizualizacja na ekranie lub wydruk ich numerów.

Przebieg badań był następujący: Każdy z wybranych do badań wyrazów wymówiony został 10-krotnie przez każdy z dwóch głosów. Równocześnie z wymawianiem wyrazu dokonywała się analiza widmowa i następował wpis jej wyniku do komputera. Tą drogą wypowiedź wyrazu została wyrażona przez ciąg liczb stanowiących rzędne 63- punktowych widm wyglądanych. Ciąg taki otrzymywał etykietę i w formie zbioru był przepisany na dyskietkę. Po zebraniu w tej postaci wszystkich 20 wypowiedzi jednym głosem



ponownie każdy zbiór przesyłano dwukrotnie do komputera. Za każdym razem następowało wyznaczenie obrazów binarnych według innej zasady. Stosując jedną zasadę binaryzacji wyznaczano dla każdej wypowiedzi 3 obrazy, odpowiednio do trzech założonych wartości współczynnika  $k$  występującego we wzorze (1). Stosując jedną zasadę binaryzacji utworzono dla każdego głosu 60 obrazów binarnych rozpatrywanych wypowiedzi. Obrazy te w zbiorach po 15 elementów zgromadzono na dyskietce. Tyle samo obrazów podobnie uporządkowanych uzyskano dla tych samych wypowiedzi stosując drugą zasadę binaryzacji. Kolejnym etapem było porównanie obrazów wyznaczonych według tej samej zasady i przy użyciu tej samej wartości współczynnika  $k$ . Spośród tych obrazów wytypowano dwa dotyczące różnych wyrazów i porównano z nimi pozostałe. Tak więc obraz wypowiedzi danego wyrazu był porównywany z obrazem innej wypowiedzi tego samego wyrazu i z obrazem wypowiedzi wyrazu podobnego. Wynikiem porównania testowanego obrazu z parą wybranych obrazów była bezwzględna odległość od bliższego z tych dwóch obrazów oraz wyrażona względem niej odległość od dalej położonego obrazu. Przykładowy wydruk tego rodzaju wyników porównań osiemnastu testowanych obrazów zamieszczono poniżej pod nazwą wydruk nr 1. Jako miarę odległości przyjęto średnią liczbę punktów na widmo testowanego obrazu, w których treść obu porównywanych obrazów była różna. Przeprowadzono 24 serie porównań. W każdej serii następowało porównanie 20 obrazów z parą obrazów wytypowanych. Wśród 24 serii połowę stanowiły porównania obrazów wypowiedzi różnymi głosami a połowę porównania obrazów wypowiedzi tym samym głosem.

Wyznaczenie punktów założonego podziału widma zredukowanego do części przypadającej powyżej umownego poziomu maskowania przeprowadzono dla trzech wypowiedzi każdego z dwóch wyrazów przez każdy z dwóch głosów. Ustalono wpiery szacunkowo, jakiej wartości podział widma 63-punktowego (zredukowanego) przypada w dających się określić miejscach i zrezygnowano z niektórych spośród ośmiu założonych wartości podziału.



Wydruk nr 1. Przykładowe odległości pomiędzy obrazami wypowiedzi. Obrazy wyznaczono według kryterium poziomu maskowania przy wartości współczynnika  $k = 1/32$ .

I1Z R		
1 -	14.	2.28
1 -	13.	1.92
1 -	13.	2.38
1 -	12.	2.08

II1Z R		
1 -	12.	2.33
1 -	12.	2.33
1 -	18.	1.83
1 -	31.	1.22
1 -	11.	2.9

I7Z R		
7 -	19.	1.57
7 -	23.	1.43
7 -	16.	1.87
7 -	19.	1.47

II7Z R		
7 -	11.	2.36
7 -	21.	3.8
7 -	19.	1.57
7 -	18.	1.61
7 -	20.	1.5

W częściach wydruku otoczonych ramką w poszczególnych kolumnach podane są:

w kolumnie pierwszej - etykieta wyrazu, którego obraz porównawczy Op był najbliższy do obrazu testowanego Ot,

w kolumnie drugiej - odległość  $d$  Ot od Op,

w kolumnie trzeciej - stosunek odległości obrazu testowanego od dalej położonego obrazu porównawczego do odległości  $d$ .

Poszczególne wiersze odnoszą się do porównań obrazów różnych wypowiedzi wyrazu jeden (dwie pierwsze ramki) i wyrazu siedem (następne ramki) z obrazami porównawczymi. Przykład dotyczy wypowiedzi głosem żeńskim.

#### 4. WYNIKI I ICH OCENA

Na podstawie porównania jednego typu obrazów dziewięciu wypowiedzi danego wyrazu z obrazem dziesiątej wypowiedzi tego wyrazu i jednej z wypowiedzi drugiego wyrazu określono:

- liczbę przypadków negatywnych PNg, w których odległość od obrazu wypowiedzi tego samego wyrazu jest większa niż odległość od obrazu wypowiedzi drugiego wyrazu,
- liczbę przypadków nieokreślonych PNo, w których odległość testowanego obrazu od obrazu porównawczego tego samego wyrazu i wyrazu obcego jest identyczna lub większa od wartości dopuszczalnej,
- zakres odległości  $\Delta d$  oraz odległość średnia  $d_{sr}$  pomiędzy obrazami wypowiedzi tego samego wyrazu z wykluczeniem przypadków określonych jako negatywne i nieokreślone,
- zakres względnej odległości  $\Delta D$  oraz średnia odległość względna  $D_{sr}$  pomiędzy obrazami wypowiedzi różnych wyrazów z pominięciem przypadków negatywnych i nieokreślonych.

Dane uzyskane dla poszczególnych serii porównań zamieszczono w tablicach 1 i 2. Znaczenia symboli użytych w tych tablicach są następujące:

$k$  - współczynnik maskowania ze wzoru (1),



MB - metoda binaryzacji, której dotyczą wyniki,

pm - metoda binaryzacji na podstawie kryterium poziomu maskowania,

pm + wo - metoda binaryzacji na podstawie kryteriów poziomu maskowania i wypukłości obwiedni widma,

W - wyraz,

T - głos wypowiedzi testowanej,

P - głos wypowiedzi porównawczej,

M - męski, Z - żeński.

Tablica nr 1. Wyniki porównań obrazów wypowiedzi tym samym głosem.

T	P	MB	W	k	PNg	FNc	Δd	d <sub>śr</sub>	ΔD	D <sub>śr</sub>
M	M	pm + wo	1	1/32	0	0	15-30	25.2	1.03-2.66	1.46
				1/25	0	0	14-30	23.2	1.08-3.19	1.71
				1/16	2	2	16-26	20.4	1.1 -3.8	1.96
		7	1/32	0	0	20-29	24.7	1.04-1.28	1.2	
			1/25	2	0	18-27	22	1.03-1.44	1.26	
			1/16	1	0	13-18	15.5	1.17-6.15	2.64	
M	M	pm	1	1/32	0	0	14-25	20.3	1.2 -1.64	1.44
				1/25	0	0	11-21	17.5	1.31-5.33	2.60
				1/16	1	0	13-21	18.1	1.1 -4.44	1.86
		7	1/32	1	0	17-25	21.3	1.04-3.19	1.56	
			1/25	1	0	17-30	20.3	1.04-2.66	1.42	
			1/16	1	0	10-21	14	1.2 -8.	3.56	
Z	Z	pm + wo	1	1/32	0	0	14-21	18.6	1.4 -2	1.8
				1/25	0	0	13-26	19.6	1.54-2.19	2.03
				1/16	0	0	13-24	18.7	1.25-2.26	1.72
		7	1/32	0	0	15-27	23.5	1.18-1.86	1.36	
			1/25	0	0	18-27	23.5	1.2 -4.44	1.92	
			1/16	0	0	13-26	18.5	1.11-2.38	1.72	
Z	Z	pm	1	1/32	0	0	11-31	15.1	1.22-2.9	2.14
				1/25	0	0	10-31	16.1	1.25-2.5	1.96
				1/16	0	0	9-29	16.5	1.62-2.75	2.04
		7	1/32	0	0	11-23	18.4	1.5 -3.8	1.90	
			1/25	0	0	14-22	18.4	1.45-5.71	2.32	
			1/16	0	0	12-22	15.7	1.35-2.25	1.82	



Tablica nr 2. Wyniki porównań obrazów wypowiedzi różnymi głosami.

T	P	MB	W	k	PNg	PNo	Δd	d <sub>śr</sub>	ΔD	D <sub>śr</sub>
M	Z	pm + wo	1	1/32	1	4	40-63	45.4	1.26-2.22	1.82
				1/25	2	4	31-51	42	1.56-2.58	1.97
				1/16	4	6	-	-	-	-
			7	1/32	3	0	30-40	33.4	1.12-1.38	1.37
				1/25	2	0	27-36	29.7	1.02-2.35	1.33
				1/16	0	0	21-35	26.4	2.28-3.63	3.1
M	Z	pm	1	1/32	1	1	19-58	37.6	1.28-2.28	1.84
				1/25	0	0	16-50	40.8	1.59-2.66	1.86
				1/16	0	0	20-43	29	1.09-2.16	1.68
			7	1/32	4	1	28-39	33.5	1.05-1.28	1.14
				1/25	2	0	29-38	32.6	1.05-1.2	1.13
				1/16	0	0	18-29	22.7	1.22-4.44	1.83
Z	M	pm + wo	1	1/32	0	0	28-43	32.9	1.11-2.5	1.73
				1/25	0	3	26-34	30.6	1.15-3.07	1.88
				1/16	0	1	27-36	30.6	1.05-2.75	1.33
			7	1/32	2	0	38-44	42	1.04-1.9	1.65
				1/25	0	0	32-52	38.2	1.2 -2.35	1.89
				1/16	0	1	26-37	33.6	1.3 -2.35	2.20
Z	M	pm	1	1/32	0	1	30-38	33	1.02-2.16	1.24
				1/25	1	2	28-37	31.2	1.09-2.16	1.32
				1/16	1	1	26-35	30.8	1.03-2.75	1.43
			7	1/32	0	0	31-43	35.5	1.02-2.35	1.80
				1/25	0	0	30-47	34.4	1.5 -2.66	2.24
				1/16	0	1	22-37	29.5	1.19-3.63	2.58

Zbrane wyniki pozwalają ocenić siłę dyskryminacyjną obrazów wyznaczonych przy użyciu dwóch różnych metod i trzech wartości parametru związanego z pojęciem funkcji wagi widma. Przyjęto zasadę, według której najlepiej reprezentują wyraz te obrazy, które najliczniej wykazały mniejszą odległość od obrazów innych wypowiedzi tego samego wyrazu niż od obrazów wypowiedzi innego wyrazu, niezależnie od tego, czy wypowiedzi pochodziły od tego samego, czy też od różnych głosów.

W tablicy nr 3 zestawiono liczby przypadków negatywnych oraz nieokreślonych w wynikach porównania obrazów tego samego typu. Typ obrazu określony jest przez rodzaj bina-



ryzacji i wartość współczynnika  $k$ . Symbole WRG i WTG oznaczają odpowiednio, że wyniki dotyczą porównania obrazów wypowiedzi różnymi głosami i tym samym głosem.

Najgorszy wynik wyrażający się 7-oma przypadkami negatywnymi i 10-oma nieokreślonymi uzyskano dla obrazów wyznaczonych metodą  $pm + wo$  przy wartości współczynnika  $k$  równej  $1/16$ . Wyniki najlepsze z liczbami przypadków negatywnych i nieokreślonych równymi odpowiednio 3 i 3 oraz 4 i 2 uzyskano dla metody  $pm$  przy wartościach współczynnika  $k$  równych odpowiednio  $1/25$  i  $1/16$ . Bliższa analiza innych danych zawartych w tablicach 2 i 3 wykazuje, iż średnie odległości obrazów wypowiedzi tego samego wyrazu są dla trzech spośród 4-ech kombinacji głosowych mniejsze w przypadku wartości współczynnika  $k$  równej  $1/16$ . Jednocześnie oddalenie obrazu wypowiedzi drugiego wyrazu jest wówczas średnio nieco większe. Obie te okoliczności łącznie z faktem, że przy wartości współczynnika  $k$  równej  $1/16$  liczba negatywnych przypadków była o jeden mniejsza, przemawiają za wyborem wariantu  $pm$  z  $k = 1/16$ . Metoda  $pm$  wykazuje zasadniczą przewagę nad metodą  $pm+wo$  w przypadku obrazów wypowiedzi różnymi głosami. Klasycznym przykładem tej przewagi są wyniki porównania obrazów 10 wypowiedzi wyrazu jeden głosem męskim z obrazami wypowiedzi wyrazów jeden i siedem głosem żeńskim. Przy wartości współczynnika  $k$  równej  $1/16$  dla metody  $pm+wo$  uzyskano dla wszystkich 10 obrazów testowanych wyniki negatywne lub nieokreślone. Dla metody  $pm$  wszystkie wyniki z analogicznego porównania były poprawne. Pod względem liczby przypadków negatywnych i nieokreślonych obie metody nie różnią się zasadniczo, gdy obrazy pochodzą z wypowiedzi tym samym głosem. Nieco korzystniej dla metody  $pm$  przedstawiają się wówczas średnie odległości obrazów wypowiedzi tego samego i różnych wyrazów. Średnie odległości obrazów pochodzących z wypowiedzi różnymi głosami jedynie po części wypadły korzystniej dla metody  $pm$ .



Tablica nr 3. Liczebności wyników negatywnych i nieokreślonych uzyskane dla każdej z badanych metod, przy trzech wartościach współczynnika  $k$  oraz dla różnych wariantów głosowych.

MB		k=1/32		k=1/25		k=1/16	
		PNg	PNo	PNg	PNo	PNg	PNo
pm+wo	WRG	6	4	4	7	4	8
	WTG	0	0	2	0	3	2
	WRG+WTG	6	4	6	7	7	10
	RNg+PNo	10		13		17	
pm	WRG	5	3	3	2	1	3
	WTG	1	0	1	0	2	0
	WRG+WTG	6	3	4	2	3	3
	PNg+PNo	9		6		6	

Osobna część badań dotyczyła reprezentatywności wypowiedzi wyrazu przez ciąg liczbowy wyrażający punkty założonego podziału kolejnych, odpowiednio zredukowanych widm sygnału mowy, tworzącego tę wypowiedź. Spodziewano się, iż ciągi liczbowe uzyskane dla tego samego podziału widma i dla wypowiedzi tego samego wyrazu będą zbliżone. Przypuszczano, iż dzięki temu poszczególne widma da się wyrazić za pomocą par liczb będących numerami punktów różnych podziałów. Rozpatrywano 8 wartości podziału, rozłożonych liniowo od 1/8 do 8/8. Poniżej zamieszczono wydruki przykładowych ciągów dla trzech wypowiedzi wyrazu jeden głosem męskim i trzech wypowiedzi wyrazu siedem głosem żeńskim, dla trzech wartości współczynnika  $k$  równych 1/32, 1/25, 1/16 oraz czterech wartości współczynnika podziału widma zredukowanego równych: 1/4, 1/2, 3/4, 1. Wydruki te ponumerowano liczbami od 2 do 5. Każdy wydruk dotyczy jednej z czterech wyżej wymienionych wartości współczynnika podziału widma. Powyżej ciągu figuruje jego oznaczenie zawierające kolejno: symbol wyrazu i numer wypowiedzi, wartość współczynnika maskowania, wartość współczynnika podziału widma. Liczby ciągu oznaczają numery punktów widma, w których przypada



założony podział widma zredukowanego. Na przykład element  $n$  w ciągu wyznaczonym dla wartości  $\frac{q}{Q}$  współczynnika podziału oznacza, że  $n$  jest najmniejszą liczbą naturalną spełniającą warunek:

$$\frac{q+1}{Q} > \frac{\sum_{i=1}^n p_i}{N} \geq \frac{q}{Q}, \quad (3)$$

gdzie  $N$  - liczba punktów widma,  $Q$  - liczba podziałów widma,  $p_i$  - składowa widma dyskretnego, zredukowanego po uprzednim wygładzeniu.

$n = 99$  przyjęto umownie jako informację o braku liczby  $n$  spełniającej warunek określony wzorem (3).

Zarówno w przykładowych ciągach jak i w pozostałych badanych nie stwierdzono cech, na podstawie których możliwa byłaby identyfikacja wyrazu. Nawet ciągi dotyczące wypowiedzi tego samego wyrazu tym samym głosem nie wykazują wyraźnych cech wzajemnego podobieństwa we wszystkich rozpatrywanych wariantach współczynnika  $k$  oraz współczynnika podziału widma.



Wydruk nr 2. Ciagi liczb będących numerami punktów, w których przypadają założone podziały widm dla trzech wypowiedzi wyrazu jeden głosem męskim oraz trzech wypowiedzi wyrazu siedem głosem żeńskim, uzyskane przy wartości współczynnika podziału widma równej 1/4.

11M1/32 2/8

29 99 1 1 1 2 2 2 2 2 2 1 99 99 99 1 2 2 2 2 1 99 99 99 99 99 99 99 99

15M1/32 2/8

34 99 99 1 99 99 1 1 2 2 2 2 1 99 99 99 99 1 2 2 2 2 2 1 99 99 99 99 99 99 99 99 99 99

10M1/32 2/8

33 52 2 1 2 1 1 1 2 2 2 2 1 99 0 49 2 2 2 2 2 1 1 99 99 99 99 99 1 99 99 99 99

11M1/25 2/8

29 99 1 1 1 2 2 2 2 2 1 1 99 99 99 1 2 2 2 2 1 99 99 99 99 99 99 99 99

15M1/25 2/8

34 99 99 1 99 99 1 1 2 2 2 2 1 99 99 99 99 1 2 2 2 2 1 1 99 99 99 99 99 99 99 99 99 99

10M1/25 2/8

33 52 2 1 1 1 1 1 1 2 2 2 2 1 99 0 48 1 2 2 2 2 1 1 99 99 99 99 99 1 99 99 99 99

11M1/16 2/8

29 99 99 1 1 1 2 1 1 1 1 1 99 99 99 1 1 1 1 1 1 99 99 99 99 99 99 99 99

15M1/16 2/8

34 99 99 99 99 99 99 1 1 1 1 1 1 99 99 99 99 1 1 1 1 1 1 99 99 99 99 99 99 99 99 99 99

10M1/16 2/8

33 52 1 1 1 99 99 1 1 1 1 1 1 99 99 0 45 1 1 1 1 1 1 1 99 99 99 99 99 99 99 99 99 99

7121/32 2/8

28 56 45 41 40 37 37 33 1 3 5 4 4 2 99 99 0 0 99 2 2 2 2 99 99 99 99 99 99

7521/32 2/8

28 47 38 36 35 33 35 31 99 2 3 3 2 99 99 99 51 2 2 2 2 99 99 99 99 99 99 99

7021/32 2/8

28 40 36 37 36 37 36 30 2 3 3 3 1 0 0 99 2 3 2 2 99 99 99 99 99 99 99

7121/25 2/8

28 56 45 41 40 37 36 32 1 2 4 3 3 2 99 99 0 0 1 2 2 2 1 99 99 99 99 99 99

7521/25 2/8

28 46 38 35 35 32 34 31 1 2 2 2 2 99 99 99 51 1 2 2 1 99 99 99 99 99 99 99

7021/25 2/8

28 40 36 36 36 36 34 30 2 3 3 2 1 0 0 99 2 2 2 2 99 99 99 99 99 99 99

7121/16 2/8

28 99 41 40 37 34 35 31 1 2 3 3 2 2 99 99 0 0 1 1 1 1 1 99 99 99 99 99 99

7521/16 2/8

28 46 37 34 34 31 33 30 1 2 2 2 1 99 99 99 51 1 2 1 1 99 99 99 99 99 99 99

7021/16 2/8

28 38 35 34 35 36 31 29 2 2 2 2 1 0 0 52 1 2 2 1 99 99 99 99 99 99 99

Wydruk nr 3. Ciagi liczb będących numerami punktów, w których przypadają założone podziały widm dla trzech wypowiedzi wyrazu jeden głosem męskim oraz trzech wypowiedzi wyrazu siedem głosem żeńskim, uzyskane przy wartości współczynnika podziału widma równej 1/2.

15M1/32 4/8  
29 2 3 29 29 27 24 6 6 5 4 2 99 99 1 3 4 6 5 5 3 99 99 99 99 99 99 1 99 99  
15M1/32 4/8  
34 99 99 3 99 99 2 3 5 5 4 5 3 1 99 99 99 3 5 5 5 5 4 2 99 99 99 99 99 99 99 99 1 99  
10M1/32 4/8  
33 53 51 29 29 3 4 3 4 5 5 5 4 2 99 0 99 4 5 5 4 4 3 2 99 99 99 99 2 2 99 1 99

11M1/25 4/8  
29 99 3 28 28 25 23 5 5 4 3 2 99 99 99 99 4 5 4 5 99 99 99 99 99 99 1 99 99  
15M1/25 4/8  
34 99 99 2 99 99 2 3 4 4 4 4 3 1 99 99 99 3 4 4 4 4 4 2 99 99 99 99 99 99 99 99 1 99  
10M1/25 4/8  
33 53 51 28 27 3 3 3 3 4 4 4 4 2 99 0 56 3 4 4 4 4 3 2 99 99 99 99 2 2 99 1 99

11M1/16 4/8  
29 99 2 2 2 3 4 3 3 3 3 2 99 99 99 2 3 3 3 4 2 99 99 99 99 99 99 99 99  
15M1/16 4/8  
34 99 1 99 99 1 99 2 3 3 3 3 99 99 99 99 1 2 3 3 3 3 3 99 99 99 99 99 99 99 99 1 99 1 99  
10M1/16 4/8  
33 53 50 3 2 99 99 2 2 3 3 3 3 2 99 0 99 2 3 3 3 3 2 2 99 99 99 99 2 99 99 1 99

7121/32 4/8  
28 57 49 50 47 43 42 38 3 5 24 9 11 4 1 99 0 0 3 4 6 7 6 1 1 1 1 1 99  
7521/32 4/8  
28 51 48 46 43 41 43 38 3 6 6 7 5 99 99 99 99 4 6 6 5 1 1 1 1 1 1 99  
7021/32 4/8  
26 51 44 42 43 44 41 35 4 6 6 5 99 0 0 53 5 8 7 6 1 1 1 1 1 99 99

7121/25 4/8  
28 57 49 50 46 42 41 38 99 5 9 7 8 4 1 99 0 0 3 4 5 6 4 1 1 1 1 1 99  
7521/25 4/8  
28 51 48 46 41 40 42 37 3 5 5 6 4 99 99 99 52 4 5 5 4 1 1 1 1 1 1 99  
7021/25 4/8  
26 51 42 41 42 42 40 34 4 5 5 5 99 0 0 53 5 7 6 5 1 1 1 1 1 99 99

7121/16 4/8  
28 57 48 49 44 39 38 36 2 99 7 7 5 3 1 99 0 0 3 3 3 4 3 99 1 1 1 1 99  
7521/16 4/8  
28 48 46 40 38 35 36 35 99 4 4 4 3 99 99 99 52 3 4 4 3 1 1 1 1 1 1 99  
7021/16 4/8  
26 51 38 38 38 38 34 99 4 4 4 99 0 0 99 4 5 5 4 1 1 1 1 1 99 99



Wydruk nr 4. Ciagi liczb będących numerami punktów, w których przypadają założone podziały widm dla trzech wypowiedzi wyrazu jeden głosem męskim oraz trzech wypowiedzi wyrazu siedem głosem żeńskim, uzyskane przy wartości współczynnika podziału widma równej 3/4.

11M1/32 6/8

29 4 50 37 34 32 29 25 23 10 6 99 99 99 99 4 7 20 8 9 5 99 99 99 99 99 99 99 99

15M1/32 6/8

34 99 99 35 3 99 4 5 21 8 7 9 6 2 99 99 3 5 8 8 7 8 7 3 99 99 99 99 99 99 99 3 3 2 99

10M1/32 6/8

33 54 53 52 37 36 38 28 24 10 10 8 20 3 99 0 99 23 8 8 7 7 6 99 99 99 99 99 99 3 3 2 99

11M1/25 6/8

29 4 50 32 33 30 28 24 22 6 5 99 99 99 99 4 6 8 7 8 4 99 99 99 99 99 99 99 99

15M1/25 6/8

34 99 2 29 3 2 99 4 7 6 6 7 5 2 99 99 3 5 7 7 7 7 6 3 99 99 99 99 99 99 99 3 2 99

10M1/25 6/8

33 99 53 52 36 30 36 27 5 6 7 7 7 3 99 0 99 6 7 6 6 6 5 99 99 99 99 99 99 3 3 3 2 99

11M1/16 6/8

29 3 30 29 30 28 25 6 6 99 4 3 1 99 99 99 5 5 6 6 3 1 1 1 1 1 1 99 1 99

15M1/16 6/8

34 99 99 3 2 2 3 4 5 5 5 5 4 99 99 1 2 99 5 5 5 5 5 3 99 99 99 99 99 99 99 2 3 2 99

10M1/16 6/8

33 99 52 51 28 3 3 99 99 5 5 5 5 3 1 0 99 4 5 5 5 5 4 99 1 99 99 99 99 3 3 99 99 1

71Z1/32 6/8

28 58 53 54 52 49 48 47 49 28 30 27 27 6 99 99 0 0 48 8 9 11 10 99 99 99 99 99 2

75Z1/32 6/8

28 56 53 51 49 49 50 45 6 27 23 25 23 99 99 99 99 6 8 9 8 2 2 99 99 99 99 2 2

70Z1/32 6/8

26 55 51 50 49 51 48 38 6 25 8 22 99 0 0 99 22 22 10 10 99 99 99 99 99 2 2

71Z1/25 6/8

28 58 52 54 51 48 47 41 49 27 28 25 26 5 99 99 0 0 6 6 8 10 9 99 99 99 99 99 2

75Z1/25 6/8

28 56 52 51 48 48 49 40 5 26 9 9 8 99 99 99 99 5 7 8 7 99 2 99 99 99 99 99 2

70Z1/25 6/8

26 54 49 48 48 50 47 38 6 9 8 7 99 0 0 99 7 10 9 9 99 99 99 99 99 2 2

71Z1/16 6/8

28 58 51 52 49 43 41 38 49 5 27 28 8 5 99 99 0 0 5 4 6 8 7 99 99 99 99 99 2

75Z1/16 6/8

28 56 49 47 42 40 43 38 4 6 6 6 5 99 99 99 53 99 6 6 6 99 99 99 99 99 99 2

70Z1/16 6/8

26 53 43 40 43 43 41 38 5 6 6 6 99 0 0 99 5 8 7 8 99 99 99 99 99 2 2

Wydruk nr 5. Ciągi liczb będących numerami punktów, w których przypadają założone podziały widm dla trzech wypowiedzi wyrazu jeden głosem męskim oraz trzech wypowiedzi wyrazu siedem głosem żeńskim, uzyskane przy wartości współczynnika podziału widma równej 1.

10M1/32 B/B

29 54 53 54 51 37 35 31 29 29 24 5 2 2 2 21 23 26 23 23 22 2 2 2 2 2 2 3 2 2

15M1/32 B/B

34 54 53 53 29 4 28 25 28 28 21 30 23 5 2 2 52 24 24 23 21 22 20 5 2 2 2 2 2 2 2 5 5 5 2

10M1/32 B/B

33 55 55 55 52 51 54 55 40 26 26 25 25 5 2 0 59 31 24 24 22 22 22 5 2 2 2 2 2 2 5 5 5 4 2

11M1/25 B/B

29 54 52 53 36 35 33 29 27 23 22 5 2 2 2 22 23 24 22 22 21 2 2 2 2 2 2 2 3 2 2

15M1/25 B/B

34 54 53 53 28 4 6 24 24 22 20 28 22 5 2 2 52 22 22 21-21 21 20 5 2 2 2 2 2 2 2 2 4 5 5 2

10M1/25 B/B

33 55 55 55 51 51 54 55 28 24 23 23 23 5 2 0 59 30 22 22 21 21 20 5 2 2 2 2 2 2 5 5 5 4 2

11M1/16 B/B

29 53 51 36 35 32 30 25 23 7 6 5 2 2 2 5 21 20 20 9 5 2 2 2 2 2 2 2 3 2 2

15M1/16 B/B

34 54 3 52 27 3 4 5 7 7 7 7 6 4 2 2 52 5 7 7 8 8 8 5 2 2 2 2 2 2 2 4 5 5 2

10M1/16 B/B

33 55 55 54 38 30 30 28 5 7 7 7 21 5 2 0 59 22 7 7 7 7 5 5 2 2 2 2 2 2 5 5 4 3 2

7121/32 B/B

26 60 57 57 57 56 56 55 51 33 37 34 34 8 3 2 0 0 50 36 24 25 24 3 3 3 3 3 3 3

7521/32 B/B

29 59 58 57 56 56 56 53 30 33 29 32 33 2 2 2 55 24 26 24 20 3 3 3 3 3 3 3 3 3

7021/32 B/B

26 55 57 57 57 56 40 31 32 27 28 5 0 0 55 36 35 24 21 3 3 3 3 3 3 3 3

7121/25 B/B

26 60 57 57 56 55 54 53 51 32 33 32 32 7 3 2 0 0 50 35 23 24 23 3 3 3 3 3 3 3

7521/25 B/B

29 59 57 56 55 55 55 50 30 31 28 30 32 2 2 2 55 8 23 23 11 3 3 3 3 3 3 3 3

7021/25 B/B

26 56 56 56 55 56 40 30 31 26 27 5 0 0 55 25 25 23 21 3 3 3 3 3 3 3 3

7021/16 B/B

26 56 56 56 54 49 47 40 51 28 55 51 26 6 3 2 0 0 50 7 8 23 10 3 3 3 3 3 3 3

7521/16 B/B

29 58 55 51 49 49 50 41 29 28 9 9 7 2 2 2 55 6 8 9 9 3 3 3 3 3 3 3 3 3

7021/16 B/B

26 55 52 48 47 51 47 39 29 26 8 25 5 0 0 55 8 22 10 20 3 3 3 3 3 3 3 3



## 5. WNIOSKI

Pierwszy z rozpatrywanych w tej pracy sposobów opisu wypowiedzi wyrazu różni się od sposobu użytego w pracy [6] zasadą tworzenia obrazu. Obecnie zastosowano dodatkowe kryterium wywodzące się z pojęcia wagi widma [7]. Drugi sposób korzysta jedynie z tego kryterium. Oba sposoby prowadzą do wyrażenia wypowiedzi wyrazu w formie ciągu wektorów dwubajtowych. Praca dostarczyła danych, na podstawie których przeprowadzono ocenę porównawczą obu sposobów wyrażenia wypowiedzi wyrazu. Pierwszy z rozpatrywanych sposobów otrzymał gorszą notę, mimo iż stanowi bardziej precyzyjne ujęcie cech widmowych w porównaniu ze sposobem drugim, który jedynie bardzo pobieżnie uwzględnia konfigurację obwiedni widma. Z tego faktu wpływa praktyczny wniosek, iż operując jedynie umownym pojęciem wagi widma można reprezentatywnie opisać wypowiedź wyrazu w postaci ciągu wektorów dwubajtowych. Procedura wyznaczania obrazu wypowiedzi według kryterium poziomu maskowania jest znacznie prostsza od procedury wyznaczania obrazu na podstawie obu kryteriów rozpatrywanych w pracy. Wynik pracy pozwala na uproszczenie jednego z ważnych elementów procesu automatycznego rozpoznawania wyrazów. Uzyskane rezultaty potwierdzają wcześniej zauważoną dla obrazów o szerokości 64 punktów pozytywną funkcję pojęcia wagi widma w tworzeniu reprezentacji wypowiedzi wyrazu.

Negatywny wynik próby przedstawienia wypowiedzi wyrazu za pomocą ciągu liczb wyrażających punkty założonego podziału widma dowodzi, że współzależności ilości energii w pasmach częstotliwości sygnału mowy nie są cechami jednoznacznie charakteryzującymi poszczególne segmenty fonetyczno-akustyczne. Wynik ten utwierdza w przekonaniu iż cech takich należy poszukiwać w parametrach częstotliwościowych, które wyrażają między innymi rozpatrywane w tej pracy obrazy binarne.

Na marginesie tej pracy należy wspomnieć o zagadnieniu wyznaczania obrazu wzorcowego wyrazu. Postępowanie z tym związane jest uzależnione od formy obrazu. Próby tworzenia wzorca wyrazu dla typu obrazu ocenionego w tej pracy najkorzystniej zostaną podjęte w najbliższym czasie.



6. BIBLIOGRAFIA.

- [1] KUBZDELA, H. : Techniczna realizacja formantowej metody rozpoznawania samogłosek polskich, Prace IPPT 90/1975, Warszawa 1976.
- [2] MYTKOWSKI, K. : Kanał funkcji analogowych typ KF-01 do wprowadzania i wyprowadzania informacji w systemie "ON-LINE" do/z pamięci minikomputera MOMIK 8B/100, Prace IPPT 39/76, Warszawa 1976.
- [3] KUBZDELA, H. : Weryfikacja i optymalizacja metody rozpoznawania wyrazów w skończonych zbiorach hasłowych w oparciu o spektrogramy binarne, Prace IPPT 10/1982, Warszawa 1982.
- [4] KUBZDELA, H. : Badania nad udoskonaleniem spektrogramów binarnych, Prace IPPT 24/1983, Warszawa 1983.
- [5] KUBZDELA, H. : Próby automatycznego rozpoznawania wyrazów wymawianych przez różne głosy w oparciu o grupowe zbiory wzorcowych spektrogramów binarnych, Prace IPPT 47/83, Warszawa 1983.
- [6] KUBZDELA, H. : Metoda globalnego rozpoznawania wyrazów na podstawie spektrogramów binarnych, Prace IPPT 28/1986, Warszawa 1986.
- [7] KUBZDELA, H. : Udoskonalenie reprezentacji sygnału mowy w formie obrazów binarnych, Prace IPPT 24/1987, Warszawa 1987.