



Lutęslawa Richter

ROZPOZNAWANIE WZORCÓW WIZUALNYCH  
100 NAJCZĘSTSZYCH WYRAZÓW POLSKICH  
PRZEZ OSOBY  
Z GŁĘBOKIMI UPOŚLEDZENIAMI SŁUCHU

14/1987

P. 269



WARSZAWA 1987

Praca wpłynęła do Redakcji dnia 8 grudnia 1986 r.



56771



Na 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. 0,99 Ark.druk. 1,5

Oddano do drukarni w kwietniu 1987 r.

Nr zamówienia 250/87.

---

Warszawska Drukarnia Naukowa, Warszawa,  
ul. Śniadeckich 8

Lutosława Richter  
Pracownia Fonetyki Akustycznej  
IPPT PAN

ROZPOZNAWANIE WZORCÓW WIZUALNYCH  
100 NAJCZĘSTSZYCH WYRAZÓW POLSKICH  
PRZEZ OSOBY Z GŁĘBOKIMI UPOŚLEDZENIA-  
MI SŁUCHU.

Streszczenie.

Doświadczenie przeprowadzone z trzema osobami niesłyszącymi polegało na wzrokowym rozpoznawaniu spektrogramów komputerowych 100 najczęstszych wyrazów polskich. Spektrogramy sporządzono dla jednego głosu męskiego. Testowanie zostało poprzedzone fazą treningu w rozpoznawaniu cech fonetyczno-akustycznych segmentów. Odczyty przebiegały na zasadzie wyboru wymuszonego spośród 10 lub 20 zamieszczonych na liście wyrazów lub w oparciu o identyfikację segmentów na zasadzie wolnego wyboru spośród fonemów istniejących w systemie językowym. Rezultaty i przebieg doświadczenia wskazują, iż percepcja wzrokowa sygnału mowy u osób głuchych może przebiegać w oparciu o rozpoznawanie segmentalne.

1. Wstęp.

Badania nad wizualnym rozpoznawaniem mowy przeprowadza się w zależności od założonych celów z osobami wyszkolonymi w zakresie czytania spektrogramów [1, 5, 6, 7, 11], bądź z osobami niewyszkolonymi fonetycznie [2, 3, 8, 9, 10]. Fachowe przygotowanie uczestników doświadczenia wykorzystywane jest głównie w przypadkach, gdy autorzy spodziewają się dojść poprzez analizę percepcji wzrokowej wszelkich możliwych cech fonetyczno-akustycznych zakodowanych w sygnale mowy do określonego systemu reguł. Reguły te zastosowane w układach automatycznego rozpoznawania mowy powinny pomóc w usunięciu zasadniczych barier utrudniają-  
Praca wykonana w ramach CPBR 11.9

cych znaczny postęp w tej dziedzinie, tj. ograniczenie się do zamkniętego słownika haseł oraz głosu konkretnego mówcy. Doświadczenia przeprowadzane z osobami niewykształconymi są bardzo cennym źródłem informacji o możliwości wykorzystania wizualnego rozpoznawania mowy dla celów rewalidacji głuchych.

Klatt i Stevens [5] badali umiejętność rozpoznawania segmentów opierając się na wynikach osobiście przez nich przeprowadzonych odczytów spektrogramów nieznanych wypowiedzi. W celu zminimalizowania możliwości rozpoznawania wyrazów (a więc częściowo odrzadywania niektórych segmentów), odczytów dokonywano w okienku o szerokości 300 ms, podczas gdy pozostała część zapisu była zasłonięta. Zidentyfikowano poprawnie 33 % segmentów, częściowo poprawnie, tzn. podano tylko klasę głosek - 40 %, co łącznie daje 73 % dobrych rozpoznań.

Lindblom, Svensson [7] badali rolę prozodycznej organizacji wypowiedzi w procesie dekodowania sygnału mowy (zmiany  $F_0$ , intensywność, czas trwania). W doświadczeniu wzięli udział autorzy oraz osoby po ukończonym kursie czytania spektrogramów. Według autorów uwzględnienie informacji o cechach prozodycznych stanowi ułatwienie przy czytaniu spektrogramów, gdyż dostarczają one pewnych danych na temat aspektów gramatycznych wypowiedzi.

Praca Cole, Rudnicki, Zue, Reddy [1] zawiera szczegółowy opis metody stosowanej przez Zue przy czytaniu spektrogramów. Materiał językowy stanowiły 23 zdania (z tego 6 wypowiedzi zawierało błędy językowe: fonetyczne, syntaktyczne lub semantyczne) oraz 45 wyrazów umieszczonych we frazie nośnej. Zue przeprowadził poprawną segmentację w 97 % w odniesieniu do zdań, a w 100 % w odniesieniu do wyrazów, zaś poprawnie zidentyfikował 86 % segmentów w zdaniach, a 93 % w wyrazach. Autorzy przewidują zastosowanie umiejętności rozpoznawania wzrokowego spektrogramów przy optymalizacji układów do automatycznego rozpoznawania mowy oraz w procesie rewalidacji głuchych (terapia mowy, rozpoznawanie mowy ze spektrogramów dokonywane w czasie rzeczywistym).

Zue [11] na podstawie własnych, bogatych doświadczeń z odczytywaniem spektrogramów reprezentuje pogląd, iż wysoki poziom wiedzy fonetycznej pozwoli wyekstrahować z sygnału mowy

parametry określające relewantną charakterystykę akustyczną segmentów dla celów zastosowania w systemach automatycznego rozpoznawania mowy.

Johnson, Connoly, Edmonds [5] stwierdzają, iż formalizacja wiedzy lingwistycznej, z której korzysta człowiek w procesie dekodowania mowy, bazująca na segmentalnych cechach akustycznych widocznych na spektrogramie, powinna być wykorzystana w procesie automatycznego rozpoznawania. Autorzy prezentują próbkę działania systemu reguł opracowanego przez wielodyscyplinarny zespół, służącego do interpretowania spektrogramów w zakresie fonemów i wyrazów.

Prekursorami badań nad rozpoznawaniem wzrokowym mowy byli Potter, Kopp, Greene [8]. Od czasu ich wystąpienia pojawiło się w fonetyce nowe pojęcie : Visible Speech. Autorzy przewidywali, że optyczne wzorce mowy okażą się przydatne dla celów rewalidacji głuchych. W przeprowadzonym przez nich doświadczeniu oprócz osób o normalnym słuchu wzięła również udział jedna osoba niesłysząca. Żaden z uczestników nie posiadał fachowego przygotowania fonetycznego. Spektrogramy uzyskiwano w czasie rzeczywistym za pośrednictwem urządzenia zwanego Direct Translator. Osoby badane uczono globalnego rozpoznawania wzorców wyrazów i krótkich fraz umożliwiających prostą konwersację, a dodatkowo zapoznano je z cechami fonetyczno-akustycznymi segmentów. Osoba głucha opanowała w trakcie treningu 800 wyrazów. Pomimo dość zachęcających wyników metoda ta nie znalazła praktycznego zastosowania w pracy z głuchymi ze względu na wysokie koszty urządzenia.

Greene, Pisoni, Carell [3] również oparli czytanie spektrogramów na globalnym rozpoznawaniu wyrazów przez osoby niewykszkolone. Uczestników eksperymentu nauczono rozpoznawać 50 wyrazów jednosylabowych, po czym zlecono przeprowadzić odczyty 50 spektrogramów nieznanych wyrazów. W nowym materiale zidentyfikowano poprawnie ok. 30 % głosek, 66 % segmentów zostało poprawnie zaliczonych do właściwej klasy głosek. Odczytano prawidłowo 6 % wyrazów. Prześledzenie opisów werbalnych towarzyszących przeprowadzanym odczytom nowych wyrazów wykazało, że uczestnicy eksperymentu dzielili wypowiedzi na segmenty, następnie identyfikowali każdy segment w oparciu o cechy, które zdołali wyekstrahować ze

spektrogramów w trakcie wzrokowego przyswajania wzorców wyrazów objętych treningiem.

W innej pracy ci sami autorzy [2] porównali wyniki rozpoznawania spektrogramów przez dwie grupy osób : grupa I, wyszkolona fonetycznie, miała za sobą specjalistyczny kurs czytania spektrogramów, grupa II, niewykształcona, przyswoiła sobie wizualne wzorce 50 wyrazów. Doświadczenie przeprowadzono na 50 nowych wyrazach. Grupa I - zidentyfikowała poprawnie 40 % segmentów, grupa II - 30 % segmentów. Liczba poprawnych odpowiedzi w odniesieniu do klas spółgłosek wynosiła dla obu grup 66 %. Zdaniem autorów sugeruje to, że możliwa jest ekstrakcja podstawowych cech segmentalnych bez względu na rodzaj treningu. Jednakże dalsze wydzielenie konkretnych spółgłosek spośród klas wymaga bardziej szczegółowej wiedzy fonetycznej.

Opracowanie metody wizualizacji sygnału mowy w Pracowni Fonetyki Akustycznej IPPT PAN pozwoliło bliżej zająć się zagadnieniem rozpoznawania wzrokowego mowy polskiej dla celów rewalidacji głuchych. Pierwsze z tej dziedziny prace autorki [9, 10] opisują doświadczenia przeprowadzone z dziećmi o normalnym słuchu. Ponieważ zakładano możliwość uzyskania negatywnych wyników, nie angażowano na tym etapie osób głuchych. Doświadczenia obejmowały każdorazowo dwie fazy : treningu i testowania. W trakcie treningu zapoznawano uczestników z charakterystyką fonetyczno-akustyczną segmentów, po czym w fazie testowania odczytywano spektrogramy nieznanymi wyrazów. Wyniki doświadczeń doprowadziły do następujących wniosków : (1) rozpoznawanie wzrokowe mowy można przeprowadzać w oparciu o uproszczone spektrogramy komputerowe, (2) z trzech typów spektrogramów uzyskiwanych w PFA najlepiej nadają się do odczytów spektrogramy z kilku stopniami kwantyzacji amplitudy, (3) rozpoznawanie wyrazów może opierać się na identyfikowaniu poszczególnych segmentów akustyczno-fonetycznych, (4) czytanie spektrogramów komputerowych nie wymaga fachowego przygotowania fonetycznego.

W wyniku doświadczenia przeprowadzonego na materiale obejmującym wszystkie fonemy języka polskiego, na spektrogramach 30 nowych wyrazów rozpoznano poprawnie 85 % segmentów (w tym

jednoznacznie 34 %) oraz 55 % wyrazów. Sporządzono matryce błędów oraz wyróżniono głoski poddające się najłatwiej oraz najtrudniej rozpoznawaniu wzrokowemu.

Uzyskanie pozytywnych rezultatów w eksperymentach przeprowadzonych z osobami słyszącymi zachęciło do podjęcia badań nad percepcją wzrokową sygnału mowy przez osoby głuche.

Należało w nowych warunkach doświadczalnych zweryfikować opracowaną metodologię polegającą na przyswajaniu optycznych wzorców segmentów odpowiadających określonym głoskom. Przebieg doświadczenia powinien uzmysłwić podstawowe trudności występujące w pracy z osobami niesłyszącymi.

## 2. Opis doświadczenia.

### 2.1. Osoby uczestniczące w doświadczeniu.

Do eksperymentu zaangażowano trzy uczennice (w wieku 17 lat) Szkoły Zawodowej przy Ośrodku Szkolno-Wychowawczym dla Dzieci Głuchych w Poznaniu. Wszystkie trzy utraciły słuch w pierwszym roku życia, a więc w wieku, w którym nie zdążyły jeszcze wykształcić mowy. Między sobą posługiwały się językiem migowym.

Kontakt z eksperymentatorem, który nie znał języka migowego, napotykał na dość duże przeszkody. Dziewczęta odbierały treści komunikatów w oparciu o czytanie mowy z ust, zaś przekazywały komunikaty posługując się mową. Na skutek znacznych zaburzeń artykulacji mowa ta nie była w pełni zrozumiała dla osoby prowadzącej doświadczenie. Dodatkową trudność stanowił fakt, że był to pierwszy kontakt eksperymentatora z głuchymi, w związku z czym brakowało "osłuchania" z mową upośledzoną.

### 2.2. Materiał doświadczalny.

W fazie treningu wykorzystano spektrogramy 17 logatomów oraz 51 wyrazów dwu- lub trzysylabowych. Logatomy zawierały wszystkie samogłoski i spółgłoski języka polskiego. Wyrazy mieściły się w kilku seriach ułożonych w sposób umożliwiający korzystanie z nich na różnych etapach treningu, co ilustruje następujące zestawienie.

Seria	Materiał językowy	Liczba wyrazów	Nowowprowadzona klasa głosek
T 1	Spółgłoski trące i zwarte, samogłoski. Liczebność poszczególnych samogłosek wyrównana.	18	Spółgłoski trące i zwarte, samogłoski.
T 2	Spółgłoski trące i zwarte, samogłoski. Liczebność poszczególnych samogłosek wyrównana.	15	
M	Spółgłoski nosowe, zwarte i trące, samogłoski.	5	Spółgłoski nosowe.
L	Spółgłoski /l/, /r/, nosowe, zwarte, samogłoski.	4	Spółgłoski /l/, /r/.
J	Głoski /j/, /w/, /r/, nosowe, zwarte, trące, samogłoski.	5	Głoski /j/, /w/.
C	Spółgłoski zwarto-trące, /j/, /w/, /l/, /r/, nosowe, zwarte, trące, samogłoski.	7	Spółgłoski zwarto-trące.

Spośród łącznie 51 wyrazów - 23 należy uznać za bardzo rozpowszechnione w języku (rangI od 1 do 1000 według "Słownika częstotliwości subiektywnej wyrazów w języku polskim" [4]). 21 wyrazów posiadało rangI od 1000 do 5000, zaś 10 wyrazów - powyżej 5000.

Na etapie testowania wykorzystano spektrogramy 100 najczęstszych wyrazów według słownika [4].

Jak wykazały rezultaty wstępnych badań, częstość używania wyrazów w języku odgrywa istotną rolę przy rozpoznawaniu wizualnym. W części doświadczenia obejmującej właściwe rozpoznawanie poddano testowaniu 100 wyrazów najczęstszych, które były znane uczestniczkom eksperymentu. Natomiast wśród wyrazów



wykorzystanych w trakcie treningu znalazły się trzy wyrazy, które były im nieznane.

Wszystkie wyrazy i logotomy zostały wymówione przez jeden głos męski. Z utrwalonych na taśmie magnetofonowej wypowiedzi sporządzono spektrogramy 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. Dla całego materiału doświadczalnego wykonano spektrogramy z kwantowaną amplitudą, stanowiące ciąg następujących po sobie co 20 ms sekcji widmowych sygnału o zmiennej szerokości zaczerpnienia. Każdy obraz z monitora był fotografowany, a sporządzone zdjęcia posłużyły jako bezpośredni materiał do przeprowadzenia badań.

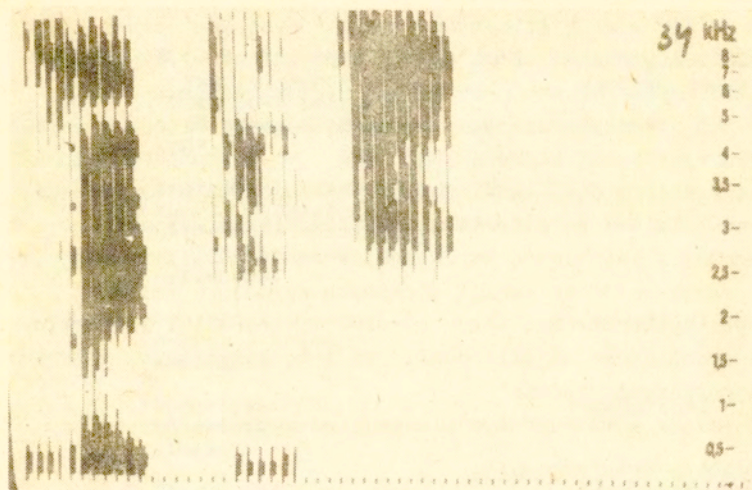
Przykłady spektrogramów widnieją na rycinach 1 - 4.

### 2.3. Przebieg doświadczenia.

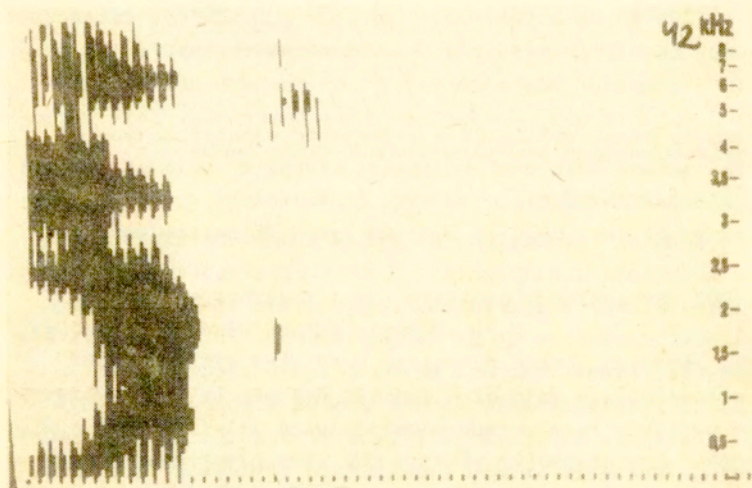
Zanim przystąpiono do właściwego doświadczenia, które w całości przeprowadzono w oparciu o zdjęcia spektrogramów, zaprezentowano trzem uczestniczkom eksperymentu działanie zestawu służącego do uzyskiwania wzorców wizualnych. Dziewczęta obserwowały spektrogramy wyrazów izolowanych bezpośrednio na monitorze zapoznając się wstępnie z ich cechami wizualnymi. Właściwe doświadczenie, obejmujące fazę treningu i fazę testowania, przeprowadzono indywidualnie z każdą osobą.

#### 2.3.1. Faza treningu.

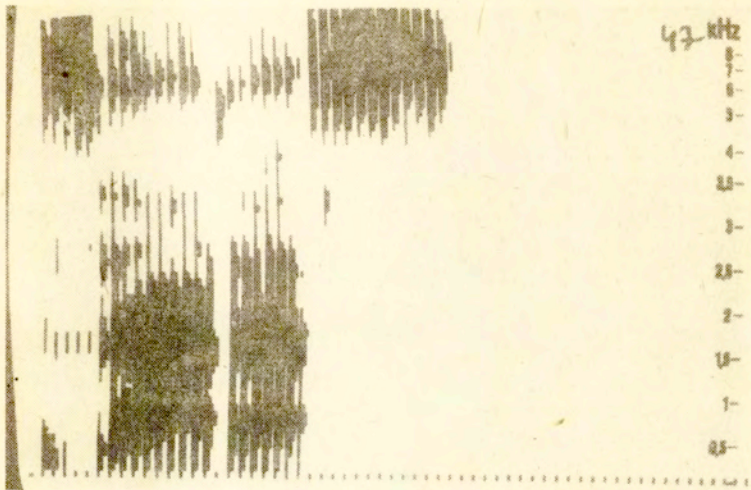
Fazę treningu podzielono na kilka etapów różniących się zakresem materiału fonetycznego. W celu usystematyzowania wiedzy uczestniczek eksperymentu wprowadzono stopniowo nowe klasy głosek, demonstrując je na logatomach. Po zapoznaniu się z charakterystycznymi cechami segmentalnymi klasy głosek uczestniczki identyfikowały żadaną głoskę w zaprezentowanym wyrazie. Następnie otrzymywały kilkuwyrazową listę, przy pomocy której na zasadzie wyboru wymuszonego dokonywały rozpoznania zaprezentowanego spektrogramu.



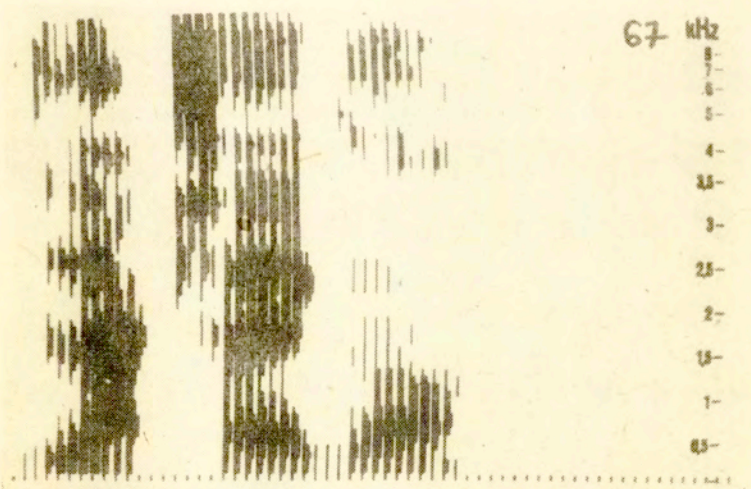
Ryc.1. Spektrogram wyrazu "wypić" /vɨpʲiɕ/.



Ryc.2. Spektrogram wyrazu "jak" /jak/.



Ryc.3. Spektrogram wyrazu "zaraz" /zarás/.



Ryc.4. Spektrogram wyrazu "dlaczego" /dlatʃego/.

Na wstępie wprowadzono wszystkie samogłoski ustne oraz spółgłoski zwarte i trące. Po zapoznaniu się z nimi w logatomach, prezentowano po kilka (4 do 6) spektrogramów z serii T 1 z poleceniem, aby osoby biorące udział w doświadczeniu przeprowadziły w oparciu o podpisane logatomy rozpoznanie segmentów, umieściły pod nimi odpowiednie podpisy, po czym podały cały rozpoznany wyraz. Następnie rozpoznawano spektrogramy z serii T 2 już bez posługiwania się podpisanymi logatomami, natomiast z pomocą list obejmujących po pięć wyrazów z tej serii.

Kolejno wprowadzono klasę spółgłosek nosowych i przeprowadzono rozpoznawanie spektrogramów z serii M (w oparciu o listę wyrazów), dalej spółgłoski /r, l/ - rozpoznawanie z serii L, głoski /j, w/, rozpoznawanie wyrazów z serii J, spółgłoski zwarto-trące, rozpoznawanie wyrazów z serii C, a na końcu wprowadzono samogłoski nosowe.

Na początku każdego spotkania przeprowadzano ćwiczenie polegające na identyfikowaniu określonej głoski na demonstrowanym spektrogramie, przy czym dotyczyło to wszystkich poznanych do tego momentu fonemów.

W tej fazie eksperymentu odbyto z różnymi osobami 7, 8 oraz 9 spotkań trwających od 0,5 do 1 godziny. Nie ograniczono czasu potrzebnego na dokonanie odczytu. W razie popełnienia błędów należało rozpoznawanie przeprowadzić powtórnie, a jeśli zachodziła potrzeba nawet więcej razy, aż do uzyskania właściwej odpowiedzi.

### 2.3.2. Faza testowania.

Sto najczęstszych wyrazów ułożonych w kolejności rangowej zostało podzielonych na 10 bloków 10-wyrazowych. Trzy osoby przeprowadzały rozpoznawanie spektrogramów na zasadzie wyboru wymuszonego w sześciu pierwszych blokach, dwie osoby w ośmiu blokach, jedna osoba w dziesięciu blokach. Odczyty przebiegały w ten sposób, że rozpoznająca otrzymywała listę wyrazów stanowiących blok, po czym dokonując między nimi wyboru, identyfikowała wyraz, którego wzorzec wizualny widniał na spektrogramie.

Poczynając od bloku IV niektóre spektrogramy podawano dwukrotnie do odczytu. Zapobiegało to ułatwianiu sobie przez

uczestniczki zadania poprzez eliminowanie z listy wyrazów już rozpoznanych.

Osoba MS na etapie czterech ostatnich bloków przeprowadzała rozpoznawanie w połączonych blokach 20-wyrazowych, tzn. dokonywała wyboru z listy obejmującej wyrazy z bloku VII i VIII oraz z listy obejmującej wyrazy z bloku IX i X.

Z dwiema osobami przeprowadzono doświadczenie w oparciu o metodę wyboru wymuszonego w połączeniu z utrwalaniem pamięciowym spektrogramów, które przebiegało w następujący sposób: pięć pierwszych wyrazów z bloku rozpoznawano za pomocą listy, następnie te same wyrazy (ułożone w innej kolejności niż poprzednio) rozpoznawano bez listy. Kolejne pięć wyrazów z bloku odczytywano w taki sam sposób, po czym testowano cały blok (10 wyrazów) bez korzystania z listy. Osoba MP odczytywała tą metodą blok IX i X, osoba AW - blok I, II, VI, VII, VIII. AW ponownie dokonywała odczytów bez korzystania z listy w odstępach kilkudniowych (od 1 do 6 dni).

Dwie osoby rozpoznawały wybrane wyrazy na zasadzie wolnego wyboru w oparciu o identyfikację segmentów, z tego osoba MP odczytywała 35 spektrogramów, a osoba AW - 10 spektrogramów.

Faza testowania objęła u osoby MS 4 spotkania, u MP - 5 spotkań, u AW - 8 spotkań.

### 3. Wyniki doświadczenia.

#### 3.1. Faza treningu.

Cwiczenia polegające na identyfikowaniu żądanej głoski w ciągu segmentów, które miały na celu utrwalenie wzorców wizualnych, wykonywane były przez uczestniczki niemal bezbłędnie. Trudności pojawiły się, gdy przystąpiono do identyfikowania całych wyrazów.

Z obserwacji eksperymentatora wynika, że uczestniczki kolejno "doposowywały" wyraz umieszczony na liście do obrazu spektrograficznego. Rozpoznawanie przebiegało kilkustopniowo. Najpierw przeprowadzano segmentację zapisu, która pozwalała wstępnie wyeliminować wyrazy znacznie odbiegające liczbą głosek składowych od liczby wyróżnionych segmentów, następnie w

oparciu o przyswojcną wiedzę o cechach wizualnych fonemów porównywano zgodność zapisu literowego z jego odpowiednikiem akustycznym.

Rozpoznawanie wyrazów rozpoczęto od serii T 2. W całości liczy ona 15 wyrazów, lecz podano ją w trzech zestawach 5-wyrazowych (należało identyfikować spektrogram z jednym z pięciu wyrazów). W tab. 1 podano rezultaty w odniesieniu do poszczególnych serii.

Tab. 1. Wyniki rozpoznawania wyrazów w fazie treningu.

seria wyrazowa	łączna liczba wyrazów rozpoznawanych przez 3 osoby	liczba poprawnych odpowiedzi w %
T 2	15	33
	15	67
	15	80
M	15	93
L	12	100
J	15	93
C	21	67

Największa liczba błędów wystąpiła w pierwszym prezentowanym zestawie z serii T 2 (33 % poprawnych rozpoznań), lecz już przy rozpoznawaniu następnego zestawu nastąpiła znaczna poprawa - poczynając od niego do serii J uzyskano 80-100 % dobrych odpowiedzi. Większa ilość błędów pojawiła się w serii ostatniej (67 % poprawnych rozpoznań), która zawierała spółgłoski zwarto-trące. Przyczyna tego może tkwić z jednej strony w wydłużeniu listy wyrazowej (z 4 lub 5 wyrazów do 7 wyrazów), z drugiej strony w zwiększeniu stopnia trudności przy identyfikacji (ma-

teriał doświadczalny obejmował już wszystkie fonemy języka polskiego). Z 7 wyrazów tylko 1 został właściwie rozpoznany przez wszystkie osoby.

Zaznaczyły się pewne różnice pomiędzy osobami. Na 36 wyrazów rozpoznawanych przez każdą z nich MS uzyskała 75 %, MP - 81 % a AW - 92 % poprawnych odpowiedzi. Większa ilość błędów u osoby MS nie została zapewne spowodowana mniejszą rzetelnością w wykonywaniu ćwiczeń, ale większymi trudnościami w przyswojeniu sobie umiejętności odczytywania spektrogramów. U MS najdłużej trwało podejmowanie decyzji, co spowodowało, że w tej fazie eksperymentu odbyto z nią o jedną lub dwie godziny więcej spotkań niż z pozostałymi uczestniczkami.

### 3.2. Faza testowania.

#### 3.2.1. Rozpoznawanie wyrazów metodą wyboru wymuszonego.

Rozpoznawanie spektrogramów przeprowadzono dokonując wyboru na listach wyrazowych utworzonych ze 100 najczęstszych (subiektywnie) wyrazów w języku polskim. Listy te liczyły 10, 13, lub 20 wyrazów. Listy 10-wyrazowe (podstawowe) utworzono dzieląc liczbę 100 wyrazów na 10 bloków. W listach 13-wyrazowych trzy wyrazy występowały dwukrotnie, listy 20-wyrazowe powstały z połączenia dwóch bloków.

Łączne wyniki dla wszystkich bloków zamieszczono w tab. 2. Bloki I-VI były testowane przez trzy osoby, bloki VII i VIII przez dwie osoby, bloki IX i X przez jedną osobę. Bloki VII-X były również testowane w zestawach 20-wyrazowych. Najwięcej błędów popełniono przy rozpoznawaniu wyrazów z I bloku (63 % poprawnych odpowiedzi). W dalszym ciągu utrzymywał się przybliżony poziom błędów (ponad 80 % poprawnych odpowiedzi), jedynie dla bloku V wyniki uległy pogorszeniu (75 %). Najlepsze wyniki uzyskano dla bloków IX i X testowanych przez osobę AW - ponad 90 % poprawnych rozpoznań. Odegrał tu zapewne rolę wpływ długości treningu. 100 % poprawnych identyfikacji uzyskano tylko w jednym przypadku - podczas rozpoznawania przez AW wyrazów z bloku IV. Opólkem spośród wszystkich 100 wyrazów zidentyfikowano poprawnie 81 % spektrogramów.

Tab. 2. Wyniki rozpoznawania wyrazów w fazie testowania.

Blok	Osoby rozpoznające	Łączna liczba wyrazów rozpoznawanych	Łączna liczba poprawnych rozpoznań	Liczba poprawnych rozpoznań w %
I	MS, MP, AW	30	19	63,3
II	MS, MP, AW	30	24	80,0
III	MS, MP, AW	30	25	83,3
IV	MS, MP, AW	33	27	81,8
V	MS, MP, AW	36	27	75,0
VI	MS, MP, AW	39	32	82,0
VII	MP, AW	26	22	84,6
VIII	MP, AW	26	23	88,5
IX	AW	13	12	92,3
X	AW	13	12	92,3
VII + VIII	MS	20	18	90,0
IX + X	MS	20	15	75,0
OGÓŁEM		316	256	81,0

Kolejne bloki były rozpoznawane tego samego dnia (na jednym spotkaniu testowano jeden do trzech bloków) lub po przerwie trwającej od jednego do dziesięciu dni. Nie zaobserwowano zależności pomiędzy ilością popełnionych błędów a odstępem czasowym izmierzającym odczyty.

Z osobą MS przeprowadzono doświadczenie z rozpoznawaniem wyrazów w blokach 10- oraz 20-wyrazowych. Zwiększenie bloków nie



spowodowało trudności w identyfikacji wyrazów. W blokach 10-wyrazowych MS udzieliła łącznie 79 % poprawnych odpowiedzi, w blokach 20-wyrazowych - 83 %, co wskazuje na nieznaczną poprawę. Dla stwierdzenia, czy różnice w rozkładzie wyników w tych dwóch grupach są statystycznie istotne, przeprowadzono test  $\chi^2$ . Uzyskana wartość

$$\chi^2 = 0,216 < \chi^2 = 3,841$$

nie pozwala odrzucić hipotezy zerowej, że oba rozkłady są jednakowe. Takie same proporcje odpowiedzi poprawnych i niepoprawnych w obu częściach materiału pozwalają uznać, że zwiększenie bloków z 10 do 20 wyrazów nie odbiło się ujemnie na identyfikacji.

W tab. 3 zamieszczono wyniki rozpoznawania uzyskane przez poszczególne osoby. Osoba MS została uwzględniona dwukrotnie (osobno dla bloków 10-wyrazowych i 20-wyrazowych).

Tab. 3. Wyniki rozpoznawania wyrazów dla poszczególnych osób.

Osoby	Liczba rozpoznawanych bloków	Łączna liczba rozpoznawanych wyrazów	Łączna liczba poprawnych rozpoznań	Liczba poprawnych rozpoznań w %
MS	6	66	52	78,8
MS	4 (VII + VIII, IX + X)	40	33	82,5
MP	8	95	75	79,0
AW	10	115	96	83,5
OGÓŁEM		316	256	81,0

Liczba poprawnych identyfikacji dla 3 osób wyniosła od 79 % do 84 %. Przyjęto hipotezę zerową, że rozkłady odpowiedzi dobrych i

złych są jednakowe u wszystkich osób. W rezultacie przeprowadzonego testu  $\chi^2$  obliczona wartość

$$\chi^2 = 0,989 < \chi^2 = 7,815$$

nie pozwala odrzucić hipotezy zerowej. Na tej podstawie można uznać, że wszystkie uczestniczki doświadczenia opanowały w jednakowym stopniu umiejętność odczytywania spektrogramów, a ogólna średnia - 81 % poprawnych identyfikacji wskazuje na pewien pułap możliwości, jaki udało się osiągnąć na danym etapie.

### 3.2.2. Rozpoznawanie wyrazów metoda wyboru wymuszonego w połączeniu z ich utrwalaniem pamięciowym.

Doświadczenie z pamięciowym przyswajaniem spektrogramów przeprowadzono z osobą AW na pięciu blokach oraz z osobą MP na dwóch blokach.

Powtórny odczyt spektrogramów, który przeprowadzono bez posługiwania się listą przebiegał niemal bezbłędnie (na 70 wyrazów 67 rozpoznano poprawnie). Odczyt następował bezpośrednio po pierwszym (przeprowadzonym w oparciu o listę), dzięki czemu wzorce wizualne wyrazów pozostały w pamięci wzrokowej uczestniczek doświadczenia. Dodatkowym ułatwieniem było ograniczenie wyboru do zestawów 5-wyrazowych.

Kolejny odczyt bez listy tych samych spektrogramów wymagał dokonania wyboru spośród 10 zapamiętanych wyrazów. Poszerzenie listy wyboru zwiększyło liczbę popełnionych błędów - poprawnie zidentyfikowano 54 spektrogramy.

Osoba AW jeszcze dwukrotnie odczytywała te same spektrogramy w odstępach kilkudniowych. Wyrazy z I bloku były łatwo przez nią identyfikowane - liczba popełnionych błędów nie przekraczała dwóch. W bloku tym, zawierającym 10 najczęstszych wyrazów języka polskiego, przeważały wyrazy jednosylabowe, o wyrazistej segmentacji i łatwo czytelnym wzorcach. W przypadku pozostałych bloków odstęp czasowy odbijał się niekorzystnie na wynikach. Liczba poprawnych odpowiedzi wynosiła od 1 do 6 na 10 wyrazów z bloku. Rozpoznająca nie zawsze pamiętała, które wyrazy podlegały testo-

waniu, gdyż w niektórych przypadkach podane odpowiedzi nie należały do badanych bloków. Błędne odczyty świadczą o tym, że w percepcji nastawiała się raczej na identyfikowanie segmentów, gdyż na ogół zaproponowane wyrazy były zbliżone pod względem fonetycznym do faktycznie wymówionych - zawierały częściowo takie same fonemy lub fonemy o bardzo podobnych wzorcach wizualnych. Poniżej podano kilka przykładów błędnych odczytów

Wyraz wymówiony		Wyraz odczytany
/bĩtĩ/	—————>	/dźiĩ/
/mjētĩ/	—————>	/mĩtĩ/
/robiĩ/	—————>	/muvitĩ/
/mĩtĩ/	—————>	/jĩtĩ/
/umĩtĩ/	—————>	/mjētĩ/
/mjeĩkajĩ/	—————>	/jĩstetĩ/

Wydaje się, że w przypadku odwołania się do pamięci krótkotrwałej, gdy testowanie bloku następowało bezpośrednio po pierwszych odczytach z listą, percepcja przebiegała drogą globalnego rozpoznawania wzorca wizualnego widocznego na spektrogramie. Gdy natomiast testowanie miało miejsce po co najmniej jednodniowej przerwie, a obraz wzrokowy wzorca ulegał zatarciu w pamięci, osoba odczytująca próbowała identyfikować wyrazy częściowo na drodze rozpoznawania segmentalnego.

### 3.2.3. Rozpoznawanie wyrazów na zasadzie wolnego wyboru.

W tej części doświadczenia korzystano z wyrazów należących do grupy 100 najczęstszych, które zostały już uprzednio poddane rozpoznawaniu w blokach. Ich percepcja przebiegała tak samo, jak w przypadku wyrazów nieznanymi, gdyż po jednorazowym odczycie uczestniczki nie były w stanie zapamiętać wszystkich wyrazów, a tym bardziej utrwalić ich wzorców wizualnych.

Osoba MP rozpoznawała 35 wyrazów, w tym 22 jednosylabowych i 13 dwusylabowych. Odczyty przeprowadzano drogą rozpoznawania segmentalnego - dla każdego segmentu w wypowiedzi uczestniczka podawała pisemnie odpowiadającą mu głoskę lub więcej niż jedną głoskę, jeśli nie potrafiła jej jednoznacznie określić. W razie

poprawienia błędu prowadząca doświadczenie polecała wnieść poprawkę do zapisu. Po zidentyfikowaniu głosek składowych podawano cały wyraz reprezentowany przez spektrogram.

Odczytywanie wzorców wizualnych tą metodą okazało się najtrudniejszą częścią doświadczenia. Uzyskanie poprawnych odpowiedzi (poprawna identyfikacja wyrazu) wymagało bardzo częstych interwencji eksperymentatora. Niemniej pewna niewielka część wyrazów (ok. 26 %) została rozpoznana całkowicie samodzielnie i bez jakichkolwiek trudności. Odczytująca nie zapisywała nawet głosek składowych, lecz podawała od razu cały wyraz. Dotyczyło to ośmiu wyrazów jednosylabowych i jednego wyrazu dwusylabowego. Pozostałe wyrazy zostały zidentyfikowane z pomocą eksperymentatora. W przypadku popełnionych błędów w rozpoznawaniu segmentów zdarzyło się, że w ponownie przeprowadzonym odczycie znów po raz kolejny popełniono błąd i należało powtórzyć odczyt. Kilkakrotnie uczestniczka nie potrafiła określić, jakiej głosce odpowiada dany segment.

Prawidłowa identyfikacja, nie wymagająca poprawek, objęła 52 % wszystkich samogłosek i 34 % spółgłosek. Najlepiej rozpoznawano /a/ oraz /j/, podobnie jak w doświadczeniu przeprowadzonym przez autorkę z osobami o normalnym słuchu.

Osoba AW odczytywała 10 wyrazów obejmujących jedną lub dwie sylaby. Całkowicie samodzielnie rozpoznała 4 wyrazy, z pomocą eksperymentatora 5 wyrazów, 1 wyraz rozpoznała błędnie. W przypadku tego odczytu wyraz /muvitç/ został odebrany jako /muçif/ - niewłaściwa identyfikacja dotyczyła więc tylko części segmentów.

Eksperyment z rozpoznawaniem wzrokowym spektrogramów dowolnych wyrazów, polegających na identyfikowaniu kolejnych segmentów, został tu potraktowany bardzo wrywkowo. Autorka zamierzała wstępnie zorientować się, czy percepcję optycznych wzorców mowy przez osoby niesłyszące można oprzeć na rozpoznawaniu segmentalnym. Pozytywne rezultaty uzyskane w doświadczeniu nawet w tak skromnym zakresie stanowią odpowiedź twierdzącą na postawione pytanie.

Porównanie wyników uzyskanych w opisywanym doświadczeniu

(MP - 26 % poprawnie zidentyfikowanych wyrazów, AW - 40 %) z wynikami pracy badającej percepcję osób o normalnym słuchu (55 % poprawnie rozpoznanych wyrazów) wypada na niekorzyść głuchych. Złożyły się na to co najmniej dwie przyczyny

- 1) brak w procesie uczenia niezwykle istotnego czynnika, jakim jest analiza popełnianych błędów. Po pierwszych próbach eksperymentator z niej zrezygnował ze względu na konieczność stosowania obszernego komentarza słownego, co dawało wątpliwe efekty w warunkach utrudnionej komunikacji między osobą uczestniczącą w doświadczeniu i eksperymentatorem.
- 2) W sytuacji, gdy nie przeprowadzano analizy błędów, czas trwania treningu oraz wielkość materiału doświadczalnego wykorzystanego w fazie uczenia okazały się niewystarczające dla pełnego uświadomienia sobie systemu optycznych wzorców mowy.

#### 4. Wnioski.

Doświadczenie przeprowadzone w niniejszej pracy miało udzielić odpowiedzi na następujące pytania :

- 1) Czy metoda wizualizacji sygnału mowy opracowana i stosowana w Pracowni Fonetyki Akustycznej okazuje się przydatna dla celów rozpoznawania wzrokowego mowy ze spektrogramów przez osoby głuche ?
- 2) Czy identyfikacja spektrogramów w przypadku osób głuchych może opierać się na rozpoznawaniu segmentalnym ?
- 3) Czy zachodzą istotne różnice w przebiegu percepcji wzrokowej mowy pomiędzy głuchymi a osobami o normalnym słuchu ?

Ze względu na pilotażowy charakter pracy wyniki eksperymentu mogą stanowić zaledwie wstępną ocenę postawionych tu zagadnień. Przebieg doświadczenia i jego rezultaty prowadzą do następujących wniosków :

- 1) Spektrogramy komputerowe o kilku stopniach kwantowania amplitudy, okazują się równie dobrym środkiem przekazywania optycznych wzorców mowy w przypadku osób głuchych, jak osób słyszących. Wydaje się, że schematyczny charakter wzorców szczególnie je predestynuje do stosowania w sytuacjach, gdy wiedzę fonetyczno-akustyczną potrzebną do rozpoznawania segmentów przekazuje się w bardzo ogólnych zarysach, a więc gdy rozpoznawania podejmują się

osoby niezwiązane profesjonalnie z badaniem mowy.

2) Zastosowana w poprzedniej pracy autorki [10] metoda uczenia cech segmentalnych, polegająca na stopniowym poszerzaniu materiału fonetycznego o kolejne klasy głosek, okazała się dobrze przyswajalna przez uczestniczki doświadczenia.

3) Rozpoznawanie mowy na spektrogramach przez osoby głuche można oprzeć na identyfikacji segmentów fonetyczno-akustycznych. Wymaga to z pewnością ze strony odczytującego większego zaangażowania intelektualnego, aniżeli globalne rozpoznawanie wyrazów wyuczonych, zapewnia jednak większe możliwości korzystania z osiągniętej umiejętności. Doświadczenie zostało zaprogramowane w taki sposób, że poprawne jego wykonanie wymagało identyfikowania kolejnych segmentów. W części eksperymentu, polegającej na rozpoznawaniu wyrazów na zasadzie wyboru wymuszonego, odczytująca musiała dokonywać analizy cech fonetyczno-akustycznych wyróżnionych przez siebie segmentów, po czym porównać własne wyobrażenie o postaci dźwiękowej badanego wyrazu z wyrazami znajdującymi się na liście.

Rozpoznawanie kilkudziesięciu wybranych wyrazów na zasadzie wolnego wyboru opierało się wyłącznie na identyfikacji poszczególnych segmentów i pomimo znacznego stopnia trudności, pozwoliło uzyskać 29 % poprawnych rozpoznań.

Doświadczenie polegające na ponownym rozpoznawaniu odczytywanych już spektrogramów sugeruje, że odczyty wyrazów dokonywane po dłuższej przerwie przebiegają częściowo na zasadzie rozpoznawania globalnego, częściowo segmentalnego.

4) Z uwagi na to, że w procesie uczenia trzeba zrezygnować z analizy błędów, przynajmniej w takim zakresie, jaki można stosować u osób słyszących, postuluje się wydłużenie fazy treningu przy równoczesnym wzbogaceniu materiału uczącego.

5) Rozpoznawanie nieznanymi wyrazów poprzez identyfikację segmentów napotyka u osób głuchych na dodatkowe trudności wynikające z nieświadomości sobie różnic pomiędzy pisownią a wymową. Ponieważ uczestniczkom znana była wyłącznie postać ortograficzna wyrazu, starały się je wiernie naśladować w swej wymowie. Problem ten dotyczył zwłaszcza wypowiedzi, w których spółgłosce bezdźwięcznej odpowiada w zapisie ortograficznym litera będąca symbolem spółgłoski dźwięcznej, np. /xlep/ - chleb.

6) W trakcie rozpoznawania dowolnych wyrazów najlepiej identyfikowano głoski /a/ oraz /j/. Taką samą tendencję stwierdzono u osób słyszących. Blizsze porównanie dotyczące percepcji poszczególnych głosek przez osoby głuche oraz słyszące nie są w chwili obecnej możliwe do przeprowadzenia z uwagi na szczupły materiał doświadczalny.

7) Znacznie większą rolę niż u słyszących odgrywa częstość występowania wyrazu w języku. Uczestniczki doświadczenia posiadały ubogi zasób pojęć, zdarzało się, że nie знаły wyrazu dość szeroko rozpowszechnionego w języku. W tej sytuacji należy szczególnie uwzględnić częstość występowania wyrazów, które miałyby być poddane odczytowi.

8) Czynnikiem bardzo ważnym dla sprawnego przeprowadzenia doświadczeń z osobami głuchymi jest ich psychiczne nastawienie do wykonywanego zadania. Niepowodzenia w silniejszym stopniu zniechęcały do kontynuowania eksperymentu, niż miało to miejsce u osób słyszących. Osoby głuche szybciej również odczuwały zmęczenie. Wydaje się, że ważnym elementem w pracy z nimi jest odpowiednio silna motywacja.

Przebieg i wyniki doświadczenia wykazały, iż proces rozpoznawania wizualnego akustycznych wzorców mowy przebiega w sposób analogiczny u osób głuchych i słyszących. Uzyskanie pozytywnych rezultatów w pracy z głuchymi wymaga jednak pokonywania specyficznych trudności, które ujawniły się w trakcie eksperymentu. Dalsze badania będą miały na celu określenie wpływu czynników odgrywających rolę w procesie percepcji wzrokowej sygnału mowy w oparciu o spektrogramy komputerowe, takich jak : wielkość zbioru słownikowego, długość rozpoznawanych wyrazów oraz czas trwania treningu.

BIBLIOGRAFIA

- [1] COLLE R., RUENICKY A., ZUE V., REDY L., Speech as patterns on paper in : Perception and Production of Fluent Speech, P.Cole, ed. Hillsdale, 1980, 3-50.
- [2] GREENE B., PISONI D., CARELL T., Identification of speech spectrograms : comparisons of naive and trained observers, Research on Speech Perception, Progress Report No 9, 1983, Indiana University, 267-283.
- [3] GREENE B., PISONI D., CARELL T., Recognition of speech spectrograms, JASA, 1985, vol. 76 No 1, 32-43.
- [4] IMIOLCZYK J., Subiektywne prawdopodobieństwo wyrazów. Podstawowy słownik frekwencyjny języka polskiego (praca w druku).
- [5] JOHNSON S., CONNOLLY J., EDMONIS E., Spectrogram analysis : a knowledge-based approach to automatic speech recognition, in : Research and Development in Expert Systems, ed. M.Bramer, Cambridge 1985, 95-103.
- [6] 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.
- [7] LINDBLOM B., SVENSSON S., Interaction between segmental and nonsegmental factors in speech recognition, IEEE, vol. AU-21 No 6, 1973, 536-545.
- [8] POTTER R., KOPP G., GREENE H., Visible Speech, New York 1947.
- [9] RICHTER L., Wizualne rozpoznawanie samogłosek polskich w prostych kontekstach spółgłoskowych na podstawie spektrogramów komputerowych, Prace IPPT 18/1985.
- [10] RICHTER L., Wizualne rozpoznawanie wybranych wyrazów w oparciu o informacje segmentalne zawarte w spektrogramach komputerowych (praca oddana do druku).
- [11] ZUE V., Acoustic-phonetic knowledge representation : implications from spectrogram reading experiments, in: Automatic Speech Analysis and Recognition, ed. J.P. Haton, 1982, 101-120.