2.23 — rozpoznawania mowy

Henryk Kubzdela

# PRÓBY AUTOMATYCZNEGO ROZPOZNAWANIA WYRAZÓW WYMAWIANYCH PRZEZ RÓŻNE GŁOSY W OPARCIU O GRUPOWE ZBIORY WZORCOWYCH SPEKTROGRAMÓW BINARNYCH

47/83



WARSZAWA 1983

ISSN 0208-5658

Praca wpłynęła do Redakcji dnia 21 listopada 1983 r.



Na prawach rękopisu

Instytut Podstawowych Problemów Techniki PAN

Nakład 140 egz. Ark. wyd. 1,25. Ark.druk. 0,9
Oddano do drukarni w grudniu 1983 r.
Nr zamówienia 18/84.

Warszawska Drukarnia Naukowa, Warszawa,
ul. Śniadeckich 8

Henryk Kubzdela
Pracownia Fonetyki Akustycznej
IPPT PAN

# PRÓBY AUTOMATYCZNEGO ROZPOZNAWANIA WYRAZÓW WYŁAWIANYCH PRZEZ RÓŻNE GŁOSY W OPARCIU O GRUPOWE ZBIORY WZORCOWYCH SPEKTROGRAMÓW BINARNYCH [1]

## 1. Wstęp

Niniejsza praca jest sprawozdaniem z rocznego etapu długofalowych badań nad automatycznym rozpoznawaniem mowy na podstawie spektrogramów binarnych. W konkluzji etap ten dotyczył rozpoznawania izolowanych wyrazów w oparciu o grupowe zbiory wzorców. Eksperyment z rozpoznawaniem poprzedziły jednak liczne prace przygotowawcze, które miały na celu dostosowanie modelu rozpoznającego do postawionego mu zadania. Polegały one na wprowadzeniu kilku zasadniczych modyfikacji podyktowanych lub zasugerowanych we wnioskach z poprzednich prac lub będących wynikiem świeżych doświadczeń i przemyśleń autora. Niektóre manipulacje strukturalne w obrębie modelu miały charakter jedynie techniczny. Wspomina się o nich jednakże, ponieważ praca z nimi związana była bardzo czasochłonna i stanowiła znaczną część całości etapu. Zmodyfikowany model poddano wpierw podstawowej próbie, która polegała na rozpoznawaniu wyrazów słownika rozszerzonego do 100 elementów , czyli dwukrotnie liczniejszego niż w poprzednim doświadczeniu, zrelacjonowanym w pracy [2]. Wnioski zawarte w niniejszej publikacji stanowią uzupełnienie wcześniejszych ocen obranej przez autora metody rozpoznawania mowy.

---

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

## 2. Modyfikacja modelu rozpoznającego

Zmiany, jakich dokonano w modelu rozpoznającym , podyktowane zostały następującymi okolicznościami:

Zredukowanie liczby widmowych parametrów binarnych z 63 do 32 we wcześniejszych eksperymentach z rozpoznawaniem wyrazów [2] nie pogorszyło istotnie wyników rozpoznawania.

W wersji dotychczas stosowanej spektrogramy binarne różnych wypowiedzi tego samego wyrazu wykazują najznaczniejsze różnice w zakresie częstotliwości powyżej około 3 kHz, co nie rokuje osiągnięcia korzystnych wyników rozpoznawania wyrazów w oparciu o wzorce grupowe. Struktura spektrogramów binarnych we wspomnianym zakresie częstotliwości wydaje się być nadmiernie szczegółowa. Badania podsumowane w pracy [2] przyniosły między innymi wniosek, iż należy poszerzyć wycinek spektrogramu binarnego, w którym poszukiwany jest fragment - obejmujący kilka kolejnych widm - najbardziej przystający do odnośnego fragmentu innego spektrogramu. W tej samej pracy wykazano, iż fragment taki może składać się jedynie z dwóch sąsiadujących z sobą widm binarnych zamiast z trzech, jak założono w pierwotnej wersji.

Do wyżej wymienionych okoliczności dodać należy jeszcze dwa inne bardzo istotne czynniki, które inspirowały do zmodyfikowania modelu rozpoznającego. Czynnik pierwszy wynikał z tematu postawionego zadania badawczego, który,jak mówi tytuł niniejszej pracy, dotyczy rozpoznawania wyrazów w oparciu o grupowe zbiory wzorców. Należało zatem zmodyfikować cyfrową reprezentację wypowiedzi wyrazu tak, aby nie odgrywały w niej znaczącej roli szczegóły odzwierciadlające indywidualne cechy głosu. Czynnik drugi stanowiło dążenie do jak najdalej idącego uproszczenia metody rozpoznawania i to zarówno poprzez zredukowanie do niezbędnego minimum ilości informacji składającej się na spektrogram binarny jak i poprzez uproszczenie operacji wykonywanych w procesie rozpoznawania.

Po tych uwagach ogólnych zostaną obecnie bliżej omówione poszczególne zmiany dokonane w modelu rozpoznającym.

Do modelu wprowadzono definitywnie przedstawioną w pracy [3] zasadę wygładzania widma wyznaczonego przez wielokanałowy analogowy analizator widma. W myśl tej zasady każda rzędna widma wygładzonego jest sumą pięciu kolejnych odpowiednio ważonych rzędnych widma pierwotnego. Okienko wagowe tworzy ciąg następujących wartości współczynnika ważącego: 0.25 , 0.75 , 1 , 0.75 , 0.25 . W modelu zastosowano także po raz pierwszy zmodyfikowane

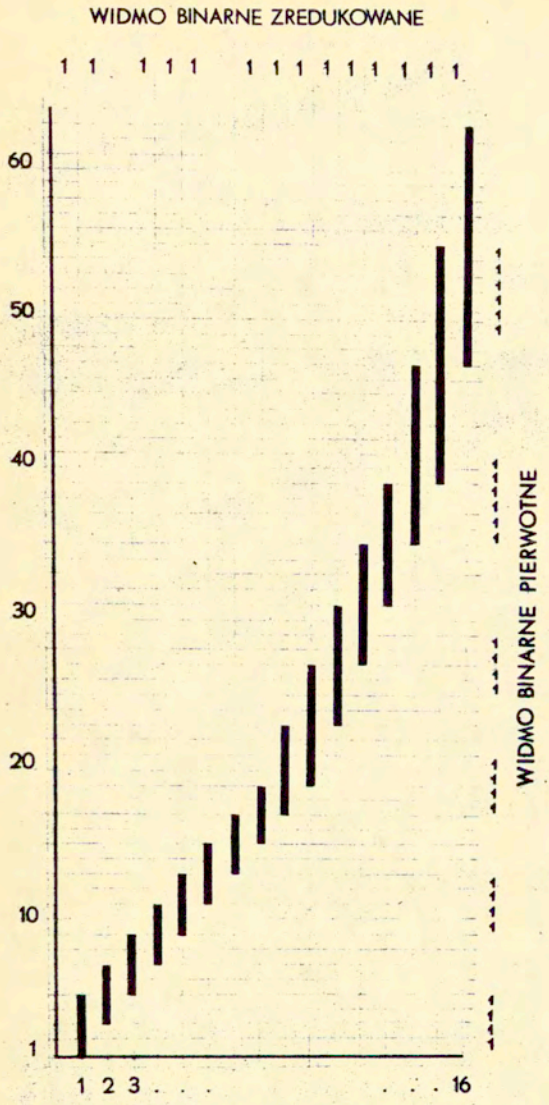
kryterium klasyfikacji binarnej parametrów widma wygładzonego w oparciu o które następuje przekształcenie tegoż widma w widmo binarne. Kryterium to wyrażone nierównościami (1) i (2) mówi, że o wartości 0 lub 1 i-tego parametru widma binarnego decyduje wklęsłość lub wypukłość obwiedni widma wygładzonego w miejscu i .

$$b_i = 1 \quad \text{dla} \quad p_i - \frac{p_{i-k} + p_{i+k}}{2} \geqslant a\left(p_{i-k} - p_{i+k}\right), \qquad (1)$$

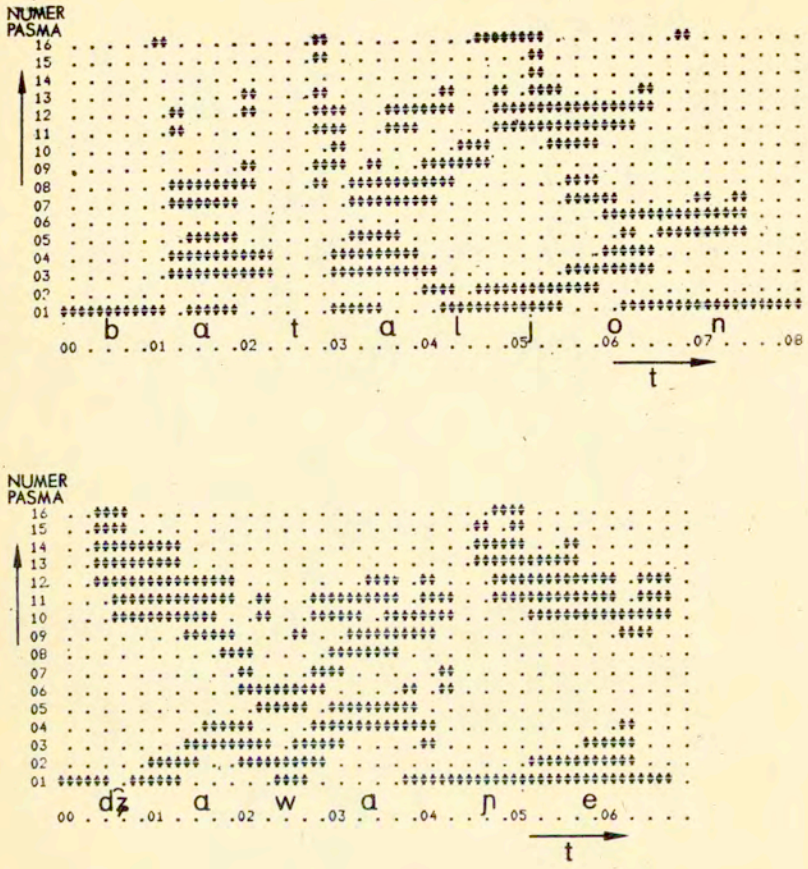
$$b_i = 0 \quad \text{dla} \quad p_i - \frac{p_{i-k} + p_{i+k}}{2} < a\left(p_{i-k} - p_{i+k}\right), \qquad (2)$$

gdzie $b_i$ oznacza parametr widma binarnego, $p_{i-k}$ , $p_i$ , $p_{i+k}$ są parametrami widma wygładzonego. Dla współczynnika a przyjęto wartość 0.25 . Kolejna zmiana w organizacji modelu rozpoznającego związana była z dalszą redukcją liczby parametrów widma binarnego. Skoro, jak przypomniano wyżej , zmniejszenie liczby parametrów widma binarnego o połowę nie spowodowało istotnego pogorszenia wyników rozpoznawania , uzasadniona była dalsza ich redukcja. Odstąpiono jednak od przeprowadzenia jej drogą pominięcia pewnych parametrów , jak to miało miejsce przy zmniejszeniu ich liczby o połowę w pracy [2] . Redukcję osiągnięto tym razem w drodze określonej kompresji widma binarnego pełnego. Parametry widma będącego wynikiem takiej kompresji wyrażają obecność lub brak par kolejnych jedynek w odnośnych zakresach widma binarnego pełnego. Istotny jest tutaj podział na owe zakresy, chociaż istnieje niewiele przesłanek , które sugerowałyby , jak należy go przeprowadzić. Wiadomo , iż szerokość takiego zakresu winna być adekwatna do stopnia dystynktywnej roli jaką w pierwotnym widmie binarnym odgrywa informacja tym zakresem objęta. W tych przedziałach częstotliwości , w których pojawia się informacja bardzo istotna i równomiernie ważna owe zakresy muszą być najwęższe i jednakowo szerokie. Odnosi się to przedew wszystkim do obszaru występowania pierwszych dwóch formantów samogłoskowych. Wyższe partie widma binarnego mogą składać się ze znacznie mniejszej liczby parametrów wyrażających np. obecność lub brak ciągów co najmniej dwóch jedynek w szerszych zakresach częstotliwości pierwotnego widma binarnego. Odbicie w obrazie widma binarnego znanych cech widmowych sygnału mowy jest zdeterminowane przede wszystkim przez rozdzielczość częstotliwościową , z jaką wykona-

na została analiza widmowa. Pełne widmo binarne uzyskane w oparciu o analizator stosowany w modelu rozpoznającym zawiera swoiste formanty , z których każdy wyrażony jest przeważnie ciągiem 3 do 5 jedynek bezpośrednio po sobie następujących. Biorąc powyższe pod uwagę , przekształcenie dolnej partii widma binarnego oparto o podział na zakresy obejmujące 4 kolejne parametry i zachodzące na siebie przedziałami dwóch parametrów. Zakresy podziału przyjęte dla przekształcenia pozostałej części widma obejmują po 6, 8 , 12 i 16 parametrów i zachodzą na siebie odpowiednio na szerokość 2, 4 i 8 parametrów. Rezultatem przekształcenia 63-parametrycznego pierwotnego widma binarnego na podstawie takiego podziału jest widmo binarne 16-parametryczne . Zasadę tego rodzaju przekształcenia zilustrowano na rys. 1 . Przekształcenie to postanowiono włączyć do modelu rozpoznającego i przeprowadzić jego ocenę na podstawie uzyskanych wyników rozpoznawania . Realizuje ono cele do jakich prowadzić miała modyfikacja modelu rozpoznającego a mianowicie : dalsza znaczna redukcja liczby parametrów widma binarnego , uproszczenie operacji rozpoznawania i zminimalizowanie roli cech indywidualnych w spektrogramie binarnym . Przybliżeniu tego ostatniego celu służy także następna innowacja , dotycząca również wyznaczania widma binarnego. Polega ona na zastosowaniu dynamicznego ograniczania zakresu spektrogramu binarnego. Granicę wyznacza stała ilość informacji składana z jedynek liczonych osobno w każdym widmie binarnym zaczynając od jego początku. W ten sposób każde widmo binarne / mowa tutaj o widmie binarnym wtórnym powstałym drogą opisanego wyżej przekształcenia widma binarnego pierwotnego / zostaje zredukowane do pewnej stałej liczby jedynek . W przypadku np. samogłoski widmo binarne zostaje tym sposobem ograniczone do najistotniejszego zakresu dwóch pierwszych formantów z pominięciem zakresu wyższego , w którym obraz spektrogramu jest najbardziej osobniczo zindywidualizowany. W przypadku natomiast spółgłosek trących do ograniczenia widma ze zrozumiałych względów zwykle nie dochodzi . Na rys. 2 zamieszczono przykład spektrogramu binarnego ze zredukowaną do 16 liczbą parametrów widmowych . Jedno 16-parametryczne widmo binarne można zapisać cyfrowo za pomocą dwóch bajtów a spektrogram binarny przeciętnej długości wyrazu przy częstotliwości próbkowania widmowego wynoszącej 50 Hz zajmuje w pamięci komputera obszar około 40 bajtów. Stwarza to dużą zachętę do wnikliwego zbadania wystarczalności takiego wyrażenia sygnału mowy w zastosowaniu do rozpoznawania wyrazów.




Rys. 1. Ilustracja zasady przekształcenia widma binarnego 63-parametrycznego w widmo binarne 16-parametryczne.




Rys. 2. Przykłady spektrogramów binarnych w wersji 16-parametrycznej .

Kolejnym zabiegiem modyfikującym model rozpoznający wyrazy było stworzenie możliwości poszukiwania podobnych fragmentów dwóch porównywanych z sobą spektrogramów binarnych w znacznie szerszym niż dotychczas przedziale. Tę innowację uznano za celową ze względu na występujące zróżnicowania w rozkładzie czasowym odpowiadających sobie fragmentów w różnych wypowiedziach tego samego wyrazu zwłaszcza , gdy w grę wchodzą różne głosy . Przedział, w którym poszukiwany jest fragment podobny do odnośnego fragmentu innego spektrogramu , poszerzono wstępnie prawie dwukrotnie pozostawiając możliwość zwężenia go w razie potrzeby .

Na zakończenie tego rozdziału wypada wspomnieć , że dokonano przeredagowania i przeadresowania bardzo wielu procedur składających się kompleks programowy realizujący wszystkie operacje adaptacji i rozpoznawania wyrazów . Miało to na celu przywrócenie ładu w układzie tegoż kompleksu , naruszanego wielokrotnie w trakcie różnych prób rozwinięcia i udoskonalenia modelu rozpoznającego . Takie uporządkowanie było konieczne ze względu na małą pamięć minikomputera Mera-303 stanowiącego główne narzędzie w realizacji modelu rozpoznającego. Nie sposób w tym miejscu nie wyrazić dygresji, że zakres i tempo relacjonowanych tutaj prac były zdeterminowane niskimi parametrami tegoż minikomputera . Programy należało układać i wprowadzać w języku wewnętrznym a ich ewentualna przebudowa powodowała na ogół bardzo kłopotliwe zmiany w rozdysponowaniu pamięci . Kompleks programów adaptacji i rozpoznawania składa się z około 4 tyś. rozkazów języka wewnętrznego minikomputera Mera-303. Tak długi ciąg rozkazów zajmuje połowę pamięci tego minikomputera. Reszta pamięci jest w całości wykorzystana w charakterze buforów przechowujących dane.

## 3. Test modelu rozpoznającego po jego modyfikacji

Rozpoznawanie wyrazów na podstawie spektrogramów binarnych opisane w pracy [2] uzyskało pozytywną ocenę . Model rozpoznający zastosowany wówczas został obecnie pod wieloma względami zmodyfikowany . Z tego powodu należało w pierwszym rzędzie zbadać, czy wprowadzone zmiany nie wpłynęły pogarszająco na jakość rozpoznawania . W tym celu przeprowadzono następujące doświadczenie testowe. Ułożono słownik złożony ze 100 słów . Dla uniknięcia ewentualnej tendencyjności przy wyborze słów sięgnięto do przypadkowo napotkanego artykułu prasowego. Artykuł ten opublikowany był przez

lokalną gazetę poznańską "Głos Wielkopolski" i nosił tytuł: " 18 dni obrony twierdzy Modlin ". Słownik ułożono z pierwszych stu rzeczowników tekstu tego artykułu sprowadzonych do pierwszego przypadku z zachowaniem ich gramatycznej liczby w jakiej zostały użyte w artykule. Poniżej zamieszczono listę w ten sposób wybranych słów w kolejności alfabetycznej . Jeden głos męski wypowiedział każde z tych 100 słów pięciokrotnie - wpierw 4-krotnie każde w kolejności występowania w słowniku a następnie jednorazowo każde w porządku przypadkowym . Podczas wypowiedzi wyznaczany był spektrogram binarny a bezpośrednio po jej zakończeniu następowało przepisanie spektrogramu do pamięci zewnętrznej w celu późniejszego wykorzystania go na etapie adaptacji lub do prób rozpoznawania. Pierwszy zestaw wypowiedzi posłużył do sporządzenia zbioru stu wzorcowych spektrogramów binarnych nazywanych też w skrócie wzorcami . Ogólna zasada,według której tworzono wzorce , była taka sama jak we wcześniejszych doświadczeniach opisanych w pracach [1],[2] . W buforze pamięciowym wydzielonym w pamięci minikomputera na przechowywanie wzorców mieści się jedynie 26 wzorców. Z tego powodu cały inwentarz wzorców podzielono na 4 części i umieszczono w pamięci zewnętrznej na dyskach elastycznych . Pierwsze 3 części liczyły po 26 wzorców a czwarta 22 . Rozpoznawaniu poddano wpierw te wypowiedzi , na podstawie których wcześniej utworzono wzorce , czyli każde słowo słownika wypowiedziane czterokrotnie co stanowiło łącznie 400 wyrazów. W trakcie rozpoznawania każdego wyrazu poszczególne grupy wzorców były przepisywane kolejno z pamięci zewnętrznej do bufora w minikomputerze. Konieczność wielokrotnego wykonywania takich przesłań stanowi przykład niekorzystnych ograniczeń z powodu niskich parametrów minikomputera. Wynik tego rozpoznawania wypadł bardzo dobrze. Wszystkie 400 wyrazów zostało rozpoznanych bezbłędnie. Wyrazy,które nie brały udziału przy tworzeniu wzorców - wszystkie słowa słownika wypowiedziane w porządku przypadkowym - rozpoznane zostały w 86 procentach w pierwszej turze rozpoznawania . Zbadano przyczyny niektórych błędnych odpowiedzi. Stwierdzono, iż powód pomyłek tkwi między innymi w niewłaściwie postawionym kryterium oceny podobieństwa poszczególnych fragmentów rozpoznawanego obiektu z odnośnymi fragmentami każdego ze wzorców. Koniecznym wydaje się przypomnieć w tym miejscu tę fazę procesu rozpoznawania , w której owo kryterium zostaje użyte.

Wynikiem porównania każdego wzorca z rozpoznawanym obiektem - spektrogramem binarnym rozpoznawanego wyrazu - jest ciąg wartości współczynnika niepodobieństwa $m_{np}$ poszczególnych fragmentów obiektu z odnośnymi fragmentami

wzorca. Rozstrzygnięcie , który z dwóch wzorców jest bardziej podobny do niewiadomego obiektu , następuje na podstawie oceny, który z tych wzorców w większej liczbie fragmentów wykazuje lepsze podobieństwo do obiektu. Przedtem jednak każdy z dwóch fragmentów wyłonionych z dwóch różnych wzorców - na podstawie tego, że wykazały lokalnie najkorzystniejsze podobieństwo z odpowiednim fragmentem obiektu - należy w oparciu o wspomniane wyżej kryterium zaklasyfikować do jednej lub każdej z dwóch klas L i G skupiających odpowiednio fragmenty lepiej i gorzej podobne do odnośnych fragmentów obiektu . Załóżmy , że zaklasyfikowane być mają fragmenty $FR_i(A)$ i $FR_j(B)$ należące odpowiednio do wzorców A i B i że współczynnik niepodobieństwa tych fragmentów do fragmentu $FR_k(OB)$ obiektu wyrażony jest odpowiednio przez $m_{np}(A)$ i $m_{np}(B)$ . Warunkiem zaliczenia fragmentu $FR_i(A)$ do klasy L a fragmentu $FR_j(B)$ do klasy G jest , aby

$$m_{npj}(B) - m_{npi}(A) \geqslant C \quad . \qquad (3)$$

Odwrotne zaklasyfikowanie fragmentów $FR_i(A)$ i $FR_j(B)$ ma miejsce , gdy spełniony jest warunek :

$$m_{npi}(A) - m_{npj}(B) \geqslant C \quad . \qquad (4)$$

Jeżeli natomiast żaden z powyższych warunków nie jest spełniony , co oznacza , że

$$\left| m_{npi}(A) - m_{npj}(B) \right| < C \quad , \qquad (5)$$

wówczas każdy z tych fragmentów zaliczony zostaje do obu klas.
C jest pewną umownie przyjętą stałą wyrażającą nieistotną różnicę bezwzględną pomiędzy wartościami $m_{npi}(A)$ oraz $m_{npj}(B)$ . Badając przypadki błędnych odpowiedzi stwierdzono , że zakres nieistotnej różnicy pomiędzy wartościami współczynnika niepodobieństwa należałoby zmodyfikować . Najwłaściwiej byłoby uczynić jego szerokość funkcją hyperboliczną mniejszej z dwóch porównywanych wartości tego współczynnika. Wiązałoby się to jednak z rozbudowaniem programu rozpoznawania i tak już ledwo mieszczącego się w pamięci minikomputera. Rozwiązano zatem tę sprawę na razie połowicznie . Zakres nieistotnej różnicy pomiędzy wartościami współczynnika niepodobieństwa nieznacznie poszerzono utrzymując jednak jego wartość na stałym poziomie. Zasadnicza zmiana polegała natomiast na wprowadzeniu dodatkowego warunku , zgodnie

z którym badanie różnicy wartości $m_{npi}(A)$ i $m_{npj}(B)$ ma miejsce jedynie wtedy , gdy przynajmniej jedna z tych wartości przewyższa pewien ustalony próg . W przeciwnym razie oba fragmenty : $FR_i(A)$ i $FR_j(B)$ zaliczone zostają zarówno do klasy L jak i G . Poziom wspomnianego progu określono empirycznie. Po wprowadzeniu obu powyższych modyfikacji poddano ponownej próbie rozpoznawania 14 wyrazów, które w poprzedniej turze nie zostały rozpoznane . 10 z tych wyrazów zostało obecnie poprawnie rozpoznanych a 4 ponownie nie . Nierozpoznanymi wyrazami były : batalion , bateria , dowódca , punkt . Zrezygnowano chwilowo z dalszych prób uzyskania jeszcze lepszego wyniku rozpoznawania , gdyż rygor planu wymagał przejścia do doświadczeń wytyczonych w temacie tegorocznego zadania badawczego , a mianowicie rozpoznawania wyrazów w oparciu o grupowe zbiory wzorców. Do kwestii innych przyczyn, które uniemożliwiły uniknięcie pozostałych jeszcze błędów w rozpoznawaniu przewiduje się powrócić niebawem .

## 4. Rozpoznawanie wyrazów w oparciu o grupowe zbiory wzorców

Dysponując zmodyfikowanym i przez to zapewne udoskonalonym modelem rozpoznającym wyrazy przystąpiono do badań określonych w tytule niniejszej pracy. Postanowiono oprzeć te badania na słowniku złożonym jedynie z 26 słów oraz na wypowiedziach czterech głosów męskich. Na decyzję taką wpłynęły różne okoliczności. Po pierwsze brak umotywowanych prognoz co do przyszłych wyników rozpoznawania wyrazów w oparciu o grupowe zbiory wzorcowych spektrogramów binarnych. Uznano, że objęcie badaniami większej liczby głosów miałoby sens jedynie wówczas, gdyby w grę wchodziło poszerzenie wiedzy o tego rodzaju rozpoznawaniu. Na przykład, gdyby wiadomo już było, że dla nielicznych grup głosów rozpoznawanie wyrazów w oparciu o grupowe zbiory wzorców jest poprawne a nieznana jest jego jakość,gdy grono głosów jest większe. Podobne motywy zadecydowały o przyjęciu niewielkiego słownika. Ograniczenia co do liczby głosów i rozmiarów słownika były także konieczne ze względu na czas , w jakim należało badania przygotować, przeprowadzić i opracować a także ze względu na wymóg racjonalnego i oszczędnego dysponowania minikomputerem stanowiącym narzędzie pracy przy realizacji niemal wszystkich zadań badawczych Pracowni Fonetyki Akustycznej IPPT PAN . Przyjąwszy słownik złożony jedynie z 26 słów uniknięto operacji wielokrotnego przepisywania inwentarza wzorców z pamięci zewnętrznej do minikomputera w trak-

cie rozpoznawania . Słownik - zamieszczony poniżej - utworzono z wybranych słów słownika użytego w próbie testowej zmodyfikowanego modelu rozpoznającego. Wyboru słów dokonywano według pewnego statystycznego klucza.
Każdy z głosów biorących udział w doświadczeniu wypowiedział czterokrotnie wszystkie wyrazy słownika za każdym razem w innej kolejności. Miało to miejsce w pomieszczeniu niewytłumionym w obecności pracującej aparatury emitującej hałasy od wirujących wentyla/torów i w warunkach sporadycznie docierających hałasów ulicznych i kolejowych . Spektogramy binarne z poszczególnych wypowiedzi zapisano w pamięci na dyskach elastycznych . W sumie zgromadzono 416 spektrogramów binarnych w nowej 16-parametrycznej wersji. Spektrogramy poszczególnych głosów połączono w grupy. Każdą grupę tworzyły spektrogramy dwóch głosów. W ten sposób powstało 6 grup. Każda składała się ze spektrogramów binarnych 204 wyrazów / każde słowo 26-wyrazowego słownika wypowiedziane 4-krotnie przez dwa głosy /. W tworzeniu wzorca grupowego danego słowa brały udział cztery spektrogramy binarne pochodzące z dwukrotnej wypowiedzi tego słowa przez dwa różne głosy. Zasada tworzenia wzorca grupowego w istocie swej niczym innym nie różniła się od tej , jaką stosowano przy tworzeniu wzorców indywidualnych.
Wyrazy wypowiedziane przez głosy należące do danej grupy rozpoznawane były w oparciu o wzorce grupowe utworzone dla tej grupy. Oddzielnie rozpoznawano te wyrazy , które brały udział w tworzeniu wzorców oraz te , które nie zostały użyte do tego celu. Poniżej zestawiono otrzymane wyniki . Poszczególne głosy oznaczono symbolicznie przez : G1 , G2 , G3 , G4 a grupy , w jakie głosy te połączono odpowiednio przez Gi/Gj . i oraz j symbolizują numer głosu wchodzącego w skład grupy . Litera A w nawiasie obok notacji głosu oznacza, że wyniki rozpoznawania dotyczą wypowiedzi , które posłużyły do zbudowania wzorców. Podobnie litera R w na/wiasie informuje , że wyniki dotyczą wypowiedzi przewidzianej wyłącznie do rozpoznawania . W kolumnie liczb podane są liczebności błędnych odpowiedzi /liczby bez nawiasów/ oraz ilości nierozpoznanych słów /liczby w nawiasach/, jeśli błąd dotyczył dwukrotnie tego samego słowa . Zapis w rodzaju : (marsz)- ludność oznacza , że wyraz marsz nie został rozpoznany a błędną odpowiedzią był wyraz ludność.
W tablicy nr 1 ujęte zostały wyniki rozpoznawania w formie bardziej syntetycznej. Tablica ta podaje mianowicie , które słowa i przez które głosy wypowiedziane nie zostały rozpoznane , ile słów wypowiedzianych przez dany głos nie zostało rozpoznanych /wiersz liczb u dołu/ oraz w przypadku ilu głosów dane słowo nie zostało rozpoznane /kolumna liczb po prawej stronie/.

## Z E S T A W I E N I E  B Ł Ę D Ó W  W  R O Z P O Z N A W A N I U  W Y R A Z Ó W  w oparciu o grupowe zbiory wzorców

GRUPA G1/G2

| | | |
|---|---|---|
| G1(A) | 1 | (karabin)- wsparcie |
| G2(A) | 0 | |
| G1(R) | 9 (8) | (noc)- amunicja ,(noc)- amunicja ,(świadek)- pirat , (marsz)- tyły ,(garnizon)- groźba ,(dowódca)- wojska , (wojska)- groźba ,(rok)- piechota ,(karabin)- batalion |
| G2(R) | 1 | (noc)- amunicja |

GRUPA G1/G3

| | | |
|---|---|---|
| G1(A) | 2 | (marsz)- ludność ,(punkt)- pirat |
| G3(A) | 0 | |
| G1(R) | 2 | (punkt)- obręcz ,(noc)- amunicja |
| G3(R) | 2 (1) | (punkt)- amunicja ,(punkt)- ludność |

GRUPA G1/G4

| | | |
|---|---|---|
| G1(A) | 3 (2) | (punkt)- piechota ,(punkt)- ludność ,(rok)- próba |
| G4(A) | 0 | |
| G1(R) | 6 (5) | (dowódca)- wojska ,(marsz)- ludność ,(pirat)- świadek , (punkt)- amunicja ,(rok)- tyły ,(rok)- lata |
| G4(R) | 2 | (pirat)- świadek ,(punkt)- ludność |

GRUPA G2/G3

| | | |
|---|---|---|
| G2(A) | 2 | (pirat)- świadek ,(batalion)- garnizon |
| G3(A) | 0 | |
| G2(R) | 1 | (pirat)- świadek |
| G3(R) | 2 | (piechota)- lata ,(punkt)- ludność |

GRUPA G2/G4

G2(A) 5 (4) (groźba)- brygada ,(groźba)- brygada ,(lata)- piechota , (pirat)- świadek ,(świadek)- odcinek

G4(A) 0

G2(R) 5 (3) (groźba)- brygada ,(pirat)- świadek ,(pirat)- świadek , (próba)- brygada ,(próba)- brygada

G4(R) 5 (4) (piechota)- brygada ,(pirat)- brygada ,(punkt)- ludność , (lata)- batalion ,(lata)- garnizon

GRUPA G3/G4

G3(A) 0

G4(A) 1 (lata)- groźba

G3(R) 1 (punkt)- pirat

G4(R) 2 (1) (noc)- marsz ,(noc)- dowódca

Tablica 1 . Sumaryczne zestawienie błędów w rozpoznawaniu wyrazów w oparciu o grupowe zbiory wzorców

| | | G1 | G2 | G3 | G4 | |
|---|---|---|---|---|---|---|
| 1 | batalion | | x | | | 1 |
| 2 | dowódca | x | | | | 1 |
| 3 | garnizon | x | | | | 1 |
| 4 | groźba | | x | | | 1 |
| 5 | karabin | x | | | | 1 |
| 6 | lata | | x | | x | 2 |
| 7 | marsz | x | | | | 1 |
| 8 | noc | x | x | | x | 3 |
| 9 | piechota | | | x | x | 2 |
| 10 | pirat | x | x | | x | 3 |
| 11 | próba | | x | | | 1 |
| 12 | punkt | x | | x | x | 3 |
| 13 | rok | x | | | | 1 |
| 14 | świadek | x | x | | | 2 |
| 15 | wojska | x | | | | 1 |
| | | 10 | 7 | 2 | 5 | |

5. Omówienie wyników i wnioski

Przedstawione powyżej wyniki dają pierwszy pogląd o możliwości rozpoznawania wyrazów w oparciu o grupowe zbiory wzorcowych spektrogramów binarnych. Dla trzech z pośród sześciu możliwych skojarzeń czterech głosów w pary uzyskano wyniki względnie korzystne. Są to pary : G3/G4 , G2/G3 , G1/G3 . We wszystkich tych parach występuje głos G3 . Ponieważ wyniki rozpoznawania wyrazów otrzymane dla grup: G1/G2 , G2/G4 , G1/G4 , w których głos ten nie występuje, są dla niektórych głosów znacznie gorsze, można sądzić, że głos G3 plasuje się w środku odległości pomiędzy pozostałymi głosami. Fakt , iż tylko dwa słowa wymówione przez głos G3 nie zostały rozpoznane , może świadczyć o tym, że w spektrogramach binarnych tego głosu nie występują wydatne cechy osobnicze. Na uwagę zasługuje też fakt , że wynik rozpoznawania wyrazów dla grup G1/G2 i G1/G4 jest znacznie zróżnicowany zależnie od głosu. Dla głosów : G2 /1 błąd/ i G4 /2 błędy/ wynik rozpoznawania uznać można za względnie dobry natomiast dla głosu G1 /9 i 6 błędów/ za zdecydowanie zły . W grupie G2/G4 rezultaty rozpoznawania wyrazów są dla obu wyrazów niekorzystne lecz dla głosu G2 gorsze .

Z otrzymanych wyników nasuwa się kilka wniosków i uwag , które poniżej zostaną przedstawione . Po pierwsze : Rozpoznawanie wyrazów w oparciu o grupowy zbiór wzorcowych spektrogramów binarnych za pomocą modelu ostatnio zmodyfikowanego jest poprawne jedynie wówczas, gdy grupę stanowią nieliczne , właściwie dobrane głosy. Po drugie : Dla pewnych głosów wchodzących w skład grupy , dla której istnieje wspólny zbiór wzorców, rozpoznawanie wyrazów może być względnie poprawne a dla innych obarczone licznymi błędami.

Okazało się, że przeprowadzone modyfikacje modelu rozpoznającego nie zapewniły jeszcze w koniecznym stopniu pożądanego ujednolicenia obrazów spektrogramu binarnego danego słowa dla pewnego szerszego grona mówców . Przy obecnym stanie modelu rozpoznającego nie zawsze możliwe będzie rozpoznawanie w oparciu o wspólny zbiór wzorców , gdyż jak wskazuje przykład grupy G2/G4, na obecnym etapie niektóre głosy nie można kojarzyć w grupy mające ten sam zbiór wzorców. Także pewne słowa nastręczać mogą szczególnych trudności w rozpoznawaniu opartym o grupowe zbiory wzorców . Dowodzą tego nierozpoznawane w doświadczeniu słowa : noc , pirat , punkt w przypadku większości głosów .

Ewentualne dalsze modyfikacje modelu rozpoznającego w celu lepszego przystosowania go do rozpoznawania wyrazów na podstawie ujednoliconych zbiorów wzor-

ców poprzedzić należy szczegółową analizą przyczyn błędów, jakie wystąpiły w dotychczasowych próbach rozpoznawania wyrazów w oparciu o grupowe zbiory wzorców. Analiza taka będzie w najbliższym czasie przeprowadzona . W zrelacjonowanym niniejszym etapie badań zwrócono jedynie uwagę na pewne przyczyny utrudniające uzyskiwanie wzorców grupowych oraz wywołujące błędy w rozpoznawaniu. Tkwią one w wydatnym zindywidualizowaniu głosów wyrażającym się np. rozległą w czasie i częstotliwości aspiracją w przypadku spółgłosek zwartych bezdźwięcznych /charakteryzował się tą cechą głos G1/, brakiem fonacji,gdy wymówiona miała być głoska zwarta dźwięczna /zdarzało się to głosowi G4/, czy też dużym zróżnicowaniem międzyosobniczym częstotliwości formantów /odnosi się to do głosów G2 i G4/. Przyszłym zadaniem będzie zminimalizowanie wpływu tych i innych osobliwości głosowych na rozpoznawanie wyrazów w oparciu o grupowe zbiory wzorcowych spektrogramów binarnych .

## B I B L I O G R A F I A

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

[2] KUBZDELA, H. : Weryfikacja i optymalizacja metody rozpoznawania wyrazów w skończonych zbiorach hasłowych w oparciu o spektrogramy binarne, Prace IPPT 10/1982, Warszawa, 1982.

[3] KUBZDELA, H. : Badania nad udoskonaleniem spektrogramów binarnych. / Przyjęto do druku w Wydawnictwie Prace IPPT w 1983 r. /

S Ł O W N I K 100-wyrazowy

/ użyty w próbie testowej zmodyfikowanego modelu rozpoznającego /

| | | | |
|---|---|---|---|
| amunicja | ga/rnizon | obrońca | rozkaz |
| armia | generał | obręcz | schron |
| artyleria | glony | oddział | siła |
| atak | godziny | odwrót | system |
| batalion | groźba | obiekt | sytuacja |
| bateria | karabin | odcinek | suma |
| bitwa | kierunek | oficer | świadek |
| bohaterstwo | kilometr | okop | świt |
| brygada | konwencja | osłona | szpital |
| brzeg | krater | pirat | tabor |
| bunkier | lata | piechota | tyły |
| czołgi | linia | ptak | twierdza |
| cel | lot | południe | uderzenie |
| determinacja | lotnictwo | pozycja | uwaga |
| dni | ludność | północ | wojska |
| dostęp | ławka | próba | wsparcie |
| dowódca | ła/ńcuch | przeszkoda | wschód |
| dowództwo | magazyn | przyczółek | wybuch |
| droga | marsz | przygotowanie | wulkan |
| drut | mech | punkt | zasiek |
| dywizja | nadwyżka | pułk | znaczenie |
| działo | natarcie | przykład | ziemia |
| dzia/łanie | noc | rejon | żołnierz |
| fort | obszar | rok | żelazo |
| fortyfikacja | obrona | rów | żar |

S Ł O W N I K 26 - wyrazowy

/użyty w próbie rozpoznawania wyrazów w oparciu o grupowe zbiory wzorców/

| | | | |
|---|---|---|---|
| amunicja | groźba | odcinek | siła |
| batalion | karabin | pirat | świadek |
| brygada | lata | piechota | tyły |
| cel | ludność | próba | wojska |
| dowódca | marsz | punkt | wsparcie |
| dywizja | noc | rok | zasiek |
| garnizon | obręcz | | |