Henryk Kubzdela

WERYFIKACJA I OPTYMALIZACJA METODY ROZPOZNAWANIA WYRAZÓW W SKOŃCZONYCH ZBIORACH HASŁOWYCH W OPARCIU O SPEKTROGRAMY BINARNE

WARSZAWA 1982

ISSN 0208-5658

Praca wpłynęła do Redakcji dnia 8 grudnia 1981 r.

Zarejestrowana pod nr 10/1982

Na prawach rękopisu

Instytut Podstawowych Problemów Techniki PAN

Nakład 130 egz. Ark.wyd. 1,2. Ark.druk. 1,75
Oddano do drukarni w marcu 1982 r.
Nr zamówienia 127/0/82 Z-87 .

Warszawska Drukarnia Naukowa, Warszawa,
ul.Śniadeckich 8

Henryk Kubzdela
Pracownia Fonetyki Akustycznej
IPPT PAN

WERYFIKACJA I OPTYMALIZACJA METODY ROZPOZNAWANIA WYRAZÓW W SKOŃCZONYCH ZBIORACH HASŁOWYCH W OPARCIU O SPEKTROGRAMY BINARNE[1].

1. Wstęp

Prace nad automatycznym rozpoznawaniem wyrazów na podstawie spektrogramów binarnych rozpoczął autor pod koniec roku 1978. Zarówno decyzja o podjęciu takiego problemu, jak i późniejszy przebieg badań z nim związanych wynikały w zasadniczym stopniu z pomyślnego kształtowania się koniecznych dla tych badań warunków technicznych. Szczególną rolę odegrały pod tym względem następujące okoliczności :

I. Wyposażenie Pracowni Fonetyki Akustycznej IPPT PAN w Poznaniu w :
   a/ zestaw minikomputerowy MERA 303 poszerzony później o urządzenie pamięci zewnętrznej na dyskach elastycznych,
   b/ oscyloskop z pamięcią typ OG2-31 RFT;

II. Skonstruowanie przez K. Mytkowskiego [6] w ramach jego pracy magisterskiej na Politechnice Poznańskiej, wykonanej pod kierunkiem autora, tzw. kanału funkcji analogowych KF-01 - urządzenia umożliwiającego automatyczne wprowadzanie do minikomputera MERA 303 informacji pochodzących ze źródeł analogowych, a także przekształcanie informacji wyjściowej z komputera do postaci analogowej;

III. Skonstruowanie przez autora wielokanałowego analogowego

---

[1] Praca wykonana w ramach problemu międzyresortowego MR.I-24

analizatora widma oraz komutatora sygnałów analogowych.

W kilku swoich dotychczasowych pracach autor przedstawił w miarę szczegółowo wielokanałowy analizator widma wraz z komutatorem [1], [2], ideę oraz jedną z metod tworzenia spektrogramu binarnego [3], [4], metodę automatycznego rozpoznawania wyrazów w oparciu o spektrogramy binarne w jej pierwotnej wersji [3] oraz w wersji zmodyfikowanej w wyniku pierwszych eksperymentów [5]. W pracy [5] przedstawiono także pierwsze wyniki rozpoznawania wyrazów należących do 13-wyrazowego słownika i wymawianych przez jeden głos męski. Etap badań, który zamyka niniejsza praca, poświęcono ponownej weryfikacji obranej przez autora metody rozpoznawania wyrazów, tym razem w poprawionych warunkach lecz jednocześnie przy zaostrzonych kryteriach. Na poprawę warunków rozpoznawania złożyło się : rozszerzenie zakresu częstotliwości spektrogramu binarnego do 8310 Hz, udoskonalenie techniki logarytmizacji danych wyjściowych z analogowego analizatora widma oraz zastosowanie korzystniejszej zasady wyznaczania początku i końca wyrazu. Na zaostrzenie kryteriów rozpoznawania złożyło się powiększenie słownika rozpoznawanych wyrazów do 48 wyrazów oraz objęcie próbami trzech głosów - 2 męskich i jednego żeńskiego. Dla przeprowadzenia badań konieczne było zatem rozbudowanie analizatora widma, a także zredagowanie kilku nowych lub przeredagowanie niektórych dotychczasowych procedur programów adaptacji i rozpoznawania. Czynności te stanowiły znaczną część całości prac. W pierwszej części badań postawiono pytanie, czy można bez istotnego pogorszenia wyników rozpoznawania dokonać następującej redukcji :

1. Zredukować do połowy liczbę parametrów widma binarnego,
2. Zmniejszyć rozciągłość czasową fragmentu porównawczego do dwóch kolejnych widm binarnych w procesach adaptacji i rozpoznawania wyrazów.

Pozytywna odpowiedź na to pytanie zadecydowała o sposobie przeprowadzenia dalszych badań. W części pracy poświęconej omówieniu wyników rozpoznawania przeanalizowano przypadki błędnych odpowiedzi próbując określić ich przyczyny. Pracę kończą wnioski i sugestie co do kierunku optymalizacji i dalszego rozwoju metody automatycznego rozpoznawania wyrazów na podstawie spektro-

gramów binarnych.

2. Nowe zespoły w układzie rozpoznającym.

Zasadniczym elementem układu rozpoznającego stosowanego przez autora jest wielokanałowy analogowy analizator widma wraz z komutatorem sygnałów wyjściowych oraz przetwornikiem liniowo-logarytmicznym. W pierwszych etapach badań zrelacjonowanych w pracach [3], [5] analizator widmowy posiadał 43 kanały, z których każdy obejmował pasmo częstotliwości 80 Hz, zaś wszystkie łącznie pokrywały zakres od 80 do 3560 Hz. W ramach obecnego zadania badawczego liczbę kanałów zwiększono do sześćdziesięciu trzech. W nowych kanałach zróżnicowano szerokości pasm przepustowych, czyniąc je liniowo zależnymi od numeru kolejnego kanału zgodnie z następującą zależnością :

$$f_p = 80 + (k - 43)\ \Delta f_d, \qquad (1)$$

gdzie k oznacza numer kolejny kanału, a Δf_d wynosi 20 Hz. Szerokość pasma przepustowego pierwszego z dodatkowych kanałów wynosi 100 Hz, a ostatniego 480 Hz. Dzięki nowym dwudziestu kanałom zakres analizy poszerzył się do częstotliwości 8310 Hz. Powstała zatem możliwość objęcia operacjami rozpoznawania nowego zakresu widmowego bardzo istotnego dla szumowych segmentów sygnału mowy. Podobnie jak kanały analizatora zbudowane wcześniej, każdy nowy kanał składa się z środkowo-przepustowego filtru czynnego o charakterystyce Butterworth'a trzeciego rzędu, wzmacniaczy, układu prostującego, dolno-przepustowego filtru czynnego RC posiadającego także charakterystykę Butterworth'a trzeciego rzędu oraz z sterowanego komutacyjnie przełącznika elektronicznego, łączącego wyjście kanału z wyjściem analizatora raz na okres komutacji i na czas ok. 100 μs.

Szerokości pasm przepustowych nowych kanałów ustalono w drodze kompromisu uwzględniającego z jednej strony dążenie do uzyskania możliwie korzystnej rezolucji częstotliwościowej analizatora, a z drugiej strony utrzymania jego rozmiarów w rozsądnych granicach. Gdyby analizator miał mieć w całym zakresie jednakową rezolucję - wynoszącą jak w części wcześniej zbudowanej 80 Hz - wówczas dla pokrycia zakresu do 8310 Hz potrzebnych byłoby ok. 50 dodatkowych kanałów. Także stosunkowo dużo, bo ok. 40 dodat-

kowych kanałów należałoby dobudować, gdyby przyjąć dla nich stałą względną szerokość pasma przepustowego, określoną przez szerokość i częstotliwość środkową ostatniego z kanałów mających jednakowe bezwzględne szerokości pasm przepustowych. Za przyjęciem stałej względnej szerokości pasm przepustowych analizatora, w zakresie częstotliwości wyższych przemawiać mogłyby powszechnie znane motywacje uzasadniające stosowanie logarytmicznej skali częstotliwości. O przyjęciu liniowo zależnej od częstotliwości szerokości pasma przepustowego dodatkowych kanałów analizatora zadecydowała też w istotny sposób okoliczność, że dane wyjściowe analizatora służą do wyznaczania widm binarnych, podczas którego binarne dane widmowe b_i wynikają z badania nierówności :

$$\frac{a_i}{\Delta f_i} > \frac{a_{i-1} + a_i + a_{i+1}}{\Delta f_{i-1} + \Delta f_i + \Delta f_{i+1}} \qquad (2)$$

przyjmując wartość 1 w przypadku, gdy nierówność (2) jest spełniona, a $\emptyset$ w przypadku przeciwnym.
Ponieważ pierwsze 43 kanały analizatora mają jednakową szerokość pasma przepustowego, tzn. :

$$\Delta f_{i-1} = \Delta f_i = \Delta f_{i+1},$$

nierówność (2) upraszcza się do postaci :

$$a_i > \frac{a_{i-1} + a_{i+1}}{2} . \qquad (3)$$

Do identycznej postaci uprościć można nierówność (2), gdy szerokości pasm przepustowych kolejnych kanałów różnią się o stały przyrost Δfd. Zatem dla wyznaczenia widma binarnego w zakresie częstotliwości, które obejmują dodatkowe kanały analizatora, stosowana może być dotychczasowa reguła.
Kolejna optymalizująca zmiana typu hardware'owego w stosunku do pierwotnej wersji układu rozpoznającego używanego w dotychczasowych badaniach polega na wprowadzeniu nowego, specjalnie skonstruowanego logarytmującego konwertera analogowo-cyfrowego. Zastąpił on dotąd stosowany uproszczony układ quasi-logarytmu-

jący oraz liniowy konwerter analogowo-cyfrowy. Nowy konwerter z logarytmowaniem wykorzystuje powszechnie znaną zależność logarytmiczną, pomiędzy czasem t rozładowywania się kondensatora a napięciem U_c, do którego nastąpiło rozładowanie, zgodnie z równaniem :

$$t = a \log U_c + b \qquad (4)$$

gdzie $a = -T$, $b = T \log U_o$, U_o - napięcie na kondensatorze w chwili t_o, T - stała czasu rozładowania kondensatora.
Od momentu t = 0 inicjującego kolejny okres konwersji analogowo-cyfrowej, napięcie podlegające konwersji jest ustawicznie porównywane z napięciem na rozładowującym się kondensatorze. W czasie oczekiwania na zrównanie się obu tych napięć następuje zliczanie impulsów o częstotliwości ok. 1 MHz. Zrównanie napięć powoduje zatrzymanie zliczania. Stan licznika będący wynikiem konwersji danej analogowej do postaci cyfrowej z równoczesnym jej zlogarytmowaniem zostaje przepisany programowo do komputera MERA 303. Dokładność konwersji wynosi ok. 0,5 dB. Czas konwersji nowego konwertera jest wielokrotnie dłuższy niż poprzedniego, co oczywiście wynika z odmienności zasady działania. Ponieważ konwersja analogowo-cyfrowa danej wyjściowej z analizatora może trwać do 100 µs, szybkość działania nowego konwertera jest wystarczająca, a o jego wyższości decyduje głównie dodatkowa funkcja jaką spełnia, a mianowicie logarytmowanie.

3. Organizacja badań.

Zwiększenie liczby parametrów widmowych z 43 do 63, oraz powiększenie słownika wyrazów z 13 do 48 wyrazów spowodowało po pierwsze wydłużenie czasów wykonywania wielu procedur wchodzących w skład programów adaptacji i rozpoznawania, a po drugie zwiększenie zapotrzebowania na pamięć. Jedynym dostępnym autorowi komputerem jest minikomputer biurowy MERA 303 operujący słowem 8-bitowym, działający stosunkowo wolno i posiadający pamięć wielkości 8K. Wynikła stąd konieczność modyfikacji procesu adaptacji i rozpoznawania. W trybie dotychczas stosowanym w trakcie wypowiedzi następowało wyznaczanie kolejnych widm binarnych, których ciąg tworzył spektrogram binarny. Równocześnie z zakończeniem wypowiedzi gotowy był już spektrogram bi-

narny i natychmiast następowało automatyczne przejście bądź do rozpoznawania, bądź do adaptacji. Komplet złożony z 13 wzorcowych spektrogramów binarnych mieścił się w pamięci operacyjnej minikomputera. W obecnych warunkach nie jest możliwe wyznaczenie spektrogramu binarnego równocześnie z trwaniem wypowiedzi, wobec czego dane widmowe napływające z analizatora po odpowiednim uśrednieniu zostają przechowane w pamięci minikomputera w celu późniejszego wyznaczenia spektrogramu binarnego. Dla tych danych z konieczności przeznaczono obszar pamięci zarezerwowany uprzednio na inwentarz wzorców. Z powyższych względów dla przeprowadzenia zamierzonych badań nad automatycznym rozpoznawaniem wyrazów przyjęto następujący 3-etapowy tok działań. Pierwszy etap obejmował zgromadzenie na dyskach elastycznych pamięci zewnętrznej zbioru spektrogramów binarnych pochodzących z wypowiedzi zadanych wyrazów, które miały bądź służyć do tworzenia wzorców, bądź też stanowić materiał do automatycznego rozpoznawania. Do realizacji tego etapu ułożono kompleks programów następujących operacji :

1. Wprowadzania, uśredniania i umieszczania w pamięci danych z analogowego analizatora widma oraz równoczesnej kontroli początku i końca wypowiedzi.
2. Wyznaczania widma binarnego.
3. Obserwacji spektrogramu binarnego na ekranie oscyloskopu z pamięcią.
4. Wydruku spektogramu binarnego.
5. Wydruku spektrogramu cyfrowego.
6. Przepisania spektrogramu binarnego do pamięci zewnętrznej na dysku elastycznym.

Uśrednianie odbywa się w okienku obejmującym każde 4 kolejne dane wyjściowe analizatora, a jego rezultatem jest nowy zbiór danych reprezentujących średnie wartości sygnału w poszerzonych pasmach częstotliwości. W tablicy 1 podano szerokości tych pasm w dodatkowo włączonym do analizy zakresie częstotliwości. Na podstawie danych średnich wyznaczane jest widmo binarne.

Tablica 1. Szerokości pasm uśredniania danych z analizatora w dodatkowo włączonym do analizy zakresie częstotliwości.

f_{-1} [Hz]	f_{+1} [Hz]	Δf [Hz]
3560	4030	470
3655	4185	530
3765	4355	590
3890	4540	650
4030	4740	710
4185	4955	770
4355	5185	830
4540	5430	890
4740	5690	950
4955	5965	1010
5185	6255	1070
5430	6560	1130
5690	6880	1190
5965	7215	1250
6255	7563	1310
6560	7930	1370
6880	8310	1430

Sposób kontroli początku i końca wypowiedzi zmodyfikowano w stosunku do poprzednio stosowanego. Opiera się on obecnie na następującej zasadzie : Początek spektrogramu sygnalizowany jest przez:

a/ jedno lub dwa kolejne widma niezerowe, po których następuje nie więcej niż 6 widm zerowych, albo

b/ co najmniej 3 widma niezerowe.

Jako koniec spektrogramu i koniec wypowiedzi przyjmuje się ostatnie niezerowe widmo, po którym następuje 8 widm zerowych czyli cisza długości 140 ms. Widmem zerowym nazwano takie, którego wszystkie dane o poziomie sygnału w poszczególnych pasmach częstotliwości nie przekraczają pewnego założonego progu. Programy operacji 3,4,5 służyły dla celów kontrolnych i mogły być użyte w razie potrzeby. Na jednym dysku elastycznym zgromadzić można spektrogramy binarne około 400 wyrazów.

Na drugi etap składały się operacje tworzenia spektrogramów wzorcowych na podstawie uprzednio zgromadzonych na dysku elastycznym spektrogramów binarnych 4-krotnych wypowiedzi każdego z wyrazów tworzących słownik wyrazów przewidzianych do rozpo-

znawania. Przy obecnym rozmiarze spektrogramu binarnego (zwiększonym przez dodatkowe parametry widmowe) w pamięci operacyjnej minikomputera MERA 303 zmieścić się może jednocześnie zaledwie 6 wzorców. Wobec tego wzorcowe spektrogramy binarne wszystkich 48 wyrazów słownika w miarę ich tworzenia przepisywano kolejno na dysk elastyczny w grupach po 6 wyrazów.
Trzeci etap stanowiło rozpoznawanie wyrazów, których spektrogramy binarne zgromadzono na dysku elastycznym w ramach etapu pierwszego. W trakcie rozpoznawania każdego wyrazu wzorce pobierane były z pamięci zewnętrznej i przesyłane grupami po 6 do pamięci operacyjnej minikomputera. Czas konieczny na otrzymanie końcowej decyzji identyfikacyjnej zdeterminowany jest wolno przebiegającą transmisją danych z pamięci zewnętrznej i wynosi 50-70 s/wyraz. Wyniki rozpoznawania były drukowane. Rozpoznawanie przebiegało w pełni automatycznie bez udziału operatora.

4. Cel badań.

Cel badań był wieloraki. W pierwszym rzędzie należało przeprowadzić automatyczne rozpoznawanie wyrazów metodą spektrogramów binarnych dla jednego głosu w celu uzyskania ogólnego poglądu o efektywności rozpoznawania w warunkach rozszerzonego zakresu częstotliwości i zwiększonego słownika wyrazów a także po wprowadzeniu kilku opisanych wyżej ulepszeń. Ta część badań nazywana będzie odtąd umownie eksperymentem nr 1. Następne dwa eksperymenty nr 2 i 3 miały na celu określenie zmiany, której ulegną wyniki rozpoznawania, jeśli najpierw (eksperyment 2) zmniejszona zostanie o połowę liczba parametrów w widmie binarnym poprzez wyeliminowanie parametrów o numerach parzystych, a następnie (eksperyment 3) zmniejszony zostanie do wielkości obejmującej dwa widma binarne rozmiar okna wydzielającego ze spektrogramów binarnych poszczególne ich fragmenty w celu określenia stopnia podobieństwa między nimi (patrz praca [3]). Wyniki eksperymentu drugiego i trzeciego miały zadecydować o warunkach, w jakich przeprowadzony będzie czwarty eksperyment, którego celem było zbadanie, jak kształtują się wyniki automatycznego rozpoznawania wyrazów metodą spektrogramów binarnych w zależności od głosu operatora. Badania miały być w sumie przeprowadzone dla trzech głosów, dwóch męskich i jednego żeńskiego, przy zastosowaniu

ich indywidualnych wzorców wcześniej wyznaczonych.

5. Materiał badawczy.

Do badań posłużono się nazwami czterdziestu ośmiu polskich miast, głównie wojewódzkich. Poniżej zamieszczono ich listę, na której zapisane są one w tej samej kolejności, w jakiej zostały uszeregowane ich wzorce. Wzorcowy spektrogram binarny każdego z tych wyrazów utworzono oddzielnie dla każdego głosu na podstawie spektrogramów binarnych czterech wypowiedzi danego wyrazu przez dany głos. Zatem materiał na podstawie którego sporządzono spektrogramy wzorcowe dla poszczególnych osób obejmował łącznie 576 wypowiedzi.

Materiał przeznaczony do rozpoznawania stanowiły co najmniej 8-krotne wypowiedzi w porządku losowym każdego z 48 wyrazów słownika przez każdego mówcę. Niektóre z wyrazów słownika wymówione zostały 9-krotnie. Niektórzy z mówców z różnych względów dostarczyli więcej wypowiedzi. Dla każdego głosu próba testowa wynosiła co najmniej 392 wypowiedzi.

Słownik haseł do automatycznego rozpoznawania

1. Gniezno	13. Kraków	25. Częstochowa	37. Sieradz
2. Gdynia	14. Krosno	26. Elbląg	38. Wadowice
3. Kutno	15. Ostrołęka	27. Gdańsk	39. Słupsk
4. Jarocin	16. Piła	28. Gorzów	40. Suwałki
5. Kościan	17. Płock	29. Legnica	41. Szczecin
6. Warszawa	18. Poznań	30. Leszno	42. Tarnobrzeg
7. Białystok	19. Przemyśl	31. Lublin	43. Tarnów
8. Kalisz	20. Radom	32. Łomża	44. Toruń
9. Katowice	21. Rzeszów	33. Łódź	45. Wałbrzych
10. Kielce	22. Bydgoszcz	34. Olsztyn	46. Włocławek
11. Konin	23. Chełm	35. Opole	47. Wrocław
12. Koszalin	24. Ciechanów	36. Siedlce	48. Zamość

6. Wyniki automatycznego rozpoznawania wyrazów, ich ocena i wnioski.

Wyniki rozpoznawania uzyskane w eksperymentach pierwszym, drugim i trzecim zestawione są poniżej w formie liczbowej w tabeli nr 2 oraz w postaci wykazu błędnych odpowiedzi. Wynik pierwszego eksperymentu przyjmuje się jako ocenę metody rozpozna-

wania użytej po raz pierwszy w danym systemie dla tak licznego słownika, złożonego z 48 wyrazów. Uzyskane wyniki dają podstawę do pozytywnej oceny przyjętej metody. Z porównania wyników trzech eksperymentów wypływają następujące korzystne wnioski : Wyeliminowanie parametrów widm binarnych o numerach parzystych, jak i następnie zawężenie okna wydzielającego z porównywanych spektrogramów binarnych poszczególne ich fragmenty w celu określenia podobieństwa między nimi nie pogarszają w istotny sposób wyniku rozpoznawania. Wnioski te zadecydowały, że kolejny tzn. czwarty eksperyment, obejmujący rozpoznawanie wypowiedzi wyrazów przez dalsze dwa głosy został przeprowadzony w takich samych warunkach jak eksperyment trzeci. Niektóre z błędów, jakie miały miejsce w eksperymencie trzecim udało się skorygować dzięki zaostrzeniu kryterium określającego istotność różnicy w podobieństwie danego fragmentu obiektu do odnośnych fragmentów poszczególnych wzorców. Wyniki uzyskane po tej korekcji figurują w pracy pod nagłówkiem eksperyment 3a. Z taką samą ostrością, jak w eksperymencie 3a wspomniane kryterium obowiązywało w eksperymencie 4.

Tablica 2. Liczbowe wyniki rozpoznawania uzyskane w eksperymentach 1,2,3 i 3a.

Eksperyment	Globalna poprawność rozpoznaw.	Liczebność haseł słownika rozpoznanych				
		bez błędu	z 1 błędem	z 2 błędami	z 3 błędami	z 4 i więcej błędami
1	96 %	42	2	1	2	1
2	97.25 %	42	3	2	-	1
3	96 %	41	4	-	1	2
3a	96.8 %	42	4	-	-	2

Wykaz błędnych odpowiedzi w eksperymentach 1,2,3 i 3a

Eksperyment 1

Nadano	Odebrano
1. Kraków	Tarnów
2. Kraków	Tarnów
3. Kraków	Ciechanów
4. Kutno	Poznań
5. Kutno	Leszno
6. Kutno	Krosno
7. Kutno	Krosno
8. Kutno	Krosno
9. Kutno	Krosno
10. Tarnów	Kraków
11. Kościan	Olsztyn
12. Kościan	Olsztyn
13. Słupsk	Suwałki
14. Wadowice	Katowice
15. Wadowice	Katowice
16. Wadowice	Katowice

Eksperyment 2

Nadano	Odebrano
1. Kraków	Ciechanów
2. Kraków	Tarnów
3. Kutno	Łomża
4. Kutno	Łomża
5. Kutno	Leszno
6. Kutno	Krosno
7. Tarnów	Kraków
8. Słupsk	Suwałki
9. Poznań	Koszalin
10. Wadowice	Katowice
11. Wadowice	Katowice

Eksperyment 3

Nadano	Odebrano
1. Kutno	Łomża
2. Kutno	Leszno
3. Kutno	Krosno
4. Kutno	Krosno
5. Kutno	Krosno
6. Tarnów	Kraków
7. Kraków	Ciechanów
8. Kraków	Tarnów
9. Kraków	Tarnów
10. Wadowice	Katowice
11. Wadowice	Katowice
12. Wadowice	Katowice
13. Wadowice	Katowice
14. Katowice	Wadowice
15. Kościan	Olsztyn
16. Słupsk	Suwałki

Eksperyment 3a

Nadano	Odebrano
1. Kutno	Poznań
2. Kutno	Krosno
3. Kutno	Krosno
4. Kutno	Łomża
5. Kutno	Leszno
6. Kraków	Suwałki
7. Tarnów	Kraków
8. Słupsk	Suwałki
9. Kościan	Leszno
10. Wadowice	Katowice
11. Wadowice	Katowice
12. Wadowice	Katowice
13. Wadowice	Katowice

W tabeli nr 3 zestawiono dla porównania uzyskane dla poszczególnych głosów globalne wyniki rozpoznawania oraz liczebności haseł rozpoznanych z określoną podaną w nagłówku liczbą błędów. Poniżej podany jest też szczegółowy wykaz błędnych decyzji identyfikacyjnych dla głosów M1, M2 i Ż1 oraz zestaw liczby błędów dla poszczególnych wyrazów i głosów.

Tablica 3. Liczbowe wyniki rozpoznawania dla poszczególnych głosów.

Głos	Globalna poprawność rozpoznaw.	Liczebność haseł słownika rozpoznanych				
		bez błędu	z 1 błędem	z 2 błędami	z 3 błędami	z 4 i więcej błędami
M1	96.8 %	42	4	-	-	2
M2	96.9 %	41	3	3	1	-
Ż1	97.1 %	41	4	2	-	1

Tablica 4. Zestaw ilości błędów dla poszczególnych wyrazów i głosów.

Hasło	Głosy		
	M1	M2	Ż1
Słupsk	1	-	1
Ciechanów	-	-	1
Toruń	-	1	2
Kraków	1	2	2
Białystok	-	-	4
Łomża	-	-	1
Jarocin	-	-	1
Wadowice	4	3	-
Tarnów	1	2	-
Szczecin	-	2	-
Wrocław	-	1	-
Siedlce	-	1	-
Kutno	5	-	-
Kościan	1	-	-
Razem	13	12	12

Wykaz błędnych odpowiedzi w eksperymencie 4

Eksperyment 4 Głos M2		Eksperyment 4 Głos Ż1	
Nadano	Odebrano	Nadano	Odebrano
1. Toruń	Konin	1. Słupsk	Siedlce
2. Wadowice	Zamość	2. Ciechanów	Rzeszów
3. Wadowice	Katowice	3. Toruń	Gorzów
4. Wadowice	Katowice	4. Toruń	Rzeszów
5. Tarnów	Kalisz	5. Kraków	Tarnów
6. Tarnów	Sieradz	6. Kraków	Tarnów
7. Szczecin	Konin	7. Białystok	Gniezno
8. Szczecin	Przemyśl	8. Białystok	Gniezno
9. Wrocław	Krosno	9. Białystok	Wrocław
10. Siedlce	Sieradz	10. Białystok	Wałbrzych
11. Kraków	Radom	11. Jarocin	Tarnobrzeg
12. Kraków	Radom	12. Łomża	Łódź

Przypadki błędnych odpowiedzi poddano szczegółowej analizie w celu ustalenia przyczyn powstania błędu. Posługiwano się podczas tych dociekań wydrukami spektrogramów binarnych błędnie rozpoznanych wypowiedzi oraz wzorców należących do błędnych i oczekiwanych odpowiedzi. Np. jeśli wypowiedź KUTNO została rozpoznana jako KROSNO, wówczas sporządzano przede wszystkim wydruk spektrogramu binarnego wypowiedzi KUTNO oraz wydruki wzorcowych spektrogramów oczekiwanej odpowiedzi KUTNO i błędnej odpowiedzi KROSNO. Ponadto posługiwano się wydrukami kontrolnymi bliżej unaoczniającymi stopień i źródło negatywnego wyniku porównania rozpoznawanego obiektu ze wzorcem należącym do oczekiwanej odpowiedzi. Wydruki takie otrzymywano następująco : Nieznany obiekt wcześniej fałszywie rozpoznany poddawano ponownie operacjom rozpoznawania, tym razem dwukrotnie. Za pierwszym razem wyznaczony został ciąg wartości współczynnika niezgodności poszczególnych fragmentów (każdego złożonego z dwóch kolejnych widm binarnych) rozpoznawanego obiektu z odnośnymi fragmentami wzorca należącego do uzyskanego wyniku rozpoznania. Za drugim razem dla zadanego wzorca, (zwykle tego, który powinien być najbardziej podobny do rozpoznawanego obiektu), otrzymywano wydruk uproszczonej numeracji fragmentów tego wzorca w kolejności ich podo-

bieństwa do poszczególnych, kolejnych fragmentów rozpoznawanego obiektu. Numeracji tej towarzyszyły znaki +, -, . . Znak informował, czy fragment danego wzorca o numerze następującym po znaku jest bardziej, mniej lub jednakowo podobny do odnośnego fragmentu obiektu niż optymalnie odpowiadający temu samemu fragmentowi obiektu fragment wzorca błędnej odpowiedzi. I tak z przykładowego wydruku zamieszczonego na rys. 1 odczytać można, że kolejnym fragmentom 1,2,3,4 itd. obiektu (w ramach normalizacji czasowej stosowanej podczas porównywania spektrogramów binarnych) odpowiadają jako najbardziej podobne fragmenty rozpatrywanego wzorca KROSNO o numerach kolejnych 1,3,3,3 oraz, że np. fragment 6 tego wzorca jest bardziej podobny do szóstego fragmentu obiektu niż jakikolwiek z możliwych fragment wzorca należącego do odpowiedzi GNIEZNO. Wydruk zawiera na końcu nazwy dwóch wzorców będące wyrazami słownika, do których te wzorce należą. Pierwszy odnosi się do wzorca, którego wydruk dotyczy, a drugi do wzorca odpowiedzi. Liczba po pierwszym z tych dwóch wyrazów równa jest łącznej liczebności znaków "+" i "." i określa globalne podobieństwo wzorca tego wyrazu do obiektu. Podobnie liczba po drugim wyrazie równa jest sumie liczebności znaków "-" i "." i wyraża globalne podobieństwo wzorca tego wyrazu z obiektem.

```
GNIEZNO
 1 2 3 4 5 6 7 0 1 2 3 4 5 6 7 0 1 2 3 4 5 6 7 0 1 2 3
-1+3+3+3+4.6.4+6+0+2+2+2.4-4.5.6-6.5-6-2.3.4.4-4-5-4-7 KROSNO    22 GNIEZNO 22
```

Rys. 1. Wydruk numeracji odpowiadających sobie fragmentów rozpoznawanego obiektu (wiersz liczbowy górny) i wzorca (wiersz liczbowy dolny) oraz znaków różnic podobieństw poszczególnych fragmentów obiektu z odnośnymi fragmentami wzorca KROSNO i należącego do oczekiwanej odpowiedzi wzorca GNIEZNO.

Posługiwano się też niekiedy wydrukami wyrażającymi podobieństwo poszczególnych wzorców z rozpoznawanym obiektem, przy czym odniesienie do porównań stanowił wzorzec wyniku rozpoznawania. Przykład takiego wydruku zamieszczono na rycinie 2.

POZNAŃ

GNIEZNO	17	POZNAŃ	22
GDYNIA	20	POZNAŃ	24
KUTNO	20	POZNAŃ	24
JAROCIN	17	POZNAŃ	26
KOSCIAN	15	POZNAŃ	32
WARSZAWA	20	POZNAŃ	26
BIAŁYSTOK	20	POZNAŃ	26
KALISZ	10	POZNAŃ	34
KATOWICE	14	POZNAŃ	32
KIELCE	13	POZNAŃ	31
KONIN	12	POZNAŃ	32
KOSZALIN	20	POZNAŃ	30
KRAKOW	10	POZNAŃ	34
KROSNO	12	POZNAŃ	30
OSTROŁĘKA	15	POZNAŃ	31
PIŁA	21	POZNAŃ	25
PŁOCK	14	POZNAŃ	33
POZNAŃ	37	POZNAŃ	37
PRZEMYSL	15	POZNAŃ	27
RADOM	13	POZNAŃ	34
RZESZOW	13	POZNAŃ	31
BYDGOSZCZ	14	POZNAŃ	33
CHEŁM	17	POZNAŃ	26
CIECHANOW	21	POZNAŃ	26
CZESTOCHOWA	17	POZNAŃ	27
ELBLAG	11	POZNAŃ	30
GDAŃSK	12	POZNAŃ	33
GORZOW	21	POZNAŃ	25
LEGNICA	21	POZNAŃ	24
LESZNO	14	POZNAŃ	27
LUBLIN	15	POZNAŃ	33
ŁOMZA	22	POZNAŃ	25
ŁODŹ	16	POZNAŃ	26
OLSZTYN	15	POZNAŃ	34
OPOLE	21	POZNAŃ	25
SIEDLCE	14	POZNAŃ	34
SIERADZ	06	POZNAŃ	34
WADOWICE	17	POZNAŃ	26
SŁUPSK	17	POZNAŃ	32
SUWAŁKI	20	POZNAŃ	21
SZCZECIN	15	POZNAŃ	31
TARNOBRZEG	16	POZNAŃ	31
TARNOW	12	POZNAŃ	32
TORUŃ	12	POZNAŃ	33
WAŁBRZYCH	14	POZNAŃ	32
WŁOCŁAWEK	22	POZNAŃ	23
WROCŁAW	14	POZNAŃ	27
ZAMOSC	11	POZNAŃ	34
POZNAŃ			

Rys. 2. Przykład wydruku ilościowych ocen podobieństwa rozpoznawanego obiektu z poszczególnymi wzorcami (Odniesienie stanowi podobieństwo tegoż obiektu ze wzorcem odpowiedzi).

Stworzono też możliwość wydruku wartości współczynnika niezgodności m_{nz} ($m_{nz} = 1 - m_p$, gdzie m_p jest współczynnikiem podobieństwa porównywanych fragmentów konfrontowanych spektrogramów binarnych - patrz praca [5]) poszczególnych fragmentów obiektu z odnośnymi fragmentami wzorców. Także ten rodzaj wydruku okazał się pomocny przy wyjaśnianiu przyczyn błędów w rozpoznawaniu wyrazów.

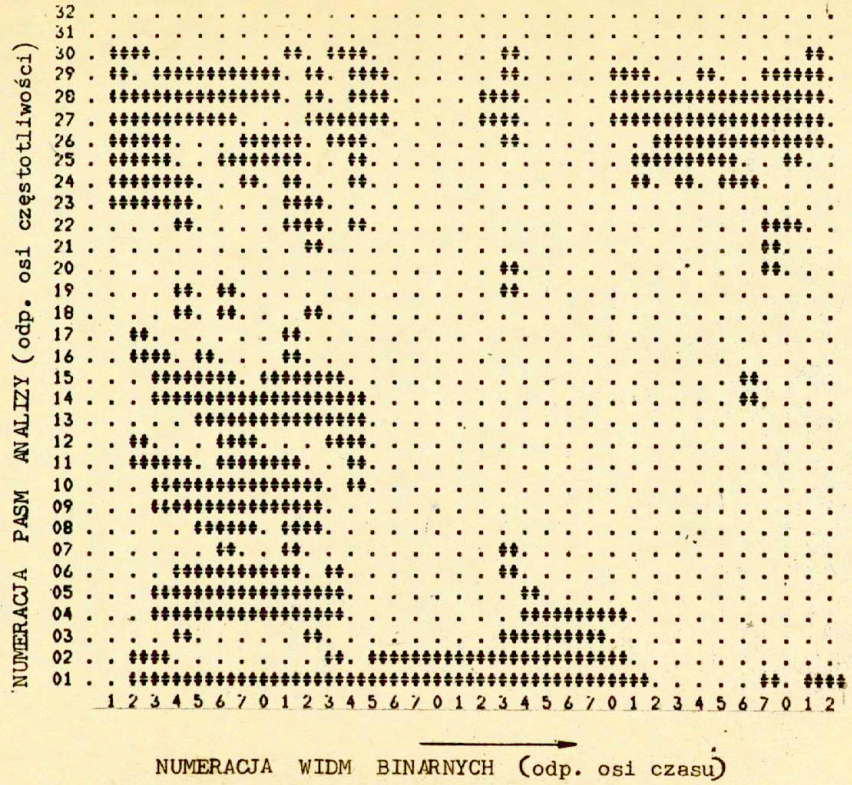
Korzystając ze wszystkich wymienionych wyżej możliwości wglądu w tok powstawania wyniku rozpoznawania, ujawniono różne przyczyny błędów. W wielu przypadkach do powstania błędu w rozpoznawaniu przyczyniły się różnice treści widmowej odpowiadających sobie fragmentów obiektu i wzorca. Treść ta niekiedy była bogatsza we wzorcu, innym razem w obiekcie. Dla przykładu zamieszczono na ryc. 3 zestaw obiektu i wzorca wyrazu TARNOW. Różnią się te spektrogramy treścią widmową w obrębie fragmentu [nuf], przy czym bogatsze jest widmo wzorca. Tak znaczne różnice jak w zacytowanym przykładzie powodują, że niektóre fragmenty obiektu są mniej podobne do odnośnych fragmentów właściwego wzorca niż innych wzorców.

Podobne przyczyny błędów zachodziły w przypadkach, gdy rozpoznawany obiekt miał dłuższy fragment identyczny lub podobny z odpowiednim fragmentem obcych mu wzorców. Np. zdarzyło się, że koniec [no] rozpoznawanego wyrazu KUTNO był bardziej podobny do końca wzorca KROSNO niż końca wzorca KUTNO. Podobnie zachodziło większe podobieństwo końca [uf] rozpoznawanego wyrazu KRAKÓW z analogiczną końcówką wzorca wyrazu TARNÓW niż wzorca wyrazu KRAKÓW. Bywały przypadki, że z wyżej wymienionych powodów w równym stopniu pretendowało do wyniku kilka wzorców, liczbowo jednakowo podobnych do obiektu, lecz dzięki podobieństwu z jego różnymi fragmentami. Wybór padał wówczas na ten wzorzec, który w inwentarzu zajmował najdalsze miejsce. Przypomnieć w tym miejscu należy, że podobieństwo do danego wzorca wyrażone jest liczbą fragmentów obiektu optymalnie przystających do odpowiednich fragmentów tego wzorca. Łącznie w 31 przypadkach przedstawione wyżej rodzaje przyczyn błędów wpłynęły po części lub w całości na błędne wyniki rozpoznawania.

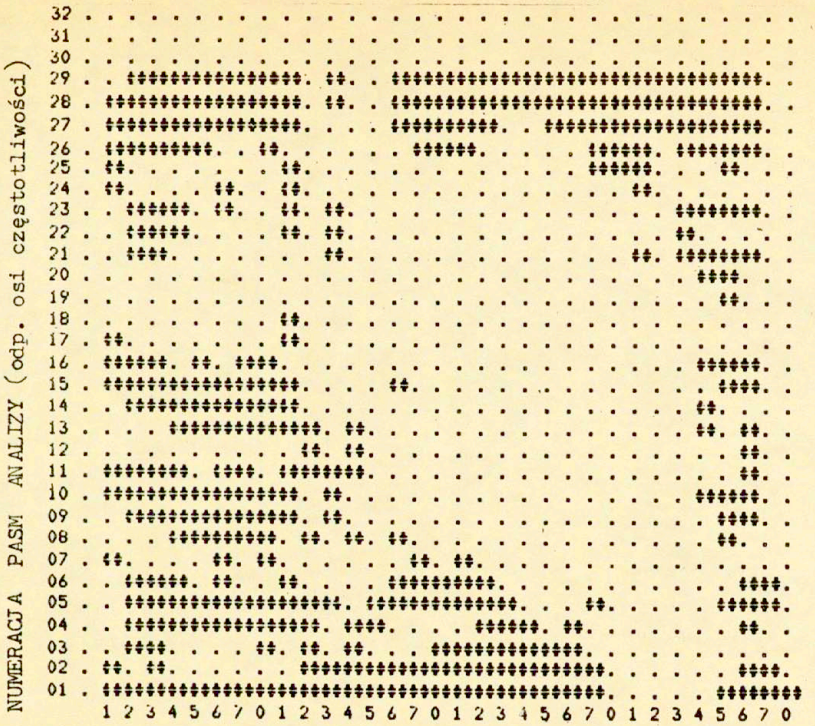
Inne przyczyny, które w 25 przypadkach zaważyły na błędach w rozpoznawaniu, polegały na dużej różnicy względnej długości czasowej odpowiadających sobie segmentów we wzorcu i obiekcie. Różnice te dotyczyły np. przerw przedplozyjnych, segmentów afrykacji i szumowych. Zauważono przypadki wydłużenia końca wypowiedzi lub braku początku wypowiedzi, gdy wyraz rozpoczynał się od głoski zwartej dźwięcznej lub dźwięcznej trącej. Oprócz tego miały miejsce przypadki pojawienia się zakłóceń bezpośrednio przed lub po wypowiedzi. Przykład dużego zróżnicowania proporcji czasowych odpowiadających sobie segmentów rozpoznawanego obiektu WADOWICE i właściwego dla niego wzorca pokazano na rycinie 4. Spektrogram wzorcowy w odróżnieniu od spektrogramu obiektu nie zawiera w tym przykładzie pierwszego fonemu [v], posiada krótką przerwę przedplozją [ts] oraz długą ostatnią samogłoskę.

Istotną cechą założeń metody rozpoznawania wyrazów na podstawie spektrogramów binarnych jest możliwość porównywania dwóch wyrazów o różnej rozciągłości czasowej. W przeważającej liczbie przypadków to założenie przynosi pozytywne rezultaty. Sama idea porównywania spektrogramów binarnych różnej długości czasowej przedstawiona w pracy [3] nie wymaga weryfikacji. Jednakże przypadek z ryciny 4, jak i inne jemu podobne wskazują na konieczność wprowadzenia pewnych ilościowych zmian. Gdyby proporcje wymiarów czasowych poszczególnych segmentów w każdym z porównywanych ze sobą spektrogramów binarnych odnoszących się do tego samego wyrzu były identyczne, wówczas po zastosowaniu liniowej normalizacji czasowej porównywane spektrogramy można by uważać za przystające. Odpowiadające sobie fragmenty jednego i drugiego spektrogramu powinny znajdować się wtedy w tych samych miejscach na osi czasu. Ponieważ wspomniane wyżej proporcje są różne, przyjęto zasadę, według której, po normalizacji czasowej każdemu fragmentowi F_i jednego spektrogramu odpowiadać powinien jeden z fragmentów $F'_{i-k}, \ldots\ldots F'_{i+k}$ wyokienkowanych z drugiego spektrogramu. Okno obejmujące fragmenty $F'_{i-k} \ldots\ldots F'_{i+k}$ powinno mieć szerokość 2-krotnie większą niż wynosi maksymalna odległość czasowa Δn pomiędzy odpowiadającymi sobie fragmentami podobnych spektrogramów binarnych znarmalizowanych czasowo. Jak się okazuje, warunek ten jest niezawsze spełniony, gdyż okno obejmuje

OBIEKT

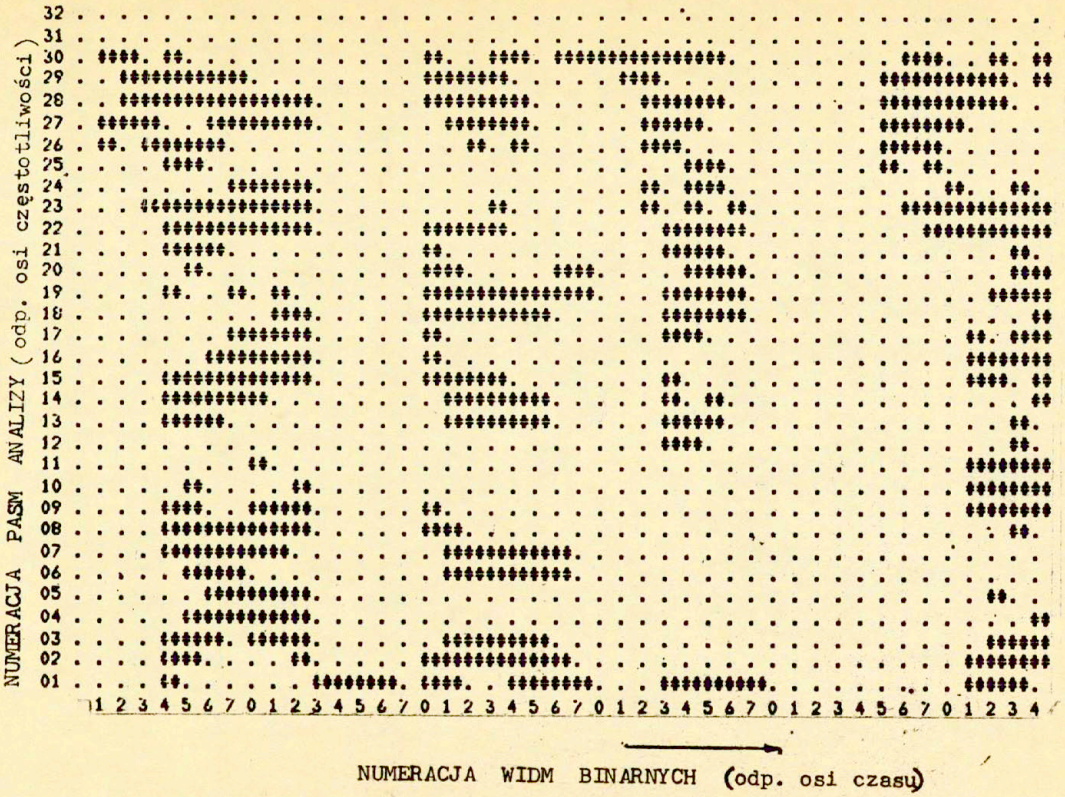


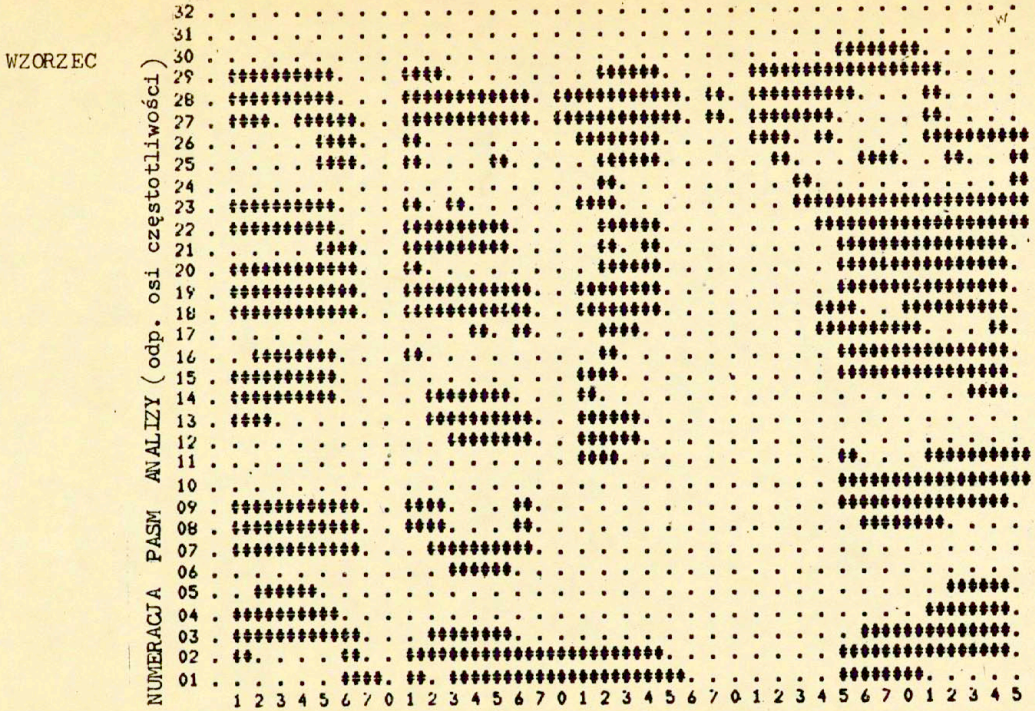
WZORZEC



Ryc. 3. Przykład spektrogramów binarnych wzorca TARNÓW i jednej z wypowiedzi tego wyrazu.

OBIEKT





Ryc. 4. Przykład zróżnicowania proporcji czasowych odpowiadających sobie segmentów rozpoznawanego obiektu WADOWICE i właściwego dla niego wzorca.

5 fragmentów, a wspomniana odległość Δ n osiągała w przypadkach błędnego rozpoznania 6 odstępów czasowych między kolejnymi fragmentami. Należy zatem poszerzyć okno 2-krotnie.

Dalszego zbadania wymagają te przypadki, w których uwidaczniają się na spektrogramie binarnym głoski zwarte dźwięczne przypadające na początku wyrazu oraz inne zróżnicowania w binarnych obrazach spektrograficznych tego samego hasła wypowiedzianego kilkakrotnie przez ten sam głos, jak w przykładzie wspomnianym wyżej na str. 15. Zjawiska te, sądząc po ich sporadycznym wpływie na wyniki rozpoznawania, nie są częste. Przyczyn ich można dopatrywać się w różnym poziomie intensywności głosu i prawdopodobnej zależności obrazu widma od tegoż poziomu. Mogą być również istotne warunki nadawania i transmisji sygnału.

Z wyników niniejszej pracy wypływają następujące końcowe wnioski :

1. 32 widmowe parametry binarne opisujące sygnał mowy w zakresie częstotliwości od 80 Hz do 8310 Hz wystarczają do niemal jednoznacznego wyrażenia w formie spektrogramu binarnego każdego z 48 różnych wyrazów wymówionych przez jeden głos.

2. Podstawowy fragment spektrogramu, którym operuje się w procesie porównywania dwóch spektrogramów binarnych, może składać się z dwóch zamiast trzech sąsiednich widm binarnych.

3. Zbieżność wyników dla trzech różnych i przypadkowo dobranych głosów świadczy korzystnie o stabilności i wiarygodności obranej metody.

4. Uzyskany procent poprawnie rozpoznanych wyrazów uznać można za dobry i skłaniający do dalszego rozwijania zastosowanej metody rozpoznawania.

5. Analiza przyczyn błędów wykazała, że istnieją możliwości uzyskania dalszej poprawy wyników rozpoznawania.

Na najbliższą przyszłość przewiduje się przeprowadzenie doświadczeń mających na celu dalszą weryfikację przyjętej metody i zbadanie możliwości jej optymalizacji.

Ukierunkowanie wdrożeniowe przedstawionej metody, zmierzające do :

a/ dalszego powiększenia słownika oraz

b/ skrócenia czasu oczekiwania na rozpoznanie do czasu

prawie rzeczywistego,
wymaga dysponowania minikomputerem spełniającym w porównaniu do MERY 303 następujące warunki :

1. Dłuższe słowo maszynowe - 32 bity
2. Większa pamięć operacyjna - co najmniej 32 K słów
3. Krótszy czas wykonywania operacji elementarnej
4. Lepsza organizacja operacji.

BIBLIOGRAFIA

[1] KUBZDELA, H. : Techniczna realizacja formantowej metody rozpoznawania samogłosek polskich, Prace IPPT 90/1975, Warszawa, 1975.

[2] KUBZDELA, H. : Wyznaczanie charakterystycznego fragmentu samogłoskowego i pomiar częstotliwości formantów dla automatycznej klasyfikacji i identyfikacji samogłosek, Prace IPPT 41/1979, Warszawa, 1979.

[3] KUBZDELA, H. : Metoda automatycznego rozpoznawania wyrazów w oparciu o spektrogramy binarne, Prace IPPT 14/1980, Warszawa, 1980.

[4] KUBZDELA, H. : Wizualizacja sygnału mowy w formie spektrogramów binarnych, Materiały Otwartego Seminarium z Akustyki, str. 167-171, Warszawa-Puławy, 1980.

[5] KUBZDELA, H. : Automatyczne rozpoznawanie wyrazów na podstawie spektrogramów binarnych, Prace IPPT 15/1981, Warszawa, 1981.

[6] 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 39/1976, Warszawa, 1976.