

Henryk Kubzdela

15/84

AUTOMATYCZNE  
ROZPOZNAWANIE WYRAZÓW  
NA PODSTAWIE  
SPEKTROGRAMÓW BINARNYCH



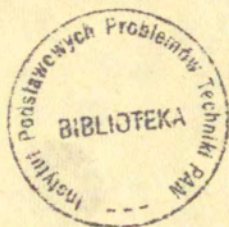
P. 269

WARSZAWA 1981

ISSN 0208-5658

Praca wpłynęła do Redakcji dnia 21 listopada 1980 r.

Zarejestrowana pod nr 15/1981



57098



Na prawach rękopisu

---

Instytut Podstawowych Problemów Techniki PAN

Nakład 140 egz. Ark. wyd. 1,3. Ark. druk. 2.

Oddano do drukarni w kwietniu 1981 r.

Nr zamówienia 258/0/81

---

Warszawską Drukarnia Naukowa, Warszawa,  
ul. Sniadeckich 8

AUTOMATYCZNE ROZPOZNAWANIE WYRAZÓW  
NA PODSTAWIE SPEKTROGRAMÓW BINARNYCH<sup>1</sup>

1. Wstęp.

Niniejsza praca relacjonuje kolejny etap badań dotyczących automatycznego rozpoznawania mowy na podstawie tzw. spektrogramów binarnych. Pojęcie spektrogramu binarnego a także jeden z konkretnych sposobów jego wyznaczania omówiono w pracach [1], [3], [4] a także w części pracy [2] będącej obecnie w druku. Pracę [2] poświęcono głównie przedstawieniu metody automatycznego rozpoznawania wyrazów poprzez rozpoznawanie ich obrazów binarnych. Podano w niej dwa zasadnicze algorytmy, pierwszy - tworzenia, dla poszczególnych wyrazów, wzorcowych spektrogramów binarnych, zwanych też krótko wzorcami oraz drugi - identyfikacji nieznanego obiektu, jakim jest spektrogram binarny rozpoznawanego wyrazu. Pracę obecną poświęcono zweryfikowaniu tej metody w oparciu o pierwsze wyniki adaptacji i identyfikacji. W rezultacie wprowadzono do pierwotnie przyjętych zasad kilka zmian i poprawek, które omówiono poniżej. Podano wyniki rozpoznawania wyrazów uzyskane dla wersji modelu jaką uznano za najkorzystniejszą spośród rozpatrywanych. Po omówieniu wyników w konkluzji sformułowano ocenę zastosowanej metody automatycznego rozpoznawania wyrazów. Należy zaznaczyć, że praca ta nie wyczerpuje zagadnienia rozpoznawania mowy na podstawie spektrogramów binarnych lecz stanowi jedno z pierwszych ogniw badań na ten temat. Badania takie będą przedmiotem zadań autora w najbliższej przyszłości.

<sup>1</sup>

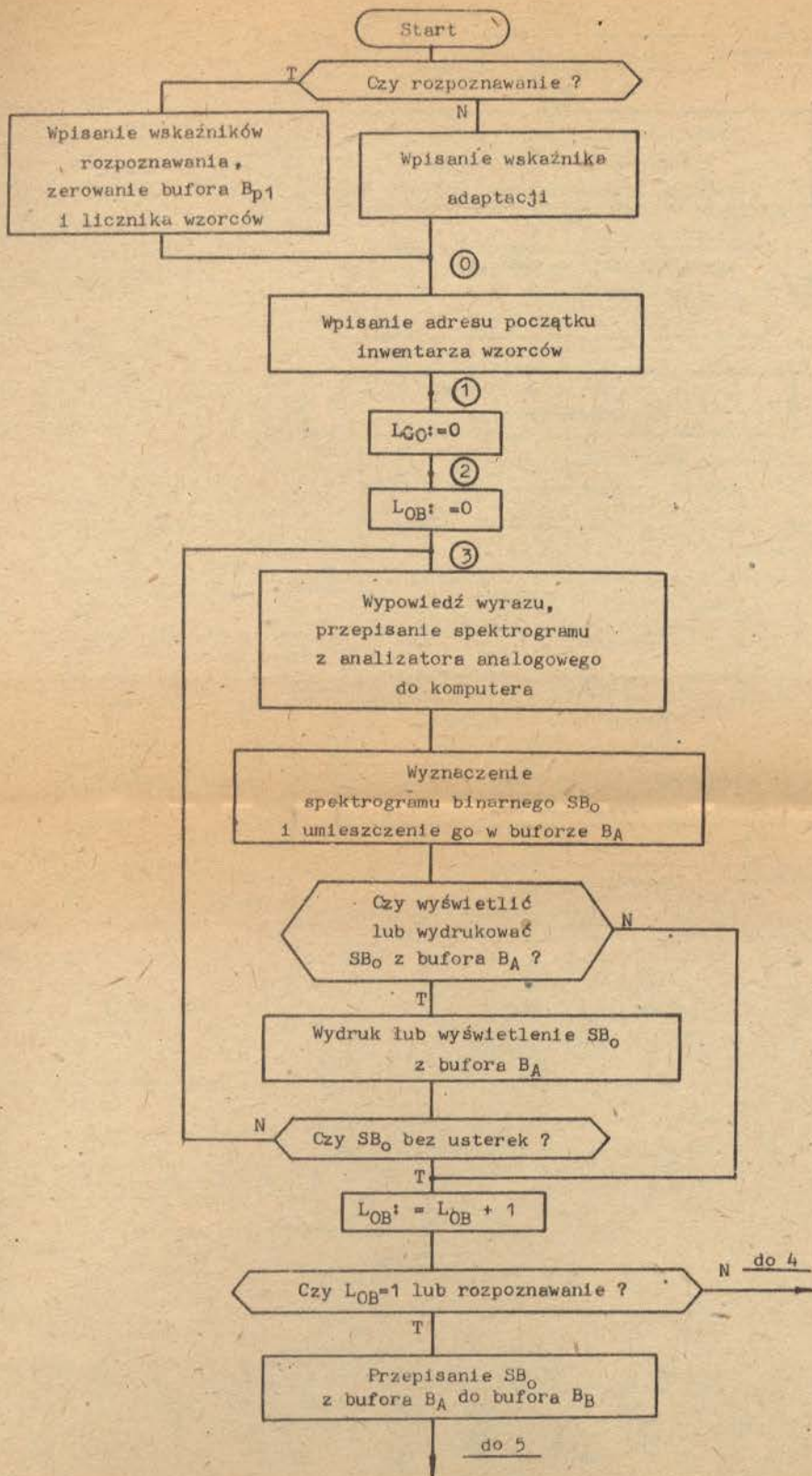
Praca wykonana w ramach problemu międzyresortowego MR.I-24

2. Realizacja koncepcji automatycznego rozpoznawania wyrazów (ARW) na podstawie ich spektrogramów binarnych (SB).

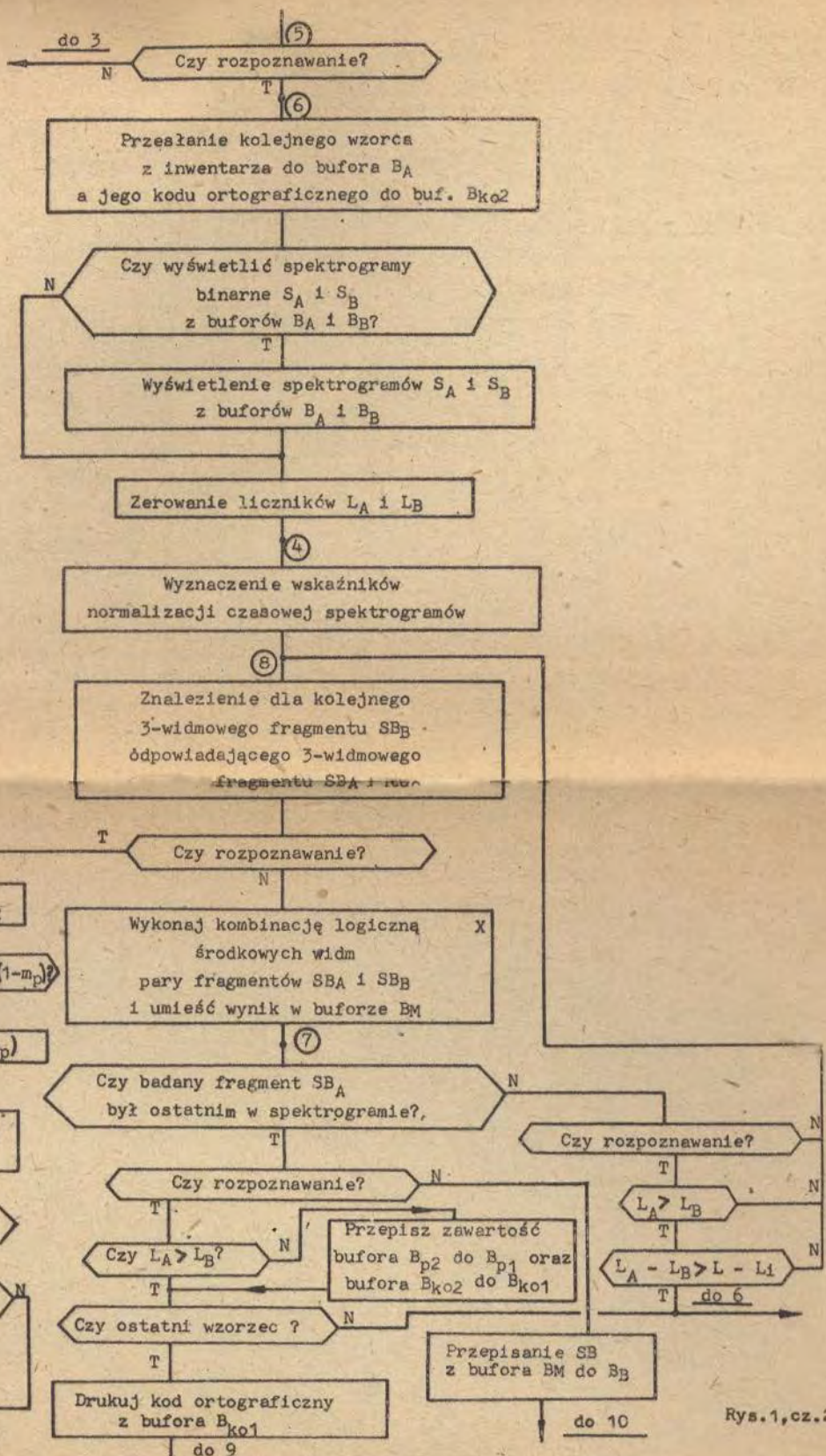
Na rys. 1 przedstawiono schemat ogólny operacji adaptacji i identyfikacji składających się na rozpoznawanie spektrogramów binarnych. Algorytm szczegółowy jest bardzo obszerny i z tego powodu nie został tutaj zamieszczony. W oparciu o ten szczegółowy algorytm ułożono programy w języku wewnętrznym systemu komputerowego MERA 303, który stanowił podstawowe narzędzie badań. Algorytm ogólny zamieszczony na rys. 1 zawiera dodatkowo operacje kontrolne uruchamiane dowolnie przez operatora, takie jak :

- a) operacja wyświetlania lub wydruku SB bieżącego wyrazu i SB będącego pośrednim lub końcowym wynikiem adaptacji,
- b) operacja wyświetlania na ekranie oscyloskopu par spektrogramów binarnych, z których jeden jest nieznanym obiektem a drugi jednym z wzorców zmagazynowanych w pamięci komputera,
- c) operacja wydruku danych o podobieństwie odpowiadających sobie fragmentów SB wzorca i SB obiektu.

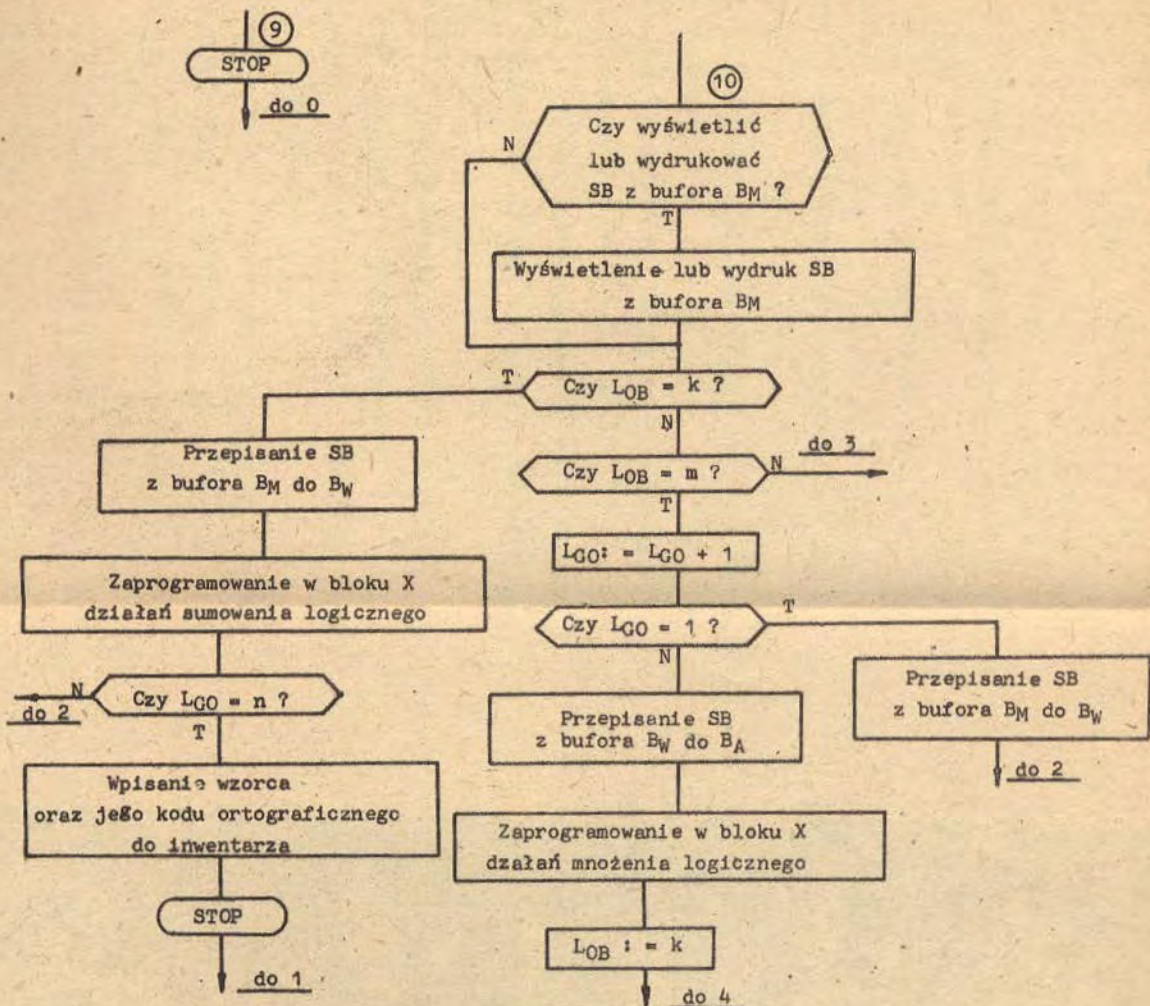
Nie ma natomiast w algorytmie zamieszczonym na rys. 1 tych układów operacji, które rozpatrywano w badaniach lecz w wyniku weryfikacji wyeliminowano i zastąpiono innymi. Cały kompleks programów realizujących operacje ARW, wraz z operacjami kontrolnymi składa się z około 3400 rozkazów zajmujących blisko połowę pamięci operacyjnej minikomputera MERA 303. W procesie adaptacji potrzebne są 4 równe bufony pamięciowe o łącznej pojemności 1024 bajtów. Podczas identyfikacji SB udział biorą 3 z tych buforów. Pozostała część pamięci operacyjnej minikomputera MERA 303 służy do przechowywania inwentarza wzorców rozpoznawanych wyrazów. Wzorcem określonego wyrazu jest spektrogram binarny wyznaczony w procesie adaptacji na podstawie spektrogramów binarnych kilku wypowiedzi tegoż wyrazu przez ten sam głos lub kilka głosów bardzo podobnych. W przeznaczony na inwentarz wzorców pamięci operacyjnej minikomputera MERA 303 mieszczą się wzorce 13-tu średniej długości wyrazów. Czas operacji rozpoznawania nieznanego wyrazu zależy od długości tego wyrazu, miejsca jego wzorca w inwentarzu, stopnia podobieństwa do innych wzorców oraz od kilku innych czynników.



Rys. 1, cz. 1.



Rys. 1, cz. 2.



Rys. 1, cz. 3.

Rys. 1. Uproszczony algorytm operacji automatycznego rozpoznawania wyrazów na podstawie spektrogramów binarnych. Rys. 1, którego ten podpis dotyczy składa się z trzech części.

Objaśnienia oznaczeń występujących w schemacie algorytmu zamieszczonym na rys. 1.

- SB - spektrogram binarny,
- SB<sub>A</sub>, SB<sub>B</sub> - spektrogramy binarne A i B,
- L<sub>OB</sub> - licznik obiektów i bufor kodu operacji logicznej
- L<sub>GO</sub> - licznik grup obiektów,
- m - założona liczba obiektów odpowiednio sumowanych logicznie,
- n - założona liczba sum obiektów odpowiednio mnożonych logicznie,
- k - kod deklarujący operację logicznego mnożenia spektrogramów binarnych,
- m<sub>p1</sub> - wartość miary podobieństwa danego fragmentu obiektu z odnośnym fragmentem porównywanego wzorca,
- m<sub>p</sub> - wartość miary podobieństwa danego fragmentu obiektu z odnośnym fragmentem jednego z wcześniej porównywanych wzorców,
- L<sub>A</sub> - liczebność kolejnych fragmentów obiektu wykazujących większe podobieństwo z odnośnymi fragmentami najodpowiedniejszego wzorca spośród wcześniej porównanych niż aktualnie porównywanego,
- L<sub>B</sub> - liczebność kolejnych fragmentów obiektu wykazujących większe podobieństwo z odnośnymi fragmentami aktualnie porównywanego wzorca niż wzorca najodpowiedniejszego spośród wcześniej porównanych,
- L - liczba wszystkich 3 - widmowych fragmentów spektrogramu binarnego obiektu,
- L<sub>i</sub> - liczebność fragmentów porównanej części obiektu.



### 2.1. Adaptacja.

Zakładając, że czytelnikowi dostępny jest opis podstawowej zasady adaptacji zawarty w publikacji [2] podane zostaną tutaj jedynie pewne innowacje wprowadzone do pierwotnej metody adaptacji w wyniku weryfikacji. Zasady dwóch operacji w łańcuchu działań adaptacyjnych są szczególnie ważne i wymagały eksperymentalnego wyboru, Pierwsza z nich dotyczy znajdowania podobnych wśród odpowiadających sobie fragmentów spektrogramów wzorca i obiektu. Określenie podobieństwa wymaga zastosowania odpowiedniej miary. Wzięto pod uwagę 3 wersje takiej miary i zweryfikowano je eksperymentalnie. Porównaniu podlegają każdorazowo 2 fragmenty spektrogramów binarnych, każdy złożony z trzech widm binarnych. Każdy taki fragment można uznać za trójkolumnową macierz zero-jedynkową. Liczba wierszy tej macierzy równa jest liczbie parametrów widmowych i wynosi 40. Wszystkie 3 zbadane miary podobieństwa wyrazić można jedną formułą matematyczną :

$$m_p = 2 \cdot \frac{k_1 z_1 + k_0 z_0}{2k_1^2 z_1 + k_0^2 (2z_0 + n_0) + n_1} \quad (1)$$

w której  $z_1$  i  $z_0$  są odpowiednio liczebnościami par jedynek i zer występujących zgodnie w obu macierzach natomiast  $n_1$  i  $n_0$  odpowiednio liczebnościami jedynek i zer występujących niezgodnie.  $k_1$ ,  $k_1'$ ,  $k_0$  i  $k_0'$  są współczynnikami wagowymi.

Dla pierwszej wersji miary podobieństwa założono :

$$k_1' = k_0' = k_1 = k_0 = 1.$$

Mianownik w wyrażeniu (1) staje się w tym przypadku stały, gdyż wszystkie porównywane ze sobą fragmenty spektrogramów są macierzami zero-jedynkowymi jednakowych rozmiarów. Wyznaczenie podobieństwa porównywanych macierzy sprowadza się do policzenia, w ilu odpowiadających sobie miejscach obie macierze mają zgodne liczby binarne.

Następna wersja miary podobieństwa różni się od poprzedniej jedynie tym, że współczynnik  $k_1 = 2$ , co oznacza nadanie większej wagi zgodnie występującym jedynekom w odpowiadających sobie miejscach obu macierzy.

W trzeciej wersji miary podobieństwa nie mają udziału li-

czebności zgodnie występujących zer, natomiast liczebność zgodnie występujących jedynek odniesiona zostaje do liczebności wszystkich jedynek w porównywanych dwóch macierzach. Jest to równoznaczne z podstawieniem we wzorze (1) :  $k_0 = k'_0 = 0$ . Wyznaczenie podobieństwa wymaga więc w tym przypadku operacji dzielenia. Spośród tych trzech przedstawionych wersji miary podobieństwa najkorzystniejszą okazała się wersja trzecia. Porównywanie macierzy zero-jedynkowych ma miejsce zarówno w procesie adaptacji jak i identyfikacji i odbywa się seriami. W jednej takiej serii porównań bierze zawsze udział jedna macierz pochodząca z jednego spektrogramu np. A oraz pięć macierzy będących wycinkami drugiego spektrogramu np. B. W przypadku adaptacji A jest pośrednią formą wzorca a B jest spektrogramem obiektu natomiast w przypadku identyfikacji A jest obiektem a B wzorcem. Wśród pięciu macierzy jednego spektrogramu znaleziona ma być macierz najbardziej podobna do macierzy pochodzącej z drugiego spektrogramu. Przy zastosowaniu pierwszej i drugiej wersji miary podobieństwa stwierdzono przypadki błędów. Miały one miejsce zwykle wtedy, gdy macierz, dla której szukano podobnej w grupie pięciu macierzy, miała mniej jedynek od najbardziej do niej podobnej wśród tych pięciu oraz gdy równocześnie jedna z tych pięciu macierzy była zerowa. Przypadek taki zilustrowano na rys. 2. Można by ograniczyć tego rodzaju błędy przez powiększenie współczynnika wagowego  $k_1$ . Tym właśnie, że  $k_1 > 1$  wariant drugi miary podobieństwa różni się od pierwszego. Stwierdzono, że także i ten wariant nie zapewnia bezbłędnych wyników. Gdy zachodzi np. tego rodzaju przypadek, jak pokazany na rys. 3 następuje błędne określenie macierzy  $M_{B(j+2)}$  jako najbardziej podobnej do macierzy  $M_{A_i}$ , ponieważ macierz  $M_{B(j+2)}$  posiada więcej jedynek, między innymi we wszystkich tych miejscach co macierz  $M_{A_i}$ . Bardziej podobna wizualnie do macierzy  $M_{A_i}$  jest macierz  $M_{B(j-2)}$  która nie posiada jednak jedynek we wszystkich tych miejscach co macierz  $M_{A_i}$ . Błąd spowodowała właśnie ta okoliczność, że zgodnie występującym jedynkom nadano większą wagę. Mimo, że pierwsze dwa warianty podobieństwa byłyby bardzo korzystne ze względu na prostotę obliczeń, to jednak z przyczyn wyżej zilustrowanych przykładami należało oba te warianty odrzucić.

$M_{Ai}$	$M_{B(j-2)}$	$M_{B(j-1)}$	$M_{Bj}$	$M_{B(j+1)}$	$M_{B(j+2)}$
1 1 1	1 1 1	1 1 0	1 0 0	0 0 0	0 0 0
0 0 0	1 1 1	1 1 0	1 0 0	0 0 0	0 0 0
0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
= = =	= = =	= = =	= = =	= = =	= = =
0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
0 0 0	1 1 1	1 1 0	1 0 0	0 0 0	0 0 0
1 1 1	1 1 1	1 1 0	1 0 0	0 0 0	0 0 0
0 0 0	1 1 1	1 1 0	1 0 0	0 0 0	0 0 0
0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
0 0 0	1 1 1	1 1 0	1 0 0	0 0 0	0 0 0
1 1 1	1 1 1	1 1 0	1 0 0	0 0 0	0 0 0

Macierz  $M_{Ai}$  jest podobna : a/ wizualnie do macierzy  $M_{B(j-2)}$ ,

b/ na podstawie  $m_p$  (wg 1-szej wersji)  
do macierzy  $M_{B(j+2)}$ .

Rys. 2. Przypadek, w którym ma miejsce błędne określenie podobieństwa macierzy przy zastosowaniu pierwszej wersji miary podobieństwa.

$M_{Ai}$	$M_{B(j-2)}$	$M_{B(j-1)}$	$M_{Bj}$	$M_{B(j+1)}$	$M_{B(j+2)}$
0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
= = =	= = =	= = =	= = =	= = =	= = =
0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
0 0 0	0 0 0	0 0 0	0 0 1	0 1 1	1 1 1
0 0 0	0 0 0	0 0 1	0 1 1	1 1 1	1 1 1
1 1 1	0 0 0	0 0 1	0 1 1	1 1 1	1 1 1
1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1
1 1 1	0 0 0	0 0 0	0 0 1	0 1 1	1 1 1
0 0 0	0 0 0	0 0 0	0 0 1	0 1 1	1 1 1
0 0 0	0 0 0	0 0 1	0 1 1	1 1 1	1 1 1
1 1 1	0 0 0	0 0 1	0 1 1	1 1 1	1 1 1
1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1
1 1 1	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0

Macierz  $M_{Ai}$  jest podobna : a/ wizualnie do macierzy  $M_{B(j-2)}$ ,

b/ na podstawie  $m_p$  (wg 2-giej wersji)  
do macierzy  $M_{B(j+2)}$ .

Rys. 3. Przypadek, w którym ma miejsce błędne określenie podobieństwa macierzy przy zastosowaniu drugiej wersji miary podobieństwa.

Wariant trzeci miary podobieństwa okazał się najpewniejszy. Posługiwanie się taką właśnie miarą podobieństwa wymaga jednak bardziej złożonych działań maszynowych, co wydłuża nieco czasu adaptacji i identyfikacji.

Drugim działaniem, które odgrywa zasadniczą rolę w procesie adaptacji i które poddać należało weryfikacji jest modyfikacja spektrogramu binarnego będącego pośrednią formą wzorca wyłonioną w poprzednich krokach adaptacji. Oznaczmy ten spektrogram umownie symbolem  $SB_{WP}$ . Na modyfikację całego spektrogramu  $SB_{WP}$  składają się modyfikacje jego poszczególnych widm. Modyfikację każdego widma binarnego poprzedza omówione wyżej porównywanie macierzy zero-jedynkowych, czyli porównywanie fragmentu  $F_{Wi}$  spektrogramu  $SB_{WP}$  z pięcioma fragmentami  $F_{O(j-2)}, \dots, F_{O(j+2)}$  spektrogramu binarnego obiektu  $SB_O$ . Gdy liczby widm  $L_W$  i  $L_O$  w spektrogramie  $SB_{WP}$  oraz w spektrogramie obiektu są równe lub różnią się co najwyżej o dwa, wówczas  $j = i$ . W przeciwnym razie wartość  $j$  nie wzrasta zgodnie z  $i$  lecz na określony przyrost wartości  $i$ ,  $j$  wzrasta o 1 mniej lub o 1 więcej niż  $i$ , zależnie od tego, który ze spektrogramów  $SB_O$  i  $SB_{WP}$  jest dłuższy. Regułę tę można zapisać następująco :

Gdy  $|L_W - L_O| > 2$ , wówczas dla  $i = k \Delta L_W$   
 $j = i - k$ , jeżeli  $L_W > L_O$   
 oraz  $j = i + k$ , jeżeli  $L_W < L_O$ ,

przy czym  $k = 1, \dots, \left( \frac{L_W}{\Delta L_W} + \delta_1 \right)$  oraz

$$\Delta L_W = \frac{L_W}{L_W - L_O} + \delta_2,$$

gdzie  $\delta_1$  i  $\delta_2$  są odpowiednio zaokrągleniami ilorazów :

$$\left| \frac{L_W}{L_W - L_O} \right| \quad \text{i} \quad \frac{L_W}{\Delta L_W} \quad \text{do najbliższej liczby całkowitej.}$$

W wyniku porównania  $F_{Wi}$  z  $F_{O(j-2)}, \dots, F_{O(j+2)}$  wiadomo, który z pięciu fragmentów obiektu jest najbardziej podobny do fragmentu  $F_{Wi}$  spektrogramu  $SB_{WP}$ . Środkowe widma obu tych

fragmentów biorą udział w modyfikacji. Polega ona na logicznej kombinacji odpowiadających sobie parametrów jednego i drugiego widma. W pracy [2] założono, że kombinacją tą będzie suma logiczna. W pierwszych eksperymentach adaptacyjnych poddano to założenie weryfikacji. Stwierdzono, że wzorcowy SB otrzymany w ten sposób zawiera zbyt wiele szczegółów nie reprezentujących istotnych cech spektrogramów binarnych uczestniczących w adaptacji. Wskutek sumowania logicznego kumulują się w wynikowym spektrogramie binarnym wszystkie cechy przypadkowe. Zastąpienie sumowania mnożeniem logicznym zapobiega temu, lecz wywołuje inny negatywny skutek. Przypadkowy zanik którejkolwiek z istotnych cech choćby w jednym ze spektrogramów formujących wzorzec wywołuje brak tej cechy także w wynikowym spektrogramie wzorcowym. Przyjęto więc rozwiązanie kompromisowe oparte na założeniu, że cechą istotną jest ta, która pojawi się co najmniej w jednym z  $m$  kolejnych spektrogramów binarnych  $SB_{O_i}$  formujących wzorzec. Wzorcowy spektrogram binarny  $SB_W$  powstaje zatem jako iloczyn logiczny  $n$  sum logicznych, przy czym na każdą sumę składa się  $m$  spektrogramów binarnych  $SB_{O_j}$ . Zasadę tę wyraża następująca formuła :

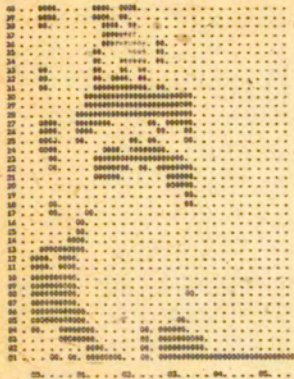
$$SB_W = \prod_{i=0}^n \sum_{j=1}^m SB_{(j+mi)} \quad (2)$$

w której symbole  $\prod$  i  $\sum$  oznaczają mnożenie i sumowanie logiczne. Przyjęto wstępnie  $n = 1$ ,  $m = 2$ .

Działania logiczne odnoszące się do spektrogramów binarnych w istocie wykonywane są oddzielnie na kolejnych parach odpowiadających sobie widm pochodzących z tych spektrogramów i dobranych w sposób wyżej opisany. Porządek tworzenia wzorcowego spektrogramu binarnego w procesie adaptacji ujęty jest w algorytmie na rys. 1. Na rys. 4 zamieszczono komplet spektrogramów binarnych składających się na wynik adaptacji wyrazu [osiem]. Spektrogramy 1 i 2 dotyczą pierwszych dwóch wypowiedzi tego wyrazu. Spektrogram I jest wynikiem logicznego zsumowania spektrogramów 1 i 2. Z następnych dwóch wypowiedzi wyrazu [osiem] pochodzą spektrogramy 3 i 4 a ich sumą logiczną jest spektrogram II. Iloczyn logiczny spektrogramów I i II jest



I



II



WZORZEC



Rys. 4, cz. 2.

Spektrogramy binarne składające się na wzorzec wyrazu [osiem]. Skala pozioma dotyczy wymiaru czasowego spektrogramu binarnego i wycechowana jest w dziesiątych częściach sekundy. Skala pionowa odnosi się do wymiaru częstotliwościowego SB, który wyrażony jest numerami poszczególnych pasm częstotliwości.

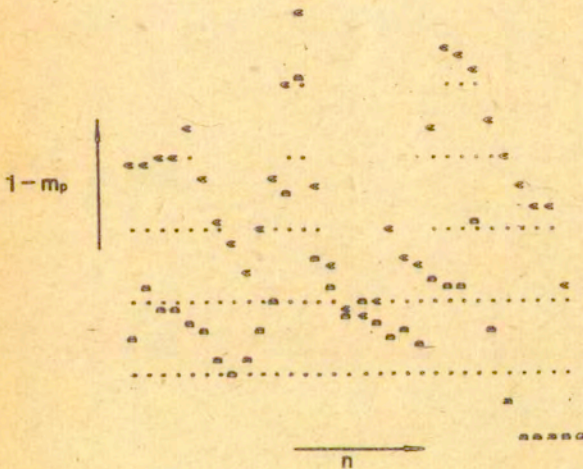


wynikiem adaptacji i stanowi wzorcowy spektrogram binarny wyrazu [osiem]. Wzorzec ten zostaje umieszczony w inwentarzu wzorców wraz z kodem symbolizującym wyraz [osiem], którym jest zapis ortograficzny tego wyrazu.

## 2.2. Identyfikacja spektrogramu binarnego.

Jak wiadomo z pracy [2] i wyjaśnień podanych w części 2.1. a także z algorytmu zamieszczonego na rys. 1, procesy adaptacji i identyfikacji przebiegają w znacznej części wspólnym torem. Jedno z istotnych rozgałęzień tego toru występuje w miejscu, w którym kończy się porównywanie kolejnego fragmentu jednego spektrogramu binarnego z pięcioma odpowiednimi fragmentami drugiego. Jeżeli w grę wchodzi adaptacja, następnym działaniem jest modyfikowanie widma, jeżeli zaś odbywa się identyfikacja nieznanego obiektu, kolejne działanie jest inne. Według opracowania przedstawionego w publikacji [2] miało to działanie polegać na obliczeniu osobno liczebności wszystkich jedynek oraz jedynek występujących zgodnie na odpowiadających sobie miejscach w środkowym widmie binarnym fragmentu obiektu i jednego z pięciu fragmentów wzorca. Te dwie dane wyliczane dla kolejnych 3-widmowych fragmentów  $SB_0$  i podobnych do nich odpowiednich fragmentów  $SB_W$  należało kumulować osobno. Ze stosunku wyników kumulacji obu tych danych obliczanego po każdym kroku identyfikacji proponowano wnioskować o podobieństwie nieznanego obiektu do danego wzorca. Ten szczegół metody identyfikacji obiektu okazał się nieodpowiedni głównie z dwóch powodów. Pierwszym powodem było występowanie dzielenia liczb 2-bajtowych, które wymagałoby dodatkowych programów kosztem uszczuplenia obszaru pamięci operacyjnej przeznaczonej na inwentarz wzorców oraz wydłużenia czasu oczekiwania na wynik rozpoznawania. Akumulator w minikomputerze MERA 303 jest bowiem jednobajtowy. Drugim powodem było to, że jak się okazuje nie we wszystkich przypadkach spodziewać by się można trafnych wyników. Przy takim sposobie identyfikacji fragmenty obiektu i wzorca, w których ilość jedynek jest bardzo mała, w niewielkim stopniu rzutują na wartość współczynnika określającego podobieństwo spektrogramów obiektu i wzorca. Wprowadzono więc szereg innowacji do pierwotnej metody identyfikacji, które usunęły wymienione wady. Podobieństwo  $SB_0$  i  $SB_W$  wyrażono ciągiem wartości

miary podobieństwa  $m_p$  kolejnych fragmentów  $SB_0$  i odpowiadających im fragmentów  $SB_w$ . Zasada kompletowania podobnych fragmentów obu tych spektrogramów jest podobna jak w przypadku adaptacji. Formalna różnica polega jedynie na tym, że w trakcie identyfikacji dla kolejnych fragmentów obiektu poszukiwane są odpowiadające i podobne fragmenty wzorca odwrotnie niż podczas adaptacji. Na rys. 5 zamieszczono 2 wykresy wartości wyrażenia  $1 - m_p$  pochodzących z porównania kolejnych fragmentów spektrogramu binarnego wyrazu X z właściwymi fragmentami spektrogramów wzorcowych wyrazów A i B. Decyzja o tym, do którego z dwóch wzorców A lub B obiekt X jest najbardziej podobny, opiera się na ocenie, ile fragmentów obiektu X wykazuje na podstawie wartości  $m_p$  lepsze podobieństwo do wzorca A niż do B a ile lepsze podobieństwo do wzorca B niż do A.



Rys. 5. Wykresy wartości wyrażenia  $1 - m_p$  pochodzących z porównania kolejnych fragmentów  $m_p$  spektrogramu binarnego wyrazu X z właściwymi fragmentami spektrogramów wzorcowych wyrazów A i B.

- $m_p$  - miara podobieństwa dwóch odpowiednich fragmentów porównywanych spektrogramów binarnych.
- $n$  - numery kolejne fragmentów obiektu porównywanych z odpowiednimi fragmentami wzorca.

Założono przy tym, że jeśli różnice pomiędzy wartościami  $m_p$  odnoszącymi się do tego samego fragmentu obiektu mieszczą się w określonych arbitralnie ustalonych granicach ( $-a$ ,  $+a$ ) to uznaje się je za nieistotne. Wyliczyć zatem należy, jaka jest liczebność  $L_A$  przypadków, w których spełniona jest nierówność :

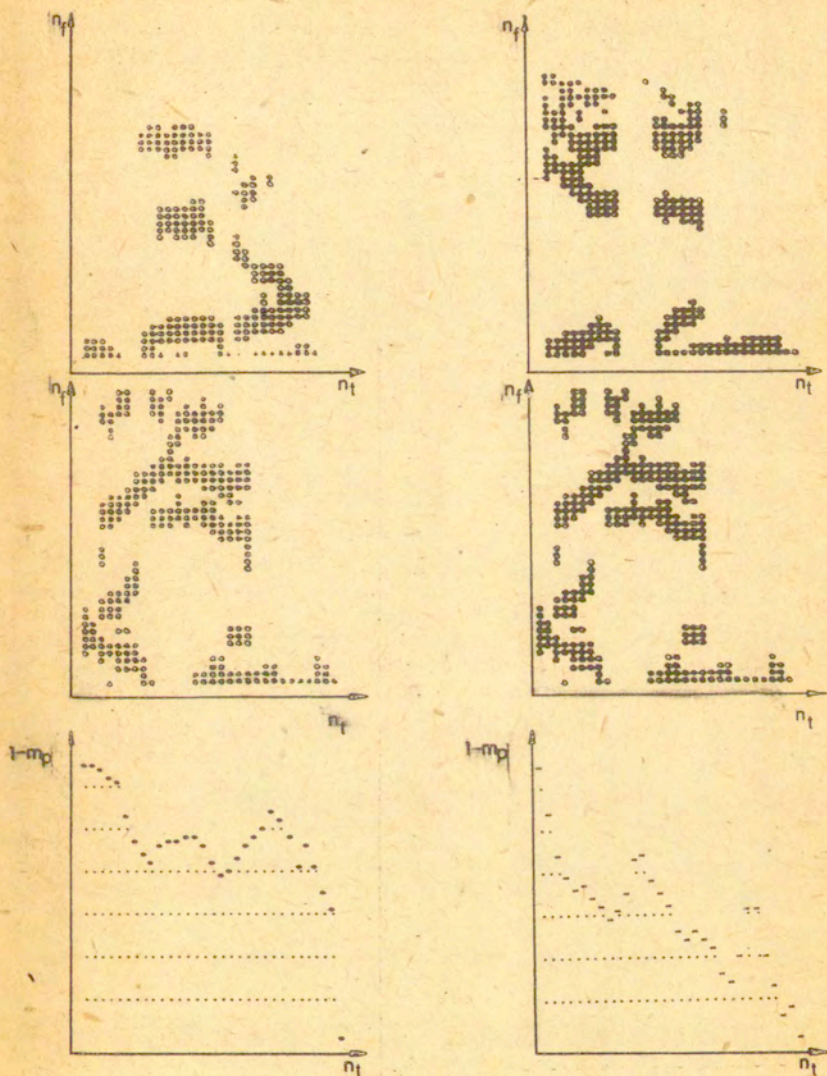
$$m_{pA} - m_{pB} > a \quad (3)$$

oraz jaka jest liczebność przypadków  $L_B$ , w których spełniona jest nierówność odwrotna :

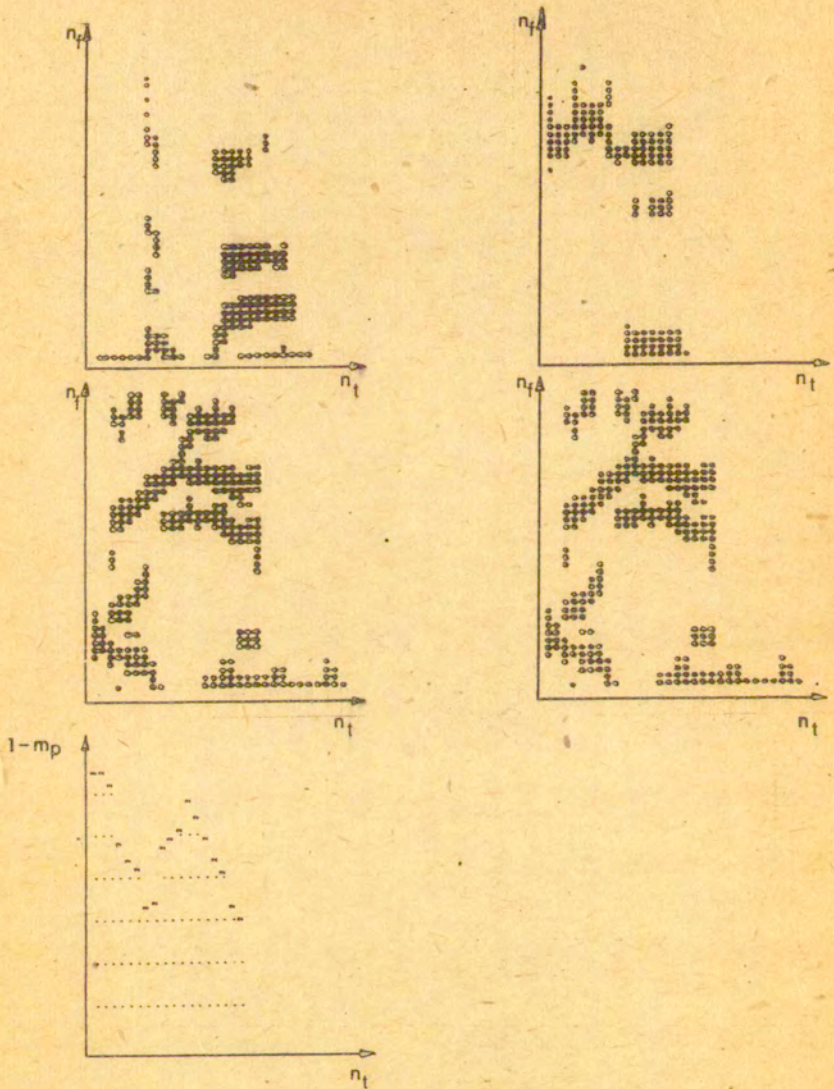
$$m_{pB} - m_{pA} > a \quad (4)$$

Jeżeli  $L_A - L_B \geq 0$ , wówczas uznaje się, że obiekt X jest bardziej podobny do wzorca A niż do B. Ciąg wartości miary podobieństwa  $m_{pA}$  reprezentujący podobieństwo kolejnych fragmentów obiektu X z odpowiednimi fragmentami wzorca A zostaje przechowany w celu skonfrontowania go z wynikiem porównania obiektu z następnym wzorcem inwentarza. Jeżeli nierówność  $L_A - L_B \geq 0$  nie jest spełniona, wówczas uznaje się, że obiekt X jest bardziej podobny do wzorca B niż do A i przechowany zostaje ciąg wartości  $m_{pB}$  otrzymanych z porównania obiektu X ze wzorcem B. Liczebności  $L_A$  i  $L_B$  są uzupełniane i porównywane po uzyskaniu każdej nowej danej o  $m_p$ . Porównanie obiektu z danym wzorcem zostaje przerwane, jeżeli wiadomo, że kontynuowanie porównania nie zmieni przewagi  $L_A$  nad  $L_B$  lub  $L_B$  nad  $L_A$ . Przedstawiony tutaj fragment metody identyfikacji obiektu został eksperymentalnie zweryfikowany i zaakceptowany.

Na rys. 6 zamieszczono serię obrazów par spektrogramów binarnych. Każdy obraz przedstawia spektrogram tego samego obiektu oraz spektrogram jednego z wzorców. Na podstawie tych ilustracji można ocenić, do którego wzorca obiekt jest najbardziej podobny. Poniżej każdego obrazu pary spektrogramów zamieszczony jest wykres wartości wyrażenia  $1 - m_p$ , ilustrujący podobieństwo tych spektrogramów. Im bliżej osi odciętych ten wykres jest usytuowany tym większe jest podobieństwo obiektu do wzorca. Brak takiego wykresu dotyczy przypadku istnienia dużej różnicy w długościach spektrogramów obiektu



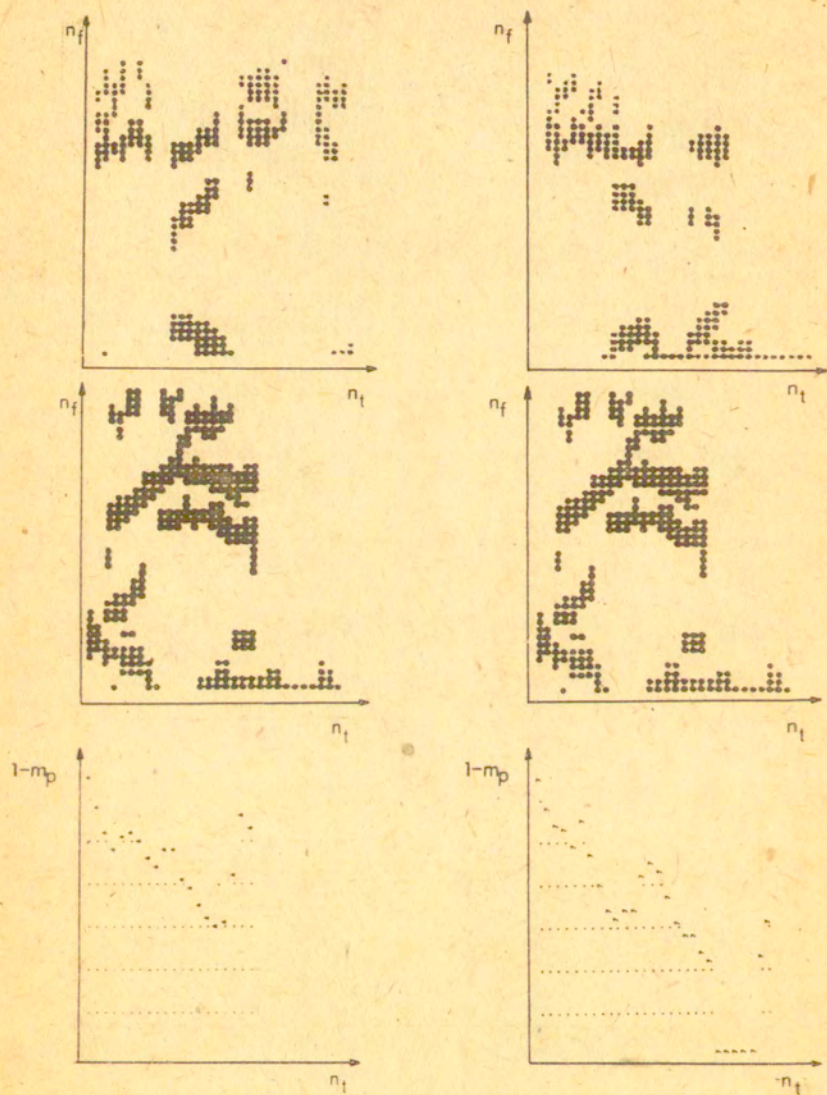
Rys. 6, cz. 1 .



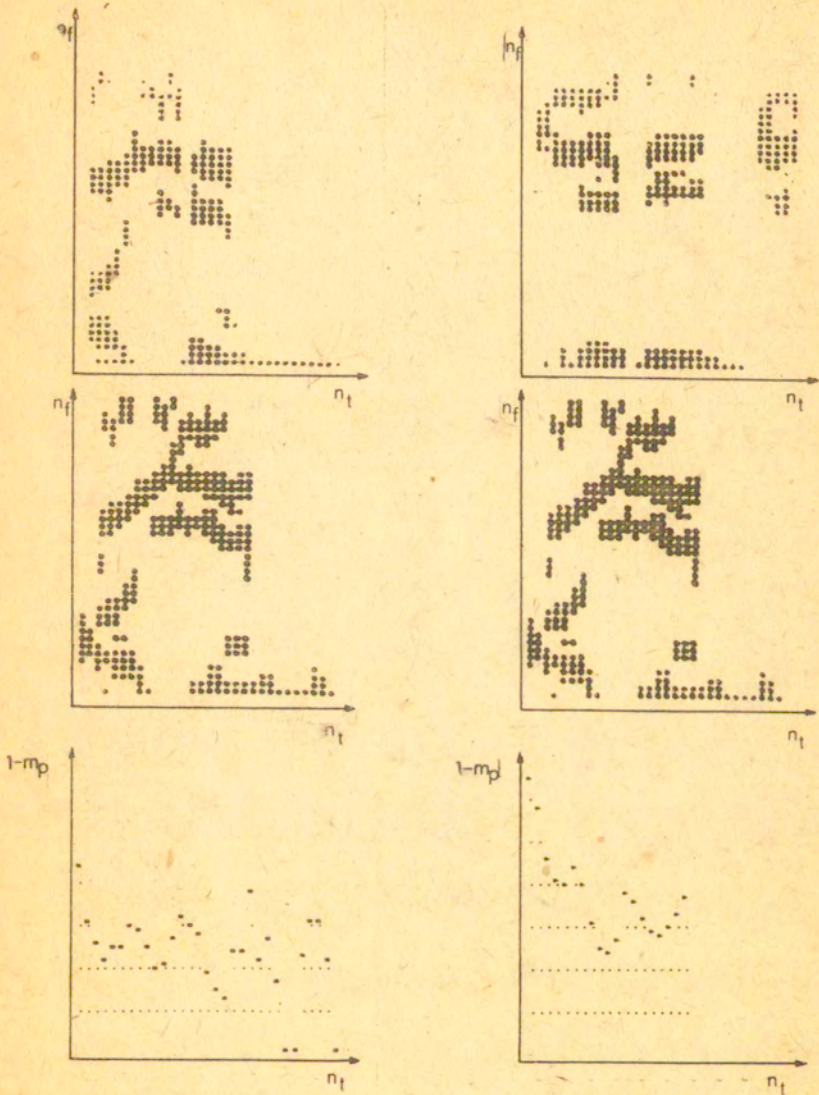
Rys. 6 , cz. 2 .



Rys. 6 , cz. 3.

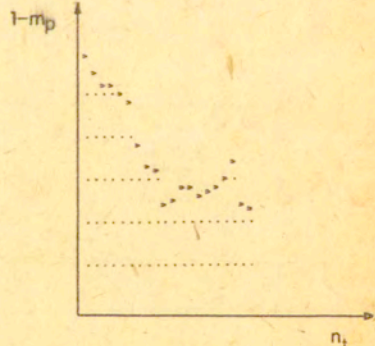
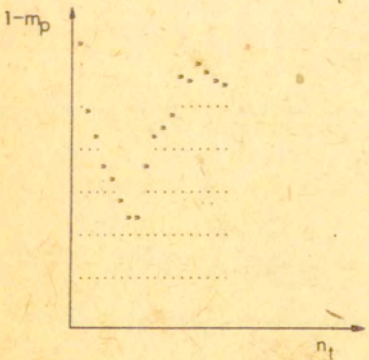
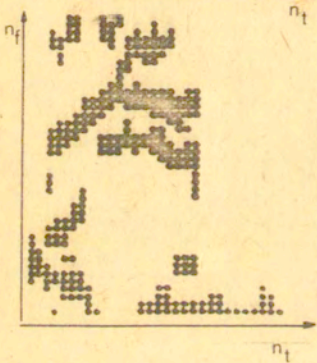
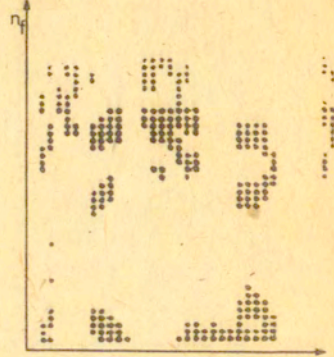
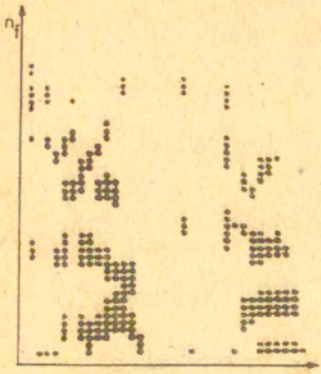


Rys. 6 , cz. 4 .

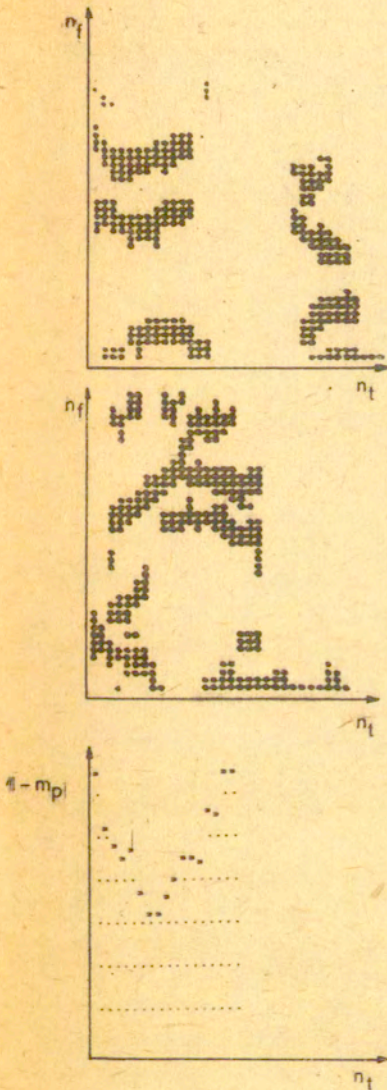


Rys. 6 , cz. 5 .





Rys. 6 , cz. 6 .



Rys. 6, cz. 7.

Obrazy par spektrogramów binarnych obiektu i poszczególnych wzorców oraz wykresy ilustrujące podobieństwo między spektrogramami każdej pary. Podpis ten odnosi się także do części 1 do 6 rysunku 6.

i wzorca. Jeśli bowiem przypadek taki ma miejsce, obiekt nie jest porównywany z danym wzorcem, gdyż bliższe podobieństwo takich dwóch spektrogramów zostaje wykluczone. Niektóre wykresy wartości wyrażenia  $1 - m_p$  na rys. 6 są krótsze od innych. Krótszy wykres świadczy o tym, że porównanie obiektu ze wzorcem zostało przerwane ponieważ jego kontynuacja nie zmieni już decyzji o tym, że wzorzec aktualnie rozpatrywany jest mniej podobny do obiektu niż jeden ze wzorców wcześniej porównywanych.

### 3. Opis i wyniki prób automatycznego rozpoznawania wyrazów.

Do automatycznego rozpoznawania wyrazów przyjęto słownik złożony z trzynastu następujących wyrazów :

- |           |               |
|-----------|---------------|
| 1. zero   | 8. siedem     |
| 2. jeden  | 9. osiem      |
| 3. dwa    | 10. dziewięć  |
| 4. trzy   | 11. kropka    |
| 5. cztery | 12. przecinek |
| 6. pięć   | 13. kreska    |
| 7. sześć  |               |

W eksperymencie, którego wyniki podane są niżej, brał udział jeden głos męski. Mówca zajmował w czasie wypowiedzi miejsce bezpośrednio przy komputerze zainstalowanym w normalnym niewytłumionym pomieszczeniu. Brak było izolacji akustycznej od hałasów pochodzących z urządzeń zestawu komputerowego i spoza pomieszczenia. Uwolniono się od wpływu tych zakłóceń stosując mikrofon o bardzo krótkim zasięgu i wąskiej charakterystyce kierunkowej. Dla każdego z trzynastu wyrazów słownika wyznaczono wzorcowy spektrogram binarny i umieszczono go w obszarze pamięci komputera przeznaczonym na inwentarz wzorców. Wzorzec był tworzony w procesie adaptacji na podstawie czterech wypowiedzi danego wyrazu. Wzorzec etykietowano jego kodem ortograficznym. Wyznaczenie spektrogramu wzorcowego nie trwa długo. Jeżeli podczas adaptacji obserwowane są spektrogramy binarne, na podstawie których tworzony jest wzorzec oraz spektrogramy będące formami pośrednimi powstającego wzorca, wówczas adaptacja jednego wyrazu trwa około 5 minut a wszystkich trzynastu wyrazów słownika około 1 godziny. Obrazy spektrogramów wzorcowych wszystkich wyżej podanych wyrazów słownika figurują na rys. 5. Ten sam głos, który brał udział w adaptacji wymawiał

rozpoznawane wyrazy w porządku losowym, każdy wyraz słownika 20-krotnie. W sumie więc poddano rozpoznawaniu 260 wypowiedzi. Wyniki rozpoznawania zamieszczone są w tabeli nr 1.

R	N	0	1	2	3	4	5	6	7	8	9	.	,	-
0		20												
1			17						3					
2				18	2									
3					20									
4			1			18			1					
5							20							
6								20						
7			1						19					
8										20				
9											20			
.												18		2
,													20	
-														20

N - nadano, R - rozpoznano

Tabela 1. Wyniki automatycznego rozpoznawania wyrazów na podstawie spektrogramów binarnych.

Na 13 wyrazów słownika 8 rozpoznawanych było ze 100-procentową poprawnością, pozostałe 5 z poprawnością w granicach od 85 do 95 %. Czas oczekiwania na wynik rozpoznania oscylował w granicach kilku sekund a najczęstsza jego wartość przypadła pomiędzy 5 a 7 sekund. Większość błędów rozpoznawania można wytłumaczyć, jak następuje : wyraz [jeden] w trzech przypadkach rozpoznany był jako [siedem] a wyraz [siedem] w jednym przypadku także jako [jeden]. Te błędy uzasadnia okoliczność, że górna granica częstotliwości spektrogramu binarnego wynosi obecnie 3500 Hz. Różnica obwiedni widm fonemów [ɕ] oraz [j] jest poniżej tej częstotliwości nieduża, szczególnie gdy pierwszy formant fonemu [j] jest słaby, co bardzo często ma miejsce. Spodziewać się można, że pomyłki tego rodzaju nie będą zachodzić, gdy górna granica częstotliwości spektrogramów

binarnych zostanie podniesiona. Podobnie tłumaczy się 2-krotne rozpoznanie wyrazu [kropka] jako [kreska]. W zakresie częstotliwości poniżej 3500 Hz różnica między spektrogramami binarnymi tych wyrazów jest mała, szczególnie gdy słaba jest ploszja spółgłoski [p] w wyrazie [kropka]. Dwukrotnie błędnie rozpoznany był wyraz [dwa]. Pomiędzy spektrogramami binarnymi różnych wypowiedzi tego wyrazu istnieją różnice wynikające z różnych sposobów artykulacji głoski [d]. Periodyczny segment tej spółgłoski poprzedzający ploszję jest niekiedy tak słaby, że pominięty zostaje na spektrogramie binarnym.

Wspomnieć przy tej okazji należy, że w przypadku, gdy sygnał posiada w pewnym zakresie widmowym jeden tylko parametr różny od zera, wówczas odpowiadający mu parametr widma uśrednionego w procesie wczytywania go z analizatora analogowego do komputera jest 4-krotnie słabszy. Z kolei, jeżeli dany parametr  $a_i$  widma uśrednionego ma wartość poniżej pewnego progu, wtedy pochodny parametr  $b_i$  w widmie binarnym otrzymuje z góry wartość zero. Tak więc uśrednianie widma skądinąd konieczne może przyczynić się do wyzerowania dolnych części spektrogramu binarnego w segmentach odnoszących się do periodycznej części głosek zwartych dźwięcznych i niektórych dźwięcznych trących. Jeżeli taka głoska jest na początku wyrazu, nastąpić może skrócenie spektrogramu binarnego. Miało to po części miejsce dla tych wypowiedzi wyrazu [dwa], które układ błędnie rozpoznał. W celu wyeliminowania błędów tego rodzaju należy zwiększyć czułość układu na słabe segmenty w mowie. Zauważono także, że gdy mówca stara się wypowiedzieć głoskę [d] w wyrazie [dwa] zbyt starannie, bywa po niej artykułowany krótki segment o charakterze samogłoskowym, mający swoje niepożądane odbicie w obrazie spektrogramu binarnego. Autor nie znajduje właściwego wyjaśnienia na okoliczność, że wyraz [cztery] był w jednym przypadku rozpoznany jako [jeden] a w innym jako [siedem]. Przypuszcza się, że przyczyną tego mógł być ukryty błąd w programie, dający o sobie znać w pewnych sytuacjach. Sprawa ta wymaga wyjaśnienia. Szczegółowo analizowane będą wszelkie nowe przypadki błędnego rozpoznania.

#### 4. Uwagi końcowe i wnioski.

Przedstawione wyżej próby automatycznego rozpoznawania wyrazów stanowiły początek długiej serii podobnych prób zaplanowanych na najbliższą przyszłość. Praca niniejsza stanowiła zadanie badawcze autora w roku 1980. Ramy czasowe tego zadania nie pozwoliły na wyczerpujące zbadanie wszystkich cech metody automatycznego rozpoznawania wyrazów na podstawie spektrogramów binarnych. Praca bieżąca koncentrowała się głównie na oprogramowaniu, uruchomieniu i pierwszym zweryfikowaniu tej metody. Weryfikacja będzie kontynuowana w miarę pojawiania się różnych wątpliwych wyników rozpoznawania, które zapewne wskażą na potrzebę wprowadzenia dodatkowych zmian. Przedstawione w tej pracy pierwsze wyniki pozwalają twierdzić, że metoda rozpoznawania wyrazów na podstawie spektrogramów binarnych zasługuje na akceptację i warta jest dalszych badań. Metoda ta przy swojej prostocie, daje dobre rezultaty. Prosta jest w tej metodzie procedura adaptacji i krótkie są czasy oczekiwania na wynik rozpoznania. W ramach tej metody możliwe jest stworzenie biblioteki inwentarzy wzorców różnych słowników wyrazowych dla różnych grup głosów i dla różnych celów. Biblioteka taka mogłaby egzystować na jakichkolwiek pamięciowych nośnikach informacji. Adaptację sprowadzić by można wtedy do przetransmitowania z nośnika do pamięci operacyjnej gotowego inwentarza wzorców, który w określonym momencie miałby być użyty. Rozpoznawanie na podstawie spektrogramów binarnych obecnie wyrazów może w przyszłości dotyczyć także innych jednostek lingwistycznych lub też segmentów fonetyczno-akustycznych. Metodę rozpoznawania wyrazów na podstawie ich spektrogramów binarnych można by ogólniej określić jako metodę rozpoznawania obrazów binarnych sygnału mowy i jest rzeczą obojętną jakie segmenty tego sygnału obrazy te reprezentują. Rozpoznawanie segmentów fonetyczno-akustycznych będzie jednak możliwe dopiero wtedy, gdy istnieć będzie niezawodna metoda automatycznej segmentacji. Dotychczas metody takiej brak. Na podkreślenie zasługuje okoliczność, że mimo iż spektrogramy binarne, na podstawie których rozpoznawano wyrazy obejmowały jedynie zakres częstotliwości do około 3500 Hz, pierwsze wyniki rozpoznawania są ogólnie korzystne. W przyszłości zakres częstotliwości spektrogramów binarnych zostanie poszerzony.

Zamierza się także zredukować liczbę parametrów widma binarnego i zbadać jak wpłynie to na poprawność rozpoznawania.

## BIBLIOGRAFIA

- [1] KUBZDELA, H. : Wizualizacja sygnału mowy w formie spektrogramów binarnych, Sprawozdanie z XXVII Otwartego Seminarium z Akustyki, Warszawa - Puławy, 1980.
- [2] KUBZDELA, H. : Metodą automatycznego rozpoznawania wyrazów w oparciu o spektrogramy binarne, Prace IPPT nr 14/80, Warszawa, 1980.
- [3] RUSKE, G. : An Efficient binary representation of Sonograms, *Acustica*, vol. 34, No. 4, s. 234-239, Stuttgart, 1976.
- [4] RUSKE, G. : Real-time information reduction in digital sound spectrograms of Speech, *IEEE Int. Conf. on Acoust., Speech and Sign. Process.*, Philadelphia, 1976.