

Grażyna Demeńko

KLASYFIKACJA PRZEBIEGÓW
CZĘSTOTLIWOŚCI PODSTAWOWEJ

23/1985

WARSZAWA 1985

ISSN 0208-5658

Praca wpłynęła do Redakcji dnia 14 listopada 1984 r.

1092 312 II



Na prawach rękopisu

Instytut Podstawowych Problemów Techniki PAN

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

Oddano do drukarni w maju 1985 r.

Nr zamówienia 323/85

Warszawska Drukarnia Naukowa, Warszawa,
ul. Śniadeckich 8

Bibl. UAM

85 EO 687

Grażyna Demenko
Pracownia Fonetyki Akustycznej
IPPT PAN

KLASYFIKACJA PRZEBIEGÓW CZĘSTOTLIWOŚCI PODSTAWOWEJ¹.

Streszczenie

W pracy przedstawiono metodę analizy i klasyfikacji przebiegów częstotliwości podstawowej. Przyjęto materiał językowy składający się z ośmiu wymówień pewnej frazy oraz ich powtórzeń przez 15 osób. Dokonano analizy dyskryminacyjnej przebiegów parametru F_0 w materiale składającym się z 10 replikacji każdego wzorca intonacyjnego dla trzech osób. Na podstawie wyników analizy określono trzy cechy charakterystyczne badanych klas. Na mikrokomputerze Sinclair zrealizowano algorytm klasyfikacji przebiegów częstotliwości podstawowej. W celu przetestowania algorytmu wybrano losowo materiał składający się z trzech powtórzeń przez 15 osób każdego wzorca intonacyjnego. Wyniki klasyfikacji wskazują na możliwość analizy przebiegów parametru F_0 w przestrzeni trójwymiarowej.

1. Wstęp.

Problem automatycznej analizy parametrów sygnału mowy zajmuje w ostatnich latach centralne miejsce w badaniach fonetycznych. W zakresie cyfrowej ekstrakcji częstotliwości podstawowej istnieje bardzo dużo opracowań. Istotne problemy analizy tego parametru przez maszynę cyfrową pozostają nadal jednak nie rozwiązane. Klasyfikacja przebiegów częstotliwości podstawowej oparta jest zwykle na analizie percepcyjnej i ściśle dotyczy danego języka (por. np. [4], [6]). Najbardziej szczegółowa klasyfikacja intonacyjnych wzorców występujących w języku polskim przedstawiona jest w pracy M. Steffen-Batogowej [12]. Konieczne

Jest znalezienie odpowiednich reguł, za pomocą których przebiegi częstotliwości podstawowej będzie można sprowadzić do inwariantnych postaci dla każdego wzorca intonacyjnego. Z uwagi na specyficzny charakter parametru F_0 analiza automatyczna powinna być powiązana z badaniami percepcyjnymi, ponieważ uwzględnienie wyłącznie cech fizycznych sygnału może prowadzić do klasyfikacji nienaturalnej. Poszczególne przebiegi intonacyjne wytwarzane przez określonego mówcę powinny być zaklasyfikowane względem siebie i względem percepcyjnie ekwiwalentnych, ale różnych fizycznie przebiegów parametru F_0 , w wypowiedziach innych mówców.

2. Dobór materiału eksperymentalnego.

Na taśmie magnetofonowej zapisano osiem wypowiedzi "dobrze" wymówionych przez fonetyka tak, aby wszystkie były różne intonacyjnie. W eksperymencie wzięła udział grupa złożona z 10 mężczyzn i 5 kobiet, w której 5 osób posiadało przygotowanie fonetyczne. Pozostałe osoby w większości z wykształceniem politechnicznym, zawodowo nie były związane z fonetyką. Przed doświadczeniem podano do odsłuchu wszystkim osobom wypowiedzi wzorcowe, które stanowiły uprzednio przygotowane zapisy frazy "dobrze". Każdą z osób poproszono o jak najwierniejsze powtarzanie poszczególnych wzorcowych wypowiedzi.

Podobny eksperyment w zastosowaniu do analizy częstotliwości podstawowej przedstawiono w pracy [7]. Otrzymano nagranie następującego typu : Głos 1 : Wzorzec 1 - odtworzenie 1 ..., wzorzec 8 - odtworzenie 8. Głos 2 : Wzorzec 1 - odtworzenie 1... itd. Doświadczenie powtarzano kilkakrotnie uzyskując po 10 replikacji poszczególnych wypowiedzi wzorcowych dla każdej osoby. Ekstrakcji parametru F_0 dokonano w układzie składającym się z tonometru, przetwornika analogowo-cyfrowego, minikomputera MERA 303, memoskopu oraz oscylografu pętlicowego. Szczegółowy opis układu pomiarowego przedstawiono w pracy [2].

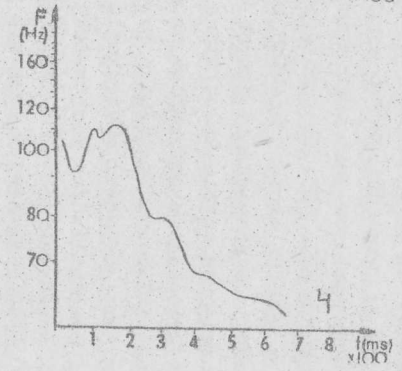
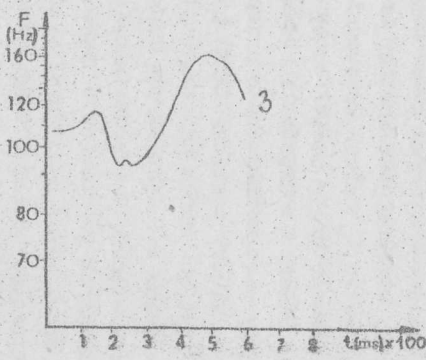
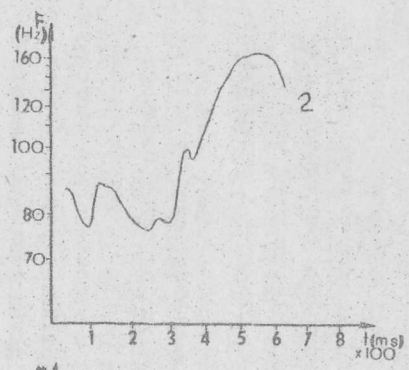
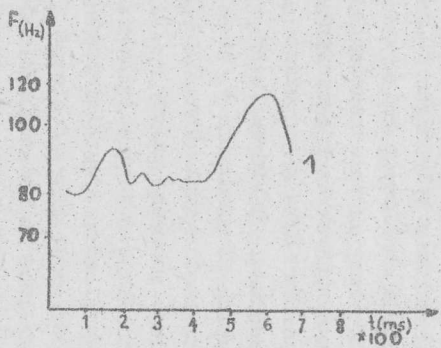
Do pamięci maszyny cyfrowej wprowadzono wyniki pomiarów długości każdego okresu. W celu uzyskania liniowej skali czasu oraz redukcji danych, każdą wypowiedź podzielono na osiem odcinków czasowych w których obliczono średni okres. Parametr F_0 może być definiowany różnie, wybór określonego sposobu pomiaru zależy od indywidualnej metody. W przypadku regularnych przebiegów dokonano

automatycznego uśredniania danych. Opracowanie jednak elastycznego algorytmu uwzględniającego wszystkie możliwości pojawienia się błędnych pomiarów jest zbyt skomplikowane do realizacji na minikomputerze MERA 303. Dlatego w przypadku ekstrakcji parametru F_0 w niskich głosach męskich lub niestarannej wymowy konieczne było wyprowadzanie danych na drukarkę i przeprowadzenie obliczeń przy użyciu kalkulatora.

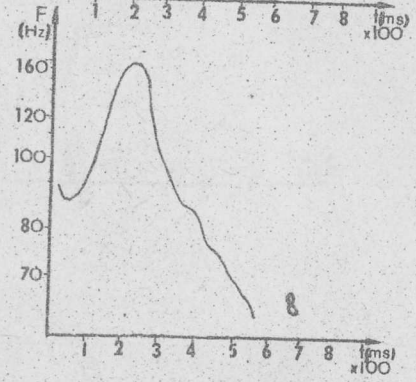
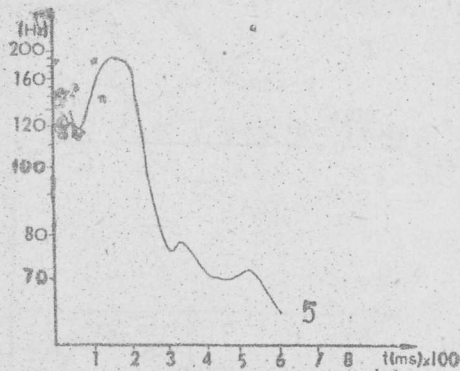
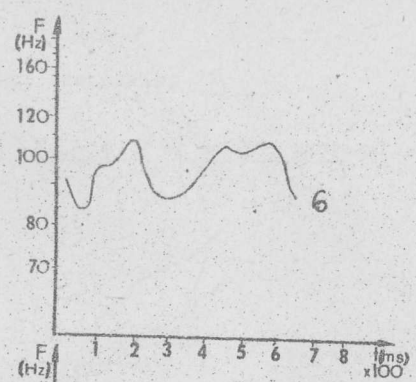
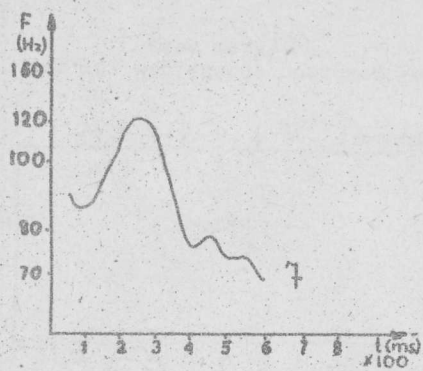
Wszystkie przebiegi wyprowadzono na memoskop, część z nich na oscylograf pętlicowy. Umożliwiło to wizualną ocenę badanego materiału jak i długości ekstrakcji. Na ryc. 1 przedstawiono przebiegi częstotliwości podstawowej w wypowiedziach wzorcowych. Na ryc. 2 przedstawiono po cztery replikacje poszczególnych wypowiedzi wzorcowych otrzymanych dla głosu męskiego, na ryc. 3 przedstawiono replikacje dla głosu żeńskiego.

Ustalenie właściwej liczby danych jest jednym z ważniejszych problemów w analizie parametru F_0 . Przyjęto liczbę ośmiu danych głównie ze względów technicznych, z założeniem jednak że będzie można przy ich pomocy zbadać określone cechy przebiegów. Ponieważ średnia długość wypowiedzi "dobrze" wynosiła około 700 ms odcinki czasowe w których dokonano uśredniania były rzędu 80 ms. W przypadku konieczności analizy krótkich chwilowych zmian wartości parametru F_0 należy wprowadzić krótszy okres uśredniania. W eksperymencie brały udział osoby o różnych skalach głosowych, najniższa zmierzona częstotliwość wynosiła 60 Hz a najwyższa 500 Hz. Aby więc możliwe było przeprowadzenie jakiegokolwiek porównania między poszczególnymi przebiegami parametru F_0 konieczne jest dokonanie normalizacji częstotliwości. Normalizację przeprowadzono stosując zasadę iż należy zachować zmiany względne występujące w przebiegu, informację o jego położeniu względem innych przebiegów w tym samym głosie, a wyeliminować różnice spowodowane odmiennymi wysokościami głosów. Ponieważ osobom biorącym udział w eksperymencie dano za zadanie jak najwierniejsze odtworzenie wzorcowych intonacji, można spodziewać się, że replikacje powinny być podobne do wzorców pod względem charakteru i zakresu zmian oraz położenia na znormalizowanej skali częstotliwości.

Zakres zmian wartości parametru F_0 w głosie, który podawał

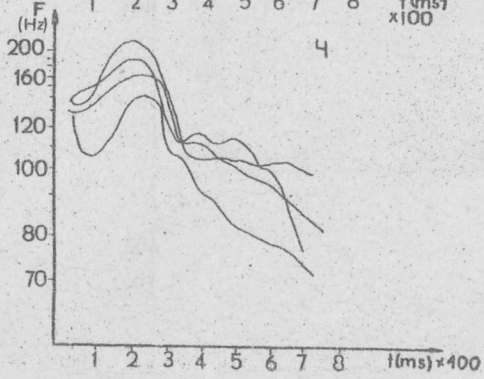
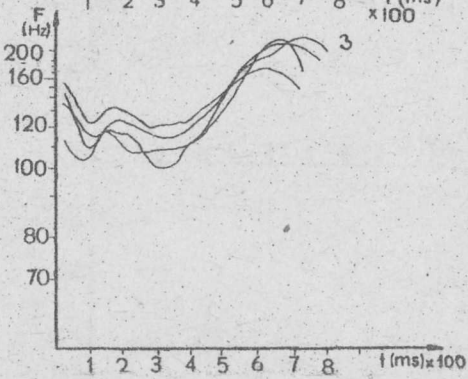
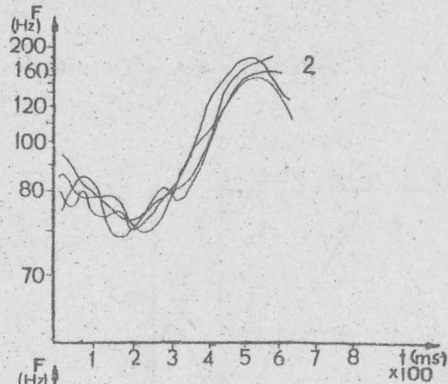
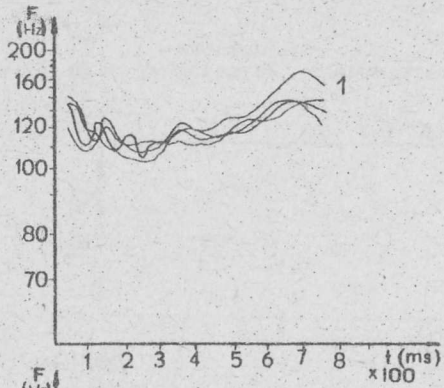


Ryc. 1. Przebiegi częstotliwości podstawowej w wypowiedziach wzorcowych.

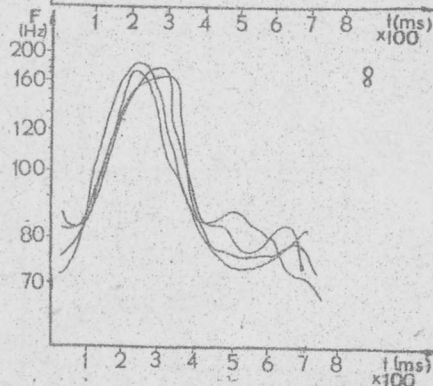
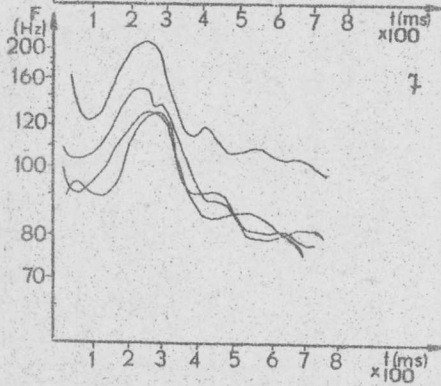
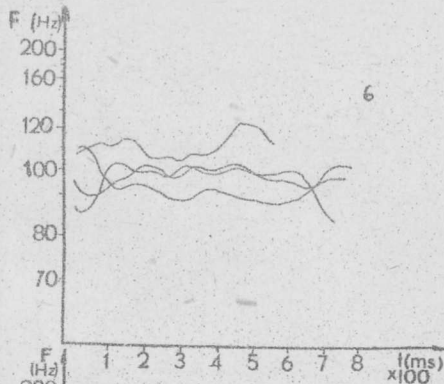
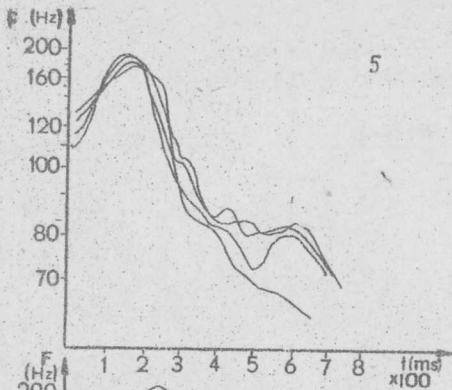


Ryc. 1. Przebiegi częstotliwości podstawowej w wypowiedziach wzorcowych.

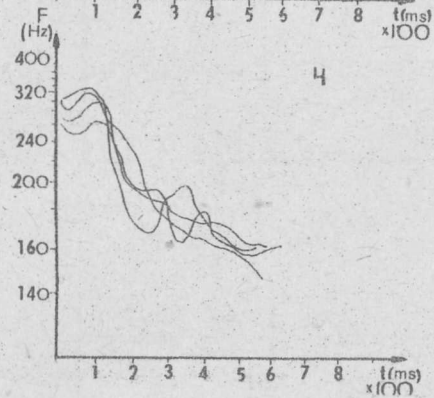
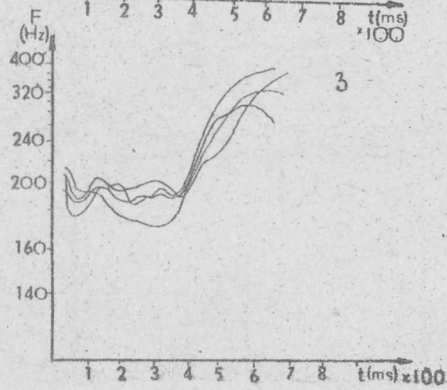
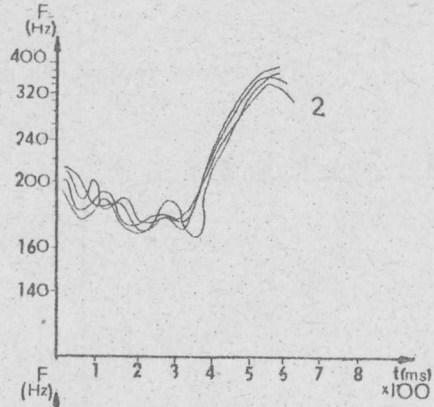
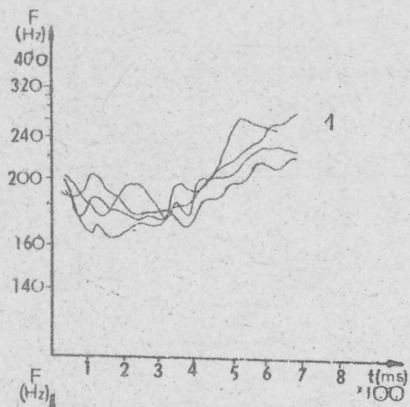




Ryc. 2. Replikacje poszczególnych wypowiedzi wzorcowych
głos męski .

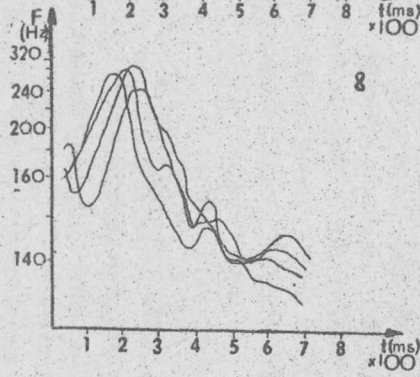
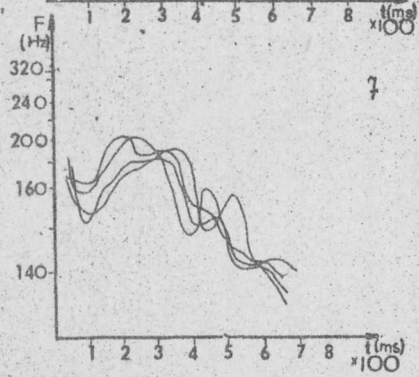
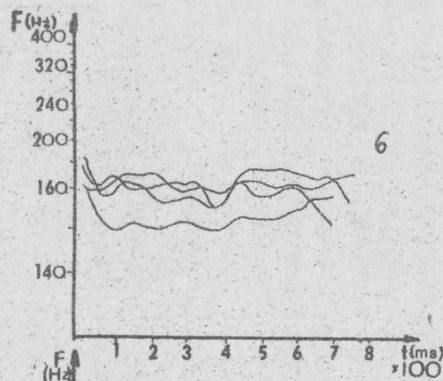
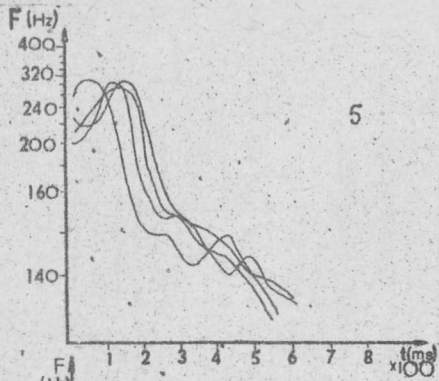


Ryc. 2. Replikacje poszczególnych wypowiedzi wzorcowych
głos męski .



Ryc. 3. Replikacje poszczególnych wypowiedzi wzorcowych (głos żeński).

- 10 -



Ryc. 3. Replikacje poszczególnych wypowiedzi wzorcowych (głos żeński).

- 11 -

intonacje wzorcowe określono następująco :

$$\Delta F = \ln F_{\max} - \ln F_{\min}$$

Jako częstotliwość minimalną (F_{\min}) przyjęto wartość średnią z trzech najniższych częstotliwości występujących dla danej osoby w całym analizowanym materiale. Odpowiednio dla F_{\max} przyjęto wartość średnią z trzech najwyższych częstotliwości. Normalizacji dokonano odejmując od kolejnych logarytmów poszczególnych wartości, logarytm wartości najniższej (F_{\min}). Następnie dodając lub odejmując różnicę wynikającą z położenia środka zakresu zmian częstotliwości występujących w danym głosie, względem środka zakresu zmian F_0 w głosie który podawał intonacje wzorcowe, uzyskano możliwość porównywania przebiegów względem siebie i wzorców.

3. Analiza dyskryminacyjna.

W przypadku analizy częstotliwości podstawowej w wypowiedzi "dobrze" przyjęto, iż osiem danych wystarczająco poprawnie reprezentuje poszczególne przebiegi. Należy liczyć się jednak z koniecznością przeanalizowania dłuższych i bardziej złożonych wypowiedzi. Metody klasyfikacji oparte na dużej liczbie danych są nieekonomiczne. Wygodniej jest opisać własności każdego klasyfikowanego obiektu za pomocą pewnej liczby cech. Selekcji cech należy dokonać transformując pierwotną przestrzeń danych w nową przestrzeń optymalną pod względem własności dyskryminacyjnych. W zależności od sposobu konstrukcji macierzy transformującej istnieje kilka metod selekcji cech.

Wymienić tu należy trzy podstawowe grupy metod. W grupie pierwszej macierz transformującą konstruuje się przy użyciu rozwinięcia Karhunen-Loevego, w grupie drugiej korzysta się z rozwinięcia w szeregi funkcji ortogonalnych, w grupie trzeciej wykorzystuje się analizę dyskryminacyjną. W zastosowaniu do analizy sygnału mowy stosowano wszystkie podane wyżej metody, lub ich odmiany (por. np. [8],[9],[10]). Stosując rozwinięcie Karhunen-Loevego otrzymuje się najlepszą przestrzeń w sensie opisu obiektu. Ponieważ celem rozpoznawania jest rozróżnianie klas, większe potencjalnie możliwości daje podprzestrzeń

skonstruowana na podstawie wektorów dyskryminacyjnych.

Algorytm analizy dyskryminacyjnej oprogramowano na maszynie cyfrowej RIAD-32. Opis algorytmu przedstawiony jest w pracy [10]. Założono, iż przy pomocy analizy dyskryminacyjnej można będzie zbadać zróżnicowania międzyosobnicze oraz zróżnicowania między poszczególnymi wzorcami intonacyjnymi. Zagadnienie klasyfikacji niektórych parametrów sygnału mowy przy zastosowaniu analizy dyskryminacyjnej przedstawiono w pracach [8],[9]. Przyjęto, iż replikacje wypowiedzi wzorcowych będą stanowiły badane klasy. Wybrano wstępnie materiał składający się z 10 powtórzeń każdej wypowiedzi wzorcowej, otrzymanych dla trzech osób z wykształceniem fonetycznym. Skale tych osób były zróżnicowane, dla głosów męskich najniższe częstotliwości wynosiły 65 Hz i 110 Hz, dla głosu żeńskiego 160 Hz.

Zbadano, jakie istnieją różnice między trzema osobami w poszczególnych ośmiu klasach. Wyniki analizy dyskryminacyjnej przedstawia tabela 1. W niektórych przypadkach widać, iż wartość statystyki T^2 przewyższa wartość krytyczną dla przyjętego poziomu istotności. Świadczy to o istniejącym zróżnicowaniu międzyosobniczym. Istotne jest stwierdzenie jaki jest jego wpływ na klasyfikację poszczególnych przebiegów.

Tabela 1.

Wartości statystyki T^2 charakteryzujące różnice między poszczególnymi osobami.
Wartość krytyczna statystyki T^2 dla $\alpha = 0.05$ wynosi 56.77.

		K l a s y							
		1	2	3	4	5	6	7	8
O s o b y	1 i 2	95.8	40.4	32.6	95.6	80.8	35.9	140.3	88.3
	1 i 3	62.7	244.4	230.4	63.3	88.1	29.8	111.6	56.4
	2 i 3	139.5	52.9	51.3	42.8	31.3	41.2	33.4	46.8

W celu zbadania różnic między klasami przeprowadzono analizę dyskryminacyjną dla osoby, która wymawiała wypowiedzi wzorcowe. W tym przypadku poszczególne klasy reprezentowane były przez 10 replikacji otrzymanych dla tego głosu. Wyniki analizy przedstawiono w tabelach 2 i 3. O rzeczywistym

wymiarze przestrzeni dyskryminacyjnej wnioskujemy na podstawie pierwiastków charakterystycznych istotnie różnych od zera. Z tabeli 2 widać, iż suma dwóch pierwszych pierwiastków równa się 92 %, zaś trzech pierwiastków 99,9 % ogólnej sumy. Możliwy jest więc opis przestrzeni dyskryminacyjnej za pomocą dwóch lub trzech zmiennych. W tabeli 3 przedstawiono wartości średnie ośmiu klas w układzie dwóch pierwszych zmiennych dyskryminacyjnych.

Tabela 2.
Niezerowe pierwiastki charakterystyczne.

Nr	λ_1	$(\lambda_1 / \sum \lambda_1) \cdot 100 \%$
1	93.369	80.16
2	13.258	11.99
3	9.305	7.99
4	0.323	0.28
5	0.164	0.14
6	0.039	0.03
7	0.0145	0.01

Tabela 3.
Średnie wartości ośmiu klas w układzie dwóch zmiennych dyskryminacyjnych.

	K l a s a							
	1	2	3	4	5	6	7	8
d_1	-10.6	-17.4	-16.5	2.3	5.8	-7.7	3.4	5.5
d_2	-8.5	-10.9	-15.6	-12.6	-18.1	-12.4	-6.5	-10.5

Na ryc. 4a przedstawiono dendryt rozpięty na wartościach średnich badanych klas w układzie dwóch pierwszych zmiennych. Nad liniami łączącymi poszczególne klasy naniesiono wartość statystyki T^2 . Z porównania obliczonych wartości statystyki z wartością krytyczną równą 136 na poziomie istotności

$\alpha = 0.05$ wynika, że wszystkie odległości między badanymi klasami są statystycznie istotne. Analogiczny wniosek wynika z analizy dyskryminacyjnej przeprowadzonej łącznie dla trzech osób. Poszczególne klasy w tym przypadku reprezentowane były przez 30 replikacji danego wzorca. Wyniki analizy przedstawiono w tabelach 4 i 5. Na ryc. 4b wykreślono dendryt w układzie dwóch pierwszych współrzędnych. Różnice międzyosobnicze nie spowodowały więc odmiennego układu odległości między klasami w przestrzeni dyskryminacyjnej (por. ryc. 4a i 4b).

Tabela 4.

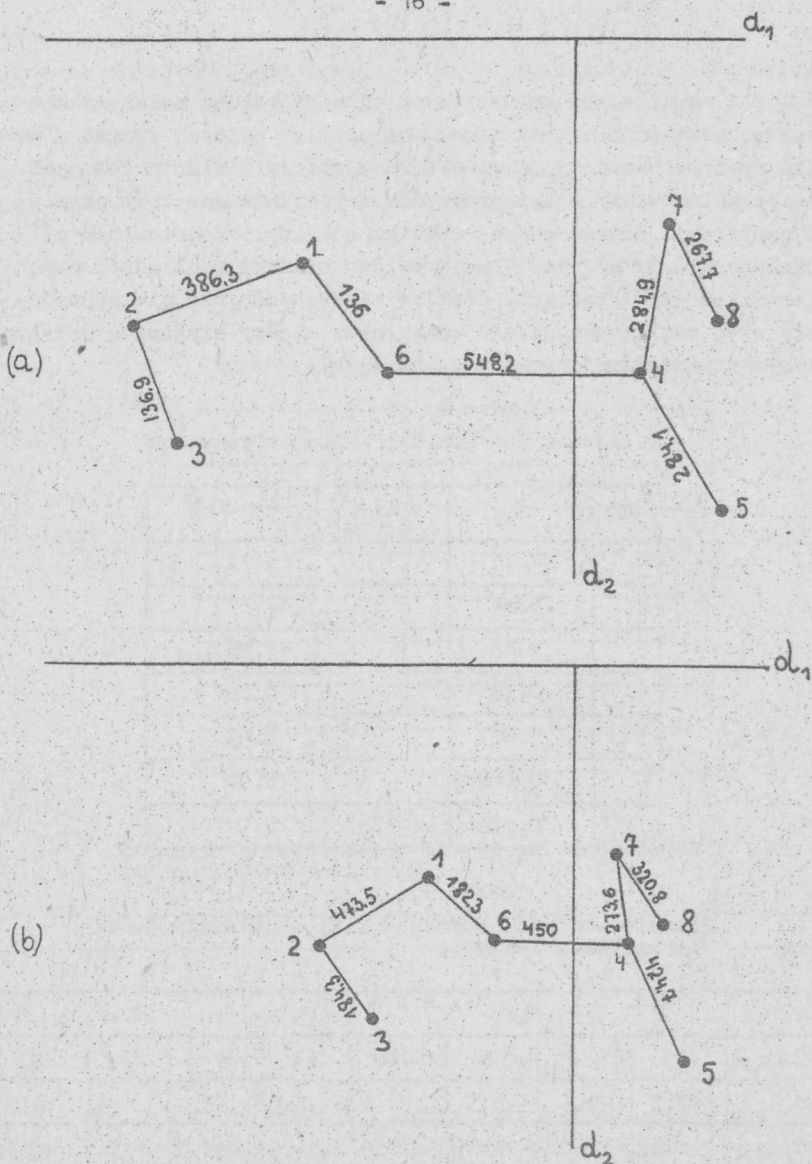
Niezerowe pierwiastki charakterystyczne.

Nr.	λ_i	$(\lambda_i / \sum \lambda_i) \cdot 100\%$
1	24.11	71.71
2	5.84	17.37
3	3.45	10.26
4	0.18	0.55
5	0.027	0.08
6	0.00816	0.02
7	0.0054	0.01

Tabela 5.

Średnie wartości badanych klas w układzie dwóch pierwszych zmiennych.

	K l a s y							
	1	2	3	4	5	6	7	8
d_1	-5.3	-9.4	-8.3	1.9	3.9	-3.2	1.6	2.4
d_2	-8.2	-10.8	-13.6	-10.3	-14.9	-10.7	-7.2	-10.2



Ryc. 4. Średnie ośmiu klas w układzie dwóch pierwszych zmiennych dyskryminacyjnych
4(a) - w przypadku analizy jednego głosu
4(b) - w przypadku analizy trzech głosów łącznie.

Z przeprowadzonej analizy nasuwają się następujące wnioski:

- badane klasy można scharakteryzować w przestrzeni dwuwymiarowej z zachowaniem odległości między nimi w 90 % lub trójwymiarowej z zachowaniem odległości w 99 %,
- różnicowania między poszczególnymi klasami są statystycznie istotne,
- istnieje znaczne różnicowanie międzyosobnicze w niektórych klasach,
- przyjęta liczba ośmiu danych wystarczająco dobrze pozwoliła na określenie różnicowań między klasami.

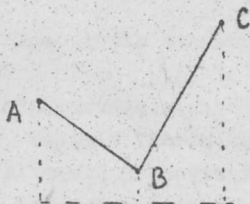
Dla przeprowadzenia statystycznej klasyfikacji należy dokonać podobnej analizy dla większej grupy osób. Można spodziewać się, iż różnicowania międzyosobnicze będą większe niż w wybranej grupie fonetyków. Aby ustalić, które przebiegi będą reprezentować klasy należy zbadać, czy poszczególne replikacje są ekwiwalentne percepcyjnie z wzorcami.

4. Interpretacja wyników analizy dyskryminacyjnej.

Na ryc. 4a i ryc. 4b można zauważyć, iż po jednej stronie wykresu wzdłuż osi poziomej umieszczone są punkty odpowiadające przebiegom rosnącym 2,3,1, po drugiej przebiegom opadającym 4,7,8. Między nimi znajduje się punkt 6 odpowiadający przebiegowi równemu (praktycznie przebieg ten nie był równy lecz lekko rosnący (por. ryc. 1,2,3)). Wzdłuż osi pionowej wyodrębnić można trzy grupy punktów: punkty 3 i 5 odpowiadające przebiegom, których punkt początkowy jest najwyższy względem punktów początkowych pozostałych przebiegów, punkty 1 i 7 odpowiadające przebiegom z najniższym punktem początkowym. Pomiędzy nimi umieszczona jest grupa punktów 2,4,6,8 w których punkt początkowy jest w pobliżu środka zakresu (por. ryc. 1,2,3). Można spodziewać się, że istnieje korelacja między pierwszą zmienną dyskryminacyjną i kierunkiem przebiegu oraz drugą zmienną i położeniem punktu początkowego przebiegu. Określenie przebiegu wyłącznie przez podanie położenia punktu początkowego i kierunku jest bardzo niejednoznaczne. Wydaje się celowe wprowadzenie trzeciego wymiaru charakteryzującego stopień wklęsłości lub wypukłości przebiegu. W pracy [3] stosując transformację Kar-

hunnena-Loevego wykazano, iż po eliminacji wartości średnich o dwóch najważniejszych wymiarach decyduje kierunek przebiegu i jego wklęsłość lub wypukłość.

Na ryc. 5 przedstawiono sposób interpretacji omawianych wyżej trzech wymiarów.



Ryc. 5. Interpretacja przyjętych wymiarów.

- jako pierwszy wymiar przyjęto różnicę między punktem końcowym i początkowym przebiegu (C-A).
- jako drugi wymiar przyjęto wartość punktu początkowego (A).
- jako trzeci przyjęto wartość wyrażenia (C-B) - (B-A), charakteryzującego stopień wklęsłości lub wypukłości przebiegu.

Istnieje analogia między wyrażeniem (C-A) i pierwszą pochodną

$(f_{nt}' \sim \frac{a_{n+1} - a_n}{t})$, oraz analogia między wyrażeniem (C-B) - (B-A)

i drugą pochodną $(f_{nt}'' \sim \frac{a_{n+1} - 2a_n + a_{n-1}}{t^2})$.

Do uzyskania precyzyjniejszego opisu należałoby wprowadzić czwarty wymiar decydujący o miejscu wystąpienia ekstremum.

Na podstawie szeregu prac można jednak przypuszczać, iż wymiar ten ma percepcyjnie najmniejsze znaczenie (por. [4], [6]). W celu zbadania, czy za pomocą wybranych trzech cech można przeprowadzić klasyfikację dla wszystkich osób, wybrano losowo po trzy replikacje wypowiedzi wzorcowych i w każdej z nich określono wartości (C-A), (A), (C-B) - (B-A). Wyniki przedstawiono w tabeli 6. Obliczeń dokonano na mikrokomputerze Sinclair. Jako wartość ekstremum przyjmowano wartość ekstremum globalnego.

Niektóre przebiegi różnią się znacznie między sobą wszystkimi trzema cechami (np. "1" i "8") imie różnią się tylko dwoma (np. "1" i "2") przebieg zaś "3" można odróżnić od "1" tylko za pomocą jednej cechy. Należy sprawdzić, czy zróżnicowanie między tymi cechami pozwoli na dokonanie klasyfikacji poszczególnych przebiegów.

5. Automatyczna klasyfikacja przebiegów parametru F_0 .

Matematyczne metody klasyfikacji mogą być podzielone na dwie kategorie: deterministyczną i statystyczną. Podejście deterministyczne jest oparte na regułach, których ustalenia nie wymagają znajomości statystycznych własności badanych klas. Jednym z podstawowych algorytmów stosowanym w deterministycznym podejściu jest algorytm określany w literaturze jako "perceptron algorithm" (por. np. [5], [11]). Decyzyjne funkcje są generowane z wzorców podanych maszynie cyfrowej przy pomocy iteracyjnego uczącego się algorytmu. W celu określenia współczynników funkcji decyzyjnej zastosowano następujący algorytm. Założono, że istnieje M decyzyjnych funkcji, które mają tę własność, że jeżeli $\underline{x} \in \Pi_i$ to

$$d_i(\underline{x}) > d_j(\underline{x}) \text{ dla wszystkich } j \neq i \text{ gdzie:}$$

\underline{x} - rozpoznawany wektor
 Π_i - klasa i

Rozważmy M klas $\Pi_1 \dots \Pi_M$. Załóżmy, że w k -tym kroku iteracyjnym podczas uczenia się wzór \underline{x}_k należący do klasy Π_i jest podany maszynie. Obliczone zostają funkcje decyzyjne

$$d_j[\underline{x}_k] = \underline{w}_j(k) \cdot \underline{x}_k$$

Jeśli

$$d_i[\underline{x}_k] > d_j[\underline{x}_k] \quad j=1 \dots M \quad j \neq i$$

to wektor wagowy \underline{w}_j pozostaje nie zmieniony w następnym kroku iteracyjnym

$$\underline{w}_j(k+1) = \underline{w}_j(k) \quad j = 1, 2, \dots, M$$

W sytuacji przeciwnej następuje zmiana wektora wagowego zgodnie z zależnościami:

$$\underline{W}_i(k+1) = \underline{W}_i(k) + C \underline{X}(k)$$

$$\underline{W}_i(k+1) = \underline{W}_i(k) - C \underline{X}(k)$$

$$\underline{W}_j(k+1) = \underline{W}_j(k) \quad j=1,2,\dots,M, \quad j \neq i, \quad j \neq l$$

gdzie C jest stałą. Jeśli klasy są liniowo rozłączne, to algorytm jest zbieżny w skończonej liczbie iteracji dla arbitralnie wybranego początkowego wektora wag. Można wykazać iż w tym przypadku algorytm jest optymalny. Kiedy klasy nie są rozłączne, interesujące jest wiedzieć, jak duża liczba wzorów może być poprawnie rozpoznaną. Algorytm nawet w tym przypadku wyszukuje rozwiązanie optymalne.

Nie wszystkie przebiegi przyjęte do analizy reprezentujące poszczególne klasy w układzie trójwymiarowym są dobrze separowalne. Wynika to z ich geometrycznego usytuowania. Np. klasę nr "1" trudno będzie oddzielić od klasy nr "2" i "3". Numerację klas przeprowadzono zgodnie z numeracją przebiegów na ryc. 1. Proces iteracyjny nie był zbieżny, algorytm zatrzymano w momencie kiedy otrzymane funkcje decyzyjne zapewniały w badanym układzie optymalny podział. Algorytm oprogramowano na mikrokomputerze Sinclair.

Otrzymano następujący zbiór decyzyjnych funkcji:

$$d_1(x) = 8.24x_1 - 10.005x_2 + 13.82x_3 - 4$$

$$d_2(x) = 33.315x_1 - 19.275x_2 + 28.24x_3 - 14$$

$$d_3(x) = 15.61x_1 + 5.26x_2 + 20.41x_3 - 13$$

$$d_4(x) = -28.055x_1 - 2.62x_2 - 20.305x_3 - 4$$

$$d_5(x) = -35.1x_1 + 19.14x_2 - 40.645x_3 - 26$$

$$d_6(x) = -3.22x_1 + 2.88x_2 + 8.37x_3 - 5$$

$$d_7(x) = -10.995x_1 - 11.435x_2 - 34.615x_3 - 2$$

$$d_8(x) = -11.55x_1 + 0.08x_2 - 53.515x_3 - 19$$

Na ryc. 6 przedstawiono wydruk wyników algorytmu. Maszynie cyfrowej podano wstępnie do rozpoznania obiekty, które stanowiły wypowiedzi wzorcowe. Z lewej strony wydruku podano wartość funkcji decyzyjnych, z prawej numer funkcji, literką P oznaczono numer rozpoznawanej klasy.

I tak w przypadku klasyfikacji wzorca intonacyjnego nr "1" otrzymano następujące wartości funkcji decyzyjnych :

$$d_1 = -0.3264$$

$$d_2 = -1.776025$$

$$d_3 = -0.6521$$

$$d_4 = -18.8197$$

$$d_5 = -42.769$$

$$d_6 = -2.2675$$

$$d_7 = -19.594$$

$$d_8 = -39.429$$

Maksimum równe -0.3264 osiąga funkcja decyzyjna d_1 co oznacza, iż dokonano poprawnego rozpoznania. W przypadku klasyfikacji wzorca intonacyjnego nr "5" otrzymano następujące wartości (por. ryc. 6) :

$$d_1 = -34.5304$$

$$d_2 = -88.42105$$

$$d_3 = -43.9768$$

$$d_4 = 38.9073$$

$$d_5 = 63.735$$

$$d_6 = -9.4909$$

$$d_7 = 36.1956$$

$$d_8 = 50.6636$$

W tym przypadku osiąga maksimum funkcja d_5 co oznacza, że wzorzec nr "5" został poprawnie rozpoznany. W przypadku klasyfikacji wzorca nr "7" widać, iż maksimum osiąga funkcja d_7 ($d_7 = 21.356$). Również bliska maksimum jest wartość funkcji d_8 ($d_8 = 21.0389$).

Na ryc. 7 przedstawiono wydruk przykładu, w którym algorytm nie dokonał poprawnej klasyfikacji. W przypadku rozpoznawania wzorca nr "5" (dla głosu 5, patrz tabela 6, kolumna 5) otrzymano maksimum dla funkcji decyzyjnej d_8 , co oznacza, iż wzorzec nr "5" należy do klasy 8, a nie do klasy 5.

Na ryc. 8 przedstawiono wydruk przykładu, w którym algorytm dokonuje poprawnej klasyfikacji w przypadku wszystkich wzorców dla głosu 2 (patrz tabela 6, kolumna 2).

Ponieważ należy się spodziewać pewnych błędów z uwagi na nierozłączność klas wyprowadzono dwie największe wartości funkcji decyzyjnych. W tabeli 7 zestawiono wyniki klasyfikacji dla przebiegów scharakteryzowanych poprzez trzy wymiary podane w tabeli 6.

W tabeli 7 podano, które dwie klasy są najbardziej podobne do rozpoznawanego obiektu.

W większości przypadków można było rozpoznać badany obiekt za pomocą jednej funkcji decyzyjnej. Dalszą decyzję w określeniu, czy badany przebieg należy do którejś z dwóch klas, czy do żadnej z nich, należałoby podjąć na podstawie analizy percepcyjnej.

Wnioski.

Przeprowadzona analiza dyskryminacyjna wykazała, że istnieje statystyczne zróżnicowanie między badanymi klasami. Możliwe jest więc dokonanie usystematyzowania przebiegów częstotliwości podstawowej. Dalsza statystyczna analiza wymaga jednak poszerzenia materiału językowego. Istnieje możliwość przedstawienia przebiegów parametru F_0 za pomocą trzech zmiennych. Przyjęty opis okazał się wystarczający dla zbadania różnic między poszczególnymi klasami. W 80 % przypadków imitacje rozpoznano zgodnie z odpowiadającym jej wzorcem. Teoretycznie pozostałe 20 % rozpoznania niezgodnych z wzorcem może być skutkiem bądź błędnej imitacji, bądź błędnej klasyfikacji. Wstępna ocena słuchowa oraz wizualna ocena przebiegów F_0 w imitacjach sugeruje, że w tych pozostałych 20 % przypadków wynik rozpoznania przypisać należy niewłaściwej imitacji. Ostateczne rozstrzygnięcie może się opierać na wynikach analizy percepcyjnej. Rezultaty klasyfikacji trudno jest więc podsumować z uwagi na brak weryfikacji wyników przez analizę percepcyjną. Interesujące byłoby również zbadanie dłuższych i bardziej skomplikowanych przebiegów. W przypadku negatywnych rezultatów podejścia deterministycznego należy zastosować statystyczną metodę klasyfikacji.

Tabl. 6. (c.d.)

Wartości współczynników x,y,z, dla wszystkich osób w poszczególnych klasach. Oznaczono $x = C-A$, $y = A$, $z = (C-B)-(B-A)$.

		Osoby												
		5 replikacje			6 replikacje			7 replikacje			8 replikacje			
		1	2	3	1	2	3	1	2	3	1	2	3	
Klasy	1	x	0,226	0,175	0,26	0,192	0,204	0,137	0,222	0,187	0,229	0,08	0,107	0,042
	y	0,245	0,254	0,2	0,356	0,255	0,308	0,301	0,278	0,244	0,458	0,41	0,486	
	z	0,458	0,401	0,312	0,479	0,313	0,258	0,402	0,388	0,34	0,365	0,24	0,38	
	2	x	0,568	0,647	0,606	0,474	0,49	0,51	0,351	0,046	0,406	0,378	0,367	0,522
	y	0,2	0,206	0,239	0,38	0,28	0,35	0,361	0,69	0,327	0,486	0,466	0,346	
	z	0,687	0,728	0,878	0,864	0,617	0,765	0,542	1,23	0,751	0,602	0,463	0,699	
	3	x	0,373	0,369	0,335	0,374	0,457	0,514	0,217	0,253	0,155	0,387	0,346	0,371
	y	0,34	0,374	0,392	0,396	0,369	0,304	0,403	0,365	0,447	0,464	0,511	0,469	
	z	0,513	0,563	0,423	0,603	0,569	0,588	0,43	0,369	0,456	0,566	0,447	0,464	
	4	x	-0,17	-0,23	-0,245	-0,453	-0,448	-0,516	-0,45	-0,331	-0,485	-0,186	-0,313	-0,288
	y	0,345	0,357	0,34	0,575	0,511	0,505	0,52	0,412	0,508	0,43	0,516	0,525	
	z	-0,59	-0,508	-0,56	-0,602	-0,514	-0,56	-0,56	-0,42	-0,628	-0,33	-0,326	-0,327	
	5	x	-0,566	-0,407	-0,38	-0,63	-0,72	-0,667	-0,54	-0,623	-0,371	-0,296	-0,324	-0,345
	y	0,69	0,528	0,511	0,733	0,76	0,704	0,63	0,753	0,519	0,599	0,619	0,572	
	z	-0,56	-0,93	-1,07	-0,89	-0,835	-0,83	-0,88	-0,752	-0,967	-0,82	-0,709	-0,94	
	6	x	0,034	0,057	0,0004	-0,019	0,012	-0,015	0,06	0,129	0,011	-0,001	0,005	-0,03
	y	0,335	0,29	0,33	0,36	0,282	0,38	0,35	0,274	0,301	0,408	0,44	0,43	
	z	0,064	0,14	0,09	0,205	0,07	0,141	0,14	0,254	0,189	0,09	0,164	0,097	
	7	x	-0,037	-0,117	-0,16	-0,176	-0,269	-0,308	-0,33	-0,304	-0,26	-0,08	0,013	-0,115
	y	0,171	0,23	0,266	0,34	0,273	0,328	0,45	0,447	0,398	0,354	0,27	0,35	
	z	-0,63	-0,707	-0,403	-0,55	-0,72	-0,608	-0,5	-0,745	-0,48	-0,389	0,112	-0,613	
	8	x	-0,099	-0,213	-0,212	-0,29	-0,35	-0,636	-0,46	-0,524	-0,34	-0,145	0,047	-0,196
	y	0,234	0,364	0,309	0,35	0,381	0,684	0,625	0,577	0,47	0,381	0,293	0,513	
	z	-1,43	-1,329	-1,215	-0,82	-1,24	-0,8	-0,81	-0,906	-1,04	-1,133	-1,244	-1,35	

Tabl. 6. (c.d.)

Wartości współczynników x,y,z, dla wszystkich osób w poszczególnych klasach. Oznaczono $x = C-A$, $y = A$, $z = (C-B)-(B-A)$.

		Osoby												
		9 replikacje			10 replikacje			11 replikacje			12 replikacje			
		1	2	3	1	2	3	1	2	3	1	2	3	
Klasy	1	x	0,282	0,342	0,257	0,158	0,269	0,132	0,224	0,242	0,167	0,179	0,09	0,103
	y	0,175	0,158	0,187	0,198	0,251	0,208	0,352	0,34	0,326	0,325	0,301	0,424	
	z	0,677	0,613	0,573	0,395	0,483	0,276	0,368	0,363	0,246	0,348	0,278	0,43	
	2	x	0,633	0,885	0,84	0,452	0,619	0,366	0,329	0,294	0,42	0,46	0,495	0,404
	y	0,107	0,116	0,12	0,219	0,223	0,243	0,45	0,415	0,4	0,37	0,36	0,407	
	z	0,891	1,2	0,93	0,73	0,811	0,545	0,533	0,58	0,58	0,572	0,56	0,717	
	3	x	0,625	0,477	0,46	0,564	0,579	0,545	0,089	0,147	0,185	0,369	0,2	0,22
	y	0,251	0,426	0,259	0,276	0,307	0,351	0,59	0,51	0,537	0,538	0,56	0,512	
	z	0,778	0,727	0,47	0,737	0,65	0,787	0,251	0,283	0,246	0,549	0,444	0,51	
	4	x	-0,564	-0,402	-0,39	-0,34	-0,5	-0,375	-0,42	-0,34	-0,35	-0,17	-0,411	-0,096
	y	0,496	0,5	0,46	0,54	0,63	0,55	0,57	0,45	0,456	0,528	0,57	0,43	
	z	-0,954	-0,719	-0,53	-0,62	-0,916	-0,83	-0,51	-0,51	-0,55	-0,35	-0,57	-0,23	
	5	x	-0,679	-0,742	-0,565	-0,507	-0,73	-0,44	-0,58	-0,45	-0,48	-0,607	-0,78	-0,29
	y	0,76	0,746	0,61	0,659	1,01	0,76	0,713	0,547	0,608	0,89	0,97	0,59	
	z	-0,707	-0,79	-0,599	-1,17	-1,29	-0,867	-0,733	-0,78	-0,75	-0,64	-0,96	-0,76	
	6	x	-0,05	0,011	0,08	-0,046	0,017	0,02	-0,02	-0,02	0,05	-0,04	-0,06	-0,03
	y	0,244	0,17	0,119	0,274	0,272	0,263	0,506	0,47	0,42	0,407	0,431	0,488	
	z	-0,109	0,186	0,14	0,265	0,219	0,122	0,124	0,183	0,18	0,105	-0,15	-0,06	
	7	x	-0,519	-0,394	-0,358	-0,129	-0,39	-0,145	-0,242	-0,203	-0,204	0,044	-0,048	-0,04
	y	0,538	0,408	0,342	0,35	0,57	0,474	0,33	0,319	0,47	0,357	0,34	0,416	
	z	-0,534	-0,514	-0,851	-1,254	-0,59	-0,79	-0,57	-0,513	-0,25	-0,489	-0,575	-0,55	
	8	x	-0,201	-0,613	-0,943	-0,267	-0,237	-0,102	-0,235	-0,207	-0,237	-0,065	-0,12	-0,04
	y	0,251	0,673	0,956	0,442	0,538	0,3	0,341	0,364	0,386	0,394	0,38	0,451	
	z	-0,584	-0,71	-0,973	-1,398	-1,18	-1,714	-0,803	-0,88	-0,78	-0,78	-0,972	-0,76	

Tabl. 6. (c.d.)

Wartości współczynników x, y, z, dla wszystkich osób w poszczególnych klasach.
Oznaczono x = C-A, y = A, z = (C-B) - (B-A).

		Osoby														
		13 replikacje			14 replikacje			15 replikacje			15 replikacje			15 replikacje		
		1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
1	x	0,258	0,29	0,268	0,234	0,255	0,233	0,124	0,153	0,169	0,124	0,153	0,169	0,124	0,153	0,169
	y	0,399	0,31	0,294	0,434	0,466	0,414	0,451	0,416	0,352	0,451	0,416	0,352	0,451	0,416	0,352
	z	0,47	0,55	0,458	0,364	0,321	0,557	0,13	0,209	0,49	0,13	0,209	0,49	0,13	0,209	0,49
2	x	0,462	0,503	0,47	0,413	0,514	0,607	0,365	0,4	0,35	0,38	0,4	0,35	0,38	0,4	0,35
	y	0,356	0,355	0,35	0,356	0,35	0,261	0,38	0,4	0,35	0,38	0,4	0,35	0,38	0,4	0,35
	z	0,699	0,678	0,67	0,598	0,68	0,79	0,494	0,44	0,48	0,494	0,44	0,48	0,494	0,44	0,48
3	x	0,31	0,37	0,36	0,48	0,46	0,46	0,274	0,29	0,23	0,274	0,29	0,23	0,274	0,29	0,23
	y	0,55	0,53	0,445	0,375	0,39	0,41	0,5	0,464	0,4	0,5	0,464	0,4	0,5	0,464	0,4
	z	0,379	0,53	0,45	0,79	0,88	0,554	0,45	0,365	0,43	0,45	0,365	0,43	0,45	0,365	0,43
4	x	-0,221	-0,414	-0,48	-0,24	-0,21	-0,197	-0,21	-0,19	-0,14	-0,21	-0,19	-0,14	-0,21	-0,19	-0,14
	y	0,46	0,596	0,6	0,413	0,405	0,356	0,585	0,55	0,56	0,585	0,55	0,56	0,585	0,55	0,56
	z	-0,71	-0,625	-0,88	-0,42	-0,414	-0,38	-0,49	-0,24	-0,8	-0,49	-0,24	-0,8	-0,49	-0,24	-0,8
5	x	-0,33	-0,709	-0,33	-0,28	-0,54	-0,35	-0,236	-0,18	-0,26	-0,236	-0,18	-0,26	-0,236	-0,18	-0,26
	y	0,47	0,705	0,51	0,502	0,78	0,49	0,65	0,53	0,58	0,65	0,53	0,58	0,65	0,53	0,58
	z	-1,21	-1,29	-0,96	-1,14	-0,73	-1,3	-0,5	-0,415	-0,62	-0,5	-0,415	-0,62	-0,5	-0,415	-0,62
6	x	0,107	-0,016	0,14	-0,02	0,017	0,185	0,07	0,02	0,15	0,07	0,02	0,15	0,07	0,02	0,15
	y	0,445	0,57	0,427	0,52	0,45	0,417	0,48	0,47	0,29	0,48	0,47	0,29	0,48	0,47	0,29
	z	0,23	0,321	0,215	0,33	0,25	0,411	0,09	0,2	0,24	0,09	0,2	0,24	0,09	0,2	0,24
7	x	-0,35	-0,65	-0,16	-0,15	-0,28	-0,19	-0,09	-0,05	-0,15	-0,09	-0,05	-0,15	-0,09	-0,05	-0,15
	y	0,58	0,23	0,35	0,43	0,426	0,36	0,42	0,411	0,4	0,42	0,411	0,4	0,42	0,411	0,4
	z	-0,98	-0,504	-0,66	-0,8	-0,7	-0,827	-0,48	-0,43	-0,5	-0,48	-0,43	-0,5	-0,48	-0,43	-0,5
8	x	-0,235	-0,207	0,24	-0,008	-0,05	-0,19	0,22	-0,19	-0,185	0,22	-0,19	-0,185	0,22	-0,19	-0,185
	y	0,34	0,364	0,386	0,365	0,36	0,4	0,52	0,545	0,484	0,52	0,545	0,484	0,52	0,545	0,484
	z	-0,8	-0,881	-0,78	-1	-1,16	-1,125	-0,94	-0,73	-0,03	-0,94	-0,73	-0,03	-0,94	-0,73	-0,03

Tabl. 7.

Rozpoznawanie przebiegów wyszczególnionych w tabeli 6.

Numer rozpoznawanego przebiegu	Klasy rozpoznane									
	osoba 1		osoba 2		osoba 3		osoba 4			
	replikacje	replikacje	replikacje	replikacje	replikacje	replikacje	replikacje	replikacje		
1	1	2	1	2	1	2	3	1	2	3
2	1/2/	1/2/	1/3/	1/3/	1/3/	1/3/	1/3/	1/3/	1/6/	3/1/
3	2/3/	2/3/	2/3/	2/3/	2/3/	2/3/	2/3/	2/3/	2/3/	2/3/
4	3/2/	3/6/	3/6/	3/2/	3/1/	3/6/	3/1/	3/2/	1/3/	3/6/
5	5/4/	5/8/	5/4/	5/4/	5/4/	5/4/	5/4/	4/5/	5/4/	5/4/
6	6/1/	6/1/	6/1/	6/1/	6/1/	6/1/	6/1/	6/1/	6/7/	4/6/
7	8/7/	8/7/	7/8/	7/8/	7/4/	7/4/	4/7/	7/8/	7/8/	7/8/
8	8/7/	8/5/	8/7/	8/7/	8/7/	8/7/	8/7/	8/7/	8/7/	8/7/

Numer rozpoznawanego przebiegu	Klasy rozpoznane								
	osoba 5		osoba 6		osoba 7		osoba 8		
	replikacje	replikacje	replikacje	replikacje	replikacje	replikacje	replikacje	replikacje	
1	1	2	1	2	1	2	1	2	3
2	1/2/	1/3/	1/2/	1/3/	3/1/	1/3/	1/2/	2/1/	2/3/
3	2/3/	2/3/	2/3/	2/3/	2/3/	2/3/	2/3/	2/1/	2/3/
4	2/3/	3/2/	2/3/	2/3/	2/3/	2/3/	3/2/	3/6/	3/2/
5	7/4/	7/4/	4/5/	4/7/	5/4/	4/5/	4/5/	5/4/	5/8/
6	5/4/	5/4/	5/4/	5/4/	5/8/	5/8/	5/8/	5/4/	5/4/
7	6/1/	6/1/	6/1/	6/1/	6/1/	6/1/	6/1/	6/4/	6/3/
8	7/8/	7/8/	7/8/	7/8/	7/8/	7/8/	7/8/	8/7/	7/8/
9	8/7/	8/5/	8/7/	8/7/	8/5/	8/5/	8/7/	8/7/	8/7/

W nawiasach podano numery klas, w których funkcja decyzyjna przyjmowała drugie co do wartości maksimum.

-6.3722	1
-2.79925	0
-1.3097	0
-17.82888	0
-41.8311	0
-2.024	0
-19.27985	0
-39.5982	0
P=	0
4.91537	1
12.6261	0
7.58286	0
-29.86707	4
-60.48616	4
-1.06692	6
-30.04236	4
-54.68582	0
P=	0
-2.075426	1
-5.47789	1
1.82461	0
-20.286488	4
-39.96592	6
-6.81742	7
-24.6431	0
-43.12189	0
P=	0
-17.55438	1
-45.583805	0
-24.248055	0
11.822055	4
11.80681	0
-6.6738	7
11.692165	0
9.29772	0
P=	4
-25.68713	1
-68.598355	0
-32.96534	0
26.1933	4
37.732338	6
-6.17405	6
19.6072	7
23.477895	0
P=	0
-6.523875	1
-17.88215	0
-7.45585	0
-3.691175	4
-29.7452	6
-2.8301.	7
-21.6515	0
-25.90315	0
P=	0
-14.439116	0
-36.845749	0
-20.203139	0
4.38158459	0
9.59382	0
-7.541622	0
0.9963705	0
0.262155	0
P=	0
-23.405495	1
-58.38214	0
-35.44889	0
21.458066	4
29.69411	6
-12.25268	7
32.46236	0
39.47892	0
P=	0

Ryc. 8. Wydruk wartości funkcji decyzyjnych.

BIBLIOGRAFIA

- (1) ATAL, B., Automatic speaker recognition, JASA, vol. 52, No. 6, 1972.
- (2) DEMENKO, G., Statystyczne własności rozkładów chwilowych wartości parametru F w mowie ciągłej, Prace IPPT, 31/80, Warszawa, 1980.
- (3) DEMENKO, G., Analiza matematyczna cech osobniczych głosu w zakresie parametru F_0 , Prace IPPT, 24/84, Warszawa, 1984.
- (4) GANDOUR, J., Tone perception in Far Eastern Languages, Journal of Phonetics, vol. 11, No. 2, April, 1983.
- (5) GONZALES, R., TOU, J., Pattern Recognition Principles, Canada, 1974.
- (6) HART, J.T., COHEN, A., Intonation by rule : a perceptual quest, Journal of Phonetics, vol. 1, No. 4, 1973.
- (7) JASSEM, W., KUDELA-DOBROGOWSKA, K., Inwarianty w przebiegach parametru F_0 , Prace IPPT, 25/73, Warszawa, 1973.
- (8) JASSEM, W., KRZYSKO, M., DYCZKOWSKI, A., Klasyfikacja i identyfikacja samogłosek polskich na podstawie częstotliwości formantów, Prace IPPT, 64/72, Warszawa, 1972.
- (9) JASSEM, W., KRZYSKO, M., DYCZKOWSKI, A., Identyfikacja głosów przy zastosowaniu funkcji dyskryminacyjnych, Prace IPPT, 57/71, Warszawa, 1971.
- (10) KRZYSKO, M., Analiza dyskryminacyjna, UAM, Seria Matematyki, Nr 6, Poznań, 1982.
- (11) MENDEL, J.M., FU, K.S., Adaptive Learning and Pattern Recognition Systems Theory and Applications, Academic Press, 1970.
- (12) STEFFEN-BATOGOWA, M., Analiza struktury przebiegu melodii polskiego języka ogólnego, rozprawa doktorska, Poznań, 1963.



