Lutosława Richter

WIZUALNE ROZPOZNAWANIE SAMOGŁOSEK POLSKICH W PROSTYCH KONTEKSTACH SPÓŁGŁOSKOWYCH NA PODSTAWIE SPEKTROGRAMÓW KOMPUTEROWYCH

18/1985

WARSZAWA 1985

ISSN 0208-5658

Praca wpłynęła do Redakcji dnia 14 listopada 1984 r.

Na prawach rękopisu

Instytut Podstawowych Problemów Techniki PAN

Nakład 140 egz. Ark.wyd.1,4 Ark.druk. 2
Oddano do drukarni w kwietniu 1985 r.
Nr zamówienia 278/85

Warszawska Drukarnia Naukowa, Warszawa,
ul. Śniadeckich 8

Lutosława Richter
Pracownia Fonetyki Akustycznej
IPPT PAN

WIZUALNE ROZPOZNAWANIE SAMOGŁOSEK POLSKICH W PROSTYCH KONTEKSTACH SPÓŁGŁOSKOWYCH NA PODSTAWIE SPEKTROGRAMÓW KOMPUTEROWYCH[1].

Streszczenie.
Praca stanowi wstępny etap badań nad wizualnym rozpoznawaniem mowy ze spektrogramów komputerowych dla celów rewalidacji osób niesłyszących. Wykonano trzy rodzaje spektrogramów dla każdego z 83 wyrazów stanowiących materiał doświadczalny, w skład których wchodziły wszystkie samogłoski oraz spółgłoski zwarte i trące języka polskiego. Doświadczenia przeprowadzono z grupą 12 dzieci w wieku 9-10 lat. Odczytywanie spektrogramów przebiegało w oparciu o informacje segmentalne. W pracy zamieszczono dane liczbowe dotyczące wyników rozpoznawania oraz omówiono główne tendencje przejawiające się w procesie percepcji wzrokowej mowy.

1. Wstęp.

1.1. Percepcja wzrokowa sygnału mowy.

Jednym z zagadnień podejmowanych na gruncie fonetyki akustycznej jest wizualne rozpoznawanie mowy. Pierwsze kroki w tym kierunku zostały podjęte przez Pottera, Koppa, Greena [5]. Celem ich eksperymentów było zbadanie możliwości rozpoznawania mowy przez osoby z upośledzonym słuchem przy pomocy zmysłu wzroku. Rozpoznawanie odbywało się poprzez identyfikowanie całych wyrazów ze znanym wzorcem. Autorzy spodziewali się, że opanowanie tej umiejętności może posłużyć do nauczania mowy u osób

[1] Praca wykonana w ramach problemu węzłowego 06.9.

głuchych, do korygowania jej jakości oraz jako dodatkowy możliwy sposób komunikacji.

Badania nad wizualnym rozpoznawaniem mowy przeprowadzono również z zamiarem wykorzystania ich w systemach automatycznego rozpoznawania mowy (zwłaszcza ciągłej). Zdaniem niektórych autorów [2],[4],[6] można polepszyć te systemy uwzględniając w pełnym zakresie informacje fonetyczno-akustyczne zakodowane w sygnale mowy oraz odwzorowując model percepcji wzrokowej człowieka. Wiedza fonetyczna bazująca na cechach segmentalnych wsparta informacjami o czynnikach pozasegmentalnych (prozodycznych, syntaktycznych, leksykalnych), które odgrywają istotną rolę przy czytaniu spektrogramów, powinna być ujęta w reguły dla danego języka możliwe do wykorzystania w systemie.

1.2. Cel pracy.

Zagadnienie wzrokowego rozpoznawania mowy dla celów rewalidacyjnych zostało podjęte w Pracowni Fonetyki Akustycznej IPPT PAN jako jedna z możliwości praktycznego wykorzystania metody wizualizacji sygnału mowy opracowanej przez zespół techniczny Pracowni. Pierwsze próby czytania spektrogramów zachęciły do podjęcia szerszych badań, które pozwolą ocenić ich przydatność dla celów rozpoznawania mowy przez niesłyszących. Oparcie metody na spektrogramach komputerowych stanowi novum w tej dziedzinie, gdyż dotychczas w pracach nad rozpoznawaniem wzrokowym posługiwano się spektrogramami tradycyjnymi, uzyskanymi przy pomocy sonografu. Z góry należy zaznaczyć, że posługiwanie się w niniejszym doświadczeniu spektrogramami komputerowymi skazuje na pewien ubytek informacji w porównaniu ze spektrogramami tradycyjnymi, spowodowany choćby faktem, że kolejne widma następują w nich po sobie w odstępach 20-milisekundowych, przez co nie wszystkie szczegóły istotne dla fonetyczno-akustycznej charakterystyki segmentów zostają wychwycone w trakcie analizy. Niezaprzeczalną jednak zaletą spektrogramów komputerowych jest uzyskiwanie ich w czasie rzeczywistym, ograniczenie działań operatorskich przy ich sporządzaniu do prostych czynności manualnych oraz możliwości równoczesnego korzystania z nich przez większą liczbę osób odczytujących obraz z monitora.

Badania nad wzrokowym rozpoznawaniem mowy mogą w zależności od postawionego celu przebiegać w dwojaki sposób :

1/ ocena możliwości rozpoznawania wyrazów należących do zbioru zamkniętego, co wymagałoby ustalenia, jak wiele wyrazów i w jakim czasie można opanować wzrokowo,

2/ ocena możliwości rozpoznawania dowolnych wyrazów (a więc należących do zbioru otwartego), co zakłada rozpoznawanie kolejnych segmentów fonetyczno-akustycznych, po czym łączenie ich w wyraz.

Uzyskanie pozytywnych rezultatów w doświadczeniach drugiego typu zapewniałoby znacznie większą wszechstronność metody(nie ograniczając możliwości jej stosowania do specyficznych sytuacji, jak w punkcie 1). Stąd założeniem pracy jest sprawdzenie hipotezy o rozpoznawaniu segmentalnym.

Równocześnie ma ona udzielić odpowiedzi na pytanie, który z trzech typów spektrogramów komputerowych opracowanych w Pracowni Fonetyki Akustycznej wykazuje największą przydatność dla celów rozpoznawania wzrokowego.

Ponieważ praca ma charakter pilotażowy, a jej wyniki pozwolą nadać kierunek dalszym badaniom, uznano, że na obecnym etapie nie wprowadzi się do eksperymentów osób niesłyszących, choćby ze względu na przewidywane trudności z porozumiewaniem się. Doświadczenia postanowiono przeprowadzić z dziećmi o normalnym słuchu, reprezentującymi jednakowy w przybliżeniu poziom rozwoju psycho-intelektualnego. Badaniem objęto grupę 12 dzieci w wieku 9-10 lat.

2. Materiał doświadczalny.

2.1. Wizualizacja materiału doświadczalnego.

Dla wszystkich wyrazów stanowiących materiał doświadczalny wykonano trzy typy spektrogramów komputerowych. Uzyskano je przy użyciu zestawu minikomputerowego Mera 303 współpracującego z wyspecjalizowanymi urządzeniami peryferyjnymi, na które składał się : wielokanałowy, analogowy analizator widma, kanał funkcji analogowych oraz monitor graficzny.

Spektrogramy pierwszego typu - z kwantowaną amplitudą - zostały nieco bliżej opisane w pracy [1]. Stanowią one ciąg

następujących po sobie co 20 ms sekcji widmowych sygnału o zróżnicowanym poziomie kwantowania amplitudy, wyrażającym się zmienną szerokością zaczernienia w odpowiednich miejscach ekranu (ryc. 1a).

Spektrogramy drugiego typu, tzw. binarne 63-kanałowe [3] stanowią ciąg widm binarnych oddalonych od siebie również co 20 ms. Widmo binarne będące ciągiem zer i jedynek zawiera informacje o zakresach częstotliwości, w których w określonym momencie sygnał przyjmuje wartość ekstremalną. Zaczernienie w dowolnym miejscu obrazu spektrograficznego oznacza wartość 1, brak zaczernienia - wartość 0 (ryc. 1b).

Widma składające się na spektrogramy pierwszego i drugiego typu uzyskuje się na wyjściu wielokanałowego analogowego analizatora widmowego posiadającego 63 kanały. Zmodyfikowaną postać spektrogramu binarnego 63-kanałowego stanowi spektrogram binarny 16-kanałowy (typ trzeci), który zawiera informacje o cechach akustycznych najistotniejszych dla sygnału, pomijając szczegóły niejednokrotnie zamazujące obraz na spektrogramach 63-kanałowych (ryc. 1c).

Do bezpośrednich odczytów posłużyły zdjęcia fotograficzne, na których utrwalono obrazy z monitora.

2.2. Materiał fonetyczny.

Materiał fonetyczny obejmował 9 logatomów oraz 74 wyrazy znaczące dwusylabowe (z wyjątkiem kilku trzysylabowych) wszystkie wymówione przez jedną osobę. Głoski wchodzące w ich skład stanowiły samogłoski /i ɨ a e o u/, spółgłoski zwarte dźwięczne /b d g/ i bezdźwięczne /p t k/ oraz spółgłoski trące dźwięczne /z ʒ ʐ v ɣ/ i bezdźwięczne /s ʃ ɕ f x/.

Całość materiału badawczego składa się z kilku części, które były wykorzystywane na poszczególnych etapach doświadczenia. Część pierwsza obejmuje 9 logatomów zawierających wszystkie wymienione samogłoski i spółgłoski. Część druga obejmuje 20 wyrazów składających się z sześciu zbiorów po sześć wyrazów. Zasadniczy element każdego zbioru stanowi jedna, określona samogłoska, która powtarza się w każdym z wyrazów zbioru. W pozostałej sylabie wyrazu występuje każdorazowo inna samogłoska, np. :

/ɕiva/, /ʑiva/, /ʒeka/, /ʃafa/, /ʃopa/, /suʃa/
/viɕi/, /ɕivi/, /ɕive/, /ɕiva/, /ɕito/, /sufit/

Posłużenie się wyrazami o takiej strukturze zapewniało jednakową liczebność dla wszystkich samogłosek w tej części materiału oraz umożliwiało uchwycenie różnic fonetyczno-akustycznych zachodzących pomiędzy samogłoskami, co ma szczególne znaczenie na etapie uczenia.

Część trzecia obejmuje również 20 wyrazów zbudowanych na tej samej zasadzie, co w części drugiej.

W części czwartej znalazły się 34 dowolne wyrazy zbudowane z wymienionych głosek, w tym kilka wyrazów zawierających zbitki spółgłoskowe (np. /baʃta/, /apteka/).

3. Przebieg doświadczenia.

W celu przeprowadzenia porównania pomiędzy różnymi typami spektrogramów dzieci biorące udział w doświadczeniu zostały podzielone na trzy grupy. Każda grupa obejmująca czworo dzieci odczytywała inny typ spektrogramów.

Po zapoznaniu się z charakterystycznymi cechami segmentalnymi głosek na spektrogramach logatomów przystąpiono do rozpoznawania głosek (a następnie wyrazów) w kolejnych sześciowyrazowych zbiorach stanowiących drugą część materiału. Na tym etapie odczyty były dokonywane całkowicie w oparciu o podpisane logatomy. Kolejność czynności przedstawiała się następująco : określano liczbę segmentów w wyrazach, odszukiwano segment samogłoskowy powtarzający się we wszystkich sześciu wyrazach sekwencji, identyfikowano go z określoną głoską, po czym identyfikowano pozostałe głoski. Każda prawidłowo rozpoznana głoska była podpisywana. Dzieci bez większego trudu odnajdywały powtarzające się segmenty zarówno samogłoskowe, jak spółgłoskowe, co stanowiło pierwszą przesłankę, że cechy segmentalne mogą stanowić podstawę rozpoznawania wizualnego.

W drugim etapie uczestnicy otrzymywali zbiór sześciu spektrogramów oraz po sześć karteczek z zapisanymi wyrazami. Doświadczenie polegało na dopasowywaniu gotowych podpisów do odpowiednich zdjęć. Jako materiał posłużyły tu wyrazy stanowiące trzecią część materiału, a zbiory różnych wyrazów prezentowano

sześciokrotnie. Od tego momentu pracowano bez pomocniczych wzorców polegając wyłącznie na informacjach zapamiętanych. Popełnione błędy były korygowane, a poprawki uzasadniane merytorycznie. Eksperyment ten zmuszał dzieci do samodzielnego analizowania informacji zawartych w spektrogramach w zestawieniu z posiadaną wiedzą na temat cech segmentalnych i jednocześnie tę wiedzę utrwalał.

W kolejnym etapie należało podjąć decyzję, który z dziesięciu (bądź dwunastu) proponowanych wyrazów widnieje na zaprezentowanym zdjęciu. Posłużono się wyrazami z czwartej części materiału.

Wreszcie ostatni etap stanowiło odczytywanie spektrogramów dowolnych wyrazów. Pierwszych 20 wyrazów (z trzeciej części materiału) odczytywano z pomocą prowadzącego eksperyment, który interweniował w przypadku, gdy rozpoznawanie szło w złym kierunku, sugerując ponowny odczyt źle oznaczonej głoski. Zdarzało się również, że była wymagana interwencja w przypadku błędnie przeprowadzonej segmentacji. Po odczytaniu pięciu wyrazów dokonywano ich ponownego odczytu, co w efekcie dało czterdzieści odczytów dla tej części materiału.

Kolejnych 20 wyrazów (druga część materiału) rozpoznawano całkowicie samodzielnie, a uzyskane wyniki stały się podstawą do dokonania oceny całego eksperymentu. Dodatkowo czworo dzieci odczytało po 20 wyrazów z czwartej części materiału.

Na etapie korzystania z podpisanych wzorców zajęcia przeprowadzano równocześnie z dwojgiem dzieci. Następne spotkania musiały się odbywać indywidualnie, gdyż należało wykluczyć dodatkowe źródło informacji w postaci odpowiedzi kolegi. Jedno spotkanie trwało 45 min. (tj. godzinę lekcyjną), odbyto ich z każdym dzieckiem około dziewięciu. Tempo pracy dyktowały przede wszystkim dzieci. Nie narzucano z góry czasu przeznaczonego na wykonanie zadania, pozwalając spokojnie zastanowić się nad odpowiedzią.

Zachęcającym rezultatem doświadczenia jest pozytywny stosunek dzieci do niego. Uczestnicy wykonywali zadania chętnie, proces rozpoznawania uznawali za interesujący i niezbyt trudny, wyrażali chęć kontynuowania doświadczeń. Uważamy, iż nastawienie

osoby dokonującej odczytów jest doniosłym psychologicznie czynnikiem motywacyjnym rzutującym na powodzenie metody.

4. Wyniki doświadczenia.

4.1. Wybór właściwego wyrazu ze zbioru zamkniętego.

Wstępem do odczytywania spektrogramów było doświadczenie z podejmowaniem decyzji, który wyraz należący do zamkniętego zbioru 10 lub 12 wyrazów widnieje na prezentowanym zdjęciu. Wyniki pokazały, że na tym etapie dzieci nie miały jeszcze zbyt dobrze opanowanej wiedzy z zakresu cech segmentalnych. Obok uzasadnionych podobieństwem fonetycznym pomyłek w rodzaju odczytu : /vaga/ zamiast /vaza/, /sosa/ zamiast /sova/,/deska/ zamiast /baʃta/ zdarzały się pomyłki zaskakujące w rodzaju : /zapis/ zamiast /vaga/, /beksa/ zamiast /gaza/. Należy jednak stwierdzić, że poprawność odczytu wzrastała w miarę treningu, gdyż w pierwszej serii uzyskano 60 % poprawnych wyników, a w trzeciej serii już 78 %. Natomiast nie odgrywał roli typ spektrogramu, gdyż ogólna liczba prawidłowych odczytów dla spektrogramów z kwantowaną amplitudą wyniosła 68 %, dla binarnych 63-kanałowych - 67 %, dla binarnych 16-kanałowych - 64 %. Na ryc. 1 zaprezentowano spektrogramy wyrazu /voda/, który należał do najlepiej rozpoznawanych (1 błędna odpowiedź).

4.2. Odczyty I serii spektrogramów.

Odczyt spektrogramów przebiegał w sposób następujący : dzieci zapisywały głoski odpowiadające kolejnym segmentom (w razie potrzeby prowadzący informował, czy wydzielony segment jest spółgłoską, czy samogłoską), jeśli niemożliwe było jednoznaczne określenie spółgłoski (np. /k/), zapisywano całą klasę głosek (tu zwartych bezdźwięcznych). W przypadkach wątpliwych podawano również więcej niż jedną głoskę dla danego segmentu, np.

ʒ	a	k	i
ʃ	i	p	
z		t	
v			

Powyższy zapis oznacza, że odczytujący wyróżnił 4 segmenty. Rozpoznał, że pierwszy segment należy do klasy spółgłosek trących dźwięcznych, trzeci do klasy spółgłosek zwartych bez-

dźwięcznych, czwarty określił jednoznacznie jako /i/, co do drugiego wahał się między /i/ oraz /ɨ/, wobec czego zapisał obie możliwości. Prowadzący eksperyment polecił ponownie odczytać drugi segment, gdyż został on oznaczony nieprawidłowo. Ostateczna wersja zapisu wyglądała następująco :

ʒ	ɨ	k	i
ʐ	i	p	
z	e	t	
v			

Na podstawie zapisu i w oparciu o posiadaną wiedzę leksykalną dziecko zidentyfikowało prawidłowo wyraz jako /ʒeci/.

Dzięki temu, że uczestniczący w doświadczeniu nie musieli podejmować jednoznacznych decyzji przy oznaczaniu segmentu, zapis dla większości głosek uwzględniał możliwość wyboru. Dotyczyło to zwłaszcza spółgłosek, które można było na ogół prawidłowo identyfikować jako należące do określonej klasy głosek, natomiast przypisanie mu cech jednej tylko głoski okazywało się często niemożliwe. Pozostawienie dużej swobody decyzji zmniejszało prawdopodobieństwo popełnienia błędu przy oznaczaniu segmentu, z drugiej jednak strony identyfikacja całego wyrazu przebiegała tym sprawniej, im więcej głosek wchodzących w jego skład zostało oznaczonych jednoznacznie. Zdarzały się błędne rozpoznania wyrazów, pomimo że nie popełniono pomyłki przy rozpoznawaniu segmentów, a wynikające z niejednoznacznego określenia głosek. Np. segmenty wchodzące w skład wyrazu /ʃosa/ zostały oznaczone :

s	o	s	a
ʂ	u	ʂ	
ʃ		ʃ	
f		f	
x		x	

a wyraz zidentyfikowany jako /suʃa/.

Pewną rolę przy rozpoznawaniu wyrazów odgrywała również częstość ich występowania w języku, co świadczy o znaczeniu informacji leksykalnej, z jakiej korzysta odczytujący spektrogram. Dotyczyło to np. wyrazu /fotos/, który sprawił znaczne kłopoty, nawet w przypadkach dość precyzyjnie oznaczonych segmentów.

W tab. 1 sporządzono zestawienie zapisów dla jednego z rozpoznanych wyrazów, którego spektrogramy widnieją na ryc. 2.

Tab. 2 zawiera dane dotyczące samogłosek w rozpoznawanych wyrazach. Podano łączną dla 12 osób liczbę segmentów oznaczonych prawidłowo za pierwszym odczytem, przy czym uwzględniono tu wyłącznie przypadki, gdy segmentom przypisano jedną głoskę. Najlepiej rozpoznawano samogłoskę /a/ - 82 % poprawnych odpowiedzi i to bez względu na typ spektrogramu /81 %, 84 % i 81 % dla poszczególnych typów/, dobrze rozpoznawano w porównaniu z innymi samogłoskę /i/ - 61 %. Najłatwiejsze dla odczytu okazały się więc dwie samogłoski skrajne : najniższa i najwyższa przednia. Nie można tego powiedzieć o trzeciej skrajnej samogłosce - - wysokiej tylnej /u/. Najmniej trafnych rozpoznań uzyskała samogłoska /e /. Nie stwierdzono zależności pomiędzy liczbą dobrych odczytów, a typem spektrogramu.

4.3. Odczyty II serii spektrogramów.

W tej części doświadczenia dzieci dokonywały odczytów całkowicie samodzielnie. Nie wyprowadzano ich z błędu w przypadku niewłaściwej segmentacji, mylenia samogłosek ze spółgłoskami czy błędnego oznaczenia segmentu. Ewentualna korekta błędów należała wyłącznie do odczytującego.

W tab. 3 zestawiono wyniki odczytów segmentów samogłoskowych łącznie dla wszystkich osób. Podobnie, jak w serii I, najlepsze wyniki uzyskano dla /a/, następnie /i/, najgorsze dla /e/. Na rozpoznanie samogłoski /e/ rzutował w bardzo silnym stopniu typ spektrogramu. I tak dla spektrogramów binarnych 16-kanałowych uzyskano aż 79 % jednoznacznych poprawnych odczytów, dla spektrogramów z kwantowaną amplitudą - 17 %, dla binarnych 63-kanałowych zaledwie 9 %. Np. samogłoska /e/ w wyrazie /ʃevek/ (ryc. 3), została na spektrogramach 16-kanałowych na 8 odczytów sześciokrotnie zapisana wyłącznie jako /e/, na spektrogramach z kwantowaną amplitudą jeden raz, zaś na binarnych 63-kanałowych ani razu, natomiast kilkakrotnie podawana wraz z inną zbliżoną głoską lub głoskami. Nie jest to jednak regułą dla pozostałych samogłosek, które rozpatrywane łącznie dają zbliżone wyniki bez względu na typ spektrogramu.

Istotne znaczenie dla dalszych badań nad rozpoznawaniem segmentalnym ma stwierdzenie, jakiego rodzaju błędy są popełniane w odniesieniu do określonych głosek. Zestawienia błędów zawierają tab. 5,6 i 7. Ponieważ w doświadczeniu dopuszczono zapisy dla oznaczenia jednego segmentu zawierające kilka głosek (w wypadkach wątpliwych oraz na skutek korygowania własnych błędów), liczba rozpoznawanych głosek przekraczała zawsze liczbę głosek wypowiedzianych. Np. dane z tab. 5 informują, że samogłoska /i/ została 28 razy poddana odczytom na spektrogramach z kwantowaną amplitudą (20 wyrazów z tej serii wyrazowej zawierało 7 samogłosek /i/, z których każda była rozpoznawana przez 4 osoby). W tych 28 odczytach łącznie zaproponowano 50 głosek, z czego w 25 odczytach pojawiło się /i/, w 11 /ɨ/, w 1 odczycie /e/ oraz /v/, w 4 odczytach /z/, /ʒ/ oraz /z/.

Z porównania tabel 5,6 i 7 wynika, że spośród samogłosek najmniej wątpliwości przy rozpoznawaniu wzbudzało /a/ - było mylone co najwyżej z trzema innymi samogłoskami, najczęściej z /e/, które jest do niego najbardziej zbliżone ze względu na cechy fonetyczno-akustyczne. Stanowi to kolejny dowód, że samogłoska /a/ jest głoską łatwo rozpoznawaną wzrokowo. Pozostałe samogłoski najczęściej były mylone z innymi samogłoskami, zwłaszcza wykazującymi podobieństwo fonetyczne /np. /o/ z /u/, /i/ z /ɨ/ , niekiedy odczytywano je również jako spółgłoski trące, głównie dźwięczne. Najwięcej tego rodzaju zapisów odnosi się do /i/, /ɨ/ na spektrogramach binarnych 63-kanałowych, a większość ich pochodzi od tej samej osoby.

Przy odczytywaniu spółgłosek często przypisywano jednemu segmentowi całą klasę głosek, stąd dla spółgłosek zwartych bezdźwięcznych w tab. 5,6 i 7 oraz trących dźwięcznych w tab. 7 liczebności rozpoznanych głosek (np. /p/, /k/) wykazujących podobne cechy segmentalne z głoską wymówioną (np. /t/) są niemal identyczne z liczebnościami głosek faktycznie wymówionych.

Spółgłoska /v/ na spektrogramach z kwantowaną amplitudą często była rozpoznawana jako zwarta dźwięczna. Wynika to z faktu, że spółgłoska ta wykazywała na tyle niski poziom intensywności, iż nie uwydatniały się jej cechy segmentalne właściwe dla spółgłoski trącej (patrz ryc. 1 i 5).

W tab. 8 przedstawiono rozkład wyników dotyczący całych wyrazów. Poprawnie rozpoznano 63 % wszystkich wyrazów, błędnie 23 %, nie udzielono odpowiedzi w 14 % przypadków. Przeprowadzono test χ^2 w celu zweryfikowania hipotezy zerowej mówiącej, że dla trzech typów spektrogramów zachowane są w przybliżeniu stałe rozkłady wyników odczytów. Obliczona wartość χ^2 = 1.447 mniejsza od wartości kryterialnej χ^2 = 9.488 dla 4 stopni swobody pozwoliła przyjąć, że różnice zachodzące w rozkładach są nieistotne.

Na ryc. 4 zademonstrowano spektrogramy wyrazu /boso/, dla którego uzyskano najlepsze wyniki w rozpoznawaniu (11 osób dało prawidłowe odpowiedzi, 1 nie udzieliła odpowiedzi), na ryc. 5 spektrogramy wyrazu /ɕive/ rozpoznanego najgorzej (3 odczyty poprawne, 7 błędnych, dwukrotnie brak odpowiedzi). Błędne odczyty były następstwem przede wszystkim złego rozpoznania samogłoski wygłosowej (w odpowiedziach podano 4 razy /ɕiva/, 1 raz /ɕivɨ/, 1 raz /ʃiba/.

4.4. Odczyty III serii spektrogramów.

Dwoje dzieci odczytało dodatkowo po 20 spektrogramów z kwantowaną amplitudą, dwoje po 20 spektrogramów binarnych 63-kanałowych. Błędnie rozpoznano dla pierwszego typu spektrogramów 15 % oraz dla drugiego typu 11 % wszystkich samogłosek, co się w przybliżeniu pokrywa z wartościami uzyskanymi w poprzedniej serii wyrazowej (patrz tab. 4). Znacznie więcej segmentów samogłoskowych niż w poprzedniej serii - 63 % rozpoznano prawidłowo i jednoznacznie, na co rzutuje niewątpliwie fakt, że na ogólną liczbę 80 samogłosek aż 40 stanowiło najlepiej rozpoznawane /a/ (w tej serii wyrazów liczebności samogłosek były różne).

Dane dotyczące błędów zamieszczone w tab. 9 i 10 potwierdzają tendencje zaobserwowane w przebiegu percepcji wzrokowej, omówione w rozdziale 4.3.

W tab. 11 przedstawiono wyniki rozpoznawania wyrazów II serii oraz III serii przeprowadzonego przez czworo dzieci. W jednym przypadku (PM) wyniki zdecydowanie się pogorszyły przy rozpoznawaniu III serii wyrazowej, w jednym przypadku (EK) nie

uległy zmianie, zaś w dwóch (MD, PK) polepszyły się w porównaniu z serią II. Być może, przy dalszym zwiększaniu ilości odczytywanych spektrogramów uległaby polepszeniu efektywność ich rozpoznawania. Uzyskane w III serii średnie wyniki są w każdym razie lepsze od średnich wyników dla serii II. Poprawnie rozpoznano 70 % wyrazów, błędnie 18 %, nie rozpoznano 12 % wobec odpowiadających wartości dla serii II - 63 %, 23 %, 14 %.

5. Wnioski końcowe.

Wyniki doświadczenia nad wizualnym rozpoznawaniem mowy w oparciu o spektrogramy komputerowe zachęcają do kontynuowania prac w tej dziedzinie. Szczególne znaczenie dla ustalania kierunków dalszych badań ma potwierdzenie hipotezy o możliwości rozpoznawania segmentalnego. Wyniki wstępnych doświadczeń pozwalają przypuszczać, że model percepcji wzrokowej sygnału mowy oparty jest na identyfikowaniu powtarzalnych, wykazujących zbliżone cechy akustyczne fragmentów sygnału, które znajdują swe odpowiedniki w systemie językowym.

Okazało się również, że rozpoznawanie spektrogramów komputerowych nie wymaga od osoby czytającej posługiwania się rozległą wiedzą fonetyczną. Uświadamianie sobie przez dzieci istnienia w języku jednostek fonetycznych na poziomie głoski oraz całkiem pobieżne zapoznanie się z cechami fonetyczno-akustycznymi głosek pozwalało z powodzeniem podejmować przez nie próby czytania mowy.

Zdajemy sobie sprawę, że dla uzyskania lepszych efektów należałoby wydłużyć okres przyuczania i położyć większy nacisk na rozróżnianie głosek w obrębie klas. Zredukowanie wieloznaczności zapisów odpowiadających poszczególnym segmentom zwiększyłoby prawdopodobieństwo trafnego rozpoznania całego wyrazu.

Interesującym rezultatem eksperymentu jest zbliżony rozkład wyników dla trzech typów spektrogramów, wbrew wcześniejszym przypuszczeniom o znacznie mniejszej przydatności dla naszych celów spektrogramów binarnych. Wyniki te nie przesądzają jednak sprawy, ponieważ eksperymenty przeprowadzone zostały na ograniczonym materiale fonetycznym. Ostatecznej odpowiedzi na temat przydatności określonych metod wizualizacji udzielić

będzie można po uwzględnieniu w doświadczeniach całego systemu fonologicznego.

Przebieg eksperymentu wykazał, że oprócz informacji segmentalnej wykorzystywana jest przy rozpoznawaniu wiedza o języku, jaką dysponuje osoba odczytująca. Np. dzieci kilkakrotnie korygowały błędy popełnione przez siebie, polegające na myleniu samogłosek przednich wysokich ze spółgłoskami trącymi dźwięcznymi, po zorientowaniu się, że wyraz złożony z odczytywanych głosek nie zawierałby ani jednej samogłoski. Zdarzało się, że samogłoskę inicjalną /i/ odczytywały jako /i/ albo /ɨ/, po czym po chwili decydowały się na /i/ w oparciu o regułę obowiązującą w języku polskim, że w nagłosie wyrazu nie występuje samogłoska /ɨ/.

Szczególną rolę odgrywała przy rozpoznawaniu informacja leksykalna zmagazynowana w pamięci długotrwałej osób odczytujących. Zachodziło swoiste przeplatanie się informacji segmentalnej i leksykalnej, gdyż nieudane próby utworzenia wyrazu z zapisanych głosek zmuszały do ponownego odczytu ze zwróceniem uwagi na możliwe popełnione błędy.

Kolejnym krokiem w pracy nad percepcją wzrokową mowy powinno być objęcie badaniem wszystkich głosek języka polskiego w celu pełnego poznania możliwości percepcyjnych w tym zakresie. Będzie to wymagało wydłużenia okresu, w ciągu którego czytający przyswajają sobie podstawowe wiadomości na temat cech fonetyczno-akustycznych segmentów mowy. Przy ustalaniu materiału należałoby wziąć pod uwagę częstość występowania wyrazów w języku, gdyż czynnik ten jest istotnym elementem wiedzy leksykalnej odczytującego.

Uzyskanie pozytywnych rezultatów badań nad rozpoznawaniem wzrokowym mowy mogłoby otworzyć możliwości dla ich zastosowania w procesie rewalidacji osób niesłyszących.

Tablica 1.

Zestawienie zapisów dla wyrazu /ʃɨba/.

I - spektrogramy z kwantowaną amplitudą
II - spektrogramy binarne 63-kanałowe
III - spektrogramy binarne 16-kanałowe

* - oznacza ponowny odczyt na skutek interwencji prowadzącego eksperyment

		ʃ	ɨ	b	a	rozpoznany wyraz
I	TJ	ɕʃf x	ɨ	bdgv	a	ʃɨba
I	AK	ʃɕ	iɨ	vbdg	a	ɕiva ʃɨba *
I	PK	ʃɕ	eɨ	bdg	a	ʃɨba
I	EK	ʃɕ	e ɨ *	vgdb	a	ʃɨba
II	PM	xfʃɕ	eɨ	dbg	a	xɨba ʃɨba *
II	MD	ɕʃ	ɨeu	bdg	a	ʃɨba
II	RK	sɕʃ	iɨ	bdg	a	ʃɨba
II	RZ	sɕʃfx	iɨzʒʐv	bdg	a	ʃɨba
III	AgK	sʃɕfx	ɨ	dbg	a	ʃɨba
III	KR	sʃɕfx	ɨi	vbdg	a	figa ʃɨba *
III	JM	sɕʃfx	vzʒʐ ɨ *	bdg	a	ʃɨba
III	AD	sʃɕfx	ɨ	v bdg *	a	ʃɨba

Tablica 2.

Liczba prawidłowych odczytów segmentów samogłoskowych dla I serii wyrazowej.

	i	ɨ	e	a	o	u
ogólna liczba głosek	96	84	72	96	84	48
$\bar{x}$	61 %	38 %	24 %	82 %	26 %	29 %

Tablica 3.

Wyniki rozpoznawanie segmentów samogłoskowych dla II serii wyrazowej.

		i	ɨ	e	a	o	u
ogólna liczba głosek		84	84	72	84	84	84
rozpoznanie poprawne	jedno oznaczenie na jeden segment	44 %	43 %	35 %	75 %	23 %	33 %
	kilka oznaczeń na jeden segment	49 %	39 %	37 %	14 %	67 %	52 %
	razem	93 %	82 %	72 %	89 %	90 %	85 %
rozpoznanie błędne		7 %	18 %	28 %	11 %	10 %	15 %

Tablica 4.

Wyniki rozpoznawania segmentów samogłoskowych w zależności od typu spektrogramu.

	spektrogramy z kwantowaną amplitudą	spektrogramy binarne 63-kanałowe	spektrogramy binarne 16-kanałowe
liczba głosek	164	164	164
rozpoznane poprawnie /jedno oznaczenie na jeden segment/	40 %	37 %	49 %
rozpoznanie błędne	17 %	17 %	15 %

Tablica 5.

Matryca błędów dla spektrogramów z kwantowaną amplitudą (II seria wyrazowa).

Głoski rozpoznane

Głoski wymówione	i	ɨ	e	a	o	u	b	d	g	p	t	k	v	z	ʒ	ʐ	s	ɕ	ʃ	f	x	ogólna liczba
i	25	11	1										1	4	4	4						28
ɨ	8	23	11		4	3																28
e		3	16	12	8	2																24
a		1	10	26																		28
o	3	3	3	4	23	10								1	1	1						28
u	3	3	1		10	24								1	1	1				1	1	28
b							4	4	4				2									4
d																						
g																						
p							1	1	1	7	7	7								2	2	8
t										11	12	11										12
k										4	4	6								3	3	8
v							26	26	26	1	1	1	28								1	32
z																						
ʒ													1	12	16	16						16
ʐ																						
s																	32	11	11	4	4	32
ɕ															1	1	1	20	20	4	4	20
ʃ																	2	15	16	4	4	16
f																	2			3	3	4
x										2	2	2								16	16	16

Tablica 6.

Matryca błędów dla spektrogramów binarnych 63-kanałowych (II seria wyrazowa).

Głoski rozpoznane

Głoski wymówione	i	ɨ	e	a	o	u	b	d	g	p	t	k	v	z	ʒ	ʐ	s	ɕ	ʃ	f	x	ogólna liczba
i	26	11			1								8	9	9	9						28
ɨ	5	23											9	7	7	7		1	1			28
e		3	16	4	13	12							1	1	1	1		1	1	1	1	24
a			3	27	1	1																28
o			12		26	21							2	1	1	1	1	1	1	1	1	28
u			15		15	22							1	1	1	1						28
b							4	4	4													4
d																						
g																						
p							1	1	1	7	7	7										8
t										12	12	12										12
k										8	8	8										8
v	2	2					3	3	3	2	2	2	27	13	15	15					1	32
z																						
ʒ	2	3	2		3	3							6	9	14	13		1	1			16
ʐ																						
s																	29	22	23	12	11	32
ɕ																	10	20	15	8	8	20
ʃ																	7	12	12	8	10	16
f																	3	1	1	3		4
x			1										1				4	6	6	7	12	16

Tablica 7.

Matryca błędów dla spektrogramów binarnych 16-kanałowych (II seria wyrazowa).

Głoski rozpoznane

Głoski wymówione	i	ɨ	e	a	o	u	b	d	g	p	t	k	v	z	ʒ	ʐ	s	ɕ	ʃ	f	x	ogólna liczba
i	27	11											1	2	1	1						28
ɨ	11	23	2		1	1	1	1	1				1									28
e		2	20	1	2	1															1	24
a			5	22	2	1																28
o			1	1	28	21																28
u	2	1		1	16	25							2	2	2	2						28
b							4	4	4													4
d																						
g																						
p							1	1	1	7	7	7										8
t								1	1	10	11	10										12
k										6	6	8										8
v					1	1	5	5	5				26	20	20	20						32
z																						
ʒ													14	15	16	15				1	1	16
ʐ																						
s																	32	31	31	31	31	32
ɕ																	20	20	20	20	20	20
ʃ																	15	15	16	15	15	16
f																	4	4	4	4	4	4
x																	13	13	13	13	16	16

Tablica 8.

Wyniki rozpoznawania wyrazów (seria II).

Typ spektrogramu	liczba wyrazów	rozpoznane poprawnie	rozpoznane błędnie	brak odpowiedzi
z kwantowaną amplitudą	80	51 (64 %)	19 (24 %)	10 (12 %)
binarne 63-kanałowe	80	47 (59 %)	19 (24 %)	14 (17 %)
binarne 16-kanałowe	80	53 (66 %)	16 (20 %)	11 (14 %)
razem	240	151 (63 %)	54 (23 %)	35 (14 %)

Tablica 11.

Wyniki rozpoznawania wyrazów (porównanie serii II i III).

osoba	seria wyrazowa	liczba wyrazów	rozpoznane poprawnie	rozpoznane błędnie	brak odpowiedzi
PK	II	20	16	3	1
	III	20	18	2	0
EK	II	20	12	5	3
	III	20	11	7	2
PM	II	20	17	3	0
	III	20	9	7	4
MD	II	20	11	3	6
	III	20	16	3	1

Tablica 9.

Matryca błędów dla spektrogramów z kwantowaną amplitudą (III seria wyrazowa).

Głoski rozpoznane

Głoski wymówione

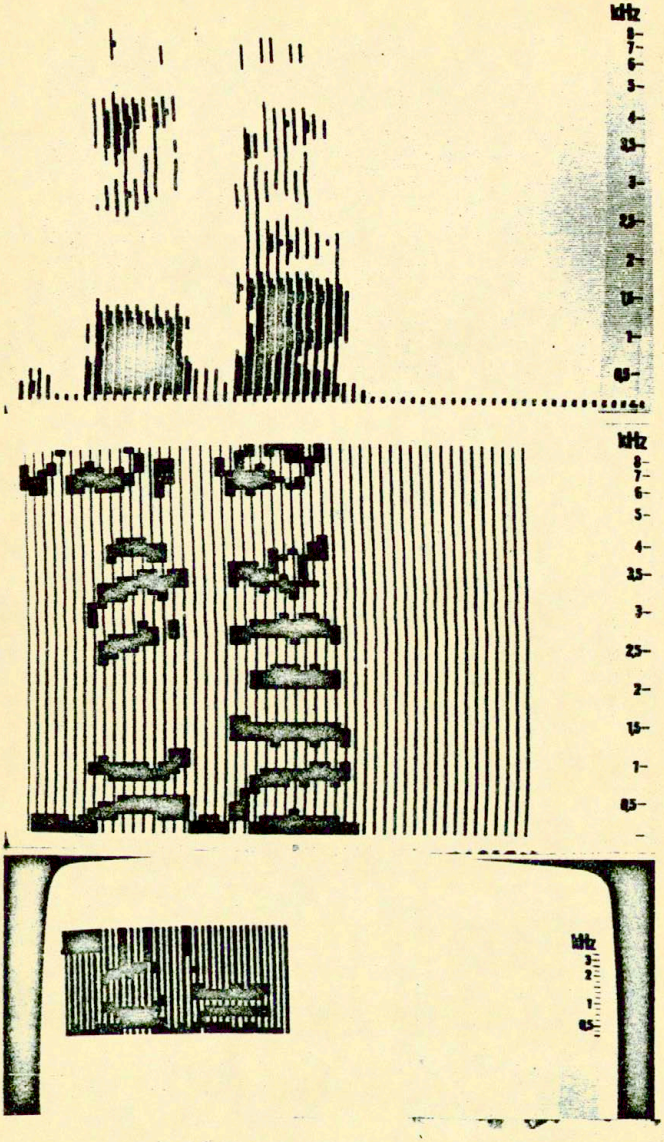
	i	ɨ	e	a	o	u	b	d	g	p	t	k	v	z	ʒ	ʐ	s	ɕ	ʃ	f	x	ogólna liczba
i	2													1	1	1						2
ɨ	4	8	1	1	1	1																10
e	.		5	1	2																	6
a	2	6		33	3					1	1	1										40
o			2		8	4																10
u					3	12																12
b							8	8	8				4									8
d							6	6	6				3									6
g							3	3	4				2									4
p										4	4	4										4
t										7	8	7								3	3	10
k										6	6	6								2	2	8
v							4	4	4				6									6
z														12	11	11						12
ʒ																						
ʐ																						
s																	15	1	1	2	2	16
ɕ																		4	4			4
ʃ																		6	6			6
f																				6	6	6
x																				2	2	2

Tablica 10.

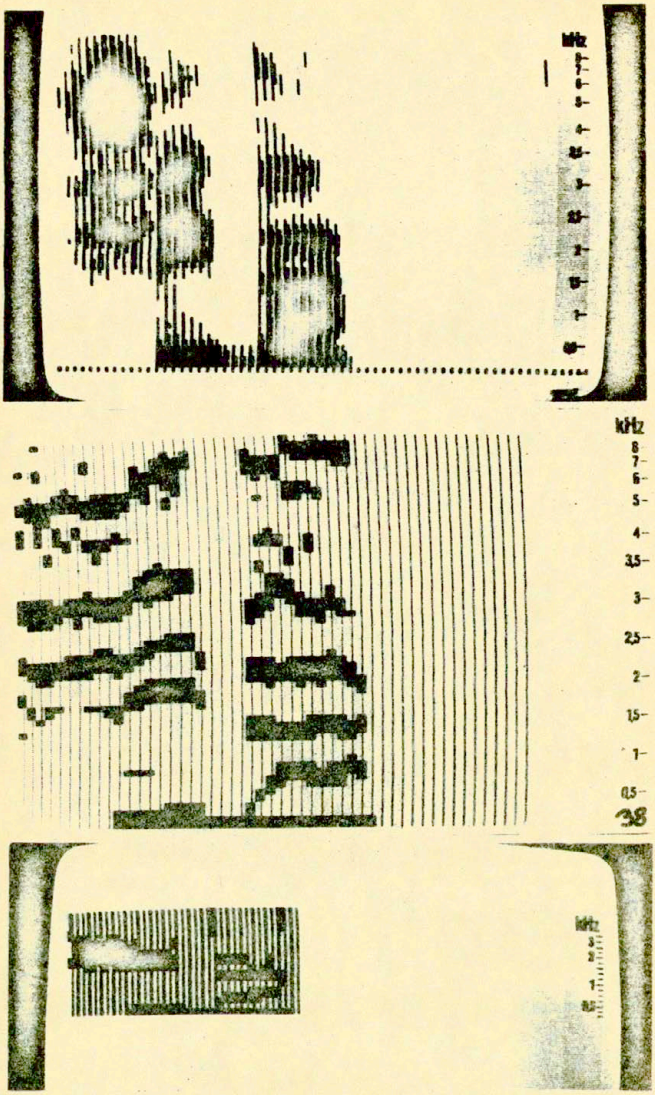
Matryca błędów dla spektrogramów binarnych 63-kanałowych (III seria wyrazowa).

Głoski rozpoznane

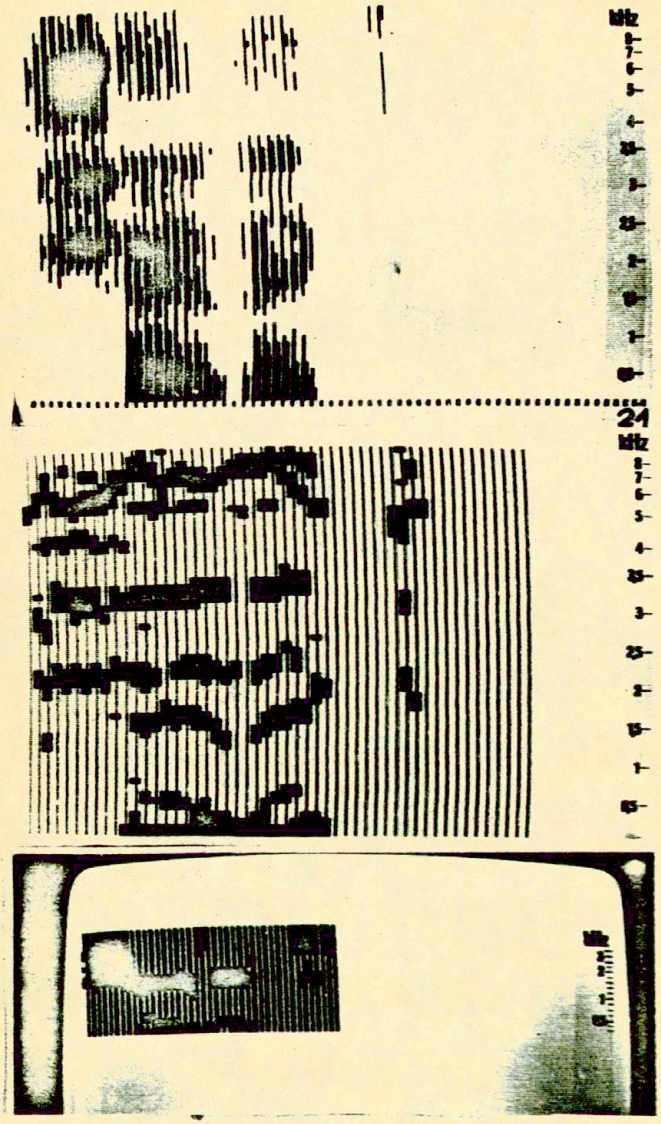
Głoski wymówione	i	ɨ	e	a	o	u	b	d	g	p	t	k	v	z	ʒ	ʐ	s	ɕ	ʃ	f	x	ogólna liczba
i	2	2																				2
ɨ	3	10	1		1									1	1	1						10
e			3		2	3																6
a			3	38	3	3																40
o			2		8	4																10
u			2		5	9																12
b							8	8	8													8
d							5	5	5				1		1							6
g							3	2	3													4
p										4	4	4										4
t										10	10	10										10
k										7	7	7								1	1	8
v													4	2	2	2						6
z					1								5	11	7	7						12
ʒ																						
ʐ																						
s																	15	3	3			16
ɕ																		4	4			4
ʃ																	1	5	5	2	2	6
f																		3	3	5		6
x																					2	2



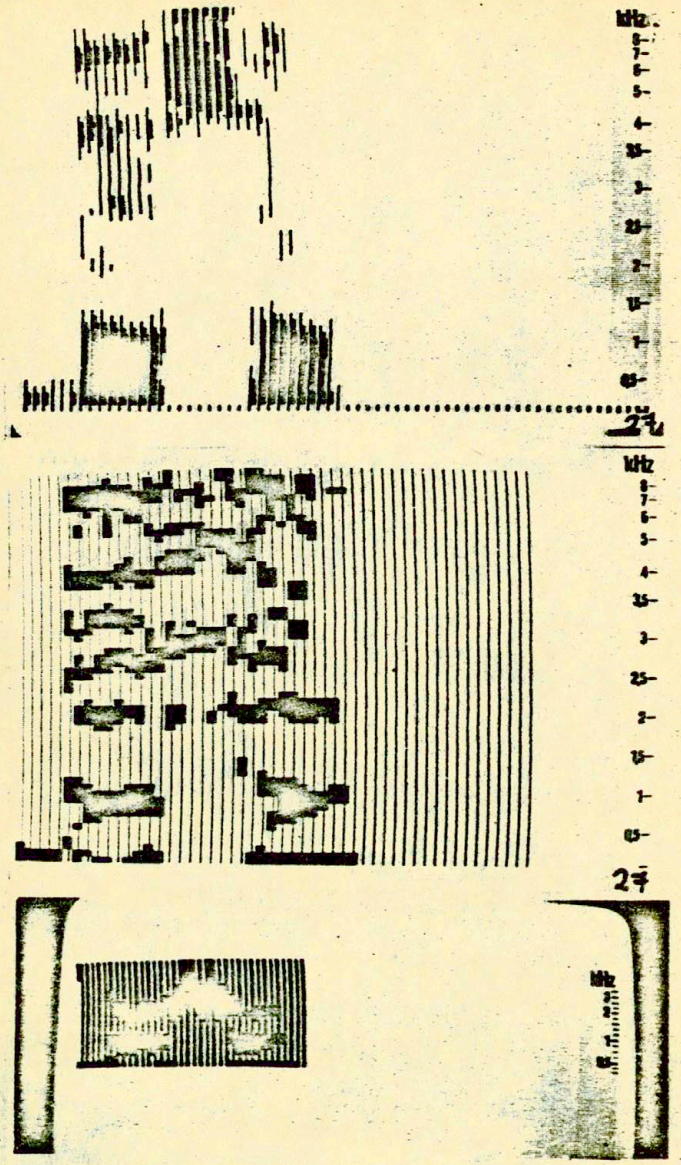
Ryc. 1. Spektrogramy wyrazu /voda/.



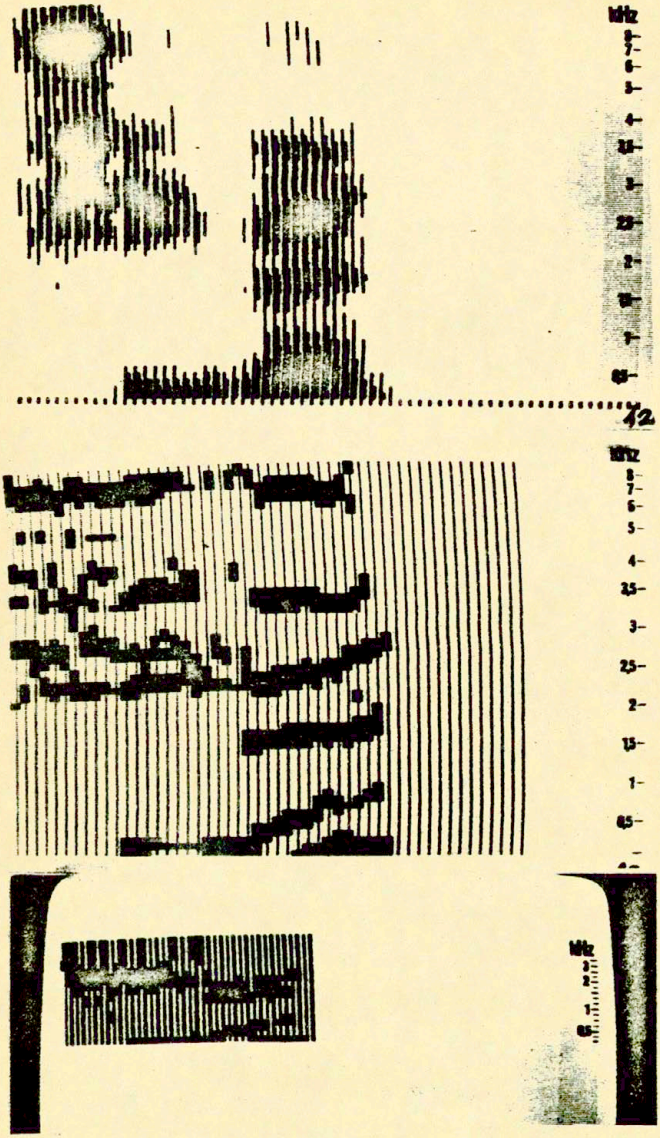
Ryc. 2. Spektrogramy wyrazu /ʃ̈ba/.



Ryc. 3. Spektrogramy wyrazu /ʃevek/.



Ryc. 4. Spektrogramy wyrazu /boso/.



Ryc. 5. Spektrogramy wyrazu /ɕive/.

BIBLIOGRAFIA

[1] CIARKOWSKI, R., Sterowana z minikomputera MERA 303 synteza wybranych diad polskich i ich percepcja (w druku).

[2] KLATT, D., STEVENS, K., On the automatic recognition of continuous speech : implications from a spectrogram - reading experiment, IEEE, vol. AU-21, no. 3, 1973, 210-217.

[3] KUBZDELA, H., Metoda automatycznego rozpoznawania wyrazów w oparciu o spektrogramy binarne, Prace IPPT, 14/1980, Warszawa, 1980.

[4] LINDBLOM, B., SVENSSON, S., Interaction between segmental and nonsegmental factors in speech recognition, IEEE, vol. AU-21, no. 6, 1973, 536-545.

[5] POTTER, R., KOPP, G., GREEN, H., Visible Speech, New York, 1947.

[6] ZUE, V., Acoustic-phonetic knowledge representation : implications from spectrogram reading experiments, in : Automatic Speech Analysis and Recognition, ed. J.P.Haton, 1982, 101-120.