

**P. Domagała, L. Richter**

**AUTOMATYCZNA KLASYFIKACJA  
SPÓŁGŁOSEK TRĄCYCH JĘZYKA POLSKIEGO  
NA BAZIE OPTYMALIZACJI  
PRZESTRZENI PARAMETRÓW**

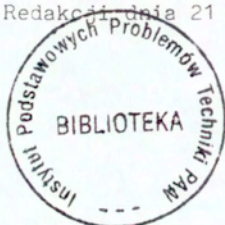
9/1994



P. 269

**WARSZAWA 1994**

Praca wpłynęła do Redakcji dnia 21 grudnia 1994 r.



56640



N a p r a w a c h r ę k o p i s u

---

Instytut Podstawowych Problemów Techniki PAN  
Nakład 100 egz. Ark.wyd. 1,5 Ark.druk. 2,00  
Oddano do drukarni w marcu 1994 r.

---

Wydawnictwo Spółdzielcze sp. z o.o.  
Warszawa, ul.Jasna 1

Piotr Domała  
Lutosława Richter  
Zakład Fonetyki Akustycznej  
IPPT PAN

## Automatyczna klasyfikacja spółgłosek trących języka polskiego na bazie optymalizacji przestrzeni parametrów

### Streszczenie.

Przedmiotem niniejszej pracy jest zbadanie możliwości automatycznej klasyfikacji i rozpoznawania głosek trących w oparciu o pewną kombinację liniową wartości funkcji autokorelacji. Przeprowadzenie automatycznej klasyfikacji spółgłosek trących stanowi jeden z etapów pracy, w ramach której dokonana zostanie klasyfikacja wszystkich głosek języka polskiego dla celów automatycznego rozpoznawania mowy niezależnie od kontekstu fonetycznego oraz od mówcy. Badania przeprowadzono na materiale fonetycznym obejmującym 166 logatomów z głoskami trącymi o strukturze CVCV i VCCV. Łączna liczba kontekstów dla każdej ze spółgłosek wyniosła około 20. Klasyfikację przeprowadzono oddzielnie dla 5 głosów kobiecych, dla 5 głosów męskich oraz łącznie, dla obu tych grup. Uzyskano odpowiednio 60%, 69% i 60% poprawnych klasyfikacji. Wyniki te są o około 10% lepsze od wyników uzyskanych w oparciu o klasyczną analizę dyskryminacyjną (tu: program CSS:Statistica 3.1). Przy zastosowaniu metod analizy skupień i skalowania wielowymiarowego uzyskano informacje o wzajemnych podobieństwach obrazów akustycznych tych głosek wraz z odniesieniem do badań percepcyjnych.

### 1. Wstęp

Przeprowadzenie automatycznej klasyfikacji spółgłosek trących stanowi jeden z etapów pracy, w ramach której dokonana zostanie klasyfikacja wszystkich głosek języka polskiego dla celów automatycznego rozpoznawania mowy niezależnie od kontekstu fonetycznego oraz od mówcy. Przedmiotem niniejszej pracy jest zbadanie możliwości automatycznej klasyfikacji i rozpoznawania głosek trących w oparciu o pewną kombinację liniową wartości funkcji autokorelacji.

Spotykane w literaturze próby klasyfikacji tej grupy głosek wykorzystują różne metody parametryzacji sygnału. Lobanov [7] zaproponował klasyfikację głosek trących języka rosyjskiego na podstawie dwóch wybranych cech akustycznych: częstotliwości środkowej widma chwilowego oraz częstotliwości początkowej

ugięcia formantu drugiego mierzonego na przejściu od spółgłoski do samogłoski. W materiale doświadczalnym uwzględniono spółgłoski bezdźwięczne w otoczeniu różnych samogłosek oraz 5 głosów. Autor wykazał, iż po wprowadzeniu określonej normalizacji wyników dla wszystkich głosów można wyznaczyć proste, które dzielą płaszczyznę na obszary przynależne różnym klasom głosek.

Pierwszą próbę klasyfikacji głosek dla języka polskiego przeprowadzili Gubrynowicz, Kacprowski, Mikiel, Skalski [3]. Parametryzację sygnału oparli na średniej gęstości przejść przez zero. Materiał fonetyczny obejmował spółgłoski trące bezdźwięczne w sąsiedztwie wszystkich samogłosek języka polskiego. W doświadczeniu uwzględniono 10 głosów męskich. Analizowano rozkłady liczebności przejść przez zero w kolejnych przedziałach częstotliwości. Ponieważ zakresy wartości dla poszczególnych głosek pokrywały się w znacznym stopniu, a średnie ważone różniły się między sobą nieznacznie (zwłaszcza w parach /ç-ʃ/ oraz /f-s/) utworzono łączne grupy dla głosek /ç ʃ/ oraz /f s/, traktując /x/ jako trzecią grupę. Aproksymacja rozkładów wyników za pomocą krzywych rozkładu normalnego pozwoliła sporządzić hipotetyczne rozkłady prawdopodobieństwa dla grup. Uwzględnienie kontekstu samogłoskowego spółgłosek trących oraz ich pozycji w sylabie ograniczyło rozrzut wartości przejść przez zero dla wszystkich rozpatrywanych głosek, co zwiększa prawdopodobieństwo ich klasyfikacji.

Jassem [4] wykorzystał dla celów klasyfikacji głosek trących języka polskiego analizę dyskryminacyjną. Parametryzacja sygnału przebiegała w oparciu o wartości poziomu widma, na podstawie których wyznaczano jego środek ciężkości, skośność oraz spłaszczenie. Materiał badawczy obejmował głoski trące bezdźwięczne w połączeniu z wszystkimi samogłoskami. Uwzględniono 3 głosy męskie. Przedstawiono wyniki klasyfikacji spółgłosek przeprowadzonej w oparciu o analizę dyskryminacyjną dla trzech wyróżnionych zmiennych. Dokładność rozpoznawania w obrębie jednego głosu wynosiła od 66% do 76%.

W niniejszej pracy zastosowano parametryzację sygnału w postaci wartości funkcji autokorelacji. Wybrany sposób opisu sygnału oraz metoda klasyfikacji zostały wykorzystane we

wcześniejszej pracy autorów, poświęconej dyskryminacji spółgłosek zwartych języka polskiego [9].

Funkcja autokorelacji zawiera w sobie w dostatecznym stopniu najważniejsze cechy sygnału mowy. Stanowi punkt wyjściowy do analizy sygnału mowy techniką LPC, dzięki której możemy otrzymać estymaty obwiedni widma sygnału mowy, funkcję przekroju toru artykulacyjnego lub współczynniki odbicia, umożliwiające resyntezę sygnału mowy. Spośród wielu technik opisywania sygnału mowy (Laface, de Mori [6]) funkcja autokorelacji nie jest stosowana w bezpośredniej postaci, ponieważ jako wektor parametrów charakteryzuje się wysokimi współczynnikami korelacji między sąsiednimi cechami. Nie jest to przeszkodą w prezentowanej metodzie, gdyż przed procesem klasyfikacyjnym wartości funkcji są poddawane odpowiedniej transformacji liniowej.

Pracę pilotażową [10] dotyczącą klasyfikacji głosek trących języka polskiego z wykorzystaniem techniki optymalizacji przestrzeni parametrów przeprowadzili autorzy na materiale jednego głosu. Badaniom poddano wszystkie głoski trące bezdźwięczne oraz dźwięczne we wszystkich kontekstach dopuszczalnych na gruncie fonotaktycznym. Na ogólną liczbę 9 głosek 7 uzyskało poprawność rozpoznawania przekraczającą 70%. Uzyskany wynik pozwalał zakładać, że zastosowana metoda klasyfikacji okaże się efektywna również w przypadku uwzględnienia w materiale doświadczalnym większej liczby głosek.

## 2. Materiał językowy

Utworzono logatomy dwusylabowe zawierające wszystkie alofony kontekstowe spółgłosek trących języka polskiego. Dla alofonów w kontekście samogłoskowym były to logatomy CVCV, w których badana spółgłoska zajmowała pozycję śródgłosową np. /sese/, zaś samogłoski poprzedzająca i następująca były identyczne. Dla alofonów występujących w sąsiedztwie spółgłosek niemożliwe było utworzenie kontekstów symetrycznych, w związku z czym dla danej spółgłoski uwzględniono kontekst poprzedzający lub następujący, umieszczając ją w logatomach VCCV, np. /s/ w /arsa/, /asfa/ itd. łączna liczba kontekstów dla każdej ze spółgłosek wyniosła około

20. O ich liczbie decydowały reguły fonotaktyczne języka polskiego, wykluczające niektóre połączenia głosek. Całość materiału doświadczalnego objęła 166 logatomów

które zostały odczytane przez 10 osób (5 głosów męskich i 5 kobiecych).

### 3. Parametryzacja sygnału mowy

Zarejestrowane na taśmie magnetofonowej logatomy zostały wprowadzone do pamięci komputera IBM PC AT. Zastosowano 13-bitowy przetwornik A/C i częstotliwość próbkowania wynoszącą 10 kHz. Sygnał został ograniczony do 5 kHz<sup>1</sup>. Filtr preemfazy zrealizowano wg zależności:

$$y(n) = x(n) - x(n-1) \quad (1)$$

Do parametryzacji sygnału wykorzystano 13 pierwszych wartości funkcji autokorelacji. Definiowana jest ona wg wzoru:

$$R(\tau) = \lim_{T \rightarrow \infty} \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t) * x(t+\tau) dt \quad (1)$$

Dla  $\tau = 0$  funkcja autokorelacji jest średnią wartością kwadratu funkcji, co dla sygnału odpowiada jego średniej energii. Dla ograniczonych czasowo, dyskretnych fragmentów sygnału (fram) wzór (2) przyjmuje postać<sup>1</sup>:

$$R(n) = \sum_{i=0}^{N-1-n} x_i * x_{i+n} \quad (3)$$

Ponieważ dla ustalonego N (odpowiadającego długości framy),  $R(0)$  jest wartością maksymalną, wykorzystano ją do normalizacji funkcji autokorelacji. Dla celów niniejszej pracy założono, że długość framy (mikrosegmentu) wyniesie 128 próbek (12.8 ms.). Krok, z jakim przesuwana była frama wzdłuż osi czasu wynosił 64

<sup>1</sup> Dla badań nad całym materiałem przyjęto pasmo 0-5000 Hz, mimo iż dla celów dokładnej analizy głosek trących jest ono absolutnie niewystarczające.

<sup>2</sup> Zakłada się, że  $x_i = 0$  dla  $i < 0$  oraz  $i > N-1$

próbki. Sygnał mowy był zapamiętywany w postaci ciągu liczb, który reprezentował po parametryzacji ciąg 12-elementowych wektorów  $(R(1) \dots R(12))$  wyznaczonych dla kolejnych mikrosegmentów.

#### 4. Baza danych

Utworzono bazę danych obejmującą pliki tekstowe (po jednym dla każdej głoski trącej). Każdy wiersz pliku zawiera 12 wartości funkcji autokorelacji  $(R(1) \dots R(12))$ , nr osoby, nr logatomu oraz nr framy w ramach logatomu. Pliki tworzył specjalnie do tego przygotowany program "narzędziowy" do wycinania wybranych fragmentów sygnału. Segmentację przeprowadzano w oparciu o zwizualizowane wartości funkcji autokorelacji dla kolejnych fram, zsynchronizowane na ekranie z postacią czasową sygnału. Z uwagi na znaczną liczebność populacji (18183 obiektów dla głosów kobiecych i 16172 obiektów dla głosów męskich) napisano kilka drobnych programów narzędziowych, umożliwiających sekwencyjną obróbkę danych w mniej licznych partiach.

#### 5. Analiza statystyczna.<sup>3</sup>

Obliczenia statystyczne przeprowadzono niezależnie dla trzech grup danych: (1) głosów kobiecych, (2) głosów męskich, (3) łącznie wszystkich głosów. W tabelach 1 - 3 zamieszczono podstawowe parametry statystyczne dla każdej ze zmiennych w określonej grupie głosów. Obejmują one wartość minimalną, maksymalną oraz średnią dla zmiennej, błąd standardowy, odchylenie standardowe oraz przedział ufności.

Tabele 4 - 6 zawierają macierze korelacji dla zmiennych. Puste miejsca w tabelach odnoszą się do przypadków, gdy wyliczone wartości okazały się nieistotne na zadanym poziomie  $p=0.01$ . Z nielicznymi wyjątkami uzyskano wartości istotne współczynników, co świadczy o silnej współzależności pomiędzy zmiennymi w przestrzeni pierwotnej.

<sup>3</sup>Obliczenia statystyczne przeprowadzono z wykorzystaniem programu CSS: Statistica 3.1

W tabelach 7 -9 podano wartości kwadratów odległości Mahalanobisa pomiędzy wektorami średnimi poszczególnych populacji, stanowiące podstawę do przeprowadzenia klasyfikacji obiektów metodą analizy dyskryminacyjnej.

Dane z tabeli 10 dotyczą wartości własnych (por. też rys.1) świadczą, iż wystarcza 3- wymiarowa przestrzeń utworzona na bazie zmiennych dyskryminacyjnych odpowiadających największym wartościom własnym, aby w 95% zachować podstawowe odległości w przestrzeni parametrów.

W tabelach 11 -13 przedstawiono procentowo wyniki klasyfikacji. Ostatnia kolumna w tabeli podaje liczbę poprawnych klasyfikacji dla danej głoski. Generalnie lepszy wynik uzyskano dla głosów męskich - 58% poprawnych rozpoznań wszystkich obiektów, wobec 52% dla głosów kobiecych. łącznie dla wszystkich głosów liczba ta wyniosła 49%. Zdecydowanie gorzej przebiegała klasyfikacja spółgłosek dźwięcznych - z wyjątkiem /v/. W zasadzie można uznać głoski /z ʒ ʒ/ za nierozpoznawalne - niejednokrotnie zdarzały się wskazania na inną głoskę niż przewidywana, np. w tab.13 znacznie częstsze były rozpoznania /z/ jako /ʒ/ lub /ʒ/, niż oczekiwane /z/, a /ʒ/ jako /ʒ/ lub /ʒ/ niż oczekiwane /ʒ/.

Rysunek 2 przedstawia lokację wektorów średnich dla 9 populacji głosek trących w przestrzeni dwuwymiarowej, uzyskaną w wyniku skalowania wielowymiarowego. Pomimo pewnych przesunięć obserwowanych w różnych grupach można mówić o ogólnej tendencji dotyczącej rozmieszczenia głosek. Skrajne pozycje na każdym z wykresów przyjęły spółgłoski /v s x/, te same, które uprzednio uzyskały najlepsze wyniki rozpoznawania (patrz tabele 11 - 13). Z kolei najbliższe skupienie wykazują głoski /ʒ ʒ ʒ/ często mieszane w procesie klasyfikacji, jak również wykazujące znaczne podobieństwo w zakresie charakterystyki widmowej.

Dla wektorów średnich przeprowadzono analizę skupień w oparciu o odległości Mahalanobisa. Historię powstawania skupień przedstawia rysunek 3. Procedura łączenia się głosek w coraz liczniejsze grupy przebiegała w ośmiu krokach. W początkowych etapach połączeniu podlegały głoski o najmniejszych odległościach (wykazujące największe podobieństwo). Dotyczyło to w każdej grupie głosów spółgłosek /ʒ ʒ ʒ/, które sytuowały się blisko



siebie również w wyniku skalowania wielowymiarowego. (rys. 2). W ostatnim kroku dołączały do rozrastającego się skupienia spółgłoski /v/ lub /s/, które w rezultacie skalowania wielowymiarowego przyjmowały skrajne pozycje wobec pozostałych głosek.

Podjęto próbę porównania wyników analizy statystycznej dla materiału doświadczalnego, traktującego głoski na płaszczyźnie akustycznej z analizą statystyczną odnoszącą się do ich podobieństwa percepcyjnego. W tym celu wykorzystano dane zamieszczone w pracy Łobacz ([8]-str. 129), poświęconej subiektywnemu podobieństwu fonemów języka polskiego. W oparciu o wartości odległości między głoskami trącymi wyrażone przez prawdopodobieństwo niepodobieństwa przeprowadzono skalowanie wielowymiarowe (rys. 4) oraz analizę skupień (rys.5). Rozmieszczenie średnich wektorów w przestrzeni dwuwymiarowej przybrało odmienną postać, niż na rysunku 2. Wydaje się, że na płaszczyźnie percepcyjnej można mówić o tendencji do łączenia się w pary głosek różniących się jedynie cechą dźwięczności. Również analiza skupień doprowadziła do utworzenia sekwencji dwuelementowych głosek różniących się wyłącznie dźwięcznością. W odniesieniu do uwzględnionego materiału doświadczalnego można stwierdzić, iż podobieństwa między poszczególnymi głoskami trącymi przebiegały w nieco odmienny sposób na płaszczyźnie akustycznej i percepcyjnej.

#### 6. Podstawy teoretyczne klasyfikacji z wykorzystaniem techniki optymalizacji przestrzeni parametrów

Wektor zbudowany z  $p$  elementów wygodnie jest przedstawić jako punkt umieszczony w  $p$ -wymiarowej przestrzeni. Zbiór wielu takich punktów tworzy w tej przestrzeni skupienie o pewnym kształcie i koncentracji. Jeśli ważny byłby postulat o normalności rozkładu takich punktów, to obszar ufności o zadanym poziomie procentowym stanowi  $p$ -wymiarową elipsoidę. Pomimo, iż taki postulat na ogół nie bywa spełniony, często jego założenie prowadzi do wyników, które można uznać jako suboptymalne. Równanie wielowymiarowego rozkładu normalnego przedstawia się wzorem:

$$N(X, M, \Sigma) = (2\pi)^{-\frac{1}{2}n} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(X-M)^T \Sigma^{-1} (X-M)\right) \quad (4)$$

Wyrażenie (5) jest równaniem kwadratowym p wymiarowej powierzchni - elipsoidy ufności, obejmującej zadany procent populacji:

$$(X-M)^T \Sigma^{-1} (X-M) = \text{const} \quad (5)$$

Dla wielowymiarowej elipsoidy można wyznaczyć kierunki jej osi. Ich znalezienie przynosi podstawową korzyść, jaką jest znalezienie nowych, wzajemnie nieskorelowanych cech (będących liniowymi kombinacjami cech pierwotnych). Oś związana z największą wariancją wyznacza kierunek, na którym rozrzut elementów jest największy. Wprowadźmy nową zmienną  $Z = X - M$ . Oznacza to przesunięcie środka układu współrzędnych do wektora średniego populacji. Macierz kowariancji dla nowej zmiennej nie ulegnie zmianie (elipsoidala zostanie tylko przesunięta). Kwadratowa funkcja przyjmie wyrażenie:

$$Z^T \Sigma^{-1} Z = \text{const} \quad (6)$$

Poszukiwany wektor z maksymalizujący (3) przy dodatkowym warunku

$$z z^T = 1 \quad (7)$$

można obliczyć z równania

$$\frac{\delta}{\delta z} [z^T \Sigma^{-1} z - \mu (z^T z - 1)] = 0 \quad (8)$$

gdzie  $\mu$  jest czynnikiem Lagrange'a.

$$2\Sigma^{-1} z - 2\mu z = 0 \quad (9)$$

a więc

$$\Sigma z - \frac{1}{\mu} z = 0 \quad (10)$$

Równanie (10) można rozpoznać jako zagadnienie własne macierzy  $\Sigma$ . Będzie ono miało nietrywialne rozwiązanie, gdy

$$\det\left(\Sigma - \frac{1}{\mu} I\right) = 0 \quad (11)$$

Oznacza to, że  $1/\mu$  jest wartością własną macierzy kowariancji,

a z jej wektorem własnym.

Niech dla nieujemnie określonej macierzy  $\Sigma$ ,  $\Phi$  jest macierzą zbudowaną z wektorów własnych,  $\Phi_i$  oraz  $\lambda_i$  będą wektorem i odpowiadającą mu wartością własną. Wektory własne macierzy kowariancji są ortogonalne ( $\Phi_i^T \Phi_j = 0$ ). Niech scentralizowana populacja zostanie przedstawiona w nowym układzie współrzędnych:

$$Y = \Phi^T Z \quad (12)$$

Macierz kowariancji ulega kolejnym przekształceniom:

$$\Sigma_Y = E(YY^T) = E(\Phi Z (\Phi Z)^T) = E(\Phi Z Z^T \Phi^T) = \Phi \Sigma_Z \Phi^T = \Phi \Sigma_Z \Phi^T = \Lambda \quad (13)$$

gdzie  $\Lambda$  jest macierzą diagonalną. Dokonując podstawienia:

$$Y_i = \Lambda^{-1/2} \Phi_i^T Z \quad (14)$$

otrzymujemy jednostkową macierz kowariancji:

$$\Sigma_{Y_i} = E(Y_i Y_i^T) = 1 \quad (15)$$

Takie przekształcenie oznacza, że elipsoida ufności została przekształcona w kulę. Przedstawione powyżej operacje zilustrowano graficznie na rys 6.

Rozpatrzmy dwie populacje, posiadające różne wektory średnie i różne macierze kowariancji  $\Sigma_1$  i  $\Sigma_2$ . Niech  $\Phi_1$  będzie macierzą, której kolumny stanowią będą wektory własne pierwszej macierzy kowariancji. Niech  $\Lambda_1$  będzie macierzą diagonalną, na której głównej przekątnej znajdują się wartości własne. Przekształcenie (15) zamienia elipsoidę związaną z  $\Sigma_1$  na sferę. Macierz kowariancji drugiej populacji ulega przekształceniu:

$$K = \Lambda_1^{-1/2} \Phi_1^T \Sigma_2 \Phi_1 \Lambda_1^{-1/2} \quad (16)$$

gdzie  $K$  jest nową macierzą kowariancji, która na ogół nie jest diagonalna. Można znaleźć macierz jej wektorów własnych  $\Psi$  i obie populacje poddać ortonormalnej transformacji, czyli takiemu obrotowi, ażeby kolejny układ współrzędnych miał kierunki zgodne z jej osiami. Ponieważ macierz kowariancji pierwszej populacji jest jednostkowa (poszczególne cechy są nieskorelowane i ich wariancja wynosi 1), to po takim przekształceniu nie ulegnie zmianie. Macierz kowariancji  $K$  ulegnie diagonalizacji:

$$\Psi^T K \Psi = \Lambda_1 \quad (17)$$

Powyższe rozumowanie pozwala sformułować następujące twierdzenie o istotnym znaczeniu dla analizy dyskryminacyjnej: Jeżeli  $\Sigma_1$  jest macierzą symetryczną określoną dodatnio i  $\Sigma_1$  jest nieujemnie określoną macierzą symetryczną, to istnieje taka nieosobliwa macierz  $\Psi = [\Psi_1 \ \Psi_2 \ \Psi_3 \ \dots \ \Psi_j]$ , że

$$\Psi^T \Sigma_1 \Psi = \Lambda \quad (18)$$

oraz

$$\Psi^T \Sigma_1 \Psi = 1 \quad (19)$$

gdzie  $\Lambda = \text{diag}(\lambda_1 \ \lambda_2 \ \lambda_3 \ \dots \ \lambda_j)$ ;  $\det(\Sigma_1 - \lambda_i \Sigma_1) = 0$ ; natomiast  $\Psi_1$  spełnia równanie:

$$(\Sigma_1 - \lambda_i \Sigma_1) \Psi_1 = 0 \quad (20)$$

Dowód, jak i szerokie przedstawienie omawianej problematyki można znaleźć w literaturze [2], [5]. Powyższe równanie przedstawia uogólnione zagadnienie własne dla dwóch macierzy. Operacje te zilustrowano na rysunku 7.

Rozpatrzmy teraz  $k$   $p$ -wymiarowych populacji normalnych o równych macierzach kowariancji  $\Sigma$  i różnych wektorach średnich  $M_i$ . Jako  $T$  oznaczymy macierz sum kwadratów i iloczynów wszystkich obiektów  $X$ , niezależnie od przynależności do poszczególnych populacji:

$$T = XX^T \quad (21)$$

gdzie  $X$  jest macierzą  $p \times N$ , zaś  $N$  liczebnością wszystkich badanych obiektów. Jako  $B$  oznaczono macierz rozprożeń międzygrupowych

$$B = \sum_{i=1}^k N_i (M_i - M)^T (M_i - M) \quad (22)$$

Jako  $W$  oznaczona zostanie macierz rozprożeń wewnątrzgrupowych:

gdzie  $X_{i1}$  oznacza  $i$ -ty wektor z populacji  $\alpha$ . Jest oczywiste, że

$$W = \sum_{i=1}^I \sum_{j=1}^{J_i} (X_{i,j} - M_i)^T (X_{i,j} - M_i) \quad (23)$$

$$T = B + W \quad (24)$$

jeżeli całą populację poddamy transformacji :

$$U = \Psi X \quad (25)$$

gdzie  $\Psi$  jest macierzą kolumnowych wektorów własnych uzyskanych z rozwiązania :

$$(B - \lambda_j W) \Psi_j = 0 \quad (26)$$

i gdzie  $\lambda_j$  jest pierwiastkiem równania:

$$\det(B - \lambda W) = 0 \quad (27)$$

to takie przekształcenie zmienia rozrzucone w przestrzeni pierwotnych parametrów elipsoidy ufności na jednostkowe kule. Otrzymujemy konfigurację kul najbardziej korzystną dla klasyfikacji, ponieważ spoglądamy na przestrzeń pod takim kątem, pod którym kule są najbardziej rozsunięte. Addytywna struktura wzorów na macierze rozprożeń wewnątrz- i międzygrupowych umożliwia ich obliczenie także w sytuacji podzielenia danej populacji na mniej liczne części składowe (co w wielu programach może być istotne).

## 7. Metoda klasyfikacji

W oparciu o powyższe podstawy teoretyczne przyjęto algorytm klasyfikacji opracowany przez Domagałę [1]. Musi on być poprzedzony etapem uczenia, który realizowany jest w następujących krokach:

1. Obliczenie macierzy BETWEEN I WITHIN.
2. Rozwiązanie uogólnionego zagadnienia własnego dla macierzy.
3. Zapamiętanie w tablicy wektora własnego odpowiadającego maksymalnej wartości własnej.
4. Dokonanie rzutowania wszystkich obiektów na kierunek zgodny z kierunkiem wektora własnego.
5. Wyznaczenie progowej wartości dyskryminacyjnej rozdzielają-

cej analizowane populacje na dwa podzbiory populacji składowych (niekoniecznie rozłączne)

6. Powrót do punktu 1 i powtórzenie całej operacji dla każdego z uzyskanych podzbiorów.

Procedura kończy się wtedy, gdy wszystkie populacje zostaną rozgraniczone lub gdy rzuty populacji wzajemnie się pokrywają. W przypadku częściowego pokrywania się populacji można przeprowadzać ich redukcję odrzucając w dalszej analizie te obiekty, które dają się sklasyfikować. Końcowym efektem przedstawionej procedury, nie zawsze możliwym do osiągnięcia, jest więc przydzielenie określonego obszaru w przestrzeni parametrów tylko jednej populacji.

Proces uczenia jest pracochłonny, ale jednorazowy. W jego wyniku otrzymujemy zbiór wektorów własnych, przyporządkowanych kolejno otrzymywanym grupom populacji i odpowiadający im zbiór wartości dyskryminujących.

Proces klasyfikacji pojedynczego obiektu reprezentującego dowolny mikrosegment przebiegać będzie następująco:

1. Skalarne mnożenie wektora parametrów pierwotnych przez odpowiedni wektor własny uzyskany na etapie uczenia.
2. Porównanie wyniku mnożenia z odpowiadającą wektorowi własnemu wartością dyskryminacyjną i na tej podstawie przyporządkowanie obiektu do właściwej grupy populacji.
3. Dobranie kolejnego wektora własnego i powrót do punktu 1.

Procedura klasyfikacyjna ulega zakończeniu, gdy obiekt zostanie zakwalifikowany do nierozdzielnej populacji.

8. Realizacja procedury klasyfikacyjnej

Na rys. 8 przedstawiono rzut wszystkich obiektów (mikrosegmentów) reprezentujących głoski trące w obu grupach na kierunku optymalny dla równoczesnej dyskryminacji wszystkich 9 populacji (oś pionowa). Każdy obiekt reprezentowany jest przez punkt, którego położenie na osi poziomej w ramach obszaru przynależnego dla danej populacji jest losowe.

Populacje głosek trących poddano procedurze klasyfikacyjnej osobno w każdej grupie głosów. Na etapie uczenia wyliczono wektory własne dla poszczególnych populacji składowych oraz wyznaczono na podstawie graficznie przedstawionego wyniku rzutowania progowe wartości dyskryminacyjne. W jednym przypadku korzystne okazało się rzutowanie na płaszczyznę wyznaczoną przez dwa wektory własne, odpowiadające dwóm największym wartościom własnym. Wówczas podział przestrzeni nastąpił w oparciu o prostą.

W czterech przypadkach, gdzie miało miejsce znaczne pokrywanie się populacji, niewystarczające okazało się dla ich rozdzielania wyznaczenie jednej wartości progowej. Wprowadzenie na danym etapie dwóch wartości progowych pozwoliło przydzielić w sposób jednoznaczny obszary przestrzeni przynajmniej dla części obiektów. Taka sytuacja zaszła między innymi przy podziale populacji składowej obejmującej głoski /ç ʃ/ oraz populacji obejmującej głoski /z ʒ/ w głosach kobiecych.

Na rys. 9-11 przedstawiono kolejne kroki prowadzące do rozdzielania zbiorów populacji na coraz mniej liczne podzbiory. Rysunki pozwalają prześledzić, pomiędzy którymi spółgłoskami zachodzi największe podobieństwo (pary rozdzielone na najniższym poziomie). Układ drzewek klasyfikacyjnych jest nieco odmienny w poszczególnych grupach głosów, nie mniej na każdym z diagramów zaobserwować można te same tendencje. Zdecydowane zróżnicowanie zachodzi pomiędzy grupą głosek bezdźwięcznych i grupą głosek dźwięcznych - wartość dyskryminacyjna wyznaczona w pierwszym kroku pozwoliła całkowicie oddzielić głoskę /v/ od pozostałych oraz częściowo oddzielić głoskę /z/, a w głosach kobiecych i głosach połączonych również głoski /z ʒ/. Niektóre pary rozdzielone na najniższym poziomie, a więc wykazujące największe podobieństwo powtarzają się w każdym z drzewek. Są to głoski /f x/, /ʃ ç/ /z v/.

Dla każdego obiektu reprezentującego dowolny mikrosegment w materiale doświadczalnym przeprowadzono klasyfikację do odpowiedniej populacji w oparciu o wartości wektorów własnych uzyskane na etapie uczenia. Tabele 14 - 16 podają procentowy udział rozpoznań poprawnych oraz błędnych dla wszystkich obiektów reprezentujących wszystkie populacje w danej grupie głosów.

Ogółem w głosach kobiecych poprawnie zaklasyfikowano 60% obiektów reprezentujących różne głoski, zaś w głosach męskich 69%. Dla wszystkich głosów liczba ta wyniosła 61%. Do najlepiej rozpoznawanych należały w każdej grupie głoski /v s x/, analogicznie jak w przeprowadzonej analizie statystycznej. Zdecydowanie lepsze rezultaty, niż uprzednio, uzyskano dla spółgłosek dźwięcznych - dla żadnej z nich nie powtórzyły się wyniki świadczące o częstszych wskazaniach na określoną, inną głoskę, niż przewidywana.

Porównanie wyników obu klasyfikacji przedstawiono graficznie na rysunku 12. Słupki zakropkowane odnoszą się do klasyfikacji przeprowadzonej z wykorzystaniem programu statystycznego CSS Statistica 3.1, słupki zaczernione do klasyfikacji opartej na rzutowaniu obiektów. Zastosowanie drugiej z metod poprawiło ogólne rezultaty w każdej grupie głosów o około 10%. Szczególnie uderzająca jest porawa w przypadku głosek /z ʒ ʒ/. W tej sytuacji należy uznać hierarchiczną metodę klasyfikacji za bardziej efektywną dla celów rozpoznawania głosek trących.

## 9. Uwagi końcowe

Uzyskane wyniki klasyfikacji można uznać za reprezentatywne dla głosek trących ze względu na wysoką liczebność badanych populacji, jak również z powodu uwzględnienia wszystkich alofonów kontekstowych tych głosek. Należy podkreślić, że wprowadzenie zróżnicowania kontekstowego do materiału doświadczalnego stanowiło znaczne utrudnienie w procesie klasyfikacji z powodu zwiększenia rozproszenia wewnątrzpopulacyjnego. Drugim czynnikiem, poważnie utrudniającym dyskryminację głosek, było uwzględnienie w materiale językowym różnych głosów i to zarówno kobiecych, jak męskich.

Uzyskane wyniki wykazały, że zaproponowana metoda hierarchicznej klasyfikacji w oparciu o analizę dyskryminacyjną może znaleźć zastosowanie w automatycznym rozpoznawaniu mowy niezależnie od kontekstu fonetycznego oraz mówcy.

Jak można wnioskować z przedstawionych wyników, uzyskanie jakościowo lepszych rezultatów klasyfikacji na płaszczyźnie



akustycznej (np. do poziomu 95 %) dla tak szerokiego materiału językowego i optymalnym nakładem kosztów (w sensie metodologii, mocy obliczeniowych i koniecznego sprzętu) jest prawdopodobnie niemożliwe. Konieczne jest filtrowanie uzyskanych wyników na wyższych poziomach - fonotaktycznym, syntaktycznym, semantycznym itd. poprzez konstruowanie funkcjonalnych modeli mechanizmów percepcji mowy, jakie stosowane są w psycholingwistyce. Taki wniosek, zdaniem autorów, stanowi ważną przesłankę, co do perspektywicznych kierunków dalszych badań.

- o -

Autorzy dziękują Panu mgr Lechowi Śmigajowi za wszelką okazaną pomoc.

Tabela 1, 2, 3. Podstawowe parametry statystyczne.

Tab.1							
Głosy kobiece N = 18183							
param.	Min	Max	Mean	St. Err.	St. Dev.	p=.9500 Up.limit	p=.9500 Dn.limit
R1	-.983	.986	-.177	.003	.521	-.169	-.184
R2	-.960	.962	.070	.003	.432	.076	.064
R3	-.940	.970	.230	.003	.406	.236	.224
R4	-.959	.931	.016	.003	.413	.022	.010
R5	-.900	.914	.137	.002	.318	.141	.132
R6	-.922	.943	.067	.002	.324	.071	.062
R7	-.934	.819	.083	.002	.297	.088	.079
R8	-.886	.920	.119	.001	.260	.123	.116
R9	-.915	.928	.006	.002	.269	.010	.002
R10	-.874	.857	.012	.001	.237	.016	.009
R11	-.855	.865	.003	.001	.224	.006	.000
R12	-.881	.903	.009	.001	.222	.012	.006

Tab.2							
Głosy męskie N = 16172							
param.	Min	Max	Mean	St. Err.	St. Dev.	p=.9500 Up.limit	p=.9500 Dn.limit
R1	-.983	.982	-.181	.003	.442	-.174	-.187
R2	-.970	.956	-.068	.003	.488	-.061	-.076
R3	-.887	.923	.106	.002	.357	.112	.101
R4	-.944	.946	.184	.002	.365	.190	.178
R5	-.971	.838	-.028	.002	.292	-.024	-.033
R6	-.937	.813	-.009	.002	.313	-.004	-.014
R7	-.855	.892	.113	.002	.260	.117	.109
R8	-.810	.915	.131	.002	.271	.136	.127
R9	-.935	.940	.057	.001	.250	.061	.053
R10	-.907	.914	-.086	.001	.253	-.082	-.090
R11	-.796	.902	.026	.001	.234	.030	.023
R12	-.782	.893	.077	.001	.227	.081	.074

Tab.3							
Głosy połączone N = 34355							
param.	Min	Max	Mean	St. Err.	St. Dev.	p=.9500 Up.limit	p=.9500 Dn.limit
R1	-.983	.986	-.178	.002	.486	-.173	-.184
R2	-.970	.962	.004	.002	.464	.009	-.000
R3	-.940	.970	.172	.002	.389	.176	.168
R4	-.959	.946	.095	.002	.400	.099	.091
R5	-.971	.914	.059	.001	.317	.062	.055
R6	-.937	.943	.030	.001	.322	.034	.027
R7	-.934	.892	.097	.001	.280	.100	.094
R8	-.886	.920	.125	.001	.265	.128	.122
R9	-.935	.940	.030	.001	.262	.033	.027
R10	-.907	.914	-.033	.001	.250	-.031	-.036
R11	-.855	.902	.014	.001	.229	.016	.011
R12	-.881	.903	.041	.001	.227	.044	.039

Tabela 4, 5, 6. Macierze autokorelacji.

Tab.4		Głosy kobiece N = 1883 p < 0.01											
param.	R1	R2	R3	R4	R5	R6	R7	R8	R9	R1	R1	R12	
R1	1	.35	.23	.57	.33	.43	.47	.29	.25	.07		-.02	
R2	.35	1	.17	.39	.56	.30	.33	.16	.21	.24	.04	-.18	
R3	.23	.17	1	.12	.29	.43		.15	.20	.12	.05		
R4	.57	.39	.12	1	.15	.21	.46	.09	.28	.12	.08		
R5	.33	.56	.29	.15	1	.05	.21	.40	-.03	.29	.05	-.12	
R6	.43	.30	.43	.21	.05	1	-.02	.06	.48	-.11	.04	.08	
R7	.47	.33		.46	.21	-.02	1	-.05	.02	.30	-.04		
R8	.29	.16	.15	.09	.40	.06	-.05	1	-.24	.06	.29	-.15	
R9	.25	.21	.20	.28	-.03	.48	.02	-.24	1	-.22	-.07	.27	
R10	.07	.24	.12	.12	.29	-.11	.30	.06	-.22	1	-.21	-.06	
R11		.04	.05	.08	.05	.04	-.04	.29	-.07	-.21	1	-.24	
R12	-.02	-.18			-.12	.08		-.15	.27	-.06	-.24	1	

Tab. 5		Głosy męskie N=16172 p < 0.01											
param.	R1	R2	R3	R4	R5	R6	R7	R8	R9	R1	R1	R12	
R1	1	.07	.35	.12	.36	.22	.36	.44	.20	.16	.03	-.03	
R2	.07	1	-.17	-.05	.46	.50	.07	-.10	.30	.44	-.05	-.26	
R3	.35	-.17	1	.32	.18	.20	.17	.08	.10	.07	.23	.21	
R4	.12	-.05	.32	1	.14	-.16	.18	.29	-.19	-.03	.29	.31	
R5	.36	.46	.18	.14	1	.25	.12	.09	.34	.13	.04	.03	
R6	.22	.50	.20	-.16	.25	1	-.06	-.14	.31	.43	-.22	-.20	
R7	.36	.07	.17	.18	.12	-.06	1	.08	-.09	.07	.22	-.20	
R8	.44	-.10	.08	.29	.09	-.14	.08	1	-.18	-.19	.12	.23	
R9	.20	.30	.10	-.19	.34	.31	-.09	-.18	1	.03	-.26	-.09	
R10	.16	.44	.07	-.03	.13	.43	.07	-.19	.03	1	-.19	-.26	
R11	.03	-.05	.23	.29	.04	-.22	.22	.12	-.26	-.19	1	.04	
R12	-.03	-.26	.21	.31	.03	-.20	-.20	.23	-.09	-.26	.04	1	

Tab. 6		Głosy połączone N=34355 p < 0.01											
param.	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R13	
R1	1	.22	.27	.38	.33	.34	.43	.35	.23	.11	.02	-.03	
R2	.22	1	.03	.15	.52	.41	.20	.02	.23	.36		-.24	
R3	.27	.03	1	.16	.28	.34	.06	.11	.14	.12	.12	.07	
R4	.38	.15	.16	1	.08	.03	.35	.18	.10		.18	.16	
R5	.33	.52	.28	.08	1	.17	.15	.24	.10	.25	.03	-.09	
R6	.34	.41	.34	.03	.17	1	-.04	-.04	.39	.17	-.09	-.07	
R7	.43	.20	.06	.35	.15	-.04	1		-.02	.18	.08	-.07	
R8	.35	.02	.11	.18	.24	-.04		1	-.21	-.07	.21	.04	
R9	.23	.23	.14	.10	.10	.39	-.02	-.21	1	-.12	-.15	.12	
R10	.11	.36	.12		.25	.17	.18	-.07	-.12	1	-.20	-.18	
R11	.02		.12	.18	.03	-.09	.08	.21	-.15	-.20	1	-.10	
R12	-.03	-.24	.07	.16	-.09	-.07	-.07	.04	.12	-.18	-.10	1	

Tabela 7, 8, 9. Kwadraty odległości Mahalanobisa między populacjami.

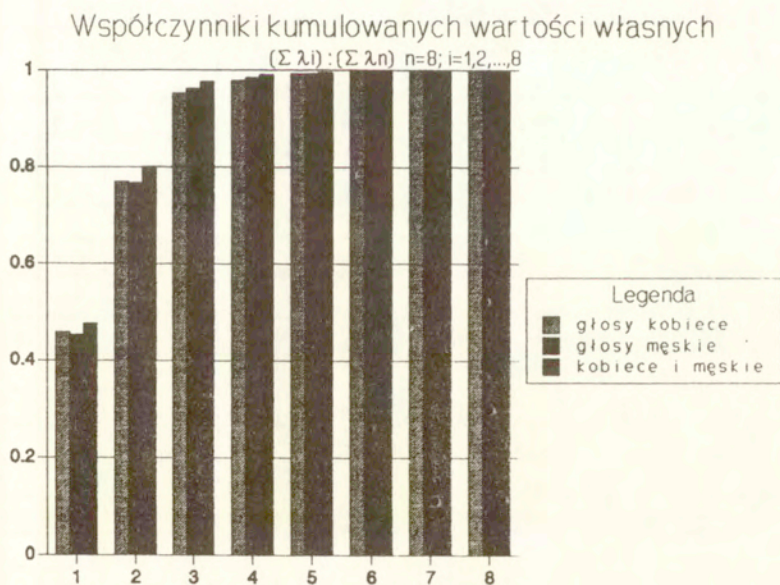
Tab. 7		Głosy kobiece								
Głoski	f	v	s	z	ϕ	ẓ	ʃ	ʒ	x	
f	0.00	18.08	9.42	14.25	5.07	6.53	5.49	5.55	3.12	
v	18.08	0.00	22.20	2.90	20.73	11.47	20.57	13.91	24.73	
s	9.42	22.20	0.00	11.21	9.27	8.99	8.54	8.40	15.46	
z	14.25	2.90	11.21	0.00	14.32	6.46	13.79	8.39	21.23	
ϕ	5.07	20.73	9.27	14.32	0.00	2.01	.76	2.00	14.03	
ẓ	6.53	11.47	8.99	6.46	2.01	0.00	2.60	1.39	15.28	
ʃ	5.49	20.57	8.54	13.79	.76	2.60	0.00	1.09	13.63	
ʒ	5.55	13.91	8.40	8.39	2.00	1.39	1.09	0.00	13.45	
x	3.12	24.73	15.46	21.23	14.03	15.28	13.63	13.45	0.00	

Tab. 8		Głosy męskie								
Głoski	f	v	s	z	ϕ	ẓ	ʃ	ʒ	x	
f	0.00	43.62	8.34	13.81	5.65	5.98	4.76	4.13	9.65	
v	43.62	0.00	46.80	21.11	50.27	37.58	48.23	37.60	57.65	
s	8.34	46.80	0.00	6.22	13.53	13.40	8.24	7.89	23.67	
z	13.81	21.11	6.22	0.00	19.42	14.04	14.10	9.61	29.29	
ϕ	5.65	50.27	13.53	19.42	0.00	1.44	2.03	3.20	26.82	
ẓ	5.98	37.58	13.40	14.04	1.44	0.00	2.93	2.38	26.54	
ʃ	4.76	48.23	8.24	14.10	2.03	2.93	0.00	.98	24.68	
ʒ	4.13	37.60	7.89	9.61	3.20	2.38	.98	0.00	22.87	
x	9.65	57.65	23.67	29.29	26.82	26.54	24.68	22.87	0.00	

Tab. 9		Głosy połączone								
Głoski	f	v	s	z	ϕ	ẓ	ʃ	ʒ	x	
f	0.00	23.47	7.64	12.76	4.28	4.93	3.86	3.80	4.75	
v	23.47	0.00	26.46	6.94	27.20	17.35	26.29	19.04	32.40	
s	7.64	26.46	0.00	7.83	8.51	7.99	6.12	6.15	17.08	
z	12.76	6.94	7.83	0.00	14.03	7.45	12.06	7.60	22.79	
ϕ	4.28	27.20	8.51	14.03	0.00	1.54	.45	1.40	16.70	
ẓ	4.93	17.35	7.99	7.45	1.54	0.00	1.45	.38	17.02	
ʃ	3.86	26.29	6.12	12.06	.45	1.45	0.00	.83	15.65	
ʒ	3.80	19.04	6.15	7.60	1.40	.38	.83	0.00	15.01	
x	4.75	32.40	17.08	22.79	16.70	17.02	15.65	15.01	0.00	

Tabela 10. Wartości własne macierzy kowariancji i iloraz kumulowanej sumy kolejnych wartości własnych do ich całkowitej sumy.

głosy	nr wart. -	1	2	3	4	5	6	7	8
kob.	w.włas.	2.04	1.39	0.81	0.12	0.06	0.03	0	0
kob.	wsp.k.	0.46	0.77	0.95	0.98	0.99	1	1	1
męs.	w.włas.	3.15	2.17	1.36	0.15	0.07	0.03	0.02	0
męs.	wsp.k.	0.45	0.77	0.96	0.98	0.99	1	1	1
poł.	w.włas.	2.15	1.46	0.79	0.07	0.02	0.02	0	0
poł.	wsp.k.	0.48	0.8	0.98	0.99	1	1	1	1



Rys.1 Analiza wartości własnych

Tabele 11,12,13. Wyniki klasyfikacji metodą analizy dyskrymin.

Tab. 11	Ni	f	v	s	z	ϕ	ϕ	∫	∫	x	popr
f	2387	58.1	2.9	1.6	0.3	13.8	0.1	3.1	0.8	19.3	58.1
v	1603	9.8	69.6	0.8	13.6	1.2	0.9	0.4	0.6	3.1	69.6
s	2168	6.5	0.2	69.4	0.2	5.6	0.2	7.0	0.2	1.2	76.7
z	1645	2.3	31.1	18.6	39.2	2.0	1.6	4.1	1.0	0.1	39.2
ϕ	2487	7.2	0.1	6.3	0.1	59.7	0.2	24.6	1.2	0.6	59.7
ϕ	1456	3.0	2.4	4.9	16.9	30.4	21.4	16.3	4.7	0.0	21.4
∫	2615	6.4	0.1	6.0	0.1	34.8	0.2	49.3	1.5	1.6	49.3
∫	1835	6.7	1.1	4.3	9.3	15.0	6.9	36.3	20.0	0.5	20.0
x	1987	24.3	0.5	3.3	0.5	5.3	0.0	4.9	0.1	61.2	61.2
Łącznie	18183										51.5

Tab. 12	Ni	f	v	s	z	ϕ	ϕ	∫	∫	x	popr
f	2056	58.3	1.41	5.06	0.73	8.81	1.22	12.4	0.49	11.5	58.3
v	1111	3.97	89.2	0	5.69	0	0.27	0.09	0.18	0.63	89.2
s	2393	7.53	0.17	81.5	0.88	3.85	1.46	4.1	0.21	0.29	81.5
z	1411	0.43	15.9	37.4	41.3	0.14	2.34	0.78	1.7	0	41.3
ϕ	2038	4.09	0.54	1.18	0.05	62.4	8.52	22.4	0.69	0.15	62.4
ϕ	1307	1.53	1.84	0.77	4.13	32.8	31.9	16.9	10.1	0	31.9
∫	2583	3.8	0.04	12.6	0.12	18.4	3.45	60.2	0.97	0.39	60.2
∫	1624	7.33	0.68	11.3	5.91	12.2	4.43	37.7	20	0.43	20
x	1649	18.2	0.59	2.03	0.72	1.38	0.39	4.27	0.66	71.8	71.8
Łącznie	16172										58

Tab. 13	Ni	f	v	s	z	ϕ	ϕ	∫	∫	x	popr
f	4443	53.7	2.25	3.13	0.36	10.2	0.05	13.4	0.43	16.5	53.7
v	2714	7.77	78.5	0.44	9.4	0.29	0.55	0.85	0.48	1.73	78.5
s	4561	7.47	0.18	74.4	0.44	5.61	0.09	10.8	0.29	0.7	74.4
z	3056	1.05	26	30.5	32.2	1.31	2	4.53	2.3	0.03	32.2
ϕ	4525	4.45	0.31	4.27	0.13	55.1	0.27	34.7	0.22	0.47	55.1
ϕ	2763	2.5	2.25	3.77	12.4	28.6	12.9	28.8	8.69	0	12.9
∫	5198	4.88	0.08	11.3	0.08	33	0.12	49.3	0.25	0.96	49.3
∫	3459	5.49	0.67	9.17	8.18	21.1	5.99	34.4	14.3	0.69	14.3
x	3636	20.9	0.56	2.78	0.53	5	0.12	5.62	0.18	64.3	64.3
Łącznie	34355										49.4

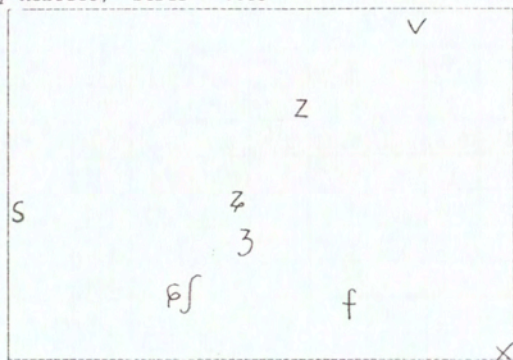
Tabela 14,15,16. Wyniki klasyfikacji metodą rzutowania obiektów

Głoska	Ni	f	v	s	z	ç	ç	ś	ż	x	
f	2387	58.1	3.6	8.4	0.1	4.1	0.1	1.7	2.5	21.3	58.1
v	1603	8	75.4	2.7	6.4	0.1	2.4	0	2.3	2.7	75.4
s	2168	8.5	1.1	81.2	0.7	2.2	0.1	4	0.7	1.7	81.2
z	1645	1	25.2	18.7	45.9	0.4	3.1	0.9	4.9	0.1	45.9
ç	2487	5.4	0.2	8.5	0	53.1	3.3	17.3	10.1	2.1	53.1
ç	1456	1.2	2.5	5.1	9.1	7.6	54.5	4.3	15.7	0.1	54.5
ś	2615	6.6	0.2	7.9	0	28.7	3.8	42.5	8.1	2.2	42.5
ż	1835	5.5	4.3	4.6	2.9	4.5	9.6	7.1	60.8	0.6	60.8
x	1987	18.4	1	5.5	0.3	1.7	0.5	2.9	1.8	68	68
Łącznie	18183										60.44

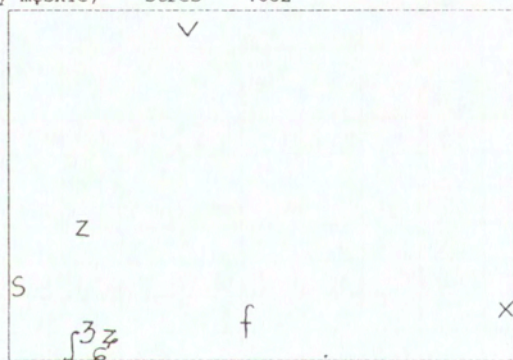
Głoska	Ni	f	v	s	z	ç	ç	ś	ż	x	
f	2056	69.6	1.5	3.3	0.5	3.2	0.4	6.7	3.4	11.4	69.6
v	1111	4.7	86.4	0.1	7.7	0	0	0	0.1	1.1	86.4
s	2393	8.7	0.2	78.4	1.7	2.2	1.2	6.3	0.8	0.6	78.4
z	1411	1.2	6.9	11.3	76.4	0	1.8	0.2	2.1	0.1	76.4
ç	2038	6.6	0.5	1.3	0	60.4	10.9	12.9	6.7	0.6	60.4
ç	1307	1.8	1.4	0.2	5.2	10	56.8	3	21.6	0	56.8
ś	2583	7.8	0	4.6	0.1	18.2	2	57.8	9.2	0.3	57.8
ż	1624	10.2	0.2	3.1	6.7	3.8	4.9	8.6	61.8	0.7	61.8
x	1649	10.6	0.5	0.5	0.4	0.3	0.4	4.1	1.2	81.9	81.9
Łącznie	16172										69.1

Głoska	Ni	f	v	s	z	ç	ç	ś	ż	x	
f	4443	69.4	2.9	6.6	0.3	3.1	0.4	2.5	1	13.9	69.4
v	2714	8.1	81.3	0.4	5.3	0	2.2	0.1	1.4	1.1	81.3
s	4561	12.6	0.7	72.5	4.6	1.3	0.1	6.3	0.9	1	72.5
z	3056	1.7	19.8	9.9	57.3	0	4.9	0.9	5.4	0	57.3
ç	4525	6.5	0.3	3.7	0.6	51.4	4.2	28.9	3.6	0.8	51.4
ç	2763	2	0.8	2.9	6	8.2	59.4	6.4	14.1	0.3	59.4
ś	5198	9.9	0.1	9.3	0.2	20.4	3.7	50.9	4.4	1.1	50.9
ż	3459	10.7	2	5.3	4.7	8	20.5	11.7	36.2	0.8	36.2
x	3636	20.7	0.9	2.3	0.3	1.7	0.6	5	0.7	68	68
Łącznie	34355										60.5

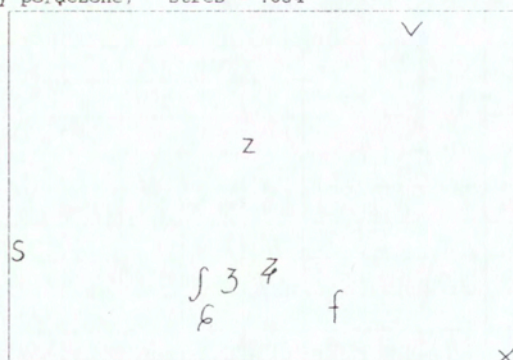
a) Głosy kobiece; stres = .035



b) Głosy męskie; stres = .062



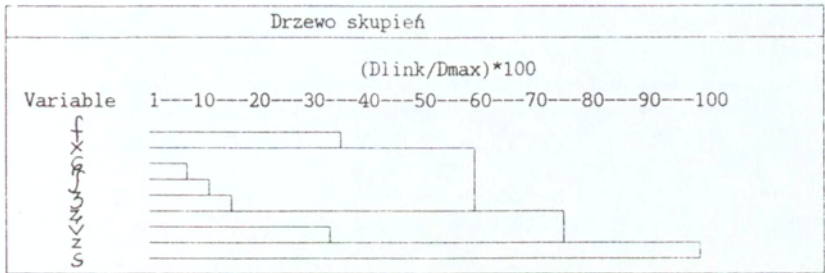
c) Głosy połączone; stres = .054



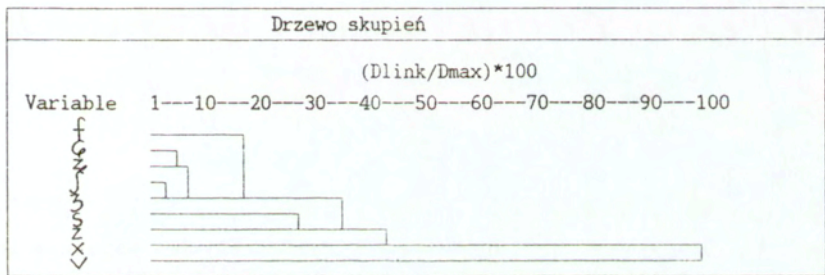
Rys. 2 Skalowanie wielowymiarowe dla głosek trących.



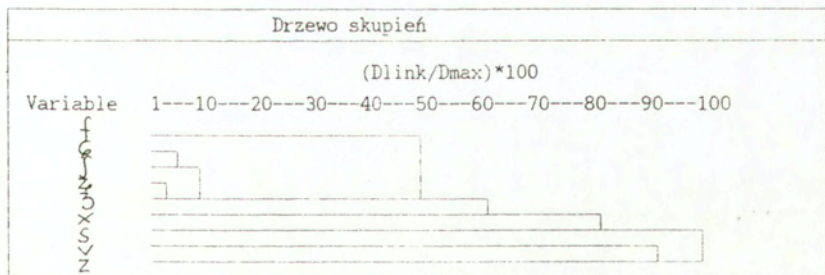
## a) Głosy kobiece



## b) Głosy męskie

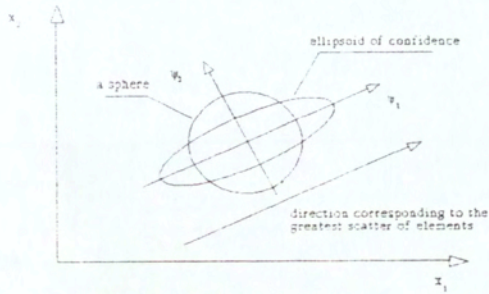


## c) Głosy połączone



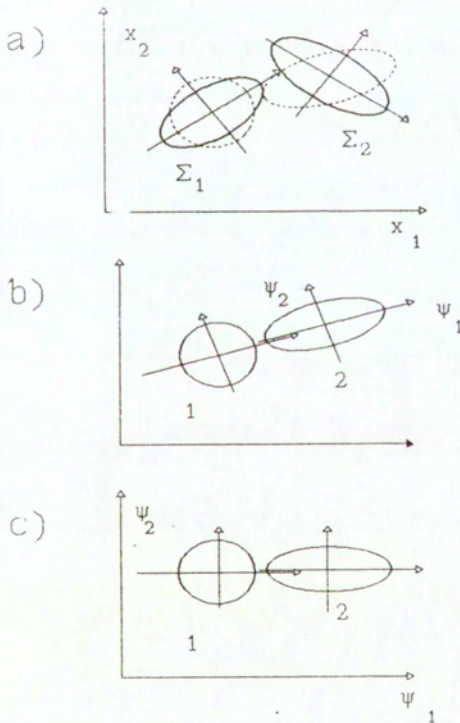
Rys. 3 Analiza skupień dla głošek trących.



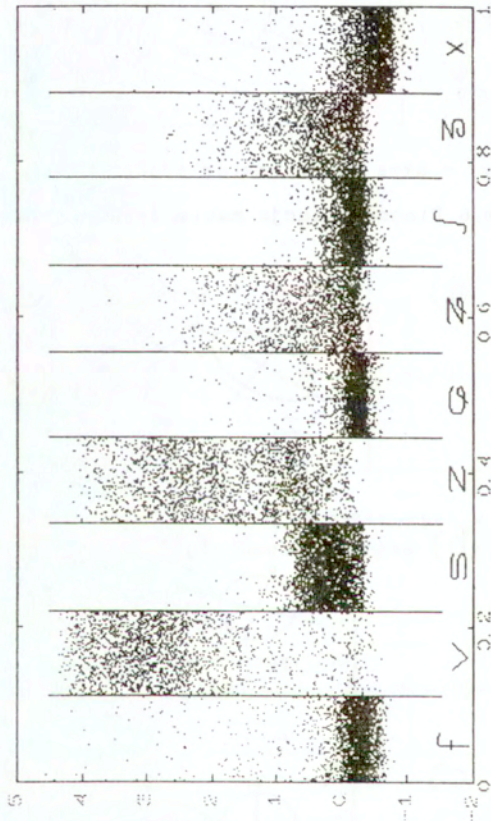


$x_1, x_2$  - axes of parameter space

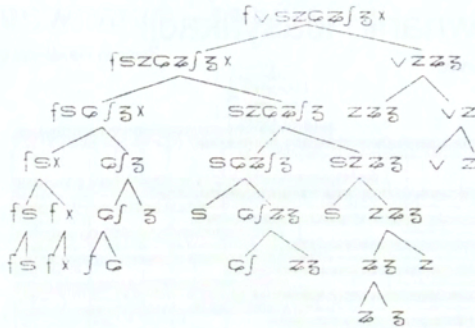
Rys.6 Diagonalizacja macierzy kowariancji.



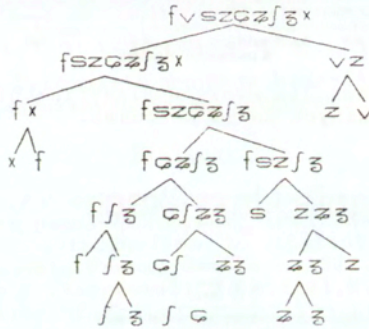
Rys. 7 Równoczesna diagonalizacja dwóch macierzy kowariancji.



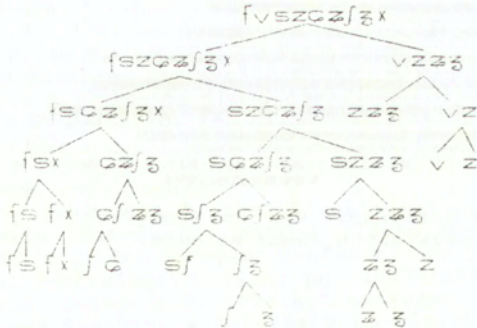
Rys.8 Rzut wszystkich obiektów na pierwszy wektor własny



Rys.9 Klasyfikacja głosek trących - g. kobiece.



Rys. 10 Klasyfikacja głosek trących - głosy męskie.

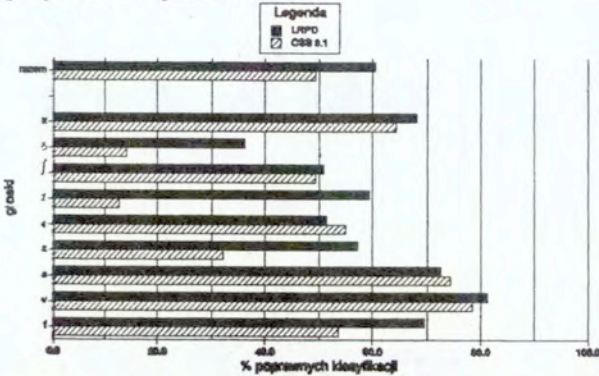


Rys. 11 Klasyfikacja głosek trących - głosy połączone.



# Porównanie klasyfikacji

głosy kobiece i męskie



Rys. 12a Porównanie wyników klasyfikacji głosek trących uzyskanych dwoma metodami.

## Bibliografia

- [1] Domagała, P., Automatyczne segmentalne rozpoznawanie wyrazów polskich, rozprawa doktorska, IPPT PAN, W-wa, 1991
- [2] Fukunaga, K., Introduction to Statistical Pattern Recognition, Academic Press, New York 1972.
- [3] Gubrynowicz, R., Kacprowski, J., Mikiel, W., Skalski, W., 1976, A Classification of Polish fricatives using the analysis of zero-crossings, in Speech Analysis and Synthesis, vol. 4, ed. W. Jassem, PWN, Warsaw.
- [4] Jassem, W., 1993, Discriminant Analysis of Continuous Consonantal Spectra, EUROSPEECH'93, Proceedings of 3rd European Conference on Speech Communication and Technology, Berlin, pp. 473-476.
- [5] Krzyśko, M., Analiza dyskryminacyjna, WNT, Warszawa, 1990.
- [6] Laface, P., de Mori, R., 1992, Speech Recognition Understanding - Recent Advances and Applications, NATO ASI series, vol 75.
- [7] Lobanov, B.M., 1971, On the Classification of Russian Fricatives in CV Syllables for Different Speakers, JASA, 49, 4, pp. 1324-1327.
- [8] Łobacz, P., 1985, Fonetyczno-leksykalne interakcje w percepcji mowy, Wyd. UAM, Poznań.
- [9] Richter, L., Domagała, P., 1993, Discrimination of Polish Stop Consonants Based on Mapped Techniques, EUROSPEECH'93, Proceedings of 3rd European Conference on Speech Communication and Technology, Berlin, pp. 1647-1650.
- [10] Richter, L., Domagała, P., 1993, Automatyczna klasyfikacja spółgłosek trących, w Materiały XL Otwartego Seminarium z Akustyki, Rzeszów-Polańczyk, pp. 325-328, Rzeszów.
- [11] Tatsuoka, M., Multivariate Analysis, John Wiley & Sons, New York 1971.