

Piotra Łobacz

**PERCEPCYJNA KLASYFIKACJA
SPÓŁGŁOSEK POLSKICH**

57 / 1979

P. 269



WARSZAWA 1979

Praca wpłynęła do Redakcji dnia 15 października 1979 r.

Zarejestrowana pod nr 57/1979



57189



N a p r a w a c h r ę k o p i s u

Instytut Podstawowych Problemów Techniki PAN

Nakład 140 egz. Ark.wyd.2,8, Ark.druk. 4.

Oddano do drukarni w listopadzie 1979 r.

Nr zamówienia 714/0/79

Warszawska Drukarnia Naukowa, Warszawa,
ul.Sniadeckich 8

Piotra Łobacz
Instytut Językoznawstwa
UAM

Praca wykonana w ramach problemu węzłowego 10.4.

PERCEPCYJNA KLASYFIKACJA SPÓŁGŁOSEK POLSKICH

Streszczenie.

Przedstawiono dwa eksperymenty psychofonetyczne, prowadzące do stworzenia percepcyjnej klasyfikacji spółgłoskowych fonemów języka polskiego. Wyniki klasyfikacji porównane zostały z deterministycznym opisem spółgłosek w kategoriach cech dystyngtywnych, sformułowanych bądź artykulecyjnie, bądź akustycznie.

Z punktu widzenia ogólnego modelu rozpoznawania mowy rozpatruje się sytuację, w której na wyjściu fonematycznego układu cech dystyngtywnych pojawia się równocześnie określona liczba znaków stanowiących cechy klasyfikacyjne.

Klasyfikacji dokonano za pomocą kilku metod statystycznych hierarchicznego grupowania.

1. Podstawy metodologiczne.

1.1. Założenia ogólne.

Badania współzależności, podobieństw, a także różnic pomiędzy przetworzeniem informacji językowej zawartej w sygnale mowy w układach technicznych z jednej strony i biologicznych z drugiej, mają na celu zarówno umożliwienie konstrukcji optymalnego wejścia fonicznego do komputera, jak również opracowanie adekwatnego modelu percepcji mowy przez człowieka. Na przestrzeni ostatnich dwudziestu pięciu lat przeprowadzono szereg eksperymentów w zakresie percepcji jednostek

Praca wykonana na zlecenie IPPT PAN nr 48.

segmentalnych i prozodycznych sygnału mowy. Te wycinkowe doświadczenia stanowiły następnie podstawę do formułowania teorii i budowania ogólnych modeli przetwarzania informacji językowej przez człowieka. Nie jest rzeczą przypadku, że najciekawsze teorie percepcji naturalnej sygnału mowy powstały w znanych ośrodkach telekomunikacyjnych, jak np. motoryczna teoria percepcji [24,25] czy model "analysis by synthesis" Stevensa, którego psycholingwistyczna interpretacja została następnie ogłoszona przez Chomsky i Halle [4].

W sformułowaniu metodologicznych założeń referowanych badań oparto się na ogólnym modelu tak biologicznego, jak i technicznego procesu rozpoznawania mowy [16]. Założenia te są następujące:

1. Percepcja sygnału mowy tak w blokach fonetycznego rozpoznania, jak i lingwistycznego przetworzenia ma charakter nielinearny.

2. Istnieje znacznie większa odpowiedniość pomiędzy dekodowaniem mowy przez człowieka a określonymi operacjami urządzenia technicznego w blokach lingwistycznego przetworzenia, niż w układach peryferyjnej i fonetycznej obróbki sygnału.

3. Przewiduje się rozpoznawanie sygnału mowy o rozciągłości co najmniej wyrazu.

4. Błędne zdekodowanie na płaszczyźnie fonetycznej może zostać skorygowane w blokach lingwistycznego przetworzenia.

5. Techniczny układ rozpoznający jest przystosowany do analizy sygnału mowy zarówno w kategoriach fonologicznych, jak i kategoriach cech dystynktywnych. Informacja na wyjściu bloku cech dystynktywnych może być od razu poddana analizie w układzie allomorficznym.

6. Na poszczególnych etapach dokonuje się rozpoznawania na zasadzie komparatora, porównującego dostarczoną informację ze zmagazynowaną zawartością układów pamięciowych.

Przeprowadzone doświadczenia psychofonetyczne miały na celu:

1. Ustalenie roli bloku cech dystynktywnych (DF) w procesie percepcji mowy poprzez próbę stworzenia klasyfikacji fonemów

spółgłoskowych w kategoriach percepcyjnych DF.

2. Zbadanie procesu korekty błędów fonematycznych w bloku wyrazowego przetworzenia informacji na podstawie zawartości pamięci długotrwałej, zgodnie z założeniem, że jeśli na wejściu układu leksykalnego pojawi się logatom, to aby na jego wyjściu pojawił się realny wyraz potrzebna jest poprawka polegająca na substytucji odpowiednich fonemów.

1.2. Cechy dystynktywne.

Sformułowane w 1952 roku przez Jakobsona, Fanta i Hallego [13], a zmodyfikowane w 1956 roku przez Jakobsona i Hallego [14] cechy dystynktywne stanowiły w sposobie kategorialnego opisu dźwięków mowy punkt zwrotny. Autorzy wprowadzili dwanaście cech (DF) sformułowanych binarnie. W fonematycznej teorii, którą sformułowali, DF nie oznaczały wprost parametrów klasyfikacyjnych, charakterystycznych dla określonych dźwięków, lecz minimalne jednostki lingwistyczne wchodzące w skład fonemów (por. [8]). Każda z cech zdefiniowana była najpierw akustycznie, a następnie "genetycznie", czyli artykulacyjnie. Nierzadko autorzy odwoływali się również do wrażeń słuchowych, co znalazło odbicie nie tylko w terminologii niektórych cech, np. *strident: mellow*, *grave: acute*, *flat: non-flat* itp., lecz także w komentarzach do opisu akustycznego. Mimo, że sami nie stworzyli adekwatnych definicji DF na płaszczyźnie percepcyjnej, to uważali opis cech w kategoriach postrzegania słuchowego za najważniejszy.

W licznych późniejszych pracach innych autorów, w których operowano pojęciem DF bądź jako najmniejszej jednostki lingwistycznej, bądź jako elementu klasyfikacyjnego dla dźwięków mowy różnych języków, zarysowują się trzy zasadnicze niezgodności, dotyczące:

1. liczby DF traktowanych jako cechy uniwersalne oraz ich terminologii,
2. binarności wszystkich cech oraz
3. podstawy ich definiowania (artykulacyjnej, akustycznej lub percepcyjnej).

Już w 1964 roku M. Halle w artykule [10] powiększył zbiór

uniwersalnych cech do 15, a Chomsky i Halle [4] (str. 299) wyliczają ich 25 (pominięto przy tym cechy prozodyczne). Ladefoged w "Preliminaries to Linguistic Phonetics" [20] wymienił 22 cechy (według jego terminologii "prime features"), ale w podręczniku z 1975 roku [21] uwzględnił ich 20. W pracy poświęconej głównym zagadnieniom percepcji mowy Stevens i House [38] stwierdzają, iż liczba cech uniwersalnych nie może być mniejsza niż 25 - 30.

Przystosowanie uniwersalnego inwentarza DF do specyficznych warunków określonych języków, nawet przy spełnieniu teoretycznego warunku minimalizacji ich liczebności, wymagało nierazkiedy wprowadzenia nowych cech lub zmian ich terminologicznego ujęcia w porównaniu z wersją pierwotną. Van den Broecke [2] zaprezentował w swojej monografii tablice (str. 43 - 44), zawierającą propozycje różnych autorów wprowadzenia nowej cechy albo zmiany już istniejącej. Łącznie zaproponowano kilkadziesiąt nowych cech.

Wspomniana wyżej rzekoma uniwersalność DF nie jest uzasadniona, albowiem przez uniwersalne rozumie się takie właściwości, które są wspólne wszystkim językom. Jest rzeczą oczywistą, że żaden język nie wykorzystuje wszystkich tych cech, lecz tylko niewielka ich część.

Wierni pierwotnej koncepcji binarności DF zostali w swych późniejszych pracach dwaj twórcy podstaw teoretycznych: Jakobson i Halle. Fant [6] dopuszcza różne możliwości definicji poszczególnych cech, jeśli zostanie zachowana główna zasada traktowania fonemu jako kombinacji cech. Ten sam autor w pracy [7] rozważa nie tylko binarne ujęcie cech, lecz także ich ekonomiczną liczebność i definicyjną (na płaszczyźnie akustycznej) niezależność. Sam pomysł powstania teorii binarnych DF jest związany z klasyczną logiką formalną, która jest oparta na dwóch możliwościach, jak również z matematycznymi korzyściami wynikającymi z algebry Boola (tworzenie kodów fonematycznych, wyznaczanie ich podstawowych miar teorio-informacyjnych).

Potrzebę binarnego opisu sygnału mowy widziano również ze względu na stworzenie metod przenoszenia i przetwarzania in-

formacji w urządzeniach telekomunikacyjnych działających na zasadzie bistabilności oraz maszynach cyfrowych operujących w systemach zero-jedynkowych lub pokrewnych (np. ósemkowym).

Teoretyczną zasadność binarnego charakteru niektórych DF podważył Ladefoged [20, 21]. Jego zastrzeżenia odnosiły się w szczególności do dwóch cech, które w prawie wszystkich językach świata stosowane są w sposób niebinarny, a mianowicie miejsca artykulacji spółgłosek oraz wysokości samogłosek. Maksymalna liczba opozycji minimalnych w jednym języku wynosi dla cechy miejsca 6 i odpowiednio dla cechy wysokości 4.

Przy tworzeniu systemów fonologicznych (fonematycznych) w kategoriach DF, stanowiących punkt wyjścia dla różnych doświadczeń percepcyjnych, jedni badacze stawiali postulat binarności wszystkich cech (np. Hanson [11], Mohr i Wang [26], Klatt [19]), inni wprowadzali obok cech binarnych również wielowartościowe. Miller i Nicely [27] na przykład wprowadzają trzystopniową skalę miejsca artykulacji: 0, 1, 2, Singh - zarówno w pracach indywidualnych [32], jak i zespołowych [35], [36], używa najczęściej czterech wartości. W systemie Wickelgrena [40] dzięki zastosowaniu pięciowartościowej cechy miejsca, 22-elementowy inwentarz fonemów spółgłoskowych amerykańskiej angielszczyzny został opisany za pomocą tylko pięciu DF. Zastosowanie zatem niektórych cech wielowartościowych prowadzi do tego, że system fonologiczny ma silniejszy związek z realną artykulacją, upraszczając przy tym cały system.

W "Preliminaries to Speech Analysis" [13] autorzy stwierdzają, że "... the specification of the phonemic oppositions may be made in respect to any stage of the speech event from articulation to perception and decoding..." (str. 13) i ustalają operacyjną hierarchię poszczególnych etapów: "... the operational hierarchy of levels of decreasing pertinence: perceptual, aural, acoustical and articulatory ... The systematic exploration of the first two of these levels belongs to the future and is an urgent duty." (str. 12).

Z punktu widzenia stanu badań w chwili obecnej, wypada stwierdzić, że najszersze zastosowanie przy opisach systemów

fonologicznych poszczególnych języków znalazły definicje DF w kategoriach artykulacyjnych.

Akustyczne cechy dystynktywne znalazły mniejsze zastosowanie. Znaczną trudnością przy ich formułowaniu było dopasowanie pomiarów zmiennych ciągłych do, z definicji skokowej, natury DF. Dotychczas nie wypracowano ogólnie przyjętego i odpowiednio prostego matematycznego modelu pozwalającego wyodrębnić inwarianty z danych kwantytatywnych.

Wielokrotnie podejmowano próby zdefiniowania DF na płaszczyźnie audytorycznej i percepcyjnej. Wykorzystywano w tym celu bardzo różne metodologie. W cytowanej monografii [2] van den Broecke w tablicy 39 (str.86), zebrał wszystkie znane sobie prace, których celem było ustalenie inwentarza percepcyjnych DF oraz ich współzależności. Z zestawienia tego wynika, że pierwszy tego typu eksperyment przeprowadzony przez Millera i Nicely w 1955 roku [27], polegający na ustaleniu matrycy błędów percepcji spółgłosek w różnych warunkach zakłócania szumem oraz zniekształcania w filtrach dolno i górnoprzepustowych znalazł kontynuatorów (por. np. Singh [33]). Oprócz badania percepcji dźwięków mowy w warunkach zniekształcenia sygnału wykonywano doświadczenia nad oceną podobieństwa głosek określonego języka w warunkach normalnego słyszenia (np. Hanson [11], Ahmed i Agrawal [1], Peters [24]), badano wpływ pamięci krótkotrwałej na ocenę audytoryczną sygnału mowy (np. Wickelgren [40], Klatt [19]), mierzono zależność DF od czasu reakcji (np. Cole i Scott [3], przeprowadzono również badania kontrastowe dla kilku języków (np. Singh i Black [35]).

Shepard [30] wykorzystał wyniki wcześniejszych prac cytowanych powyżej ([27], [29]) do zbudowania modelu hierarchii DF za pomocą skalowania wielowymiarowego - statystycznej metody grupowania, wykorzystywanej ostatnio w wielu badaniach o podobnym charakterze, np. w większości prac Singha i jego współpracowników, jak również w doświadczeniach van den Broecke [2].

Ustalenie inwentarza DF oraz ich wzajemnych współzależności jest w dużym stopniu uwarunkowane zarówno charakterem eksperymentu, jak również rodzajem analizy statystycznej.

Doświadczenia percepcyjne, zgodnie z tym, co stwierdzono poprzednio na płaszczyźnie badań artykulacyjnych, również podważają tezę o uniwersalności DF. Uzyskano różne jakościowo i ilościowo systemy cech dla jednego języka. Istnieją binarne systemy ustalonych a posteriori DF obok systemów niebinarnych.

Założenia teoretyczne twórców teorii DF dotyczące rangi percepcji w analizie fonemetycznej nie zostały, jak dotychczas, w pełni umotywowane eksperymentalnie. Niniejsza praca stanowi kolejną próbę, pierwszą przeprowadzoną na materiale języka polskiego, ustalenia, w jakim stopniu mechanizm dyskryminacji i identyfikacji segmentów opiera się na cechach dystyngtywnych.

1.3. Testy psychofonetyczne.

W większości doświadczeń percepcyjnych prowadzących bądź do ustalenia subiektywnego podobieństwa określonych elementów fonetycznych języka, bądź do sformułowania inwentarza percepcyjnych cech jednostek segmentalnych, materiał badawczy stanowiły izolowane dźwięki (najczęściej samogłoski, por. [3]), albo proste logotomy typu CV (np. [11,9,24]), VCV (np. [12]) i CVC (np. [1]). Nierzadko element nie będący bezpośrednim przedmiotem badań był dla wszystkich logotomów danego testu stały. Dobór tak ograniczonego zbioru jednostek językowych lub kwazi-językowych znacznie odbiega od warunków lingwistycznego przetwarzania elementów fonetycznych w warunkach percepcji naturalnej mowy ciągłej. Dobór ten był jednak konieczny jako eliminujący w znacznym stopniu czynniki wywierające wpływ na percepcje jednostek segmentalnych, które nie były przedmiotem bezpośredniego badania.

Przeprowadzono również wiele doświadczeń nad wyrazistością elementów fonetycznych w różnych warunkach przenoszenia sygnału, np. przy jego zmniejszaniu i maskowaniu (por. [9], [11] i in.). Celem tych eksperymentów było ustalenie, które cechy fonetyczne badanych dźwięków mowy są najbardziej odporne na różnego rodzaju zakłócenia w torach telekomunikacyjnych. Są to jednak warunki zbyt specyficzne, by można je było w sposób uogólniający wykorzystać do opisu modelu przetwarzania sygnału mowy w płaszczyźnie fonetycznej.

W niniejszej pracy przedstawiono dwa doświadczenia psycho-

fonetyczne, nie wymagające żadnego przygotowania ze strony słuchaczy. Pierwsze polegało na subiektywnej ocenie podobieństwa między wymówionymi dźwiękami przy odwołaniu się do pamięci prekategorialnej akustycznej oraz krótkotrwałej fonemicznej. Sygnały nie były w żaden sposób zniekształcane ani zakłócone.

Zastosowano jedną z odmian tzw. metody trójkątowej: ABX. Zadaniem słuchaczy było określenie, które dwa z trzech kolejno nadawanych sygnałów są do siebie bardziej podobne niż pozostały!/^{1/} Badane spółgłoski wymówione zostały w symetrycznym kontekście samogłoski /e/. Jeden bodziec składał się z trzech dwusylabowych logatomów (np. /ere, ese, ete/). Uzyskana na podstawie powyższego doświadczenia tablica odległości (podobieństwa) między badanymi spółgłoskami stanowiła punkt wyjścia przy opracowaniu eksperymentu drugiego, którego celem było zbadanie podobieństwa tych samych elementów segmentalnych mowy, co w teście I, ale w warunkach bardziej zbliżonych do tych, w jakich zachodzi normalna łączność w układzie człowiek - człowiek za pomocą mowy.

Deformacje aktualnego (rzeczywistego) sygnału, która tutaj zastosowano, określamy jako "zakłócenie czynnikiem lingwistycznym". Test logatomowo-wyrezowy, specjalnie opracowany dla celów doświadczenia, odwoływał się bądź do nieuświadamianej, bądź do częściowo uświadamianej wiedzy słuchaczy o ich języku.

Wyrez polski będący reakcją słowną na podany logatom, z założenia doświadczenia, różnił się tylko spółgłoską nagłosową lub spółgłoską interwokaliczną od sygnału nadanego. Testowano podobieństwo między nadanymi a substytuowanymi spółgłoskami w związku z subiektywnym, poprzednio ustalonym, prawdopodobieństwem występowania wyrazu stanowiącego reakcję. Przedmiotem analizy było tutaj rozstrzygnięcie, w jakim stopniu na ko-

1/ Metoda ABX została nazwana przez Thürmana [39] metodą pośrednią, ponieważ oceny słuchaczy mają charakter kategoryalny i nie opierają się na bezpośredniej skali wartościowania porównanych obiektów.

jarzenie wpływało podobieństwo subiektywne substytuowanych głosek do głosek nadanych, ustalone na podstawie poprzedniego testu, a w jakim subiektywna częstość wywołanego wyrazu.

Tego rodzaju test logotomowo-wyrazowy wprowadzał następujące czynniki lingwistycznego zakłócenia:

1. subiektywna częstość występowania wyrazu autentycznego w języku,
2. subiektywna częstość występowania wyrazu w ograniczonym eksperymentalnie zbiorze, z którego dokonywano wyboru,
3. częstość subiektywna formy gramatycznej kojarzonego wyrazu,
4. częstość występowania fonemu zastępowanego,
5. częstość fonemu substytuowanego,
6. kontekst samogłoskowy,
7. pozycja fonemu zastępowanego w logotomie i wyrazie (nagłos, śródgłos),
8. osobnicza łatwość kojarzenia usłyszanego sygnału z wyrazem występującym w języku,
9. osobnicza skłonność słuchacza do używania niektórych fonemów częściej niż pozostałe, niezależnie od obiektywnej częstości ich występowania w języku.

Test został tak skonstruowany, by wymienione wyżej zakłócenia można było scharakteryzować ilościowo.

W obu testach psychofonetyczne podobieństwo między spółgłoskami wyznaczono operując pojęciem odległości, rozumianej jako wartość probabilistyczna. Zbiór wzajemnych odległości między wszystkimi elementami stanowił podstawę do wykrycia "skupień" tych elementów. Uzyskane pogrupowania (kilkoma metodami statystycznymi hierarchicznego grupowania) potraktowano jako klasyfikacje spółgłosek w płaszczyźnie percepcyjnej i porównano je z akustycznymi i artykulacyjnymi podziałkami w kategoriach DF.

2. Test logotomowy.

2.1. Konstrukcja i przebieg testu.

Przyjęty za Jassemem [15] fonematyczny system języka polskiego zawiera 29 spółgłosek. Przedmiotem doświadczenia psycho-

fonetycznego był zbiór 28 fonemów spółgłoskowych.^{1/} Gdyby się uwzględniło każdą możliwą kombinację trzevelementową dla takiego zbioru, otrzymano by razem do zaprezentowania słuchaczom 3376 bodźców. Biorąc pod uwagę, że każdy słuchacz powinien dokonać oceny podobieństwa dla wszystkich bodźców, liczba ta okazała się zbyt duża. Postanowiono więc znacznie ograniczyć liczbę bodźców z punktu widzenia przyjętego a priori systemu cech dystynktywnych polskich fonemów.

Istnieją w literaturze przedmiotu dwa systemy fonemów polskich scharakteryzowanych w kategoriach DF. Pierwszy, opracowany przez Jassema [15] dla potrzeb takich wdrożeń, które oparte są na teorii informacji, wykorzystuje binarne, akustyczne definicje cech. Drugi, sformułowany przez Laskowskiego [23], z punktu widzenia założeń fonologii generatywnej, zawiera artykułacyjny, również binarny zespół cech.

System akustycznych DF jest co prawda bardziej ekonomiczny, jako że zawiera tylko 9 cech, jednak z uwagi na znaczną hierarchiczną współzależność pomiędzy cechami oraz fakt, że w świetle niektórych teorii percepcji naturalnej sygnał mowy jest dekodowany raczej w relacji do artykulacji niż do właściwości fizycznych, nie został wykorzystany w niniejszej pracy jako system a priori. Zrezygnowano również z inwentarza cech zaproponowanego przez Laskowskiego [23], jako że jest on zbyt rozbudowany (trzyście DF). Skonstruowano inną artykułacyjną macierzę cech dystynktywnych fonemów języka polskiego. Uwzględniono w niej 10 następujących cech (por. Tablica 1) : (1) spółgłoskowy, (2) nosowy, (3) zwarty, (4) zwarto-trący, (5) przedni, (6) peryferyjny, (7) przymknięty, (8) retrofleksyjny, (9) krótki oraz (10) dźwięczny. Liczba cech potrzebnych do rozróżnienia poszczególnych par fonemów wahała się od trzech do sześciu.

Do analizy percepcyjnej spółgłosek wybrano a priori tylko

^{1/} Fonem /ŋ/ został pominięty ze względów fonotaktycznych. Posiada on mianowicie bardzo ograniczoną dystrybucję kontekstualną.

niektóre cechy dystynktywne, a mianowicie: (1)zwarta, (2) przednia, (3) peryferyjna, (4) dźwięczna. Cpozycja zwarta:niezwarta, dla potrzeb niniejszego eksperymentu, różnicuje następujące dwie klasy fonemów spółgłoskowych:

zwarte: /p, b, t, d, c, ʒ, k, g, t̪s, d̪z, t̪j, d̪j, t̪ɕ, d̪ʒ, r/
niezwarte: /f, v, s, z, ʃ, ʒ, ɕ, ʑ, x, m, n, ɲ, l/.

Tego rodzaju podział umożliwił przebadanie całego inwentarza spółgłosek polskich. Z kolei całość 28 elementów pogrupowano na 12 podklas według następujących kryteriów:

- | | |
|----------------------------|---------------------------------|
| 1. przedni dźwięczny | 7. nieprzedni peryferyjny |
| 2. przedni bezdźwięczny | 8. nieprzedni nieperyferyjny |
| 3. przedni peryferyjny | 9. peryferyjny dźwięczny |
| 4. przedni nieperyferyjny | 10. peryferyjny bezdźwięczny |
| 5. nieprzedni dźwięczny | 11. nieperyferyjny dźwięczny |
| 6. nieprzedni bezdźwięczny | 12. nieperyferyjny bezdźwięczny |

I tak np. w podklasie 1. znalazły się następujące fonemy zwarte: /b, d, dz, d̪z, r/ oraz następujące niezwarte: /v, z, ʒ, m, n, l/.

Podklasy nie są rozłączne. W obrębie każdej podklasy utworzono triady spółgłosek w ten sposób, aby zawsze dwie zwarte były porównywane z jedną niezwartą. W podklasie 1. utworzono łącznie 60 triad (10 dwuelementowych połączeń spółgłosek zwartych skojarzonych z jedną z 6 spółgłosek niezwartych). Podobnie postąpiono w pozostałych podklasach. Jeśli jakaś triada wystąpiła w kilku podklasach, wykorzystano ją w ogólnym zestawie bodźców tylko raz. Łącznie uzyskano 501 triad. Przy takim przygotowaniu materiału porównawczego założono z góry, że nie będzie można bezpośrednio określić podobieństwa pomiędzy elementami klasy fonemów niezwartych tj. trących i nosowych. Tak z akustycznego jak i artykulacyjnego punktu widzenia są to fonemy monosegmentalne. Wysunięto hipotezę, opierając się na podobnych pracach dotyczących języka angielskiego, że dla słuchacza grupa spółgłosek /s, z, ʃ, ʒ/ będzie stanowiła zbiór elementów o znacznym wzajemnym podobieństwie. Założono również, że grupa spółgłosek nosowych jest odróżniana od pozostałych fonemów poprzez cechę nosowości.

Częstość występowania poszczególnych elementów spółgłoskowych w triadach była różna (uwarunkowana doborem cech dystyngujących).

Każdy z elementów został wymówiony w symetrycznym kontekście samogłoski /e/ np. /ere, ete, ese/. Takie trzy dwusylabowe logatomy stanowiły jeden bodziec. Obie sylaby logatomów wymówiono w przybliżeniu na jednym tonie, tak aby żadna z sylab nie sprawiała wrażenia wyraźnie akcentowanej. Przerwy pomiędzy logatomami tworzącymi jeden bodziec były bardzo krótkie. Jeden zestaw logatomów stanowiący triadę (bodziec), mianowicie /epe, ebe, eme/ nagrano dziesięciokrotnie i wmontowano w zrandomizowaną listę bodźców w sposób losowy, otrzymując łącznie zestaw 510 bodźców.

W teście wzięło udział 34 słuchaczy w następujących trzech grupach:

(a) 9 studentów pierwszego roku filologii, po odbytych zajęciach z podstaw fonetyki i fonologii polskiej (wiek około 20 lat),

(b) 13 uczniów drugiej klasy liceum ogólnokształcącego (wiek 15 - 16 lat),

(c) 12 osób z rozmaitych środowisk, innych niż poprzednie (wiek 19 - 22 lat).

Test przeprowadzono dla wszystkich słuchaczy dwukrotnie. Dla każdego trzyelementowego bodźca należało określić, które dwa elementy są do siebie bardziej podobne niż trzeci. ^{1/}

^{1/} Ponieważ słuchacze nie posługiwali się ani transkrypcją ani normalną ortografią, weliminowano możliwość sugerowania się przy ocenie podobieństwa usłyszanych sygnałów nawykami graficznymi. Fakt, że słuchacze nie dokonywali zapisu odebranych sygnałów nie pozwolił oczywiście na określenie, w jakim stopniu na wynikach oceny podobieństwa zaważyły tzw. błędy wyrazistości sygnału. Przeprowadzone przez Dukiewicza i Piele [5] badania wykazały, że znaczna większość wykorzystanych w niniejszej pracy spółgłosek jest prawie w 100% odbierana poprawnie w warunkach, gdy górna granica przenoszenia wynosi ok. 8 kHz. Obecny eksperyment psychofonetyczny przeprowadzono w standardowym laboratorium językowym, przy użyciu słuchawek, których podane przez producenta górna granica częstotliwości była nieco wyższa niż 10 kHz.

2.2. Testowanie zgodności ocen

2.2.1. Osobnicza zgodność oceny sygnałów

W pierwszym etapie zbadano zgodność oceny tego samego bodźca przez każdego słuchacza oddzielnie. Wyniki testu opracowano osobno dla każdej grupy słuchaczy. Dla celów tej analizy wykorzystano 20-krotną ocenę bodźca /epe, ebe, eme/ oraz dwukrotną ocenę wszystkich pozostałych bodźców.

W analizie zgodności osobniczej w ocenie bodźca /epe, ebe, eme/ posłużono się kryterium χ^2 , hipotezę zerową formułując następująco: rozkład decyzji jest równomierny dla każdej z trzech możliwości oceny podobieństwa. W grupie (a) nie odrzucono hipotezy zerowej dla jednej osoby, której oceny były przypadkowe: pierwsza możliwość wystąpiła 6 razy a druga i trzecia po 7 razy. Dla 7 osób z tej grupy rozkład decyzji był nieprzypadkowy na poziomie istotności $\alpha = 0,001$, a dla jednej osoby na poziomie istotności $\alpha = 0,01$. Dwie osoby dokonały identycznej oceny we wszystkich 20 przypadkach. Wyniki dla grupy (b) są bardzo podobne: dla jednej z osób hipotezę zerową przyjęto, dla pozostałych odrzucono, w tym dla 10 słuchaczy na poziomie istotności $\alpha = 0,001$, dla dwóch na poziomie $\alpha = 0,01$. Podobnie jak w grupie (a) dwie osoby dokonały takiej samej oceny wszystkich 20 bodźców.

W grupie (c) hipotezę zerową odrzucono na poziomie istotności $\alpha = 0,001$ dla 8 osób, w tym jedna dokonała identycznej oceny wszystkich 20 bodźców.

W celu porównania zgodności w ocenach takich samych bodźców w pierwszym i drugim teście przez każdego słuchacza, wykonano obliczenia, których wyniki podsumowuje Tablica 2. Liczby w Tablicy 2 wskazują na następujące fakty: między słuchaczami zachodzi znaczna rozbieżność ich ocen, przy czym największe zróżnicowanie występuje w grupie (c) (młodzież różnych środowisk) a najmniejsze w grupie (b) (uczniowie szkolni). Poza tym obliczone wartości średnie zgodności ocen dla każdej grupy są w przybliżeniu równe.

2.2.2. Zgodność ocen słuchaczy dla poszczególnych bodźców

W kolejnym etapie oszacowano zgodność ocen słuchaczy dla

każdego bodźca. Istnieją następujące możliwości oceny sygnałów: od całkowitej zgodności ocen wszystkich słuchaczy w danej grupie aż do maksymalnej rozbieżności pomiędzy słuchaczami.

Ponieważ grupy słuchaczy nie były równoliczne, teoretyczna liczba różnych możliwości stopnia zgodności oceny podobieństwa dla każdego bodźca (przy założeniu trzech wyborów) jest dla poszczególnych grup inna i wynosi dla 9 studentów 12, dla 13 uczniów 21, a dla grupy 12-osobowej odpowiednio 19.

W tablicach 3a - 3c teoretyczne możliwości ocen usystematyzowano według malejącej zgodności decyzji poszczególnych słuchaczy. W kolumnie 3 przedstawiono liczbę bodźców przypadającą na każdy stopień zgodności ocen, podając osobno wartości uzyskane w pierwszym i drugim teście. W kolumnie czwartej zamieszczono liczebności skumulowane.^{1/} Kolumna piąta zawiera wartości empiryczne χ^2 . Hipotezę zerową sformułowano podobnie jak w doświadczeniu poprzednim (rozdz. 2.2.1.), mianowicie: rozkład ocen poszczególnych słuchaczy jest dla każdego sygnału równomierny czyli w przybliżeniu równa liczba osób dokonuje wyboru 1 - 2, 1 - 3 i 2 - 3. Hipotezę zerową odrzucono na poziomie istotności $\alpha = 0,05$ dla 289 spośród 501 bodźców pierwszego testu oraz 306 bodźców testu drugiego w grupie (a), dla 375 (test I) i 398 (test II) w grupie (b) oraz dla 292 (test I) i 340 bodźców (test II) w trzeciej grupie słuchaczy. Ponieważ grupy słuchaczy różniły się liczebnością, można przypuszczalnie przyjąć, że również i różnice w wynikach, wyżej zacytowane, nie są pomiędzy grupami istotne. Natomiast porównanie rezultatów testu drugiego z wynikami testu pierwszego świadczy, że przy powtórnym przesłuchaniu oceny były mniej przypadkowe. Zachodzi zatem pewien proces uczenia, który, jak widać z porównania obu testów, jest wyraźniejszy w grupie (c), w której znalazły się osoby nie oswojone uprzednio z laboratorium językowym.

^{1/} Materiał testowy obejmował ogółem 510 bodźców, w tym 10-krotną realizację triady /epe, ebe, eme/. W dalszym ciągu operuje się wyłącznie ocenami 501 bodźców, uwzględniając pierwszą ocenę powtarzającego się bodźca.

Przedstawiona powyżej analiza zgodności ocen z jednej strony pomiędzy grupami słuchaczy, a z drugiej pomiędzy testami uzasadnia traktowanie wszystkich słuchaczy jako grupy jednorodnej oraz wykorzystanie w dalszym ciągu badań tylko ocen testu drugiego.

Dla grupy 34 słuchaczy łączna liczba wszystkich możliwych różnych ocen wynosi 114. Posłużono się ponownie testem χ^2 , tym razem wyłącznie dla wyników testu drugiego. Hipotezę zerową sformułowano tak jak poprzednio odrzucono na poziomie istotności $\alpha = 0,001$ dla 346 bodźców, a na poziomie $\alpha = 0,05$ dla 468 bodźców, czyli 93% ogólnej ich liczby.

2.3. Tablice odległości spółgłosek

Podobieństwo między dwoma dźwiękami mowy wyznaczono operując pojęciem odległości rozumianej jako wielkość probabilistyczna i wziętej wzorem:

$$d_{i,j} = 1 - \frac{f_{i,j}}{N}$$

gdzie: $f_{i,j}$ oznacza częstość reakcji preferujących skojarzenie i,j zaś N jest liczebnością wszystkich reakcji, w których możliwy był wybór i,j . Tak więc wielkość $\frac{f_{i,j}}{N}$ jest prawdopodobieństwem preferencji i,j , czyli $P_{i,j}$. Przyjęta miara odległości stanowi zdarzenie przeciwne do $P_{i,j}$, czyli $d_{i,j} = 1 - P_{i,j}$ w sensie addytywności.

Poniższy przykład ilustruje przyjętą procedurę:

Fonemy /p/ i /f/ występowały w dziewięciu triadach, a poszczególne oceny tych triad w sensie preferencji były następujące:

fonemy w triadzie:	preferencje		
	1-2	1-3	2-3
p f r	28	1	5
p f b	7	25	2
p f t	14	18	2
p f k	16	12	6
p f t̥s	29	1	4
p f d̥z	34	0	0

p f t	28	3	3
p f d	33	0	1
p f t	30	2	2

Dla przejrzystości zestawienia kolejność spółgłosek w niektórych triadach zmieniono w stosunku do ich kolejności pierwotnej, w której były prezentowane słuchaczom.

W danym przypadku $f_{i,j} = 219$, $N = 306$, $P_{i,j} = 0,7157$, zaś odległość między spółgłoską $i = /p/$ oraz $j = /f/$ wynosi $1 - 0,7157 = 0,2843$.

Tablica 4 reprezentuje trójkątną macierz odległości pomiędzy fonemami spółgłoskowymi języka polskiego. Puste miejsca w tablicy są wynikiem ograniczenia materiału eksperymentalnego.

W tablicy 5 zamieszczono częstość występowania triad z taką samą parą fonemów. Łącznie do wyznaczenia subiektywnych odległości między fonemami wykorzystano 1503 triady (potrójną liczbę bodźców testowych). W pierwszej kolumnie tablicy 5 umieszczono liczby, wskazujące z iloma różnymi spółgłoskami porównywano dany fonem.

Symetryczna (pełna) tablica odległości posłużyła do konstrukcji dendrytu taksonomicznego (por. ryc. 1). Dendryt stanowi dogodnie odwzorowanie na płaszczyźnie współzależności zachodzących między testowanymi elementami. Połączenia dendrytu wskazują na:

- (a) znaczną odrębność fonemów /r/ oraz /l/ od pozostałych spółgłosek.
- (b) podobieństwo spółgłosek tracych i zwarto-tracych o tym samym miejscu artykulacji.
- (c) małą odległość głosek tracych o słabym szumie /f, v, x/, określanych czasem jako nie-sybilantne, od zwartych o tym samym miejscu artykulacji.

Średnie odległość między elementami połączonymi na dendrycie wynosi $\bar{x} = 0,2203$, zaś odchylenie standardowe $\sigma = 0,0618$. Jeśli zastosuje się metodę Hellwiga [12], polegająca na tym, że dendryt ulega przerwaniu, jeśli odległość między elementami jest równa lub większa od $\bar{x} + 2\sigma$, czyli w rozpatrywanym przypadku $d \leq 0,3838$, otrzymać można grupy fonemów, jak na ryc.2.

Przy takim podziale tworzą się dwie wieloelementowe grupy, z których jedna zawiera głoski bezszumowe lub o słabym szumie, a druga wszystkie głoski o silnym szumie tudzież 2 grupy 2-elementowe i 2 jednoelementowe. Wydzielenie w postaci jednoelementowych grup fonemów /n/ i /ŋ/ oznacza, że cecha nosowości stanowi percepcyjną odrębność. Fakt, że obie te spółgłoski nie zostały z sobą połączone jest wynikiem ograniczeń materiału eksperymentalnego! w żadnej trójce te dwa fonemy nie stanowiły pary (por. Tablica 5).

2.4. Klasyfikacja spółgłosek za pomocą hierarchicznego grupowania

W niniejszym rozdziale przedstawionych zostanie sześć metod grupowania spółgłosek na podstawie subiektywnych odległości określających miarę ich wzajemnego podobieństwa. W pracach tego zakresu ([22], [41]) punktem wyjścia dla każdej z opisanych metod grupowania jest trójkatna macierz odległości euklidesowych. Dla celów niniejszej pracy program i algorytm opublikowany przez Karońskiego i Calińskiego [18] został przystosowany do odległości zdefiniowanych probabilistycznie.

Wspólną cechą zaprezentowanych niżej strategii skupień jest traktowanie, na początku analizy, wszystkich obiektów jako jednoelementowych grup. W każdym kolejnym kroku dokonuje się tylko jednego połączenia bądź dwu obiektów, bądź jednego obiektu z już poprzednio utworzoną grupą, bądź wreszcie dwu grup. Ostatni krok dzieli wszystkie obiekty na dwie grupy. Dla 28 spółgłosek cała procedura skupień obejmuje 26 etapów (kroków). We wszystkich metodach połączeniu w grupy podlegają te obiekty albo grupy poprzednio utworzone, dla których odległość jest najmniejsza. Metody różnią się wyłącznie statystycznymi definicjami odległości między grupami.

Definicje te są następujące:

1. W metodzie najbliższego sąsiedztwa - odległość między najbliższymi elementami (obiektami) pochodzącymi z dwu grup;
2. W metodzie najdalejzego sąsiedztwa - odległość między parą najbardziej odległych elementów każdej z grup;
3. W metodzie mediany - odległość wyrażona minimalnym dys-

tansiem pomiędzy medianami wyznaczonymi dla dwóch grup;

4. W metodzie środka ciężkości - dystans między środkami ciężkości dwu grup;

5. W metodzie średniej grupowej - odległość jako średnia wartość dystansów pomiędzy poszczególnymi elementami dwu grup;

6. W metodzie Warde - odległość między sumami kwadratów odchyleń wszystkich elementów w grupach od środków ciężkości grup, do których należą.

Na ryc. 3-6 przedstawiono graficzną ilustrację przebiegu grupowania dla poszczególnych metod. Nie uwzględniono metody najbliższego sąsiedztwa, ponieważ kryterium grupowania oparte wyłącznie na fakcie, że dwa obiekty lub grupy są bardziej odrebne niż pozostałe, dało połączenia przypadkowe i trudne do graficznego przedstawienia.

W celu umożliwienia porównania odmiennych zasad grupowania według opisanych wyżej kryteriów, wszystkie "warstwicę" nałożono na obiekty, których rozmieszczenie i wzajemne odległości zgodne są z dendrytem taksonomicznym (por. rozdz. 2.3).

Zaproponowana metoda graficznej ilustracji stanowi konieczne uproszczenie zobrazowania odległości między obiektami na płaszczyźnie, podczas gdy w właściwym odwzorowaniu obiekty znajdują się w przestrzeni wielowymiarowej.

Ryc. 6 stanowi rezultat grupowania dwiema metodami: średniej grupowej oraz Warde.

Metody mediany i środka ciężkości różnią się tym od pozostałych, że siła ich kryterium słabnie wraz z kolejnymi krokami grupowania. Każdy następny krok począwszy od dwudziestego, w naszym przypadku, polega na połączeniu do istniejącej grupy wieloelementowej jakiejś małej grupy złożonej z dwu lub trzech obiektów lub pojedynczego obiektu. Ponadto te oba rodzaje grupowania są "czułe" na ilościowe ograniczenia reprezentacji poszczególnych obiektów w teście percepcyjnym, ponieważ fonemy, które nie zostały poddane ocenie podobieństwa zostają przyłączone na końcu. Tym właśnie należy wytłumaczyć odrębność w obu metodach fonemu /j/.

Metoda najbliższego sąsiedztwa charakteryzuje się większym zróżnicowaniem grup, choć przy zastosowaniu tego kryterium zdarza się przyłączenie jednego obiektu do grupy wieloelementowej. Końcowy krok strategii daje w rezultacie dwie wieloelementowe grupy, a nie - jak metoda mediany i metoda środka ciężkości - przeciwstawienie jednego fonemu /s/ grupie 27-fonemowej.

Największym zrównoważeniem wewnątrzgrupowym charakteryzują się wyniki otrzymane metodą średniej grupowej i metodą Warda.

Im grupy wyodrębniające się w danej strategii grupowania są bardziej licznie zrównoważone, tym prostsza jest uzyskana hierarchia.

W metodzie środka ciężkości wyróżnić można 12 poziomów grupowania, w metodzie mediany 11, w metodzie najbliższego sąsiedztwa 9, zaś w metodzie średniej grupowej oraz metodzie Warda 6 poziomów. Z punktu widzenia ekonomii grupowania korzystne są te strategie, które dają w rezultacie mniejszą liczbę poziomów.

Na ryc. 7 i 8 przedstawiono diagramy binarne spółgłosek polskich takiego typu, który stosuje się zazwyczaj przy opisie fonematycznych elementów języka w kategoriach DF. Diagramy stanowią graficzne przekształcenie rycin 3 i 6. Skonstruowano je dla trzech najmniej złożonych strategii grupowania. Każdej warstwy z ryc. 3 lub 6 odpowiada jeden poziom na diagramie binarnym.

W pierwszym etapie analizy diagramów binarnych uczyniono założenie, że danemu poziomowi można przypisać jakąś określoną cechę dystynktywną. Przy takim założeniu jeden z przedstawionych podziałów jest 9-cechowy, drugi 6-cechowy. W przyjętym a priori systemie DF (por. rozdz. 2.1.) występuje 8 cech odnoszących się do polskich spółgłosek. Ponieważ nie stwierdzono nawet ilościowej odpowiedniości pomiędzy artykulacyjnym systemem cech apriorycznych a poziomami klasyfikacyjnymi opartymi na doświadczeniu percepcyjnym, dokonano z kolei próby ustalenia określonego aposteriorycznego systemu DF.

Metoda najbliższego sąsiedztwa na pierwszym poziomie odziera grupę głosek o silnym szumie od wszystkich pozostałych, co w terminologii DF odpowiada opozycji sybilant:niesybilant.

Drugiemu poziomowi podziału nie można przypisać żadnej określonej cechy. Nie udało się również skojarzyć z jakkolwiek cechą dystryktywną łącznego rezultatu podziału na poziomie 2 i 3. Powyżej poziomu 4 oddzielają się te fonemy, których udział w teście był ograniczony.

Dla rezultatów grupowanie według kryterium średniej grupowej oraz według metody Warde dokonano następującej klasyfikacji w kategoriach DF:

poziom

cechy

1		palatalne i sybilanty		niepalatalne i niesybilanty					
2		palatalne		niepalatalne sybilanty		welarne		niewelarne	
3		palatalne niesybilanty		palatalne sybilanty		zębowe		dziąsłowe	
4		zwarłe niezwarłe		dźwięczne bezdźwięczne		dźwięczne bezdźwięczne		trące zwarto-trące	
		trące		trące		trące		trące	
		nie-trące		nie-trące		nie-trące		nie-trące	
		wargowe		wargowe		wargowe		wargowe	
		nie-wargowe		nie-wargowe		nie-wargowe		nie-wargowe	

Na dalszych poziomach (powyżej 4) wydzielone zostają pojedyncze fonemy (por. wyniki metody najbliższego sąsiedztwa).

Podziałowi na pierwszym poziomie nie można przyporządkować jednej określonej cechy według tradycyjnej terminologii DF ze względu na połączenie spółgłosek /c, ʝ, ɲ/ z środkowojęzykowo-dziąsłowymi alwelopalatalnymi /t_ɕ, ʧ, dʑ, ʒ/. Na drugim poziomie podział następuje według opozycji przednia:nieprzednia

(anterior:non-anterior) ; na trzecim poziomie w dwóch grupach wyodrębniają się fonemy szumowe od nieszumowych, ale również dokonuje się dalszy podział ze względu na miejsce artykulacji, którego nie można ująć binarnie. Poziom czwarty oddziela głoski dźwięczne od bezdźwięcznych, ale tutaj dokonuje się również ostateczny podział pod względem miejsca artykulacji. Pominęto tutaj interpretacje jednoelementowej grupy /j/ (por. Tablica 5) .

W kategoriach akustycznych DF pierwszy poziom podziału następuje w zależności od obecności bądź braku charakterystycznych cech widmowych w zakresie od około 2,5 do 5,5 kHz, co odpowiada maximum czułości słuchu na krzywych równej słyszalności. Drugi poziom podziału przebiega w zależności od występowania w dolnej części zakresu największej czułości (do 4 kHz) znacznego skupienia energii. Podział na trzecim poziomie zależy bądź od czasu trwania segmentu szumowego bądź od względnie wysokiego położenia F2. W dalszej klasyfikacji odgrywa rolę kryterium obniżenia dobroci obwodów rezonansowych (na skutek powstania bocznikującego toru) oraz obecności w widmie składowej periodycznej.

Najistotniejsza różnica w grupowaniu metodami 6-poziomowymi z jednej strony, a metodą 9-poziomową z drugiej polega na sposobie przyłączenia grupy spółgłosek palatalnych-niesybilentów do istniejących już skupień. Przy podziale dendrytu taksonomicznego metodą Hellwiga (por. rozdz.2.3.) spółgłoski /j, c/ oraz /ɲ/ stanowią dwie odrębne grupy (por. ryc. 1 i 2) , ponieważ subiektywne odległości percepcyjne pomiędzy sąsiadującymi z nimi fonemami, tj. odległości /j -ɲ/ oraz /c-ɲ/ są większe od wartości krytycznej. Grupa /c,j,ɲ/ łączy się na dendrycie z pozostałymi jedynie za pośrednictwem fonemu /g/. Spółgłoski /c,j/ są spośród badanego obecnie inwentarza fonemów najbardziej kontrowersyjne.

3. Test logatomowo-wyrazowy

3.1. Słownik subiektywnego podobieństwa

Podstawę leżąca testu logatomowo-wyrazowego winien stanowić słownik subiektywnego podobieństwa wyrazów.

Dla języka polskiego istnieją dwa opracowania częstości użycia wyrazów. Pierwsze z nich obejmuje listy frekwencyjne wyrazów w tekstach pisanych, z podziałem na style (proza artystyczne, drobne wiadomości prasowe, itd. [37]). Druga stanowi próba ustalenia subiektywnego prawdopodobieństwa dla wybranego inwentarza wyrazów [17]. Wyniki pierwszej pracy, choć uzyskane z bardzo obszernego materiału, okazały się dla przeprowadzanych tutaj doświadczeń całkowicie nieprzydatne. Wykorzystanie wykazu najczęstszych subiektywnie wyrazów z drugiej z wymienionych publikacji byłoby z punktu widzenia ogólnych założeń referowanego tematu metodologicznie pożądane. Nie było ono jednak możliwe, gdyż wprowadziłoby do eksperymentu psycholingwistycznego szereg czynników nie będących przedmiotem analizy, takich jak redundancje morfofonologiczne. I tak, przy założeniu, że logotom stanowiący sygnał testowy może się różnić od wyrazu autentycznego stanowiącego reakcję słowną tylko jednym fonem spółgłoskowym, zmiana pierwszego fonemu w wyrazie o najwyższej randze w słowniku subiektywnego prawdopodobieństwa, którym jest TAK, daje np. logotom /g ak/, który z kolei dopuszcza możliwość następujących asocjacji: /pak, bak, tak, vak (wag), jak, bak, wak, lak, rak, jak/, natomiast wymiana drugiego fonemu spółgłoskowego dla otrzymania logotomu takiego jak /tax/ może wywołać skojarzenia wyrazowe /tak, tan, tam/. Tego samego rodzaju zamiana pewnego fonemu spółgłoskowego w dwudziestym z kolei najczęstszym wyrazie cytowanego słownika, mianowicie KOLEJANKA pozwoli na utworzenie następujących logotomów: /polejanke, komejanke, kojejanke, kolejanke, kolejatka, kolejanke/ itd., które nie wywołują wielu skojarzeń fonetyczno-wyrazowych polegających na substytucji. Wybór miejsca w wyrazie dla dokonania wymiany nie jest również bez wpływu na możliwości kojarzenia sygnału z autentycznymi wyrazami. Wprowadzenie liczby zamian (substytucji) spółgłosek proporcjonalnej do długości wyrazu uniemożliwiłoby w znacznym stopniu określenie roli, którą w procesie kojarzenia odgrywa subiektywne podobieństwo fonetyczne badanych dźwięków mowy.

Wymienione okoliczności spowodowały konieczność utworzenia

specjalnego, ograniczonego zbioru wyrazów, który poddany ocenie subiektywnego prawdopodobieństwa stanowiłby zarówno podstawę utworzenia odpowiednich logatomowych bodźców testowych, jak i punkt odniesienia dla wszelkich obliczeń probabilistycznych.

Założono, że słownik ten powinien spełniać następujące warunki:

1. Wszystkie wyrazy słownika powinny zawierać taką samą liczbę fonemów.
2. Badane spółgłoski powinny występować w takich samych pozycjach we wszystkich wyrazach.
3. Kontekst fonetyczny badanych spółgłosek winien być taki, by dla testowanych dźwięków zachodziły podobne współzależności koartykulacyjne.
4. Spółgłoski nie powinny się znajdować w wygłosie ze względu na zawieszenie w tej pozycji w języku polskim opozycji dźwięczna-bezdźwięczna dla fonemów zwartych, tracwych i zwarto-tracwych.
5. Zbiór wyrazów wyjściowych, stanowiących potencjalne reakcje we właściwym doświadczeniu, musi być obszerny - rzędu tysięcy - by w jego obrębie można było utworzyć co najmniej kilkaset szeregów komutacyjnych, przy czym w rozpatrywanym tu przypadku przez szereg komutacyjny rozumie się zbiór wyrazów autentycznych, którego elementy różnią się między sobą wyłącznie dzięki wymianie jednego tylko fonemu w jednej pozycji.
6. Liczba wyrazów w szeregu komutacyjnym nie może być mniejsza od 5 przy założeniu, że dla jednego szeregu komutacyjnego zostaną utworzone co najmniej 3 bodźce logatomowe.

Stwierdzono, że powyższe warunki zostaną najlepiej spełnione przez wybór zbioru wyrazów o budowie CVCV, ponieważ ogólna ich liczba w języku polskim, z uwzględnieniem wszystkich możliwych form gramatycznych, oszacować można na około 2,5 tysiąca. (W tej liczbie mieszczą się także wszystkie zadomowione w języku polskim i spolszczone nazwy imion własnych obcego pochodzenia typu Ryga, Chiny). Pominięto natomiast wyrazy, których interpretacja fonematyczna może być niejednoznaczna, np. Zofii /^xzofi, ^xzofji/.

Wyłączono również z rozpatrywanego zbioru większość praktycznie nie używanych form wołacza typu kaszo!, macho!. również zgodnie z wyłożonym wyżej założeniem, wszystkie te wyrazy, dla których nie można było ułożyć pięcioelementowego szerebu komutacyjnego.

Pozostawiono w zbiorze wszystkie wyrazy, o których przed przeprowadzeniem testu subiektywnego prawdopodobieństwa założono, że są bardzo rzadkie lub, że mogą dla części ankietowanych osób nawet być nieznanne.

Badane spółgłoski występują bądź w nagłosie wyrazu przed każdą polską samogłoską, bądź w obustronnym sąsiedztwie samogłoskowym. Niektóre wyrazy słownika wykorzystano dwukrotnie: (1) dla szeregu komutacyjnego o wymiennej spółgłosce nagłosowej i (2) dla szeregu komutacyjnego o wymiennej spółgłosce interwokalicznej. Jeżeli wśród równolicznych szeregów znalazły się takie dwa, które różniły się wyłącznie samogłoskami, wybrano tylko jeden taki szereg.

Łącznie utworzono 253 szeregi komutacyjne zawierające 2152 wyrazów. Poniżej podano zestawienie liczby wyrazów dla poszczególnych szeregów (kolumna 1), liczby szeregów komutacyjnych dla danej liczby wyrazów z wymiana nagłosową (kolumna 2) oraz z wymiana śródgłosowa (kolumna 3) :

1	2	3
5	10	17
6	20	21
7	22	20
8	17	14
9	13	13
10	12	15
11	13	9
12	7	7
13	5	4
14	4	2
15	1	2
16	0	3
17	1	1

Prawie połowe ogólnej liczby szeregów komutacyjnych stanowią zestawienia pięcio-, sześćo- i siedmioclementowe.

Częstość subiektywna wyrazów oceniono dla każdego szeregu komutacyjnego osobno. Listę 253 szeregów poddano ocenie 50 osób (20 z wykształceniem wyższym różnych specjalności, 20 studentów różnych wydziałów i 10 osób ze średnim wykształceniem).

Każda z osób otrzymała osobną ankietę i krótką instrukcję!^{1/} Przeciętny czas potrzebny na opracowanie całej ankiety wynosił dla jednego badanego od 2 do 3 godzin.

Rangę wyrazów w szeregach komutacyjnych wyznaczono według następującego wzoru:

$$R = \frac{Nf_i R_i}{f_i R_i}$$

gdzie: N - liczba wyrazów w szeregu

R_i - ranga wyrazu według oceny uczestników testu

($i = 1, 2, \dots, N$)

f_i - częstość rangi R_i

Zastosowanie rangi ważonej ze zrównoważoną liczebnością szeregu pozwoliło na:

- (a) określenie częstości wyrazu w obrębie szeregu,
- (b) porównanie odpowiednich rang w różnych szeregach.

Uznano, że wyrazy posiadają taką samą rangę, jeśli wyznaczone w powyższy sposób wartości różnią się trzecim miejscem znaczącym. Dla wyrazów o skrajnych rangach przypadków równych

1/ Treść instrukcji: Proszę ułożyć wyrazy w każdym zestawie osobno od najczęściej używanego do najrzadziej i odpowiednio ponumerować, rozpoczynając od jedynki dla wyrazu najczęstszego. Jeżeli dana postać wyrazu kojarzy się z kilku znaczeniami np. wyraz "pika": (1) materiał brzoł-nisny, (2) rodzaj broni, (3) dopełnicz l.p. rzeczownika "pik", (4) 3 os. l.p. czasownika "pikać", jak w "serce mi pika", to należy dokonać oceny częstości użycia tego wyrazu dla wszystkich znaczeń łącznie, ale tylko dla tej jednej formy gramatycznej, w której został użyty w zestawie. Każdy wyraz w szeregu powinien być oznaczony innym numerem.

frakcji jest w materiale niewiele: 11 razy wystąpiły wyrazy o równej, najwyższej randze, 18 razy przypadek taki zaszedł w rangach końcowych (wyrazy subiektywnie najrzadsze).

Znacznie trudniej było osobom ankietowanym dokonać ustalenia nieskrajnych rang wyrazów w szeregach. Trudności te były spowodowane następującymi dwoma czynnikami:

(a) Duża liczebność szeregu komutacyjnego. Już dla szeregów ośmiogłoszowych zachodziło rozchwanie ocen rang pośrednich (nieskrajnych).

(b) Częstość występowania w języku pewnych wyrazów znajdujących się w szeregu. Jeśli w danym szeregu wszystkie lub większość wyrazów miała subiektywnie równą częstość, wówczas w sposób naturalny ustalenie ich rang miało charakter przypadkowy, a zatem oceny mają znaczny rozrzut. (O podobnych trudnościach pisze Frumkina [9]).

3.2. Konstrukcja testu i przebieg doświadczenia.

Przy dobieraniu materiału testowego asemantycznego stwierdzono, że po uwzględnieniu wszystkich ograniczeń fonotaktycznych liczba logotomów uzupełniających wszystkie szeregi komutacyjne do maksymalnej liczby 28 (zbiór wszystkich wymiennych polskich spółgłosek) wynosi około 5000. Stanowiłoby to zbyt obszerny materiał testowy, ponieważ całość testu dla jednej osoby trwałaby wówczas około 70 godzin.

Ustalono zatem następujący sposób konstrukcji zredukowanego materiału logotomowego:

1. Wykorzystano jako punkt wyjściowy wyrazy, które w poszczególnych szeregach komutacyjnych posiadały rangę najwyższą (łącznie 267 jednostek leksykalnych) oraz wyrazy o najniższej randze (269 elementów).

2. Dla wyrazów subiektywnie najczęstszych utworzono zestaw logotomów:

(a) poprzez dobór spółgłosek najbardziej percepcyjnie podobnych do spółgłosek substytuowanych (por. rozdz. 2) oraz

(b) poprzez wykorzystanie możliwie najbardziej percepcyjnie różnych spółgłosek.

3. Wyrazy subiektywnie najrzadsze zostały wymienione na

logotomy przez zastąpienie spółgłosek najbardziej podobnymi dźwiękami.

Utworzono 3 listy sygnałów bodźcowych zawierające łącznie 803 logotomy. W 404 wymieniono pierwszą spółgłoskę, a w 399 druga.

Przy konstruowaniu testu wykorzystano matrycę subiektywnego podobieństwa spółgłosek wyznaczoną na podstawie doświadczenia z logotomami VCV.

Częstość występowania poszczególnych spółgłosek w wyrazach zdeterminowana była zarówno ograniczeniami strukturalnymi zbioru wyrazów, jak i wynikami testu subiektywnego prawdopodobieństwa jednostek leksykalnych i odbiega od wartości typowych dla języka polskiego (por. Łobacz i Jassem [26]). Zachowana została jednakże bardzo niska częstość fonemów /dz, dʒ, ʒ/. Natomiast w listach logotomowych fonemy te występują znacznie częściej, ponieważ ich użycie prawie zawsze powodowało utworzenie logotomu a nie innego wyrazu autentycznego o tej samej strukturze samogłoskowo-spółgłoskowej. Tak na przykład wymiana /d/ w wyrazie "duże" na najbardziej podobny fonem /t/ dała nowy wyraz autentyczny "turże", wykorzystanie następnego z kolei pod względem podobieństwa fonemu /b/ prowadziło do utworzenia wyrazu "burze", a dopiero trzecia wymiana na głoskę /dz/ spełniała warunki testu ("dzuże" jest logotomem). Dla listy logotomowej z podstawieniami w nagłosie 150 razy dokonano wymiany na najbardziej podobną spółgłoskę, 81 razy na drugą z kolei, 31 razy na trzecią z kolei oraz 6 razy na czwartą, która była teraz w niektórych przypadkach bardziej podobna do innego fonemu niż ten, który zastępowano.

Dla kilkunastu szeregów komutacyjnych wyrazy o skrajnych rangach zawierały jako fonemy wymienne tzw. samogłoski niesylabiczne /j,w/. Wyrazy takie eliminowano z dalszej procedury i wybierano wyrazy o następnych rangach. Głoski /j,w/ występują w poszczególnych szeregach ze względów strukturalnych, ale pozostają w nich w dalszej procedurze tylko w pozycji niewymienialnej np. jeśli wymianie podlegała spółgłoska nagłosowa, /j/ mogło pozostać w śródgłosie; jeśli wymianie podle-

gąsę spółgłoska interwokaliczna, /w/ mogło pozostać w nagłosie itp. .

Ustalono, że trzeci zestaw logotomów, mianowicie ten, w którym podstawia się możliwie najbardziej niepodobne spółgłoski zostanie, przy założonych warunkach testu, najbardziej zrównoważony pod względem liczebności badanych głosek. Najczęściej wykorzystywano w tym przypadku 4 lub 5 wystąpień danej spółgłoski, a w krańcowych przypadkach jedną lub osiem.

Poniżej podano dla przykładu dwa szeregi komutacyjne z rangami poszczególnych wyrazów i utworzonymi według wyżej wyłożonych kryteriów logotomami:

substytucja nagłosowa	substytucja śródgłosowa
0,51 losi̇ posi̇, poṡi̇	0,41 tata taða, taða
0,74 nosi̇	0,42 take
0,77 boṡi̇	0,75 taða
0,89 Joṡi̇	0,98 taṫsa
0,94 soṡi̇	1,09 taḋða
1,12 ṫosi̇	1,19 tama
1,16 voṡi̇	1,31 taxa
1,21 kosi̇	1,34 taða
1,31 tosi̇	1,51 tara tala
1,48 fosi̇ voṡi̇	

Logotomy zostały zapisane na taśmie magnetofonowej w wymowie jednego parlatora. Zapisy wykonano w studio bezekowym. Całość zapisano w sześciu częściach (por. kryteria doboru logotomów), na jednym śledzie. Na śledzie drugim zapisywano odpowiedzi kolejno każdego słuchacza. Test przeprowadzono z każdą osobą oddzielnie, po przygotowaniu oddzielnej kopii nagrania wzorcowego, podawanego na wysokiej klasie słuchawki.

Czas przewidziany na skojarzenie przez słuchacza bodźca z wyrazem wynosił 15s i został uprzednio ustalony po szeregu prób z kilku osobami dla kilkunastu losowo wybranych bodźców.

Całe doświadczenie przeprowadzono z każdą osobą na sześciu

osobnych sesjach. Na dodatkowej sesji zaprezentowano słuchaczowi sygnały, na które za pierwszym razem nie odpowiedział, lub na które zareagował błędnie. Jako błąd traktowano brak odpowiedzi, udzielenie odpowiedzi niezgodnie z instrukcją oraz odpowiedź niezgodna z przyjętym a priori systemem fonemetycznym.

Nawet po powtórnym nadaniu niektórych bodźców każda z badanych osób miała pewną liczbę odpowiedzi błędnych. Liczba błędów dla jednej osoby wahała się w granicach od jednego do siedemdziesięciu.

3.3. Określenie miary podobieństwa spółgłosek.

Punktem wyjścia dla wyznaczenia miary podobieństwa badanych spółgłosek były dwie niesymetryczne tablice, zawierające częstości:

1. Wystąpienia fonemu i w wyrazach określonych szeregów komutacyjnych pod warunkiem nadania fonemu j . Częstość wyznaczona wzorem:

$$P_j^1 = \frac{n_{i,j}}{N_j^1} \quad j = 1, \dots, 28$$

gdzie: $n_{i,j}$ - liczebność wystąpienia i pod warunkiem j ,
 N_j^1 - liczba wszystkich skojarzeń pod warunkiem j możliwych dla opracowanego słownika,

2. Występowania fonemu j w wyrazach szeregów komutacyjnych pod warunkiem nadania fonemu i :

$$P_i^2 = \frac{n_{i,j}}{N_i^2} \quad i = 1, \dots, 30$$

P_j^1 oraz P_i^2 traktowane są jako prawdopodobieństwa a priori.

$P_j^1 \neq P_i^2$ ponieważ:

(a) różna jest liczba wystąpień i oraz j jako fonemów nadanych, (b) różna jest liczebność i pod warunkiem j od liczebności j pod warunkiem i .

Ponadto określono analogicznie:

$$P_j^{(1)*} = \frac{n_{i,j}^*}{N_j^{(1)*}}$$

tj. oszacowane częstościowo prawdopodobieństwo skojarzenia fonemu i pod warunkiem nadania fonemu j, jak również:

$$P_i^{(2)*} = \frac{n_{i,j}^*}{N_j^{(2)*}}$$

czyli oszacowane prawdopodobieństwo skojarzenia fonemu j pod warunkiem nadania fonemu i.

$P_j^{(1)*}$ oraz $P_i^{(2)*}$ są prawdopodobieństwami a posteriori.

Jako miare podobieństwa między badanymi elementami wykorzystano prawdopodobieństwa zderzeń przeciwnych do P oraz P^* . Istnieją, zatem, dwie miary dla prawdopodobieństw a priori:

$(1 - P_j^1)$ oraz $(1 - P_i^2)$ i odpowiednio dwie miary podobieństwa dla prawdopodobieństw a posteriori: $(1 - P_j^{(1)*})$ oraz $(1 - P_i^{(2)*})$ uzyskane na podstawie eksperymentu psychofonetycznego.

Ponieważ P_j^1 , P_i^2 , $P_j^{(1)*}$ oraz $P_i^{(2)*}$ są prawdopodobieństwami zdarzeń niezależnych, otrzymujemy:

$$D = (1 - P_j^1) (1 - P_i^2) \quad 1/$$

$$D^* = (1 - P_j^{(1)*}) (1 - P_i^{(2)*}).$$

Ostateczną wartość podobieństwa między fonemami i oraz j uzyskano z iloczynu $D \cdot D^*$.

Wartość $D \cdot D^*$ w takim ujęciu j.w. można interpretować jako miarę odległości między spółgłoskami. Wartości $D \cdot D^*$ zawarte są w przedziale: $\langle 0, 1 \rangle$. Posłużyły one do konstrukcji trójkątnej macierzy podobieństwa (odległości) między spółgłoskami.

Utworzono 6 oddzielnych macierzy odległości uwzględniając następujące warunki doboru logatów:

1/ Wartość D nie reprezentuje miary podobieństwa, lecz stanowi istotną korekcję wartości D^* z punktu widzenia możliwości wystąpienia określonych skojarzeń w doświadczeniu.

1. Nagłos, fonemy a priori podobne, wyrazy o najwyższej randze,
2. Śródgłos, fonemy a priori podobne, wyrazy o najwyższej randze,
3. Nagłos, fonemy niepodobne, wyrazy o najwyższej randze,
4. Śródgłos, fonemy niepodobne, wyrazy o najwyższej randze,
5. Nagłos, fonemy podobne, wyrazy o najniższej randze,
6. Śródgłos, fonemy podobne, wyrazy o najniższej randze.

Warunki powyższe należy rozumieć jako zastąpienie określonego fonemu w określonej pozycji podanego wyrazu autentycznego w szeregu komutacyjnym przez podany fonem, dla uzyskania logotomu.

Przykład trójkątnej macierzy odległości, odpowiadający warunkowi (1) przedstawiono w Tabelicy 6.

Ponieważ na obecnym etapie badań skoncentrowano się na ustaleniu podobieństwa pomiędzy nadanymi a wymienionymi spółgłoskami niezależnie od uwarunkowań koartykulacyjnych, wyznaczono odległości między fonemami z połączonych danych dla warunków: (1) i (2), (3) i (4) oraz (5) i (6). Otrzymano w ten sposób trzy dalsze macierze odległości.

Dla wszystkich macierzy skonstruowano dendryty taksonomiczne (por. rozdz. 2.3). Ryc. 9 przedstawia dendryt dla substytucji fonemów najbardziej podobnych w nagłosie i śródgłosie najczęstszych wyrazów.

Najmniejsze odległości między badanymi spółgłoskami w nagłosie wyrazów występują dla par: /ʃ-tʃ/, /ʃ-z/, /c-tʃ/, /m-n/, w śródgłosie wyrazów dla /f-v/, /s-z/, /ʃ-z/ oraz /s-tʃ/. Małe odległości występują również między trzema badanymi fonemami nosowymi oraz /k/ i /x/.

Dendryt dla substytucji fonemów najbardziej podobnych w wyrazach o najniższej randze (ryc. 10) charakteryzuje się wyrównanymi odległościami pomiędzy obiektami

3.4. Hierarchiczne grupowanie spółgłosek.

W rozdz. 2.4 wykazano, że z 6 zaproponowanych metod hierarchicznego grupowania opisanych w [13], [22] oraz [41] największą moc eksplikacyjną posiadają trzy następujące:

- (a) metoda najbliższego sąsiedztwa,
- (b) metoda średniej grupowej oraz
- (c) metoda Warda.

Powższe trzy metody grupowania zastosowano do wszystkich 9 macierzy odległości.

W analizie hierarchicznych pogrupowań szczególną uwagę zwrócono na:

(a) stopień zgodności uzyskanych wyników w poszczególnych częściach testu z klasyfikacją spółgłosek opartą wyłącznie na subiektywnym podobieństwie (test logatomowy), przedstawioną w rozdz. 2.4.

(b) wpływ lingwistycznego zakłócenia (subiektywna częstość występowania substytuowanych wyrazów).

Stwierdzono, że dla wszystkich zastosowanych metod klasyfikacyjnych oraz wszystkich części testu, ostatnie dwa lub trzy kroki łączenia grup, uprzednio utworzonych, są przypadkowe. Oznacza to występowanie co najmniej czterech czy pięciu grup, z których każda jest prawie równo odległa od wszystkich pozostałych. Zatem, dla testu logatomowo-wyrazowego niemożliwe było, w przeciwieństwie do pogrupowania przedstawionego w rozdz. 2.4., utworzenie binarnego podziału wszystkich 28 spółgłosek i ustalenie najsilniejszej cechy klasyfikacyjnej.

Z kolei przedłożono, które grupy fonemów dla poszczególnych części testu są wspólne przynajmniej dla dwóch metod (średniej grupowej oraz Warda), ponieważ zgodność podziałów w różnych metodach świadczy o wyraźniejszych różnicach między poszczególnymi grupami. Dla większości z 9 macierzy odległości (czyli bez względu na kryterium doboru logatomów) konsekwentnie połączone zostają w jedną grupę /c, ʒ, ɲ, ɳ, z, tʃ, dʒ/ bądź w dwie grupy /c, ʒ, ɲ/ oraz /ɳ, z, tʃ, dʒ/ fonemy palatalne, analogicznie do klasyfikacji w rozdz. 2.4.

Na ryc. 11 przedstawiono grupowanie spółgłosek metodą Warda dla połączonych: (1) i (2) warunków doboru logatomów. Uzyskane w tańcie logatomowo-wyrazowym klasyfikacja jest w znacznym stopniu zbliżone do zaprezentowanej na ryc. 6. W obu doświadczeniach zachowana została odrębność wszystkich spółgło-

sek niepalatalnych i niesybilentnych. Zgodna jest również większość podziałów na mniejsze grupy. Odmiennie w obu doświadczeniach jest przesłuchanie fonemu /r/. W teście logatomowym spółgłoski nosowe nie były porównywane ze sobą nigdy w tym samym bodźcu triadowym, stąd nie mogą tworzyć odrębnej grupy.

Ryc. 12 reprezentuje klasyfikację spółgłosek metodą średniej grupowej dla połączonych części (5) i (6) testu, w których czynnik podobieństwa fonemów przy konstrukcji logatomów został zachowany, wyeliminowano natomiast wpływ częstości wyrazów. Podobnie jak dla (1) i (2) części eksperymentu wyodrębniona została duża grupa spółgłosek niepalatalnych i niesybilentnych (poza grupę zostały głoski: / f, v, m, r /).

Pomimo następujących różnic, jak przyłączenie /r/ do grupy / ʃ, ʒ, tʃ, dʒ /, odłączenie /z/ od skupienia / s, tʃ, dʒ / oraz utworzenie grupy / k, l / uzyskana klasyfikacja opiera się w znacznej mierze na podobieństwie grupowanych obiektów.

Ryc. 13 ilustruje klasyfikację spółgłosek metodą Warda dla (3) i (4) części testu, w których wyeliminowano, poprzez dobór odpowiednich logatomów czynnik fonetycznego podobieństwa (por. rozdz. 3.2). Mimo tego, w procesie hierarchicznego pogrupowania tworzą się następujące dwuelementowe grupy fonemów najbardziej do siebie podobnych: / f, v /, / p, ʒ /, / k, g /. Uzyskano również trzy grupy trójelementowe, zawierające dwie najbardziej do siebie podobne spółgłoski: / tʃ, dʒ, ʃ /, / tʃ, dʒ, x / oraz / m, n, t /. Jednakże w powstaniu większości grup, szczególnie większych, czynnik podobieństwa nie odgrywa roli.

Ustalenie 6 odrębnych kryteriów doboru logatomów do utworzonego słownika realnych wyrazów pozwoliło na weryfikację następujących hipotez:

Hipoteza zerowa:

Skojarzenia między fonemami nadanymi a substytuowanymi są całkowicie przypadkowe.

Hipotezy alternatywne:

1. Skojarzenia między spółgłoskami nadanymi a wymówionymi

przez słuchaczy są zdeterminowane wyłącznie częstością występowania wyrazów w szeregach komutacyjnych.

2. Skojarzenia między spółgłoskami w logatomach a spółgłoskami w odpowiedziach słuchaczy są uwarunkowane wyłącznie podobieństwem fonetycznym.

Na podstawie uzyskanych wyników pierwszą hipotezę zerową należy odrzucić, drugą zaś przyjąć. Warianty testu, w których uwzględniono odpowiednie hipotezy alternatywne wykazały, że tak fonetyczne podobieństwo, jak i subiektywna częstość występowania jednostek leksykalnych wywierają wpływ na odpowiedzi słuchaczy, jednakże efekt fonetycznego podobieństwa spółgłosek jest silniejszy.

4. Podsumowanie.

Ustalenie percepcyjnych odległości stanowiących miarę podobieństwa między spółgłoskami języka polskiego, jak również zastosowanie metody hierarchicznego grupowania badanych elementów segmentalnych, wykazało, że DF nie stanowią w ocenie subiektywnej jedynego kryterium podobieństwa. Niektóre cechy wykorzystywane są w procesie percepcji głosek w sposób sekwencyjny: najpierw ustale się cechy miejsca artykulacji, a potem cechy dźwięczności, ale w innych wypadkach cechy miejsca i sposobu artykulacji są wykorzystywane paralelnie. W referowanym eksperymencie nie potwierdziła się teza Petersa [29], że fonemy spółgłoskowe w procesie percepcji są sortowane najpierw według sposobu artykulacji, następnie według dźwięczności, a dopiero potem według miejsca artykulacji. Z akustycznego punktu widzenia głównym kryterium dyskryminującym jest lokalizacja energii w widmie w zakresie przypadającym w maximum czułości słuchu na krzywych równej słyszalności.

Wyniki testu logatomowego wykazały, że w ogólnym modelu rozpoznawania mowy należy uwzględnić przetworzenie zawartej w sygnale mowy informacji fonetycznej w bloku nowych, uzyskanych a posteriori cech dyskryminacyjnych. Potwierdzony został fakt wielostopniowości modelu rozpoznawania, w którym na niższym stopniu wykorzystywana jest zarówno krótkotrwała jak i długotrwała pamięć fonetyczna, a kryterium dekodowania jest okreś-

lone podobieństwo subiektywne oparte na ogólnych cechach akustycznych segmentalnych elementów sygnału mowy.

Rezultaty drugiego doświadczenia, którego celem było stwierdzenie, w jakim stopniu na wyższym poziomie rozpoznawania mowy - na dekodowanie jednostek o rozciągłości wyrazu-wpływa z jednej strony informacja fonetyczna, a z drugiej częstość subiektywna wyrazu, wykazały, że rozpoznawanie zależne jest zarówno od fonetycznego podobieństwa jak i informacji probabilistycznej charakteryzującej poszczególne jednostki leksykalne.

Wyniki obu eksperymentów psychofonetycznych mogą być wykorzystane w wielostopniowym modelu automatycznego rozpoznawania mowy. W bloku leksykalnym możliwa jest korekta błędów dekodowania bloku fonetycznego na podstawie 2 kryteriów:

1. Jeśli w pamięci leksykalnej (w której jednostki wyrazowe ułożone są w szeregi komutacyjne według podobieństwa struktury fonetycznej) nie ma żadnej informacji probabilistycznej, dotyczącej częstości elementów pamięci, to korekta odbywa się wyłącznie na podstawie fonetycznego podobieństwa (kryterium najmniejszej odległości).

2. Jeśli dane jest informacja probabilistyczna, to wówczas preferowany jest wybór skorelowany dodatkowo z podobieństwem fonetycznym i częstością wyrazu w szeregu komutacyjnym.

Tablica 1
 Artykulacyjne cechy dystynktywne fonemów polskich

fonemy	l	ł	e	a	o	u	j	w	r	m	n	ŋ	p	b	t	d	c	ʒ	k	g	tʂ	dʒ	tʃ	dʒ	ʧ	ʨ	f	v	s	z	ʃ	ʒ	x				
1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			
2	0	0	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1			
3	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
5	1	0	1	0	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
6	1	1	1	0	0	0	1	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
7	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
9	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	
10	0	0	1	1	0	1	1	0	1	1	0	1	1	0	1	1	0	1	1	0	1	1	0	1	1	0	1	1	0	1	1	0	1	1	0	1	
liczba cech	5	3	4	3	4	5	3	3	3	4	4	4	4	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6

1 - spółgłoskowy, 2 - nosowy, 3 - zwarty, 4 - zwarto-trący, 5 - peryferyjny, 6 - przedni, 7 - przyknięty, 8 - retrofleksyjny, 9 - dźwięczny, 10 - krótki.

Tabela 2

Liczba identycznych ocen wrażeń w procentach
w I i II teście dla poszczególnych słuchaczy.

L.p.	Grupa I	Grupa II	Grupa III
1.	78,8	76,5	78,0
2.	77,8	75,1	72,0
3.	73,3	74,1	71,0
4.	72,5	73,3	70,8
5.	68,2	72,9	66,5
6.	67,8	72,0	65,7
7.	67,0	71,6	63,7
8.	63,9	70,6	62,5
9.	56,9	70,6	60,4
10.		68,8	59,6
11.		66,5	53,3
12.		62,5	51,4
13.		59,6	
wartość średnia	70,0	70,3	64,6

Tablica 3a

Możliwości oceny poszczególnych bodźców dla 9 słuchaczy

/ Grupa I /

L.p.	Możliwo- ści ocen	liczeb- ność bodź- ców dla każdej mo- żliwości		liczeb- ność skumulo- wana		warto- ści /chi ² /	poziom istot- ności
		test I	test II	test I	test II		
1.	9 0 0	73	81	73	81	18,00	0,001
2.	8 1 0	95	98	168	179	12,66	0,01
3.	7 2 0	51	51	219	230	8,66	0,025
4.	7 1 1	37	44	256	274	8,00	0,025
5.	6 3 0	33	32	289	306	6,00	0,05
6.	6 2 1	55	51	344	357	4,66	
7.	5 4 0	46	35	390	392	4,66	
8.	5 3 1	48	46	438	438	2,66	
9.	5 2 2	9	13	447	451	2,00	
10.	4 4 1	21	25	468	476	2,00	
11.	4 3 2	26	20	494	496	0,66	
12.	3 3 3	7	5	501	501	0,00	

Tablica 3b

Możliwości oceny poszczególnych bodźców dla 13 słuchaczy

/ Grupa II /

L.p.	Możliwo- ści ocen	liczeb- ność bodź- ców dla każdej mo- żliwości		liczeb- ność skumulo- wana		warto- ści / χ^2 /	poziom istot- ności
		test I	test II	test I	test II		
1.	13 0 0	54	71	54	71	21,9	0,001
2.	12 1 0	48	56	102	127	19,0	0,001
3.	11 2 0	38	39	140	166	15,7	0,001
4.	11 1 1	15	26	155	192	15,5	0,001
5.	10 3 0	27	29	182	221	12,2	0,01
6.	10 2 1	34	37	216	258	11,3	0,01
7.	9 4 0	25	22	241	280	9,5	0,01
8.	9 3 1	26	25	267	305	8,1	0,025
9.	9 2 2	11	15	278	320	7,6	0,025
10.	8 5 0	16	20	294	340	7,6	0,025
11.	8 4 1	27	20	321	360	5,7	
12.	8 3 2	32	16	353	376	4,8	
13.	7 6 0	22	22	375	398	6,7	0,05
14.	7 5 1	29	21	404	419	4,3	
15.	7 4 2	19	24	423	443	2,9	
16.	7 3 3	10	4	433	447	2,5	
17.	6 6 1	19	11	452	458	3,8	
18.	6 5 2	20	17	472	475	2,0	
19.	6 4 3	21	14	493	489	1,1	
20.	5 5 3	5	6	498	495	0,6	
21.	5 4 4	3	6	501	501	0,2	

Tablice 3c

Możliwości oceny poszczególnych bodźców dla 12 słuchaczy

/ Grupa III /

L.p.	Możliwo- ści ocen	liczeb- ność bodź- ców dla każdej mo- żliwości		liczebność skumulowa- ne		warto- ści χ^2 /chi ² /	poziom istot- ności
		test I	test II	test I	test II		
1.	12 0 0	19	21	19	21	24,0	0,001
2.	11 1 0	49	53	68	74	16,5	0,001
3.	10 2 0	31	40	99	114	14,0	0,001
4.	10 1 1	24	37	120	151	13,5	0,01
5.	9 3 0	27	24	147	175	10,5	0,01
6.	9 2 1	45	58	190	233	9,5	0,01
7.	8 4 0	14	27	204	260	8,0	0,025
8.	8 3 1	48	38	249	298	6,5	0,05
9.	8 2 2	23	17	271	315	6,0	0,05
10.	7 5 0	13	19	284	334	6,5	0,05
11.	7 4 1	30	35	313	359	4,5	
12.	7 3 2	40	37	353	392	3,5	
13.	6 6 0	8	7	361	398	6,0	0,05
14.	6 5 1	46	25	407	423	3,5	
15.	6 4 2	21	8	428	431	2,0	
16.	6 3 3	16	24	444	455	1,5	
17.	5 5 2	21	11	465	466	1,5	
18.	5 4 3	33	27	498	493	0,5	
19.	4 4 4	3	8	501	501	0,0	

wyrażone prawdopodobieństwem niepodobieństwa

ć	z	x	t̂s	d̂z	t̂j	d̂j	t̂k	d̂k	m	n	ju	l	r	
		.38	.83	.93	.83	.87	.91		.40	.49		.61	.92	p
		.46	.90	.82	.89	.87		.83	.38	.49		.59	.82	b
.83			.60	.85	.58	.85	.76		.69	.49	.83	.63	.77	t
	.88		.88	.53	.87	.51		.69	.68	.46	.88	.64	.72	d
.62	.76	.60	.78		.71		.56	.74		.82	.39	.84	.88	c
.79	.68	.85		.75		.75	.80	.56		.72	.42	.84	.86	ł
.84		.28	.63		.61		.83	.88	.68		.79	.65	.78	k
	.96	.46		.68		.63	.80	.70	.65		.84	.69	.75	g
			.74	.93	.81	.90	.92						.66	f
			.90	.83	.93	.88	.94						.46	v
			.20	.55	.57	.87	.69						.81	s
			.33	.19	.73	.49	.65						.71	z
			.63	.66	.22	.43	.52						.78	ś
			.84	.57	.44	.20	.64						.65	ż
			.67		.63		.22	.40					.94	ś
				.47		.58	.29	.18					.89	z
			.78	.89	.72	.85	.91	.88					.57	x
				.21	.33	.65	.41		.91	.86	.88	.90	.90	t̂s
					.58	.25	.31		.93	.85	.87	.91	.89	d̂z
						.18	.34		.91	.88	.91	.91	.91	t̂j
								.28	.92	.89	.94	.92	.85	d̂j
								.39	.94	.92	.59	.96	.96	t̂k
									.95	.89	.69	.95	.94	d̂k
													.47	m
													.45	n
													.73	ju
													.13	l
														r

Tablica 4 c.d.

prawdopodobieństwem niepodobieństwa test logatomowo-wyrazowy

<u>6</u>	<u>z</u>	<u>x</u>	<u>tš</u>	<u>dž</u>	<u>tj</u>	<u>dź</u>	<u>tš</u>	<u>dž</u>	<u>m</u>	<u>n</u>	<u>p</u>	<u>l</u>	<u>r</u>
.89	1.0	.82	.91	.95	.92	.98	.80	.99	.72	.82	.76	.71	.76
.91	.93	.86	.98	.94	.90	.89	1.0	.98	.66	.92	.75	.81	.70
.97	.90	.85	1.0	.95	.93	.87	.89	.97	.83	.90	1.0	.88	.78
.95	.99	.90	.94	.71	.93	.97	1.0	1.0	.81	.86	1.0	.75	.78
1.0	1.0	1.0	1.0	.71	1.0	1.0	.92	1.0	.94	1.0	.78	1.0	1.0
.83	1.0	1.0	1.0	1.0	1.0	.99	1.0	1.0	.69	1.0	.83	.46	1.0
.81	.90	.29	.90	.84	.84	.92	1.0	1.0	.85	.70	1.0	.58	.64
.91	.87	.69	.91	.95	.96	.99	1.0	1.0	.90	.73	1.0	.84	.77
.95	.93	.91	.94	.97	.84	.95	.79	1.0	.80	.92	.99	.82	.87
.92	1.0	.86	.86	.98	1.0	.93	1.0	.95	.82	.89	.78	.82	.68
1.0	1.0	.92	.29	.94	.89	.94	1.0	1.0	.95	.94	1.0	.91	1.0
1.0	1.0	.94	.84	.73	.96	.93	1.0	.98	.97	.98	1.0	.81	.82
.93	1.0	.97	1.0	.98	.13	.89	1.0	1.0	.88	.96	1.0	.83	.95
.97	.93	.95	1.0	.98	.92	.56	1.0	1.0	.89	.99	1.0	.85	1.0
	.82	.97	.93	.97	.96	.98	.14	.92	.91	.97	.78	.97	.97
		.93	.93	.98	1.0	.99	.75	.44	.90	.97	1.0	.93	.90
			1.0	.95	.93	.97	1.0	1.0	.81	.92	1.0	.81	.76
				.74	.88	.95	.97	.99	.83	.95	1.0	.94	.85
					.98	1.0	.95	.98	.85	.97	1.0	.93	.91
							.79	1.0	1.0	.97	.85	1.0	.81
								.97	.98	.86	1.0	1.0	.94
									.35	.68	.98	1.0	1.0
										1.0	.97	1.0	.97
											.28	.24	.67
												.99	.85
													1.0
													1.0
													.58

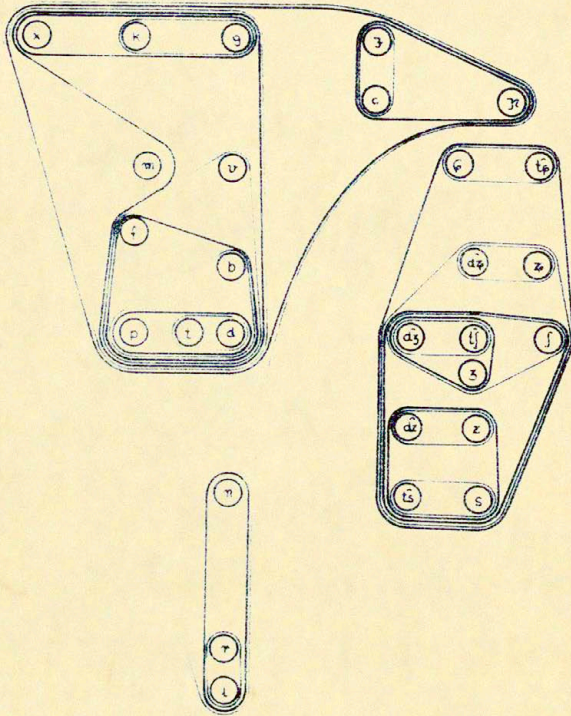
Tablica 6 c.d.

BIBLIOGRAFIA

- 1 AHMED, R. i AGRAWAL, S. Significant features in the perception of Hindi consonants, JASA, 45, 758-763, 1968.
- 2 BROECKE, VAN DEN, K.P.R. Hierarchies and rank orders in distinctive features, Van Gocum-Assen, 1976.
- 3 COLE, R.A. i SCOTT, B. Distinctive Feature Control of Decision time: Same-different judgments of simultaneously heard phonemes, Perceptual Psychophysics, 12, 91-94, 1972.
- 4 CHOMSKY, N. i HALLÉ, M. The Sound Pattern of English, New York, 1968.
- 5 DUKIEWICZ, L. i PIEM, R. Szczegółowe badania wyrazistości i rozróżnialności głosek polskich w różnych warunkach przenoszenia. Cz. 1 Wyrazistość głosek polskich w zależności od górnej granicy częstotliwości. Biuletyn Wojskowej Akademii Technicznej, 7, 45-67, 1961.
- 6 FANT, G. Models of speech perception, ZPSK 21, 5-8, 1968.
- 7 FANT, G. Speech Sounds and Features, MIT, Cambridge, Massachusetts and London, England, 1973.
- 8 FISCHER-JORGENSEN, E. Trends in Phonological Theory, A Historical Introduction, Copenhagen, 1975.
- 9 FRUMKINA, R.M. Wierojatność elementów teksta i reczoweje powiedienije, izd. Nauka, Moskwa, 1971.
- 10 HALLÉ, M. On the bases of phonology, in: The Structure of Language, Fodor, J.A. i Katz, J.J., eds., Englewood Cliffs, 324-323, 1964.
- 11 HANSON, G. Distinctive features and response dimensions of vowel perception, ZPSK 21, 99-101, 1968.
- 12 HELLWIG, Z. Elementy rachunku prawdopodobieństwa i statystyki matematycznej, PWN, Warszawa, 1967.
- 13 JAKOBSON, R., FANT G. i HALLÉ, M. Preliminaries to speech analysis. The distinctive features and their correlates, Acoustics Laboratory MIT Report No. 13, 1952 repr. 1955.
- 14 JAKOBSON, R., HALLÉ, M. Podstawy języka. Tłum. Zawadowski L., Warszawa, Kraków, Wrocław, 1964.
- 15 JASSEM, W. Mowa a nauka o łączności, PWN, Warszawa, 1974.
- 16 JASSEM, W. Założenie ogólnego modelu rozpoznawania mowy, Prace IPPT PAN, Warszawa 1977.

- 17 JASSEM, W. i SZYBISTA D. Subiektywne prawdopodobieństwo wyrazów polskich, Prace IPPT PAN Nr 71/1975.
- 18 KARONSKI, M. i CALINSKI, T. Grupowanie obiektów wielocechowych na podstawie odległości euklidesowych. Roczniki AR w Poznaniu, LXIV, Algorytmy Biometryczne i Statystyczne, zesz. 2, 117-129, 1974.
- 19 KLATT, D.H. Structure of confusions in short-term memory between consonants, JASA, 44, 401-407, 1968.
- 20 LADEFOGED, P. Preliminaries to Linguistic Phonetics, Chicago, London, 1971.
- 21 LADEFOGED, P. A Course in Phonetics, Chicago, 1975.
- 22 LANCE, G.M. i WILLIAMS, W.T. A general theory of classificatory sorting strategies: 1 Hierarchical systems, Computer Journal 9, 373-380, 1967.
- 23 LASKOWSKI, R. Studia nad morfologią współczesnego języka polskiego, Wrocław, 1973.
- 24 LIBERMAN, A.M., COOPER, F.S., HARRIS, K.S., MAC NEILAGE, P. A motor theory of speech perception, Proc. Speech Communication Seminar, Stockholm 1962.
- 25 LIBERMAN, A.M., COOPER, F.S., HARRIS, K.S., MAC NEILAGE, P.F., STUDDERT-KENNEDY, F. Some observations on a model for speech perception, AFCL Symposium on Models for the Perception of Speech in Visible Form, 2.1-2.27, 1964.
- 26 ŁOBACZ P. i JASSEM, W. Fonotaktyczna analiza mówionego tekstu polskiego, Biuletyn Polskiego Tow. Językoznawczego, XXXII, 179-197, 1974.
- 27 MILLER, G.A. i NICELY, P.E. An analysis of perceptual confusions among consonants, JASA, 27, 338-352, 1955.
- 28 MOHR, S. i WANG, S.Y. Perceptual distance and specification of phonological features, Phonetics, 18, 31-45, 1968.
- 29 PETERS, W.W. Dimensions of perception of consonants, JASA, 35, 1985-1989, 1963.
- 30 SHEPARD, R.M. Psychological representation of speech sounds, in: Human Communication. A Unified View, E.D. David, P.B. Denes, eds., New York, 67-113, 1972.
- 31 SINGH, S. Interrelationships of English consonants, Proc. 6th Intern. Congress of Phonetic Sciences, 825-828, 1970.
- 32 SINGH, S. Perceptual similarities and minimal phonetic

- differences, *JCHR*, 14, 113-124, 1971.
- 33 SINGH, S. Distinctive Features: A measure of consonant perception, in: *Measurement procedures in speech, hearing and language* Singh, S. ed. , 93-155, Baltimore 1975.
 - 34 SINGH, S. *Distinctive features. Theory and Validation.* University Park Press, Baltimore 1976.
 - 35 SINGH, S. i BLACK, J.W. Study of twenty-six intervocalic consonants as spoken and recognized by four language groups, *JASA*.39, 372-387, 1966.
 - 36 SINGH, S. i FRANK, D. A distinctive feature analysis of the consonantal substitution pattern, *Language and Speech*, 15, 209-218, 1972.
 - 37 *Stownictwo współczesnego języka polskiego praca zbiorowa* PAN, Warszawa 1974-1977.
 - 38 STEVENS, K.N. i HOUSE, A.S. Speech perception, in: *Foundations of Modern Auditory Theory* Tobias, V., ed. vol. 2, New York, 3-62, 1972.
 - 39 THÜRMAN, E. *Phonetische Ähnlichkeit, distinctive Merkmale and auditive Dimensionen. Ein Bericht.* *Hamburger Phonetische Beiträge* 13, 163-192, 1974.
 - 40 WICKENGREN, W.A. Distinctive features and errors in short-term memory for English consonants, *JASA*.39, 388-398, 1966.
 - 41 WICHART, D. An algorithm for hierarchical classifications. *Biometrics*, 25, 165-170.



Ryc. 5. Hierarchiczne pogrupowanie spółgłosek metodą najbliższego sąsiedztwa.

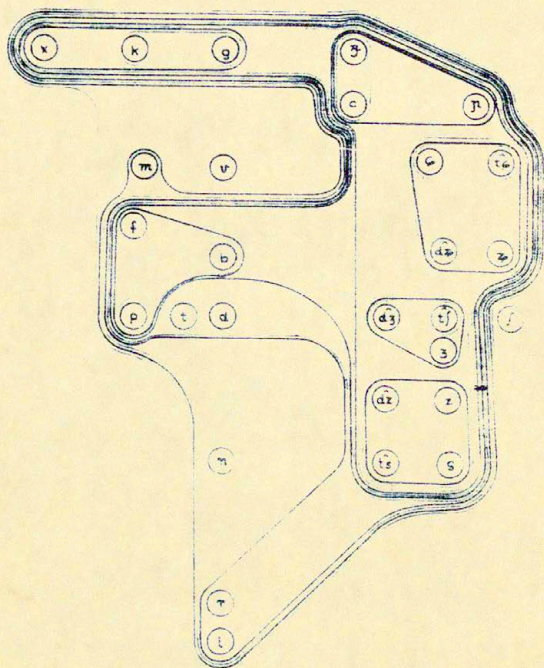
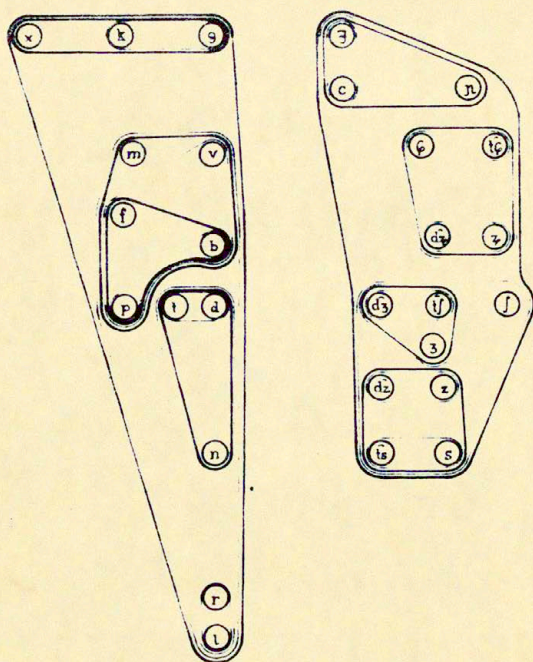
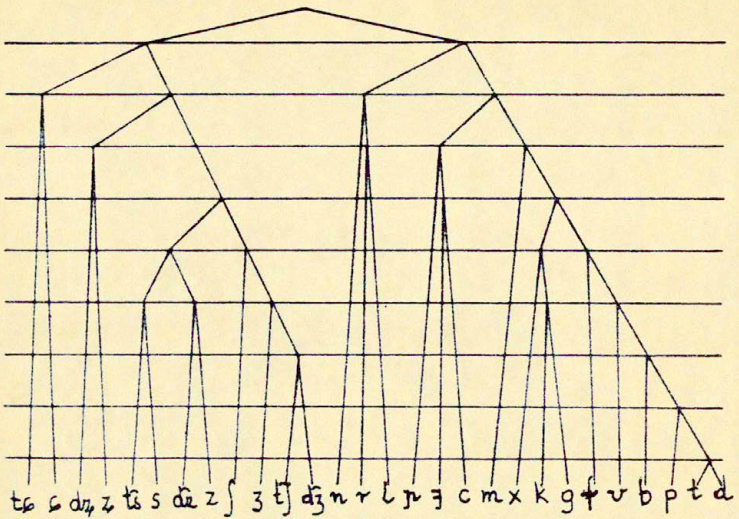


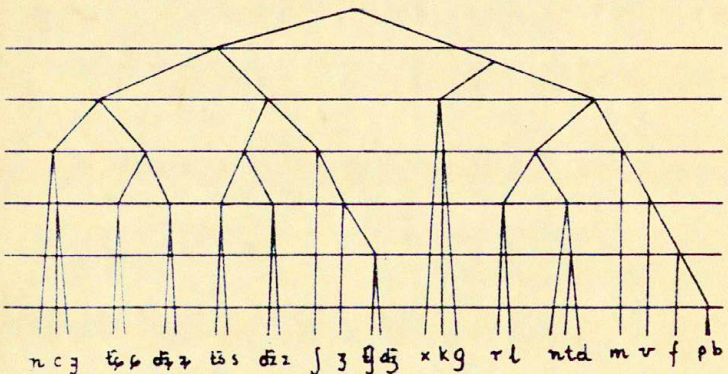
Fig. 5. Hierarchiczne pogrupowanie spółgłosek metodą środka ciężkości.



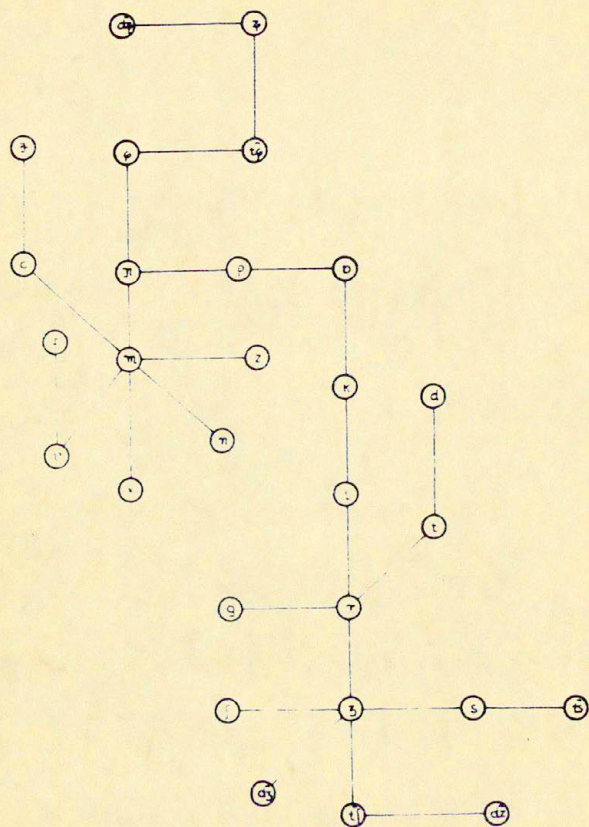
Ryc. 6. Hierarchiczne pogrupowanie spółgłosek metodą średniej grupowej oraz metodą Warda.



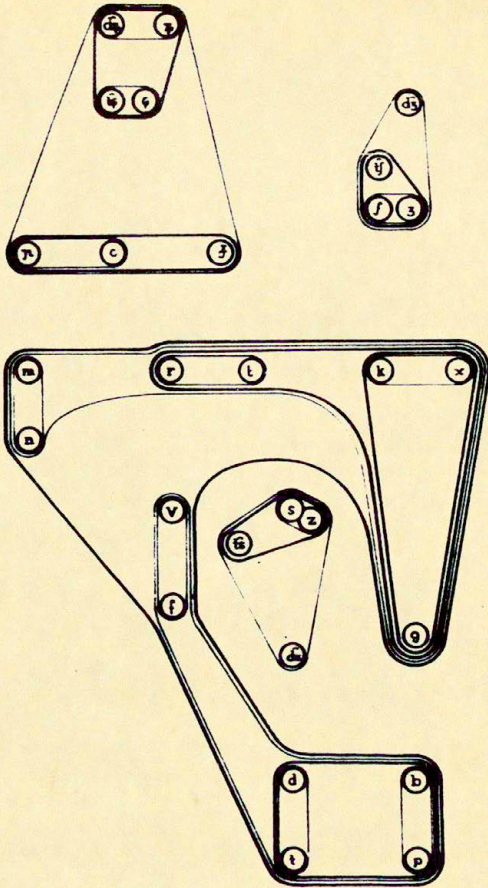
Ryc. 7. Diagram binarny spółgłosek wyznaczony metodą najbliższego sąsiedztwa.



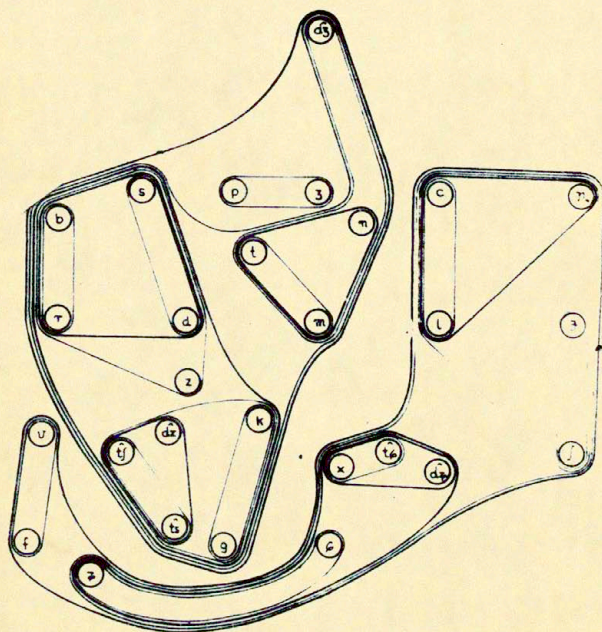
Ryc. 8. Diagram binarny spółgłosek wyznaczony metodami: średniej grupowej oraz Warda.



Ryc. 10. Diagram taksonomiczny spółgłosek dla substytucji podobnych fonemów w nagłosie i śródgłosie wyrazów o najniższych rangach.



Ryc. 11. Pogrupowanie spółgłosek metodą Warda dla substytucji podobnych fonemów w nagłosie i śródgłosie najczęstszych wyrazów.



Ryc. 13. Pogrupowanie spółgłosek metodą Warda dla substytucji niepodobnych fonemów w nagłosie i śródgłosie najczęstszych wyrazów.