3.10.13. – akustyka mowy

L. Richter, P. Domagała

WARIANTYWNOŚĆ KONTEKSTOWA SPÓŁGŁOSEK TRĄCYCH BEZDŹWIĘCZNYCH JĘZYKA POLSKIEGO

7/1995

WARSZAWA 1995

ISSN 0208-5658

Praca wpłynęła do Redakcji dnia 25 października 1994 r.

Na prawach rękopisu

Instytut Podstawowych Problemów Techniki PAN
Nakład 100 egz. Ark. wyd. 1,5 Ark. druk. 2,0
Oddano do drukarni w lutym 1995 r.

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

Lutosława Richter
Piotr Domagała
Zakład Fonetyki Akustycznej
IPPT PAN

Wariantywność kontekstowa spółgłosek trących bezdźwięcznych języka polskiego[1]

Streszczenie

Przedmiot pracy stanowi zbadanie zmian zachodzących w widmach głosek trących bezdźwięcznych pod wpływem sąsiedztwa samogłoskowego oraz spółgłoskowego. Jako podstawę parametryzacji przyjęto odczytywane w równych odstępach wartości poziomu widma, ujęte w postaci wektora 25-elementowego. Dane pomiarowe poddano analizie dyskryminacyjnej, która pozwoliła określić stopień podobieństwa pomiędzy średnimi wektorami, reprezentującymi różne konteksty. Stwierdzono, iż istotne zmiany w widmach głosek trących wywołują sąsiadujące samogłoski, a zwłaszcza /u/ oraz /i/, co jednak dotyczy nie wszystkich spółgłosek. Spośród badanych trących głoska /x/ wykazuje największą podatność na alofonizację w kontekście samogłoskowym zaś najmniejszą w kontekście spółgłoskowym.

1. Wstęp.

Celem pracy jest określenie wpływu kontekstu samogłoskowego i spółgłoskowego na charakterystykę widmową głosek trących bezdźwięcznych języka polskiego (głosek o widmie ciągłym). Uzyskane wyniki pozwolą wnioskować o możliwości generowania przez poszczególne samogłoski oraz wybrane spółgłoski alofonów kontekstowych spółgłosek trących oraz o podatności poszczególnych spółgłosek trących na alofonizację. Stwierdzone warianty kontekstowe będą stanowić tzw. alofony wewnętrzne, to jest warianty powstające w wyniku procesów koartykulacyjnych. Alofony wewnętrzne, jako uwarunkowane fizjologicznie, są wyłączone spod świadomej kontroli mówiącego.

Liczne prace dotyczące głosek trących poświęcone są badaniom ich cech widmowych, będących efektem określonych uwarunkowań artykulacyjnych, a zwłaszcza cech związanych ze zróżnicowaniem poszczególnych spółgłosek, badaniu zmian zachodzących pod wpływem

[1]Praca wykonana w ramach grantu KBN 3 P401 002 05 (kierownik projektu prof. W.Jassem).

kontekstu fonetycznego oraz próbom klasyfikacji trących dla celów ich automatycznego rozpoznawania.

W porównaniu z dotychczas spotykanymi w literaturze opisami uwzględniającymi wpływ kontekstu samogłoskowego niniejsza praca prezentuje podejście kompleksowe - poszerza zakres badań o kontekst spółgłoskowy, a prezentowane wnioski opierają się na wynikach analizy statystycznej.

2. Charakterystyka fonetyczno-akustyczna głosek trących i próby ich klasyfikacji.

Klasę głosek trących polskich bezdźwięcznych w języku polskim tworzą głoski /f s ʃ ɕ x /. Źródło pobudzenia dla spółgłosek trących stanowi szum powstający u wylotu szczeliny utworzonej przez silnie zbliżone do siebie narządy mowy, będący rezultatem turbulentnego przepływu powietrza przez tę szczelinę. Podczas artykulacji /s ʃ ɕ/ przepływający strumień powietrza trze dodatkowo o krawędzie silnie zbliżonych do siebie zębów.

Widmo akustyczne głosek trących bezdźwięcznych ma postać ciągłą. Szczyty w widmie odpowiadają rezonansom efektywnie uczestniczącej w artykulacji części toru głosowego, która w przypadku trących zawarta jest między miejscem szczeliny a wargami. Charakterystyka widma uzależniona jest od umiejscowienia i rozmiarów szczeliny oraz od stopnia zbliżenia narządów mowy, na które kierowany jest strumień powietrza (zęby) i kształtu oraz długości wylotu z jamy ustnej (wargi).

Cechy widmowe spółgłosek trących pozwalają na wydzielenie trzech grup głosek, które pokrywają się z różnymi miejscami artykulacji: przednie (wargowe, do których zalicza się polskie /f/), środkowe (polskie przedniojęzykowo-dziąsłowe /s/ dziąsłowe /ʃ/, palatalne /ɕ/) oraz tylne (tylnojęzykowe /x/) [18].

Głoski przednie charakteryzują się "długim" widmem (5-6 kHz), ze słabo zaznaczonymi szczytami energii i najniższej w trzech grupach względnej intensywności [18]. Głoski środkowe posiadają "krótkie" widmo z jednym lub kilkoma wzmocnieniami w jego środkowej

części, przy czym główną koncentrację energii posiadają w wyższych rejonach częstotliwości niż pozostałe głoski trące, a względną intensywność najwyższą. Głoski tylne charakteryzuje widmo średniej długości z zaznaczoną strukturą formantową i pośrednią intensywnością.

Zamieszczony w pracy [7] opis głosek trących bierze pod uwagę rozkład częstotliwości formantowych F_2, F_3, F_4 (F_1 widoczny jest tylko w głosce /x/). Wyróżnia dwie klasy głosek w zależności od tego, czy odległość pomiędzy F_2 i F_4 jest większa (formanty rozproszone) czy mniejsza (formanty skupione) od 1800 Hz. Do pierwszej grupy zaliczają się głoski /f s/, do drugiej /ʃ ɕ/. Grupa druga dzieli się dodatkowo na dwie podgrupy zgodnie z przyjętym kryterium, że suma $F_2 + F_3 + F_4$ jest mniejsza (głoska /ʃ/) lub większa (głoska /ʃ/) od 8000 kHz. Autor analizuje również poziomy formantów wyróżniając na tej podstawie dalsze grupy trących.

W kolejnej pracy tegoż autora [8] zastosowano cztery parametry do opisu głosek trących : (1) zakres częstotliwości szumu, który pozwala wyróżnić trzy typy głosek, (2) względny poziom intensywności formantów, które są wyznaczane przez cztery pierwsze szczyty w widmie (przy czym nie wszystkie cztery formanty są widoczne we wszystkich trących), (3) częstotliwości formantów które wykazują związek z miejscem artykulacji głoski, (4) pierwszy moment widma (środek ciężkości widma), który zależy od rozkładu energii w całym zakresie analizy.

Szczegółowy opis głosek trących języka polskiego zamieszczono w pracy [12]. Podano wartości częstotliwości i względny poziom intensywności F_1, F_2, F_3 i F_4 oraz szczytów obwiedni widma w wyższych częstotliwościach. Badano wpływ kontekstu samogłoskowego na wartości F_2 oraz niektórych wyższych formantów (tam, gdzie było to możliwe).

Interpretacja szczytów obwiedni widma jako formantów właściwych dla głosek trących budzi pewne kontrowersje. Soli [17] twierdzi, iż szczyty przypadające w rejonach drugiego i trzeciego formantu samogłoskowego są efektem koartykulacji z następującą samogłoską. W trących wymówionych w izolacji są one bardzo słabo

zaznaczone, a stopień ich uwydatnienia w kontekście samogłoskowym zależy od rodzaju sąsiadującej samogłoski. W samogłoskach częstotliwość formantu drugiego określana jest przez rozmiar i kształt tylnej komory ustnej. W głoskach trących naturalne rezonanse tylnej komory są redukowane przez powstałe w jamie ustnej przewężenie. Obecność szczytów widmowych przypadających pomiędzy 1.5 i 2 kHz wskazuje zdaniem autora na pobudzenie rezonansu drugiego formantu w konsekwencji wcześniejszego otwarcia szczeliny przez następującą samogłoskę.

Akustyczną teorię produkcji spółgłosek trących stanowiącą punkt wyjścia do badań nad ich syntezą i percepcją przedstawiono w pracy [4]. Autorzy wyznaczyli wartości biegunów i zer funkcji transmitancji, pozwalające odtworzyć w przybliżeniu naturalne widmo głosek trących.

W pracach zawierających opis głosek trących zwraca się uwagę na wpływ kontekstu samogłoskowego na ich cechy widmowe. Wpływ ten zaznacza się wyraźnie w przypadku głoski /x/. Rozrzut wartości F_2 dla /x/ w sąsiedztwie różnych samogłosek wynosi ok. 1.5 kHz, dla pozostałych spółgłosek jest niewielki [6].

W pracy [12] badano wpływ kontekstu na częstotliwości formantów F_2, F_3, F_4 głosek trących bezdźwięcznych języka polskiego. Uzależnienie kontekstowe wykazują wszystkie trące, z tym że najczęściej uwydatnia się ono w przypadku formantu drugiego /x/. Rozrzut wartości dla trzech formantów wynosi przeciętnie w każdym z trzech uwzględnionych głosów około 500 Hz podczas gdy F_2 /x/ mieści się w zakresie ok. 1500 Hz. Sąsiedztwo samogłosek tylnych generalnie obniża wartości formantów. Zależność ta jest szczególnie konsekwentnie występuje w głosce /x/, gdzie częstotliwość formantów obniża się w kolejności: $[x_i] > [x_i] > [x_e] > [x_a] > [x_o] > [x_u]$.

Silne oddziaływanie samogłosek na poprzedzające spółgłoski trące, będące przejawem koartykulacji antycypacyjnej potwierdziły testy percepcyjne opisane w pracy [19]. Podjęto próbę identyfikacji usuniętej z sygnału samogłoski na podstawie odsłuchu poprzedzającego ją fragmentu spółgłoski. Wycięty fragment obejmował końcowy odcinek spółgłoski o długości 150 ms. Zmiany w widmach prewoka-

licznych trących nie były w jednakowym stopniu efektywne w różnych kombinacjach głoskowych. Np. /a/zostało poprawnie rozpoznane w 40%, /i/, /u/ w sąsiedztwie /z/, /s/ w 75%, w sąsiedztwie /ʒ/,/ʃ/ od 65% do 82%, z wyjątkiem /ʃi/-40%. Ogólnie dobrze rozpoznawane były wysokie samogłoski: 60 -80%, z wyjątkiem /ʃi/.

Szczególną rolę w procesie koartykulacji odgrywa zaokrąglenie warg występujące przed /u/, które powoduje przesunięcie głównej koncentracji energii widmowej głoski trącej do niższych częstotliwości oraz obniżenie częstotliwości jej drugiego formantu [5], [7], [17]. Zależność ta stała się przedmiotem badań w pracy [16] poświęconej relacjom pomiędzy stopniem koartykulacji a wiekiem mówiącego. Analiza spółgłosek w sylabach /si/, /su/ wykonywana w dwóch punktach czasowych segmentu szumowego: 20 ms oraz 70 ms od jego końca wykazała różnice pomiędzy alofonami w zakresie drugiego formantu (widmo końcowe) oraz maksimum widmowego (widmo środkowe). We wszystkich przypadkach wartości przed /i/ były wyższe niż przed /u/, różnicując się w sposób istotny statystycznie.

Stopień koartykulacji trącej z następującą samogłoską zdaniem autorów prac [14], [15] zmniejsza się z wiekiem, co świadczyłoby o wrodzonym charakterze tego procesu artykulacyjnego. Jako materiał badawczy posłużyły widma głosek trących /s/,/ʃ/ w kontekście samogłosek /i/, /u/. Stwierdzono konsekwentnie wyższe wartości F_2 zarówno /s/ jak /ʃ/ przed /i/ w porównaniu z /u/, przy czym stosunek $[s_i]$ do $[s_u]$ oraz $[ʃ_i]$ do $[ʃ_u]$ był wyższy w głosach dziecięcych.

Cechy koartykulacyjne głoski /s/ uwzględniono w badaniach nad automatycznym rozpoznawaniem fonemów z wykorzystaniem "samoorganizującej się mapy" cech akustycznych działającej na zasadzie sieci neuronowej [13]. Punkty na mapie reprezentują 25-elementowe wektory odpowiadające wartościom widmowym. Lokalizacja wektorów na mapie wykazuje ścisły związek z zaokrągleniem następującej samogłoski. Wektory dla próbek pochodzących z $[s_i]$ pokrywają się w znacznym stopniu z wektorami dla próbek $[s_a]$, natomiast zdecydowanie różnią się od $[s_u]$. Różnice zachodzące pomiędzy $[s_u]$ z jednej strony, a $[s_i]$, $[s_a]$ z drugiej strony są istotne statystycznie.

W oparciu o cechy widmowe głosek trących podejmowano próby ich klasyfikacji. Za pierwszą z nich można uznać pracę [5], w której opisano metodę różnicowania głosek /f s ʃ/ polegającą na porównywaniu poziomów dla określonych fragmentów widma. Oceniono rozkład energii w paśmie powyżej 4 kHz, w paśmie do 6500 Hz oraz w środkowych rejonach widma.

Metoda klasyf
åacji polskich głosek trących przedstawiona w pracy [11] opiera się na porównaniu względnej wielkości energii (pole powierzchni pod obwiednią widma) w rozłącznych zakresach częstotliwości. Rozpoznawanie przeprowadzono z zastosowaniem funkcji dyskryminacyjnych w przestrzeni wielowymiarowej. W pierwszym etapie cały zakres analizy dzielono na trzy pasma, przyjmując takie wartości graniczne między pasmami, które pozwalały sformułować zależności między średnimi wartościami energii w poszczególnych pasmach, różnicujące w jednoznaczy widma głosek. Np. wyznaczone dla określonego głosu wartości graniczne opisują /f/ zależnością $P_1 > P_2 < P_3 < P_4$, /s/ zależnością $P_1 < P_2 < P_3 > P_4$ itd. Wyniki rozpoznawania uległy znacznej poprawie po uwzględnieniu kontekstu samogłoskowego (/i/, /a/, /u/), gdy wartości graniczne wyznaczano oddzielnie dla każdego wariantu kontekstowego. Następnie wprowadzono ujednolicone dla wszystkich głosek wartości graniczne, dzielące zakres częstotliwości na równe podzakresy (od 4 do 12). Uwzględniając konteksty już przy czterech cechach osiągnięto w większości przypadków 100% poprawnych klasyfikacji.

W pracy [9] przedstawiono porównanie wyników klasyfikacji metodą podziału zakresu częstotliwości na równe podzakresy, przeprowadzonej dla dwóch sposobów parametryzacji - wielkości pola powierzchni pod obwiednią widma (średnia energia) oraz środka ciężkości widma. Wyraźny wpływ kontekstu samogłoskowego zaznacza się w odniesieniu do pola powierzchni pod obwiednią.

Analiza dyskryminacyjna została również wykorzystana do klasyfikacji głosek trących bezdźwięcznych języka angielskiego [2] oraz języka polskiego [10] w oparciu o momenty widma: pierwszy, trzeci i czwarty (środek ciężkości, skośność oraz spłaszczenie.

Dokładność rozpoznawania w obrębie jednego głosu bez uwzględnienia zróżnicowania kontekstowego wyniosła dla materiału polskiego od 66% do 76%.

Wszystkie wspomniane powyżej prace stosowały parametryzację w oparciu o cechy widmowe głosek. Odmienny sposób parametryzacji dla celów klasyfikacji głosek trących wykorzystano w pracach [1] oraz [3]. Zastosowanie pewnej kombinacji liniowej wartości funkcji autokorelacji sygnału [1] posłużyło do automatycznej klasyfikacji głosek trących języka polskiego na bazie optymalizacji przestrzeni parametrów. Klasyfikację przeprowadzono niezależnie od kontekstu fonetycznego oraz mówcy. Dokładność rozpoznawania w obrębie dziesięciu głosów (5 głosów męskich i 5 głosów kobiecych) z uwzględnieniem zróżnicowania kontekstowego wyniosła od 60% do 70%.

W pracy [3] parametryzacja sygnału oparta została na średniej gęstości przejść przez zero. Wyznaczono rozkłady prawdopodobieństwa tych wartości dla poszczególnych głosek trących bezdźwięcznych języka polskiego. Uwzględnienie kontekstu samogłoskowego ograniczyło rozrzut wartości parametrów zwiększając prawdopodobieństwo ich poprawnej klasyfikacji. Wpływ kontekstu nastąpił najwyraźniej w przypadku /x/ oraz /f/. Wartości przejść przez zero dla spółgłosek zwiększały się przy przechodzeniu sąsiadujących samogłosek od pozycji tylnej do przedniej.

3. Materiał doświadczalny.

Sporządzono listę logatomów o budowie VCV, w których spółgłoski trące bezdźwięczne /f s ʃ ɕ x/ występowały w kontekście symetrycznym, reprezentowanym przez samogłoski / i ɨ e a o u/ oraz logatomów o budowie VCCV, w których po spółgłosce trącej następowały spółgłoski /p t k m n ɲ /. W przypadku ograniczeń fonotaktycznych dotyczących połączeń samogłoskowo-spółgłoskowych, kontekst przyjmował postać niesymetryczną, np. /ɨɕi/, wówczas wpływ samogłoski /ɨ/ badano w początkowym fragmencie spółgłoski. Logatomy VCV zostały odczytane trzykrotnie, zaś VCCV pięciokrotnie przez trzy głosy męskie i trzy głosy kobiece.

4. Parametryzacja sygnału mowy.

Analizę widmową przeprowadzono z wykorzystaniem systemu do cyfrowej analizy sygnału mowy VOLYZER, współpracującego z komputerem IBM PC AT. Przyjęto zakres częstotliwości dla spektrogramów 50-8000 Hz oraz pasmo analizy równe 320 Hz. Znaczna zmienność kolejnych widm chwilowych obserwowana w widmach o charakterze ciągłym (szumowych) utrudnia ich interpretację, w związku z czym wykorzystano opcję pozwalającą na uzyskiwanie widm uśrednionych. Optymalny dla naszych celów zakres uśrednienia (spośród oferowanych przez system) obejmował 11 kolejnych widm, co wynosi ok. 25 ms. Środek analizowanego zakresu wyznaczało położenie kursora względem osi czasu.

Zasadniczy problem stanowił wybór sposobu parametryzacji sygnału dla celów analizy statystycznej zapewniający pełny obiektywizm pomiarów oraz kompletność danych wejściowych. Nie bez znaczenia było również uwzględnienie prostoty pomiarów. Tych warunków nie spełnia parametryzacjia w postaci częstotliwości formantowych. Często zdarza się, że nie wszystkie formanty widoczne są w głosce trącej - niektóre ulegają wygładzeniu przez antyformanty. Mogą wówczas pojawić się wątpliwości co do numeracji uwydatnionych formantów. W sytuacji, gdy analizowanie rozkładów maksimów w widmie może prowadzić do błędów interpretacyjnych, a uzyskane dane są niepełne, za właściwsze uznaliśmy posługiwanie się parametrem w postaci kształtu obwiedni całego widma. Zapamiętywany przez VOLYZER-a pomiar obwiedni w postaci 25-elementowego wektora, jako odczytu poziomu widma w równych odstępach, był zamieniany na plik tekstowy i w tej postaci stanowił daną wejściową do analizy statystycznej.

Dla każdej z badanych głosek w kontekście samogłoskowym uzyskano na ogół po pięć wektorów, charakteryzujących pięć kolejno po sobie następujących uśrednionych widm wyprowadzanych w odstępach 25 ms, z wyjątkiem alofonów w kontekście niesymetrycznym, dla których zapamiętywano trzy pierwsze widma. Z kolei dla trących w kontekście spółgłoskowym zapamiętywano trzy końcowe widma. W

rezultacie każda klasa głosek wymówionych w określonym kontekście przez daną osobę reprezentowana była przez 15 wektorów.

5. Analiza statystyczna.

Dane uzyskane z pomiarów poddano obróbce statystycznej za pomocą programu CSS Statistica z wykorzystaniem opcji ANALIZA DYSKRYMINACYJNA co pozwaliło określić, w jaki sposób zmienne dyskryminacyjne (roots), będące liniowymi kombinacjami zmiennych pierwotnych pozwalają odróżnić od siebie grupy obiektów.

Praktycznie wystarczającym przybliżeniem rozmieszczenia obiektów w przestrzeni zmiennych dyskryminacyjnych jest ich rzut na płaszczyznę utworzoną przaz dwie pierwsze zmienne, które mają największą moc dyskryminacyjną, zapewniającą dostatecznie czytelną interpretację graficzną. Wszystkie zmienne, uwzględniane w obliczeniach odległości Mahalanobisa, wyjaśniają w 100% zmienność między grupami.

Sporządzono wykresy zawierające rzut obiektów reprezentujących populacje głosek trących w różnych kontekstach (osobno samogłoskowych i spółgłoskowych) na płaszczyznę wyznaczoną przez dwie pierwsze zmienne dyskryminacyjne. Zdecydowanie skrajne położenie dowolnej klasy obiektów względem klas pozostałych może stanowić podstawę do wyróżnienia określonego alofonu kontekstowego. Z kolei znaczne nakrywanie się niektórych populacji skłania do traktowania ich jako reprezentacji tego samego alofonu.

Porównanie wykresów odnoszących się do kontekstu samogłoskowego pozwala stwierdzić, że najsilniejszy wpływ na kształt widma spółgłoski trącej wywiera samogłoska /u/. Wyodrębnienie się, w bardziej lub mniej zdecydowany sposób, klasy głosek sąsiadujących z /u/ obserwowane jest we wszystkich spółgłoskach z wyjątkiem /f/ (rys.1, 2). Porównanie średnich wektorów (rys.3, 4) wskazuje na przesunięcie energii widmowej do niższych częstotliwości w spółgłosce sąsiadującej z /u/ w porównaniu z pozostałymi realizacjami tej spółgłoski. Obserwowane zmiany w widmie spowodowane są zaokrągleniem warg i wysunięciem ich ku przodowi

podczas realizaji [s_u], [$ʃ_u$], [$ɕ_u$], [x_u]. Taki układ warg nie towarzyszy artykulacji [f_u], stąd na ogół brak w widmie zmian związanych z sąsiedztwem /u/. Słaby wpływ nieznacznego zaokrąglenia warg, jakie ma miejsce przy artykulacji /o/ uwidacznia się na wykresach poprzez częste zajmowanie przez populację /o/ pozycji pośredniej pomiędzy populacjami /u/ oraz pozostałymi realizacjami.

Stwierdza się silny wpływ samogłoski /i/ na spółgłoskę /x/ (rys.5) oraz nieco słabszy na /f/ (rys.6), objawiający się przesunięciem energii widmowej do wyższych częstotliwości (rys.7, 8). Zmiany w widmie [x_i] spowodowane są całkowicie odmienną konfiguracją toru głosowego, niż przy pozostałych realizacjach /x/ - w sąsiedztwie /i/ język wysklepia się silnie ku górze , przez co miejsce szczeliny tworzącej się pomiędzy grzbietem języka a podniebieniem przesuwa się znacznie ku przodowi jamy ustnej. W widmach /s/, /ʃ/, /ɕ/ nie stwierdza się zmian spowodowanych sąsiedztwem /i/, gdyż nieznaczne uniesienie języka nie ma wpływu na lokalizację szczeliny generującej szum, a efektywna część toru głosowego pozostaje niezmieniona. Z kolei podczas artykulacji [f_i] ma również miejsce silne wysklepienie języka ku górze, co wywołuje określone zmiany w widmie w postaci przesunięcia energii akustycznej do wyższych częstotliwości. W przypadku szczeliny wargowej nie jest więc obojętna konfiguracja toru głosowego na odcinku poprzedzającym miejsce tworzenia się szczeliny.

Spośród wszystkich głosek trących największą podatność na wpływ kontekstu samogłoskowego wykazuje spółgłoska /x/, co potwierdza wcześniejsze obserwacje [6, 12]. Na wykresach zdecydowanie wyodrębniają się realizacje w otoczeniu /i/ oraz w otoczeniu /u/, a w niektórych głosach zachodzi również wyraźne zróżnicowanie pozostałych populacji (rys.9, 10).

Analiza wykresów sporządzonych dla kontekstu spółgłoskowego nie wykazuje jego wpływu na widma trących. Jedynie w odniesieniu do niektórych głosek można zauważyć bardzo słabo zaznaczoną tendencję do grupowania blisko siebie obiektów reprezentujących konteksty /p/ i /m/, a więc o takim samym, wargowym miejscu artykulacji (rys.11). Generalnie jednak wszystkie populacje pokrywają się w stopniu

uniemożliwiającym ich rozróżnienie.

Miarą oceny stopnia podobieństwa między porównywanymi populacjami jest poziom istotności *p* dla błędu w razie przyjęcia hipotezy alternatywnej mówiącej, iż pomiędzy grupami obiektów w przestrzeni zmiennych dyskryminacyjnych zachodzi różnica istotna. W tabelach 1 oraz 2 zamieszczono dane dotyczące głoski /x/ w głosie KD oraz /s/ w głosie GD. W obu głosach wartości poziomów przy porównywaniu: spółgłoska-spółgłoska są wysokie (najczęściej $p > 0.05$), co wskazuje na nierozróżnialność populacji w kontekstach spółgłoskowych. Dobrze odróżnialne są konteksty spółgłoskowe od samogłoskowych - dla /x/ wartości *p* są w zasadzie zerowe, dla /s/ $p > 0.05$ tylko w nielicznych przypadkach. Przy porównaniu: samogłoska-samogłoska uzyskano duży rozrzut wartości *p* dla /s/ i niemal wszystkie wartości zerowe dla /x/. Gdyby przyjąć wartość progową dla $p < 0.05$, to w odniesieniu do /s/ wyróżniają się w sposób istotny od wszystkich pozostałych kontekstów tylko /o/ i /u/ (z jednym wyjątkiem *p* przyjmuje wartość mniejszą od 0.001), natomiast w odniesieniu do /x/ wyróżniają się wszystkie samogłoski (również niemal we wszystkich przypadkach $p < 0.001$).

Analiza wartości *p* dotyczących wszystkich populacji głosek trących potwierdza wnioski wysunięte w oparciu o wykresy:

1) najczęściej wyróżniane są konteksty z /u/
2) w mniejszym stopniu wyróżniają się konteksty z /i/ oraz /o/
3) najbardziej podatną na wpływy jest głoska /x/ - w odniesieniu do niej stwierdza się rozróżnialność wszystkich kontekstów samogłoskowych
4) kontekst spółgłoskowy nie odgrywa roli przy różnicowaniu realizacji głosek trących, pozwala jedynie odróżniać (i to nie zawsze) głoski wymówione w sąsiedztwie spółgłoskowym od głosek wymówionych w sąsiedztwie samogłoskowym.

Dla każdej głoski trącej w każdym z głosów przeprowadzono skalowanie wielowymiarowe, dzięki czemu uzyskano graficzną reprezentację rozmieszczenia w przestrzeni wielowymiarowej średnich wektorów, reprezentujących poszczególne konteksty. Odległości pomiędzy nimi wynikają z odległości Mahalanobisa między popula-

cjami. W kilku przypadkach redukcja przestrzeni do dwóch wymiarów okazała się zbyt silna (wysoki poziom stresu) uniemożliwiając przedstawienie w sposób czytelny i wiarygodny relacji pomiędzy populacjami, np. obiekty reprezentujące średnie wektory lokowały się na płaszczyźnie wzdłuż jednej osi, lub skupiały się niemal wszystkie wokół jednego punktu. W tych przypadkach wzięto pod uwagę trzy wymiary, co pozwoliło uzyskać wyniki skalowania łatwiejsze do interpretacji. Rezultaty skalowania wielowymiarowego zaprezentowano przykładowo na rys. 12 - 16. Dla wielu głosek zaobserwowano zróżnicowanie pomiędzy kontekstami samogłoskowymi potraktowanymi łącznie, a spółgłoskowymi - obiekty reprezentujące spółgłoski skupiają się w innych rejonach płaszczyzny, niż obiekty reprezentujące samogłoski. Nie stanowi to jednak reguły dla wszystkich głosek.

Zdecydowanie zaznacza się wpływ kontekstu samogłoskowego na wszystkie badane głoski trące, czego przejawem są określone konfiguracje średnich wektorów, odpowiadających poszczególnym kontekstom samogłoskowym. Zaobserwowano następujące zależności:

1) Niemal dla wszystkich badanych spółgłosek można poprowadzić prostą, która oddziela obiekty **i**, **ɨ**, **e** od obiektów **a**, **o**, **u**. Świadczy to o zróżnicowaniu pomiędzy widmami C_i, $C_ɨ$, C_e a C_a, C_o, C_u, a więc różnym oddziaływaniu kontekstów obejmujących samogłoski przednie oraz kontekstów obejmujących samogłoski tylne oraz /a/.
2) Można poprowadzić prostą oddzielającą obiekty **i**, **u** od obiektów **ɨ**, **e**, **a**, **o**, co świadczy o wyodrębnianiu kontekstów obejmujących samogłoski wysokie /i/, /u/ względem pozostałych kontekstów samogłoskowych.
3) Największe oddalenie od pozostałych wykazują obiekty **u** we wszystkich spółgłoskach z wyjątkiem /f/ oraz obiekty **i** w spółgłoskach /f/ oraz /x/, co potwierdza szczególny wpływ, jakie wywierają samogłoski /u/ oraz /i/ na wymienione spółgłoski.

W odniesieniu do kontekstu spółgłoskowego uzyskano potwierdzenie wcześniejszych wyników - generalnie nie wywiera on wpływu na

widma spółgłosek trących. Jedynie konteksty wargowe wykazują słabą tendencję do wyodrębniania się od pozostałych - reprezentujące je obiekty **p** i **m** w wielu spółgłoskach sytuują się blisko siebie, zajmując pozycję skrajną względem pozostałych kontekstów spółgłoskowych.

Wyniki skalowania wielowymiarowego pozwalają nieco skorygować wysunięty wcześniej wniosek o szczególnej podatności /x/ na wpływy kontekstu. Wykresy sporządzone dla głoski /x/ wykazują co prawda znaczne rozproszenie obiektów odpowiadających kontekstom samogłoskowym, lecz równocześnie silne skupienie obiektów reprezentujących konteksty spółgłoskowe. Taka konfiguracja świadczy o istnieniu znacznych różnic w widmach trących występujących w różnych kontekstach samogłoskowych, ale minimalnych różnic w widmach trących w różnych kontekstach spółgłoskowych.

6. Uwagi końcowe.

Przeprowadzona analiza statystyczna danych pomiarowych pozwoliła spojrzeć szerzej na zagadnienie uwarunkowań kontekstowych głosek trących bezdźwięcznych języka polskiego. Efekty widmowe koartykulacji trących z innymi głoskami w zasadzie zauważalne są tylko w kontekście samogłoskowym. O wyraźnej alofonizacji można mówić w przypadku, gdy samogłoską sąsiadującą z głoską /s/, /ʃ/, /ɕ/, /x/ jest /u/ oraz gdy samogłoską sąsiadującą z /f/, /x/ jest /i/. W pierwszym przypadku zmiany w widmie [s_u], [$ʃ_u$], [$ɕ_u$], [x_u] wywołane są zaokrągleniem warg i wysunięciem ich ku przodowi (labializacją) w trakcie artykulacji spółgłoski, możliwym do zrealizowania we wszystkich głoskach trących z wyjątkiem /f/. W drugim przypadku, podczas wymawiania [f_i] oraz [x_i] na skutek koartykulacji z /i/ język przybiera kształt tak dalece odmienny aniżeli przy pozostałych realizacjach tych głosek, iż zmodyfikowana konfiguracja toru głosowego znajduje swe odbicie w widmie.

Udział warg w procesach koartykulacyjnych odgrywa dużą rolę w przypadku głosek trących w kontekście z /u/, w pewnym stopniu w kontekście z /o/. Minimalny wpływ artykulacji wargowej daje się

zaobserwować nawet w kontekstach spółgłoskowych - widma C_p i C_n wykazują większe wzajemne podobieństwa, niż z pozostałymi widmami.

Stwierdzone zróżnicowanie widm głosek trących w różnych kontekstach samogłoskowych wykazuje powiązanie ze strukturą formantową sąsiadujących samogłosek. Rozmieszczenie średnich wektorów w przestrzeni dwu- lub trzywymiarowej uzyskane w wyniku skalowania wielowymiarowego, przeciwstawia sobie konteksty: samogłoska wysoka (niska wartość F_1) - samogłoska średnia lub niska (wysoka wartość F_1) oraz : samogłoska przednia (wysoka wartość F_2) - samogłoska tylna lub środkowa (niska wartość F_2).

Szczególną pozycję wśród głosek trących zajmuje /x/. Wykazuje największą podatność na alofonizację w kontekście samogłoskowym, za to najmniejszą podatność spośród pozostałych trących na alofonizację w kontekście spółgłoskowym.

Zaobserwowane uwarunkowania kontekstowe dotyczące głosek trących bezdźwięcznych, uwzględnione w systemach automatycznego rozpoznawania mowy oraz syntezy mowy, mogą okazać się istotne dla podniesienia ich efektywności.

Bibliografia.

[1] DOMAGAŁA,P.,RICHTER,P.,1994,Automatyczna klasyfikacja spółgłosek trących języka polskiego na bazie optymalizacji przestrzeni parametrów, Prace IPPT 9/1994,Warszawa

[2] FORREST,K.,WEISMER,G.,MILENKOVIC,P.,DOUGALL,R.,1988, Statistical analysis of word-initial voiceless obstruents, Preliminary Data,JASA 84,1,115-123.

[3] GUBRYNOWICZ,R.,KACPROWSKI,J.,MIKIEL,W.,SKALSKI,W.,1976 A classification of Polish fricatives using the analysis of zero-crossings, Speech Analysis and Synthesis,4,147-160.

[4] HEINZ,L.,STEVENS,K.,1961,On the properties of fricative

consonants,JASA,33,5,589-596.

[5] HUGHES,G.,HALLE,M.,1956,Spectral properties of fricative consonants, JASA,28,2,303-310.

[6] JASSEM,W.,1962,The formants pattern of fricative consonants, STL-QPSR,3,6-15.

[7] JASSEM,W.,1965,The formants of fricative consonants,Language and Speech,8,1,1-16.

[8] JASSEM,W.,1968,Acoustical description of voiceless fricatives in terms of spectral parameters, Speech Analysis and Synthesis,1,189-206.

[9] JASSEM,W.,1979, Classification of fricative spectra using statistical discriminant functions, in "Frontiers of Spech Communicartions Research" (B.Lindblom and S.Öhman, ed.), Academic Press London.

[10] JASSEM,W.,1993,Discriminant analysis of continuos consonantal spectra, Eurospeech'93,Proceedings of 3rd European Conference on Speech Communication and Technology,Berlin,pp.473-476.

[11] JASSEM,W.,SZYBISTA,D.,KRZYŚKO,M.,STOLARSKI,P.,DYCZKOWSKI,A., 1976, Rozpoznawanie polskich spółgłosek trących na podstawie cech widmowych,Prace IPPT 46/1976, Warszawa.

[12] KUDELA,K., Spectral analysis of Polish fricative consonants, Speech Analysis and Synthesis,1,1968,93-188.

[13] LEINONEN,L.,HILTRUNEN,T.,TORKKOLA,K.,KANGAS,J. Self organized acoustic features maps in detection of fricative-vowel coarticulation, JASA,93,6,1993,3468-3774.

[14] MC GOWAN,R,NITTROUER,S.,1988,Differences in fricative production between children and adults; evidence from acoustic analysis of /ʃ/ and /s/ ,JASA 83,1,229-236.

[15] NITTROUER,S.,STUDDERT-KENNEDY,M.,Mc GOWAN,R.,1989,The emergence of phonetic segments: evidence from the spectral structure of fricative - vowel syllables spoken by children and adults, Journal of Speech and Hearing Research, 32, 1, 120-132.

[16] SERENO,J.,BAUM,S.,MAREAN,G.,LIEBERMAN,P.,1987,Acoustic

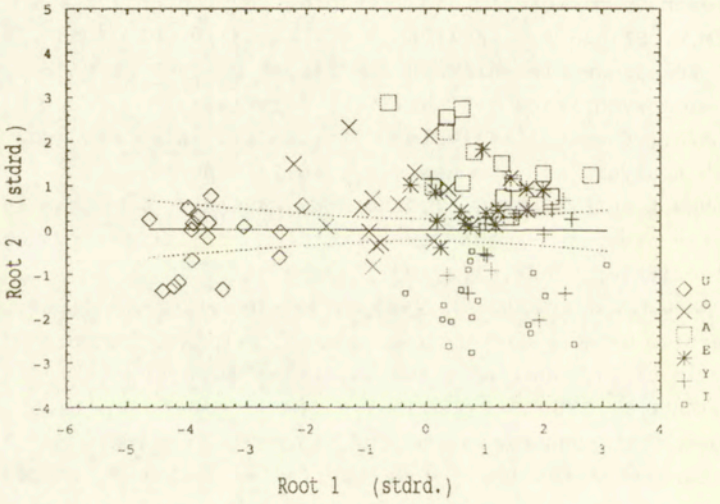
analyses and perceptual data on anticipatory labial coarticulation in adults and children,JASA,,81,2,512-519.

[17] SOLI,S.,1981,Second formants in fricatives: acoustic consequences of fricative-vowels coarticulation,JASA,70,4, 976-984.

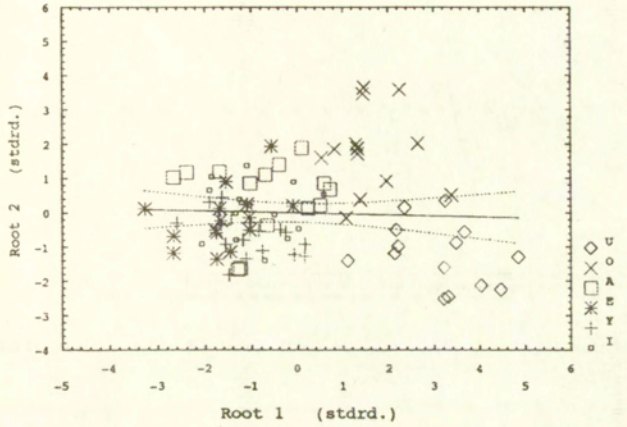
[18] STREVENS,P.,1960,Spectra of fricative noise in human speech, Language and Speech,3,1,32-49.

[19] YENI-KOMSHIAN,G.,SOLI,S.,1981,Recognition of vowels from information in fricatives: Perceptual evidence of fricative-vowel coarticulation,JASA,70,4,966-975.

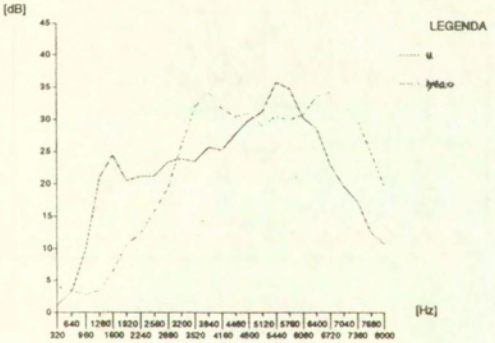
Rysunki.



Rys. 1 Rzut obiektów reprezentujących populacje głoski /ʃ/ w różnych kontekstach samogłoskowych. Głos LR

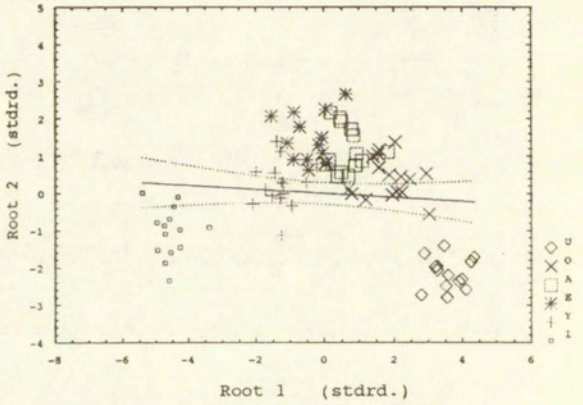


Rys. 2 Rzut obiektów reprezentujących populacje głoski /ɕ/ w różnych kontekstach samogłoskowych. Głos PD.

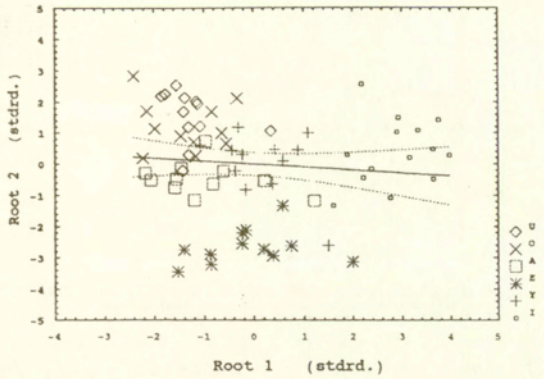


Rys. 3 Średnie wektory dla /ʃ/ w kontekście /u/ oraz w kontekście pozostałych samogłosek. Głos LR.

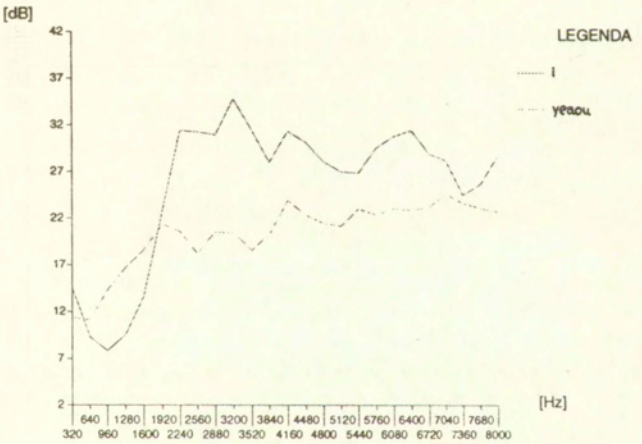
Rys. 4 Średnie wektory dla /ç/ w kontekście /i/, /a/ oraz /u/. Głos PD.



Rys. 5 Rzut obiektów reprezentujących populacje głoski /x/ w różnych kontekstach samogłoskowych. Głos MO.

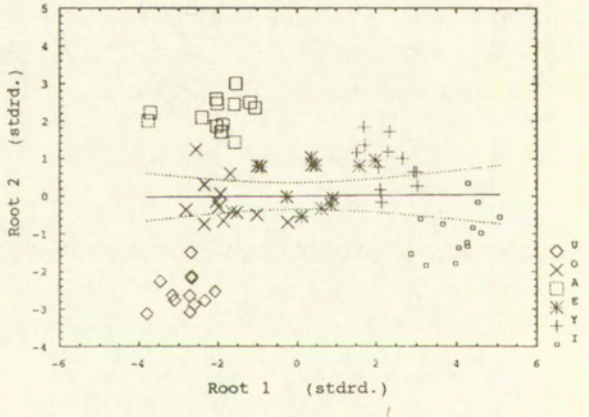


Rys. 6 Rzut obiektów reprezentujących populacje głoski /f/ w różnych kontekstach samogłoskowych. Głos KD.

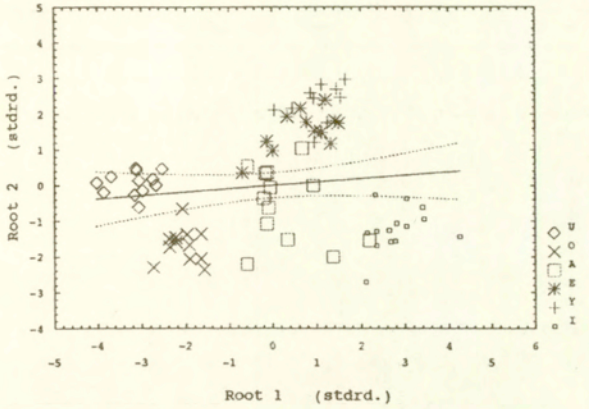


Rys. 7 Średnie wektory dla /f/ w kontekście /i/ oraz w kontekście pozostałych samogłosek. Głos KD.

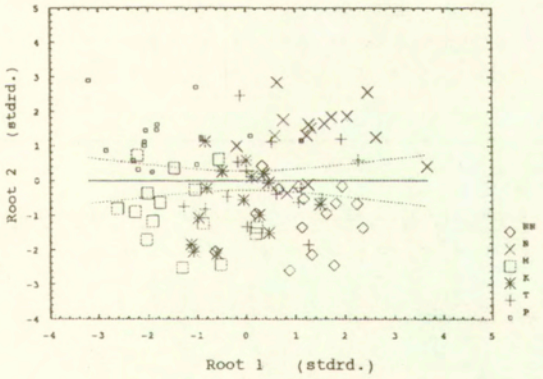
Rys. 8 Średnie wektory dla /x/ w kontekście /i/, /a/ oraz /u/. Głos MO.



Rys. 9 Rzut obiektów reprezentujących populacje głoski /x/ w różnych kontekstach samogłoskowych. Głos GD.

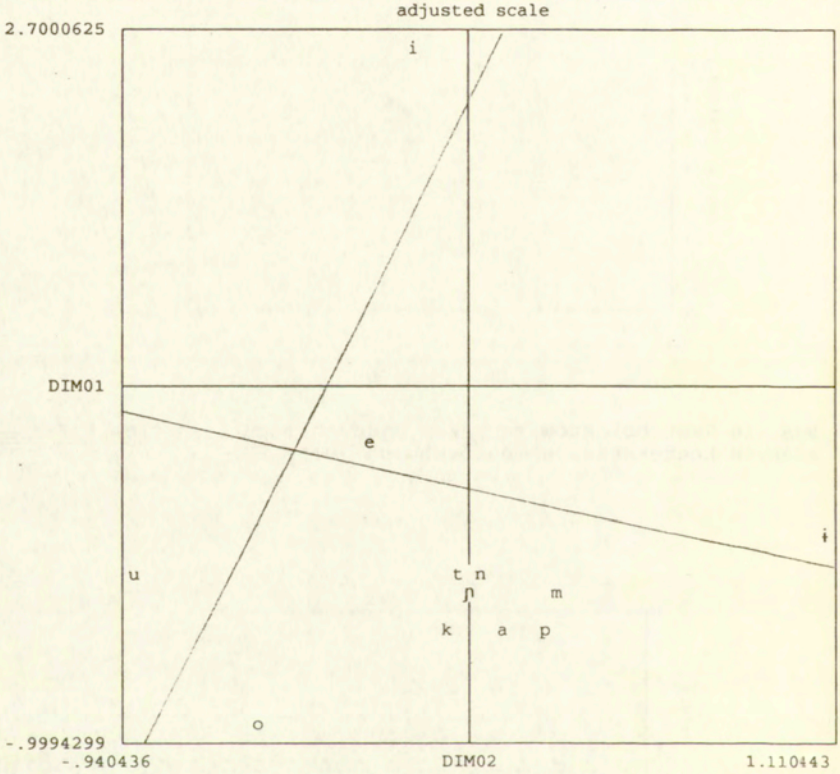


Rys. 10 Rzut obiektów reprezentujących populacje głoski /x/ w różnych kontekstach samogłoskowych. Głos KD.



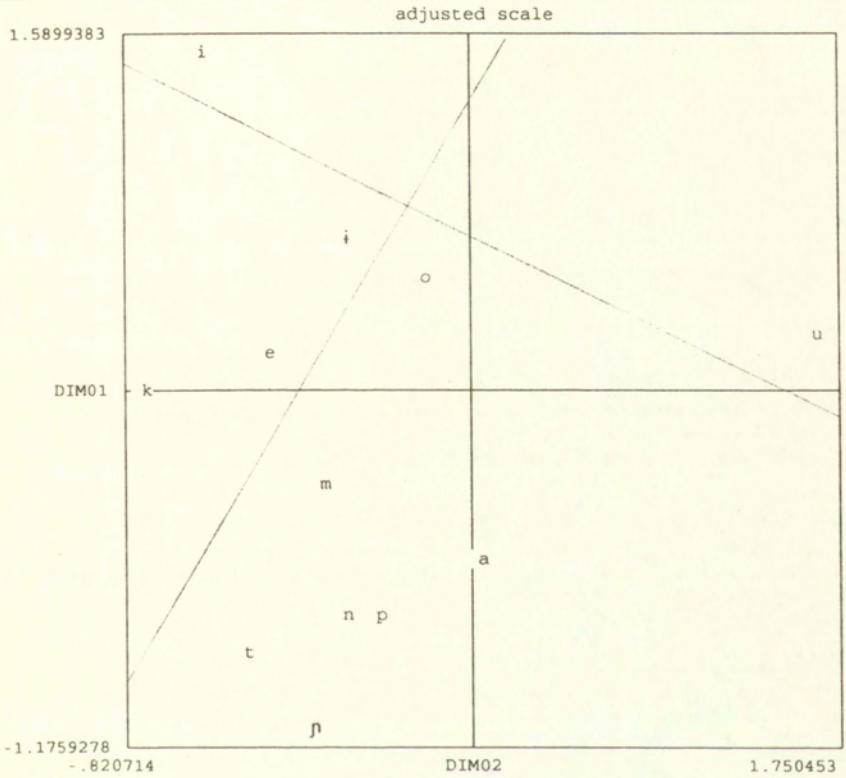
Rys. 11 Rzut obiektów reprezentujących populacje głoski /ʃ/ w różnych kontekstach spółgłoskowych. Głos IN.

css/3: multidim scaling	Final Configuration D-star: raw stress = 1.247090; alienation = .0929602 D-hat: raw stress = .4976065; stress = .0587844

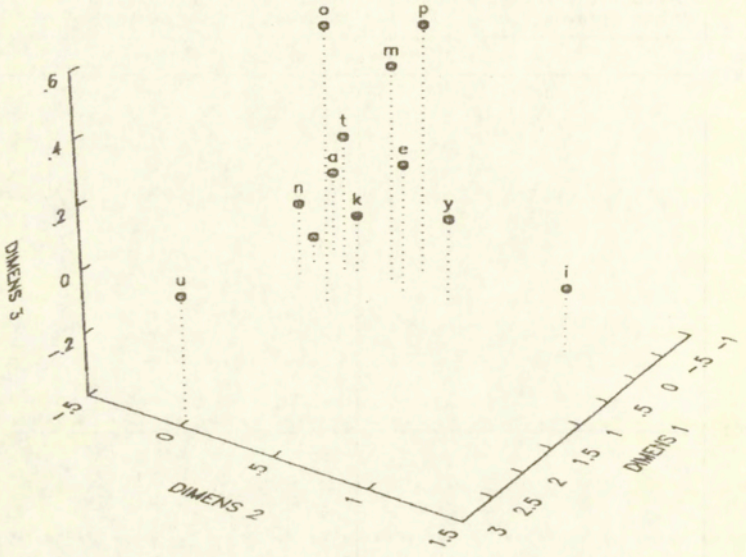


Rys.12 Wyniki skalowania wielowymiarowego. Głoska /f/, głos KD.

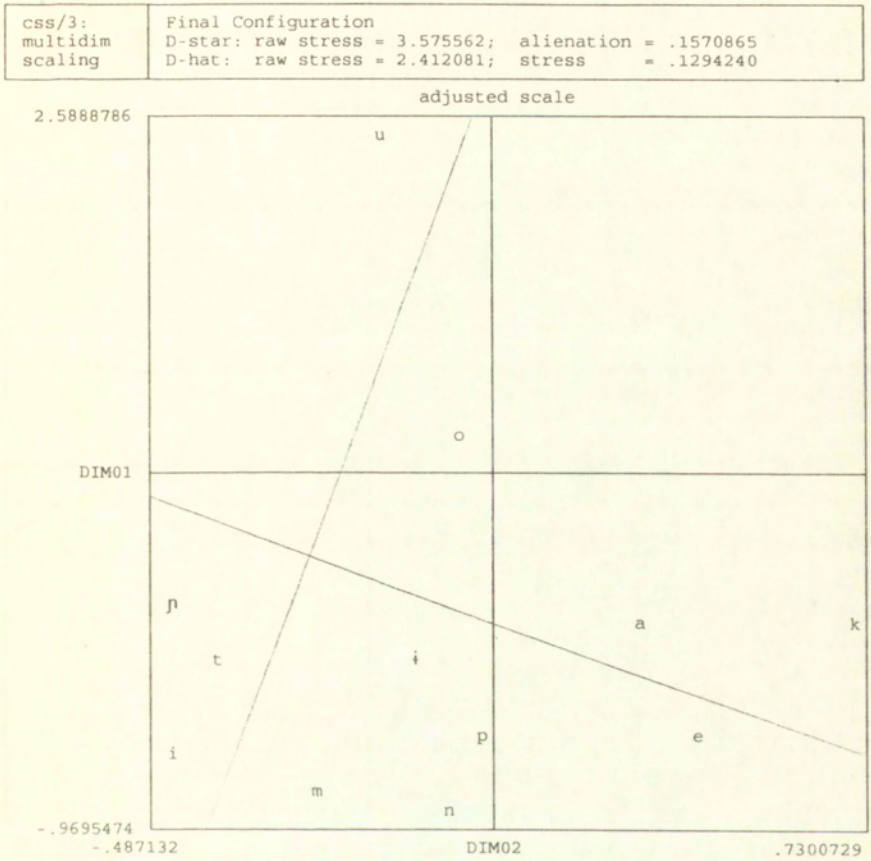
css/3: multidim scaling	Final Configuration D-star: raw stress = 1.445827; alienation = .1000763 D-hat: raw stress = .7053643; stress = .0699883



Rys.13 Wyniki skalowania wielowymiarowego. Głoska /s/, głos IN.

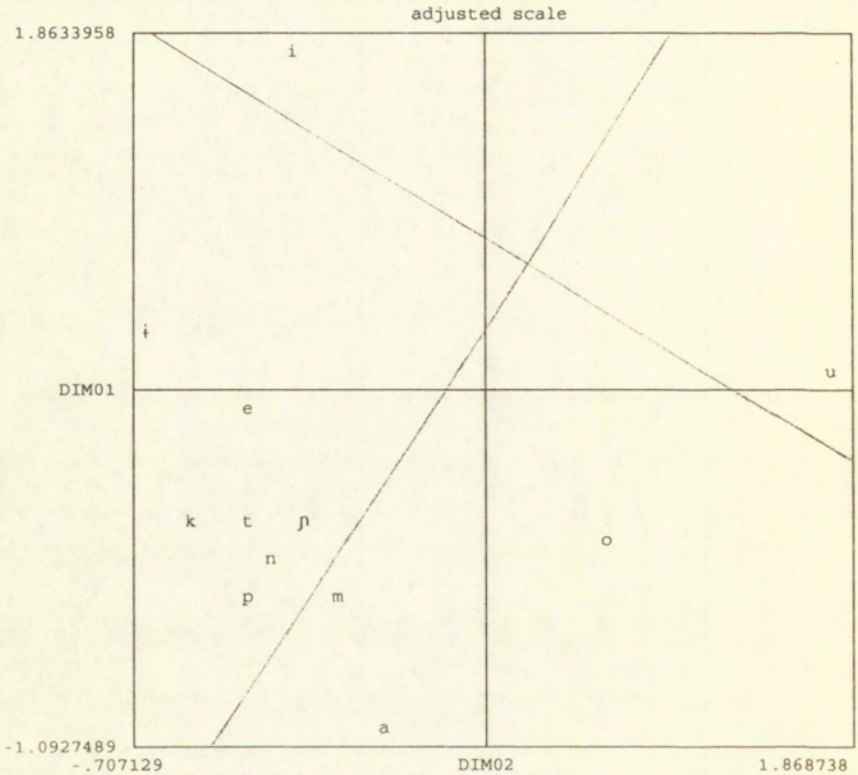


Rys.14 Wyniki skalowania wielowymiarowego. Głoska /ʃ/, głos LR.



Rys.15 Wyniki skalowania wielowymiarowego. Głoska /ɕ/, głos MO.

css/3: multidim scaling	Final Configuration D-star: raw stress = .6875154; alienation = .0690559 D-hat: raw stress = .3089119; stress = .0463165



Rys.16 Wyniki skalowania wielowymiarowego. Głoska /x/, głos GD.

Tab.1 Wartości prawdopodobieństwa p dla błędu przy założeniu braku istotnych różnic pomiędzy grupami. Głoska /x/, głos KD.

grupa	p	t	k	m	n	ɲ	i	ɨ	e	a	o	u
p		.112	.984	.822	.008	.010	.000	.000	.000	.000	.000	.000
t	.112		.031	.275	.055	.099	.000	.000	.000	.000	.000	.000
k	.984	.031		.849	.001	.003	.000	.000	.000	.000	.000	.000
m	.822	.275	.849		.098	.077	.000	.000	.000	.000	.000	.000
n	.008	.055	.001	.098		.056	.000	.000	.000	.000	.000	.000
ɲ	.010	.099	.003	.077	.056		.000	.000	.000	.000	.000	.000
i	.000	.000	.000	.000	.000	.000		.000	.000	.000	.000	.000
ɨ	.000	.000	.000	.000	.000	.000	.000		.032	.000	.000	.000
e	.000	.000	.000	.000	.000	.000	.000	.032		.000	.000	.000
a	.000	.000	.000	.000	.000	.000	.000	.000	.000		.000	.000
o	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000		.000
u	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	

Tab.2 Wartości prawdopodobieństwa p dla błędu przy założeniu braku istotnych różnic pomiędzy grupami. Głoska /s/, głos GD.

grupa	p	t	k	m	n	ɲ	a	e	i	o	u	ɨ
p		.050	.026	.052	.360	.085	.664	.007	.568	.000	.000	.044
t	.050		.004	.120	.877	.209	.139	.003	.084	.000	.000	.166
k	.026	.004		.152	.090	.436	.062	.000	.013	.000	.000	.000
m	.052	.120	.152		.232	.945	.551	.000	.002	.000	.000	.000
n	.360	.877	.090	.232		.613	.603	.070	.237	.000	.000	.036
ɲ	.085	.209	.436	.945	.613		.632	.009	.013	.015	.000	.007
a	.664	.139	.062	.551	.603	.632		.053	.500	.003	.000	.060
e	.007	.003	.000	.000	.070	.009	.053		.069	.000	.000	.076
i	.568	.084	.013	.002	.237	.013	.500	.069		.000	.000	.095
o	.000	.000	.000	.000	.000	.015	.003	.000	.000		.000	.000
u	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000		.000
ɨ	.044	.166	.000	.000	.036	.007	.060	.076	.095	.000	.000	