

R. Ciarkowski, J. Imiołczyk

SYNTEZA WYBRANYCH  
WYRAZÓW POLSKICH  
I ICH OCENA PERCEPCYJNA

1/1985

P-269



WARSZAWA 1985

ISSN 0208-5658

Praca wpłynęła do Redakcji dnia 20 listopada 1984 r.



56948



N a p r a w a c h r ę k o p i s u

---

Instytut Podstawowych Problemów Techniki PAN  
Nakład 160 egz. Ark.wyd. 2,1. Ark. druk. 3.  
Oddano do drukarni w styczniu 1985 r.  
Nr zamówienia 18/85

---

Warszawska Drukarnia Naukowa, Warszawa,  
ul.Śniadeckich 8

Ryszard Ciarkowski  
Janusz Imiołczyk  
Pracownia Fonetyki Akustycznej  
IPPT PAN

## SYNTEZA WYBRANYCH WYRAZÓW POLSKICH I ICH OCENA PERCEPCYJNA<sup>1)</sup>

### Streszczenie

W oparciu o zbiór polskich diad syntetycznych typu CV i VC powstałych z połączeń spółgłosek /j/, /l/, /w/, /m/, /n/, /ɲ/, /r/ oraz /v/ z samogłoskami /i/, /ɛ/, /e/, /a/, /o/, /u/ [2] dokonano syntezy 77 wyrazów języka polskiego. Eksperyment przeprowadzono przy wykorzystaniu zestawu minikomputerowego MERA 303 (wraz z wyspecjalizowanymi urządzeniami peryferyjnymi), pracującego pod kontrolą systemu programowego SPOS1 ([1], [2]). Uzyskany zbiór wyrazów syntetycznych poddano ocenie percepcyjnej przez 25 słuchaczy. Obok prezentacji metodologii eksperymentu i opisu otrzymanego zbioru wyrazów syntetycznych praca zawiera również próbę sformułowania ogólnych zasad stosowanych w tym eksperymencie przy syntezie wyrazów.

#### 1. Techniczne podstawy syntezy.

Syntezę realizowano przy wykorzystaniu rozszerzonego zestawu minikomputerowego MERA 303 pracującego pod kontrolą systemu programowego SPOS1 (System Programowej Obsługi Syntetyzatora). Obok jednostki centralnej MOMIK 8B (z pamięcią operacyjną 8 kbajtów) oraz typowych urządzeń peryferyjnych [1], takich jak :

- zestaw : drukarka znakowo-mozaikowa DZM 180 i klawiatura operatorska (alfanumeryczna i numeryczna),
  - moduł pamięci na dyskach elastycznych MDE-300 (4x250 kbajtów)
  - perforator taśmy (DT-105) i czytnik taśmy (CT 1001A),
- rozszerzony zestaw minikomputerowy obejmuje również wyspecjali-

<sup>1)</sup> Praca wykonana w ramach planu C-1



zowane urządzenia peryferyjne ([1], [2]):

- 63-kanałowy analogowy analizator widmowy (zakres 80-8310 Hz),
- kanał funkcji analogowych KF-01 (zawierający m.in. 8-bitowy konwerter analogowo-cyfrowy),
- formantowy syntetyzator mowy COMPUTALKER CT-1 oraz
- monitor graficzny MEMOSKOP (wielkość obrazu 64x256 punktów).

#### 1.1. Rola analizy widmowej.

Istniejąca w systemie SPOS1 możliwość wykonywania na bieżąco analizy widmowej zarówno mowy syntetycznej jak i naturalnej (rozd. 1.3. [2]) okazała się bardzo pomocna w syntezie wyrazów. Uzyskiwane w wyniku analizy spektrogramy, wyświetlane na ekranie monitora graficznego pozwalały na bieżącą, zobiektywizowaną kontrolę wyniku syntezy oraz porównanie mowy syntetycznej i mowy naturalnej (patrz Dodatek B), wspomagając tym samym ocenę słuchową. Było to szczególnie istotne dla syntezy stanów przejściowych pomiędzy stanami ustalonymi głosek w wyrazie.

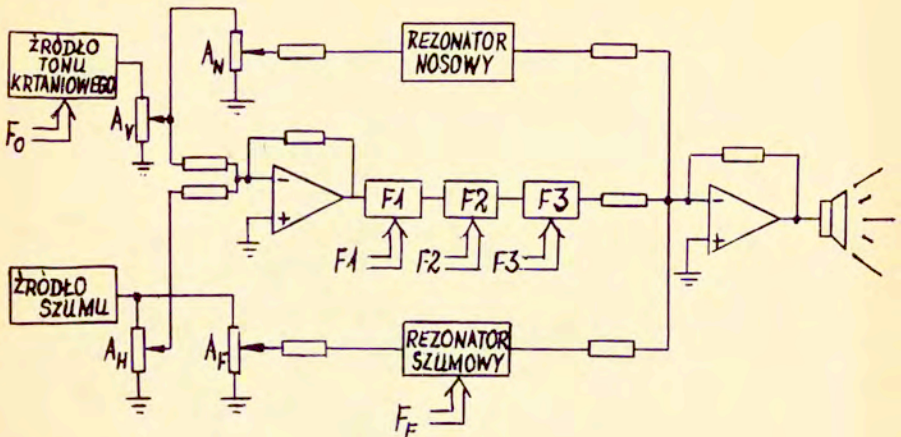
Dla oceny wyniku syntezy główną i zasadniczą podstawę - pomimo znaczącej roli analizy widmowej w procesie syntezy - stanowiła ocena słuchowa. Wynikało to z faktu istnienia rozbieżności pomiędzy wartościami parametrów charakteryzujących mowę naturalną a wartościami tych parametrów koniecznymi dla uzyskania mowy syntetycznej o takim samym brzmieniu (tzn. o brzmieniu najlepiej naśladowującym naturalne). Taki stan rzeczy należy tłumaczyć przede wszystkim niedoskonałością syntetyzatora.

#### 1.2. Użytkowe własności syntetyzatora mowy.

Opis użytego w eksperymencie formantowego syntetyzatora mowy COMPUTALKER CT-1 o strukturze przedstawionej na rysunku 1.1., podany został w rozdz. 2.1. pracy [1]. Na podstawie przeprowadzonych dotychczas eksperymentów z syntezą dźwięków i wyrazów sformułować można następujące uwagi użytkowe o własnościach tego syntetyzatora (uwagi uszeregowano w sposób nawiązujący do 3-torowej struktury syntetyzatora):

- tor formantowy

Poziom  $F_3$  jest znacznie niższy od poziomów  $F_1$  i  $F_2$ , w związku z czym wartość częstotliwości  $F_3$  nie wpływa znacząco na charakter i jakość wyniku syntezy. Oznacza to w praktyce ograniczenie



Rys. 1.1. Schemat blokowy formantowego syntetyzatora mowy COMPUTALKER CT-1.

sterowania parametrami toru formantowego do dwóch pierwszych formantów. Kolejne ograniczenie stanowi założona przez producenta syntetyzatora stała szerokość pasma dla poszczególnych formantów. Ze względu na szeregową strukturę toru formantowego i niedoskonałości konstrukcyjne, przy sterowaniu częstotliwościami poszczególnych formantów uwidaczniają się występujące między nimi niepożądane zależności.

- tor nosowy

Powstający w torze nosowym efekt akustyczny nie pogłębia wrażenia nazalizacji produkowanej głoski, a wręcz pogarsza jej brzmienie. Z tego powodu uznano tor nosowy w takiej wersji realizacyjnej za całkowicie zbędny w procesie syntezy. Należy w tym miejscu zaznaczyć, że mowa syntetyczna produkowana przy użyciu syntetyzatora COMPUTALKER CT-1 generalnie charakteryzuje się nosowością i małym odstępem od szumu.

- tor szumowy

Sterowanie formantem szumowym (czyli filtrem górnoprzepustowym składającym się na tor szumowy) jest nieefektywne. Przyczyną tego może być niewłaściwe widmo szumu i niska jego energia lub



zastosowanie nieodpowiedniego filtra górnoprzepustowego. Produkcowane przy użyciu syntetyzatora głoski szumowe (w stadium prób) różnią się od pozostałych głosek zdecydowanie innym charakterem brzmienia.

Ze względu na wykonanie syntetyzatora w technice hybrydowej oraz brak informacji szczegółowych na temat sposobu realizacji poszczególnych elementów syntetyzatora nie istnieje możliwość wprowadzenia modyfikacji układowych jak również precyzyjnego określenia przyczyn jego nieprawidłowego funkcjonowania. Dla przeprowadzenia usystematyzowanych eksperymentów z mową syntetyczną konieczne byłoby użycie syntetyzatora :

- z możliwością elastycznego sterowania amplitudą i szerokością pasma dla poszczególnych formantów,
- zawierającego w pełni sprawny kanał szumowy oraz generator produkujący szum o odpowiednim widmie i odpowiedniej energii.

### 1.3. Nagrania mowy syntetycznej.

Ze względu na ograniczony czas trwania syntetyzowanej wypowiedzi (SPOS - ok. 3 s, SPOS1 - ok. 1 s) oraz długi czas dostępu do materiału składowanego w pamięci dyskowej opracowano modyfikację systemu SPOS1 pozwalającą na przeprowadzenie syntezy dowolnie długiej wypowiedzi zapisanej na taśmie perforowanej. Zmodyfikowany system SPOS1 wykorzystywany jest do realizacji nagrań dla potrzeb oceny odsłuchowej.

Realizacja nagrania polega na :

- a/ przeniesieniu zapisu materiału dźwiękowego z dysków elastycznych na perforowaną taśmę papierową. Obok samego materiału dźwiękowego (w postaci parametrów sterujących syntetyzatorem) na taśmie dokonywany jest zapis sterujący zawierający informacje o długości wypowiedzi oraz o długościach przerw pomiędzy kolejnymi wypowiedziami.
- b/ syntezie sterowanej wartościami parametrów pobranymi bezpośrednio z taśmy lub z pamięci operacyjnej minikomputera, w której zapis wartości parametrów jest cyklicznie dokonywany na podstawie odczytu taśmy perforowanej. Drugi przypadek umożliwia jednorazowe przechowywanie wartości parametrów ste-

rujących fragmentu wypowiedzi o długości ok. 1 s i szybkie ich modyfikowanie przy użyciu taśmy perforowanej jako "pamięci zewnętrznej" o krótkim czasie dostępu.

## 2. Synteza wyrazów.

Synteza na poziomie wyrazów stanowi kolejny krok w pracach mających na celu stworzenie podstaw i opracowanie zasad całościowej, nie poddanej żadnym ograniczeniom syntezy mowy, której produkt spełniałby wszelkie istotne wymogi (zrozumiałości, naturalności) i nie odbiegałby jakościowo (pod względem percepcyjnym) od mowy naturalnej. Ta część badań stanowi logiczną kontynuację prac nad konstrukcją diad syntetycznych ([2]) i jest pierwszym krokiem dokonanym na gruncie mowy syntetycznej w dosłownym rozumieniu tego pojęcia. Opisany poniżej eksperyment jest w zamysle pracą pilotażową, mającą na celu weryfikację zasad syntezy przyjętych w poprzednim doświadczeniu oraz sformułowanie dalszych zasad ogólnych i szczegółowych, niezbędnych dla syntezy mowy ciągłej.

Na obecnym etapie badań zrezygnowano z realizacji pewnych celów, istotnych z punktu widzenia globalnej syntezy, odkładając ją do czasu, gdy stopień zaawansowania prac oraz dostępne środki techniczne uczynią ją możliwą i nieodzowną. I tak np., ze względu na brak możliwości kształtowania charakterystyki tonu krtaniowego, zrezygnowano z uzyskania efektu optymalnej naturalności brzmienia wyrazów syntetycznych, kładąc główny nacisk na ich zrozumiałość. Wychodząc z założenia, iż intonacja jest w zasadzie cechą mowy ciągłej, zrezygnowano również z różnicowania przebiegów  $F_0$ , przyjmując dla wszystkich wyrazów (poza ich absolutnym początkiem i końcem) stałą wartość tego parametru, równą 122 Hz (było to możliwe, gdyż - jak wykazała praktyka - modyfikacja  $F_0$  w obrębie poszczególnych głosek nie miała - poza przypadkiem głoski /v/ - istotnego wpływu na ich wyrazistość).

### 2.1. Dobór materiału badawczego.

Część wstępną pracy stanowił wybór zestawu wyrazów. Stopień dowolności w selekcji był ograniczony ze względu na niekompletność zbioru pozostających do dyspozycji głosek i diad syntetycznych, będących efektem poprzedniego doświadczenia ([2]). Ostateczny



wybór musiał być dokonany spośród wyrazów, w skład struktury fonetycznej których wchodziły głoski : /i, ɛ, e, a, o, u, m, n, j, l, r, ʃ, w/ oraz /v/. Zestaw pierwotny liczył ok. 150 takich wyrazów. Został on następnie zredukowany w oparciu o kilka kryteriów. Zwrócono przede wszystkim uwagę na to, aby w ostatecznym zbiorze znalazły się wyrazy o w miarę zróżnicowanej strukturze fonetycznej, zwłaszcza zawierające różne typy połączeń głoskowych. Włączono także do niego te wyrazy, w skład których wchodziły stosunkowo rzadkie połączenia głosek, odrzucając jednocześnie pewną liczbę wyrazów zawierających połączenia typowe, bardzo częste, których synteza nie nastręczałaby specjalnych trudności. W miarę możliwości starano się, aby w ostatecznym zestawie poszczególne głoski występowały zbliżoną liczbą razy. Ze względu na ograniczony wybór zamierzenia tego nie można było jednak w pełni zrealizować.

Kolejnym kryterium, które wzięto pod uwagę, była frekwencja wyrazów. Uwzględniając ewentualną przydatność zsyntetyzowanych wyrazów zdecydowano, że zestaw powinien zawierać słowa powszechnie znane i dosyć często używane. Wymóg ten był także istotny ze względu na konieczność przeprowadzenia badań odsłuchowych w celu uzyskania oceny percepcyjnej wyrazów syntetycznych. Z listy pierwotnej wyeliminowano w związku z tym wyrazy o niskiej frekwencji, które mogłyby nie być znane wszystkim respondentom (np. manuał, anomia itp.). Do zestawu doświadczalnego włączono ostatecznie 77 następujących wyrazów :

aleja	lojalny	mowa	rój
areał	ława	mój	równy
armia	maj	my	ruina
awaria	major	mylny	rura
Emilia	mała	naiwny	rywal
ile	mam	nerwy	ulewa
Iwona	mama	niania	umarł
ja	manewry	niema	wanna
jawny	Mania	norma	Wawel
jem	mania	normalny	wielna
Julia	manna	nowy	wiara
lala	Marian	numer	wina



lawa	marny	oliwa	wojna
Lenin	miał	oni	wolny
leń	miara	raj	wrona
lewy	milion	rama	wuj
lilia	miły	ranny	wy
lina	mina	rano	
linia	Mława	rola	
lniany	młyn	rower	

W zestawie zdecydowanie przeważają wyrazy dwusylabowe (52), liczby wystąpień wyrazów jedno- i trzysylabowych są natomiast zbliżone (odpowiednio : 13 i 12). Na liście znalazło się kilka nazw własnych oraz kilka wyrazów, którym nadano formę różną od podstawowej (słownikowej).

### 2.2. Sposób realizacji syntezy.

Kolejny etap pracy obejmował właściwą syntezę przedstawionego wyżej zestawu 77 wyrazów. Syntezę przeprowadzono w oparciu o bibliotekę diał syntetycznych (zob. [2]), złożoną z szeregu zapisów parametrów sterujących syntezą dla poszczególnych diał, wykonanych na dysku elastycznym zgodnie z zasadami pojętymi w rozdziale 3.5. pracy [1]. Wyrazy syntetyzowane poprzez łączenie bądź nakładanie na siebie diał (sposób wykonania tej czynności opisano w rozdziale 3.5. i 5 pracy [1]), co było zabiegiem stosunkowo nieskomplikowanym. Zasadniczy problem polegał na dobraniu odpowiednich wartości poszczególnych parametrów (amplituda i częstotliwość tonu krtaniowego, częstotliwości formantowe oraz czas trwania) w obrębie przejść między głoskami (tzn. pomiędzy końcową głoską diady poprzedzającej a początkową głoską diady następującej), a także na określeniu czasu trwania stanu ustalonego poszczególnych głosek w wyrazie.

### 2.3. Zasady ogólne syntezy wyrazów.

Na wstępie przyjęto kilka ogólnych zasad syntezy, według których miały być tworzone poszczególne "wypowiedzi". Podobnie jak w poprzedniej pracy ([2]) uznano, że wartości amplitudy w obrębie absolutnego początku i absolutnego końca każdego z wyrazów (zwykle : na przestrzeni czterech fram) powinny odpowiednio

stopniowo narastać (od wartości 1 dB<sup>1)</sup>) i stopniowo maleć (do wartości 1 dB), co miało na celu wyeliminowanie niepożądanych efektów akustycznych w postaci stuków. Typowy wzrost wartości amplitudy na początku wyrazu kształtował się więc następująco : 1, 5, 10, 16 ... [dB]. W czterech końcowych framach wyrazu wartości amplitudy były identyczne, tyle że występowały oczywiście w odwróconym porządku (16, 10, 5, 1 [dB]).

Analogiczne rozwiązanie przyjęto w odniesieniu do parametru  $F_0$ . W obrębie trzech pierwszych fram wartość tego parametru stopniowo wzrastała (odpowiednio : 80, 92, 109 [Hz]), w trzech końcowych zaś opadała, osiągając w ostatniej poziom 80 Hz. Poza obszarem początkowym i końcowym wyrazów wartość  $F_0$  nie była różnicowana i wynosiła 122 Hz. Jedyne odstępstwo od tej zasady stanowił przypadek głoski /v/, której wyrazistość była optymalna przy częstotliwości podstawowej równej 110 Hz.

Pewne ogólne reguły przyjęto również w odniesieniu do iloczasu, wykorzystując jako ich podstawę wyniki pracy Richter [8].

Główny nacisk położono na dobranie odpowiednich czasów trwania dla samogłosek akcentowanych, dostosowując do nich następnie iloczasy pozostałych głosek w obrębie poszczególnych wyrazów. Ustalono, że przeciętny czas trwania samogłoski akcentowanej powinien wynosić :

- w wyrazie 1-sylabowym : ok. 150-180 ms
- w wyrazie 2-sylabowym : ok. 130-150 ms
- w wyrazie 3-sylabowym : ok. 120-140 ms.

Przyjęte w niniejszej pracy wartości częstotliwości formantowych dla stanów ustalonych poszczególnych głosek są zasadniczo takie same, jak w poprzednim doświadczeniu (stanowi to konsekwencję wykorzystania przy syntezie wyrazów gotowych diad syntetycznych typu CV i VC). Z oczywistych względów wprowadzono natomiast pewne modyfikacje w zakresie amplitudy. W związku z rezygnacją z różnicowania przebiegów parametru  $F_0$ , percepcyjny efekt akcentu wyrazowego można było uzyskać jedynie poprzez odpowiedni dobór wartości amplitudy (oraz iloczasu) w obrębie

---

<sup>1)</sup> względem przyjętego poziomu odniesienia 0 dB .



poszczególnych głosek składających się na dany wyraz.

#### 2.4. Synteza przejść międzygłoskowych.

Zasadniczy problem syntezy polegał na ustaleniu optymalnej konfiguracji wartości poszczególnych parametrów w obrębie przejść między głoskami. Wykorzystano w tym celu wyniki uzyskane w poprzednim doświadczeniu ([2]), jak również dane kilku prac, w których analiza transientów stanowiła jedno z centralnych zagadnień ([3], [5], [6], [7]). Dużym udogodnieniem była również możliwość porównania uzyskiwanych na bieżąco i wyświetlanych na ekranie monitora spektrogramów wyrazu syntetycznego i naturalnego.

Jako dane wyjściowe przyjęto ustalenia zawarte w cytowanych wyżej publikacjach. Praktyka wykazała jednakże, iż - zapewne ze względu na specyficzne własności wykorzystywanego syntetyzatora (COMPUTALKER CT-1) - ścisłe przestrzeganie zasad fonetyczno-akustycznych nie stanowiło gwarancji uzyskania w pełni zadowalających efektów. W niektórych przypadkach zrozumiałość syntetyzowanych w myśl tych zasad wyrazów była co najmniej wątpliwa. W tej sytuacji ostatecznie zdecydowano się potraktować dane literaturowe jako pomocnicze źródło informacji, opierając się przede wszystkim na porównaniu spektrogramów wyrazów naturalnych i syntetycznych.

Wśród ogółu zsyntetyzowanych transientów wyróżnić można dwa zasadnicze typy. Pierwszy z nich, dominujący liczebnie, obejmuje takie przejścia międzygłoskowe, w obrębie których wzrost (lub spadek) częstotliwości formantowych w kolejnych 10-milisekundowych przedziałach czasowych (= framach) jest równomierny<sup>1)</sup>. Jeżeli, przykładowo, częstotliwość docelowa  $F_2$  głoski poprzedzającej wynosi 1300 Hz, a głoski następującej - 1750 Hz, to przy wymaganej długości ugięcia formantowego równej 40 ms przyrost częstotliwości w kolejnych framach przejścia będzie miał wartość 90 Hz (... 1300 || 1390 1480 1570 1660 || 1750 ... [Hz]).

Drugi typ obejmuje przejścia, na przestrzeni których zmiany częstotliwości formantowych (ściślej:  $F_2$ ) nie są równomierne. Przejścia tego typu występują w niniejszym materiale

---

<sup>1)</sup> Przy konstrukcji wszystkich przejść łączono  $F_1$ ,  $F_2$  i  $F_3$  głoski poprzedzającej odpowiednio z  $F_1, F_2$  i  $F_3$  głoski następującej.



przy połączeniach : samogłoska + /j/. Przyrost częstotliwości jest największy w obrębie 4-5 środkowych fram takiego przejścia, nieco mniejszy - w 3-4 framach początkowych, a najmniejszy - w 3-4 końcowych. Przykładowo, ugięcie  $F_2$  w wyrazie mój ma następujący przebieg : 800 || 950 1100 1250 1400 1600 1800 2000 2200 2400 2500 2600 2700 2800 || 2900 [Hz].

Należy tutaj zaznaczyć, że opisany powyżej, nierównomierny wzrost częstotliwości w obrębie przejść międzygłoskowych odnosi się wyłącznie do  $F_2$ , którego przebieg wpływał w sposób decydujący na jakość brzmienia syntetyzowanych wyrazów. W związku z tym, że ugięcia dwóch pozostałych formantów ( $F_1$  i  $F_3$ ) nie odgrywały tak istotnej roli, występujące w ich obrębie zmiany częstotliwości są we wszystkich przypadkach równomierne.

Długość przejść międzygłoskowych zależała oczywiście nie tylko od typu łączonych głosek, ale i od ich liczby (liczby sylab) w ramach poszczególnych wyrazów. Nie można w tej sytuacji dokonywać uogólnień co do średnich długości ugięć dla różnych typów połączeń (np. : /v/ + samogłoska, samogłoska + /r/ itp.). Można natomiast z pewnością stwierdzić, iż zdecydowanie najkrótsze (zwykle ok. 10-20 ms) były w niniejszym materiale przejścia typu :

$$\left. \begin{array}{l} /m/ \\ /n/ \end{array} \right\} + \text{samogłoska} \text{ oraz } \text{samogłoska} + \left\{ \begin{array}{l} /m/ \\ /n/ \end{array} \right\}$$

najdłuższe zaś - przejścia między samogłoską (poprzedzającą lub następującą) a :

$$/j/ \text{ (do 140 ms), } /w/ \text{ (do 140 ms) oraz } /j/ \text{ (do 120 ms).}$$

W obrębie najkrótszych ugięć przyrosty (lub spadki) częstotliwości były naturalnie największe:

Dodatek A niniejszej pracy zawiera zestawienie docelowych wartości poszczególnych parametrów sterujących syntezą ( $F_1$ ,  $F_2$ ,  $F_3$ , AV) oraz dane dotyczące iloczasów głoskowych. Poza przypadkiem połączeń : samogłoska + /j/ daje to możliwość ustalenia, jaką wartość chwilową miał dowolny parametr w dowolnym momencie czasowym.

W Dodatku B zamieszczone są fotografie spektrogramów kilku wybranych par wyrazów syntetycznych i naturalnych. Mimo, iż w obrębie par uwidaczniają się na spektrogramach wyraźne różnice,

uzyskane efekty akustyczne są - w ocenie percepcyjnej - bardzo zbliżone, o czym świadczy 100-procentowo poprawna identyfikacja wszystkich przedstawionych w Dodatku wyrazów syntetycznych.

### 3. Percepcyjna ocena wyrazów syntetycznych.

Podstawę do zobiektywizowanej oceny produktu syntezy stanowiły wyniki badań odsłuchowych, w których zadaniem respondentów (łącznie 25 osób) było zapisanie w wersji ortograficznej prezentowanych kolejno bodźców wyrazowych. Uzyskane w ten sposób listy wyrazów odebranych przez respondentów dały bezpośrednią i jednoznaczną odpowiedź na temat zrozumiałości poszczególnych bodźców.

#### 3.1. Metoda prezentacji wyrazów.

Zestaw doświadczalny wyrazów (por. rozdz. 1.3.) zarejestrowano na taśmie magnetofonowej w dwóch wersjach. Wersja pierwsza - będąca częścią adaptacyjną - zawierała pełny zestaw wyrazów, występujących w porządku losowym w 3-sekundowych odstępach. W wersji drugiej, stanowiącej właściwy materiał testowy, po dwukrotnej prezentacji każdego wyrazu (odstęp pomiędzy końcem pierwszej wypowiedzi a początkiem drugiej wynosił ok. 2.5 sek.) następowała 6-sekundowa pauza na wpisanie wyrazu odebranego.

Obie wersje nagrania odtwarzano bezpośrednio po sobie. Użycie nagrania "uczącego" (część adaptacyjna) miało na celu oswajanie respondentów z brzmieniem mowy syntetycznej, nieco odbiegającym od brzmienia mowy naturalnej. Badania odsłuchowe poprzedzono każdorazowo instrukcją, mającą wyjaśnić respondentom istotę eksperymentu i ich w nim zadanie.

Odsłuchy przeprowadzono w dwóch typach warunków : (a) laboratoryjnych (kabina bezdechowa) i (b) "naturalnych" (pomieszczenie charakteryzujące się pewnym pogłosem i dość znacznym tłem dźwiękowym otoczenia - przejeżdżające w niewielkiej odległości samochody i pociągi, dobiegające z zewnątrz odgłosy rozmów, itp.), przy użyciu toru akustycznego złożonego z :

- stereofonicznego magnetofonu kasetowego M601SD,
- stereofonicznego wzmacniacza PA-107 oraz
- kolumn głośnikowych ZG10C.

W warunkach (a) dokonało odsłuchu 5 osób, w warunkach (b) -



20 osób. Podział respondentów na te dwie grupy miał na celu stwierdzenie, jak dalece warunki otoczenia (tło dźwiękowe), w którym następuje odtwarzanie wpływają na zrozumiałość wyrazów syntetycznych.

### 3.2. Wyniki.

Z ogólnej liczby 77 wyrazów 46 zostało rozpoznanych błędnie w obu typach warunków odsłuchowych. Były to następujące wyrazy :

areal	lina	miły	raj
armia	lojalny	mowa	ranny
awaria	maj	mój	rola
Emilia	major	my	rój
ile	mała	-mylny	ulewa
Iwona	mam	naiwny	umarł
ja	mama	nerwy	Wawel
jem	manewry	niema	wełna
Julia	manna	norma	wojna
lala	Marian	normalny	wolny
Lenin	miał	numer	
lilia	milion	oni	

Każdy z pozostałych 31 wyrazów został błędnie rozpoznany co najmniej jeden raz, przy czym w kilku przypadkach liczba niepoprawnych odpowiedzi była znaczna (maksymalnie 13 - dla wyrazu Mania). Wykaz błędnie odebranych wyrazów zawiera tabela 3.1. W rubryce 4 tej tabeli podano błędne rozpoznania dla każdego z błędnie odebranych wyrazów. Znaleźć można wśród nich również "wyrazy" pozbawione sensu (np. dla wyrazu jawny wystąpiły rozpoznania "diabły" i "jabły"). Wynikło to z założenia przyjętego w trakcie oceny odsłuchowej : w przypadku braku sensownego wyrazu kojarzącego się respondentowi z prezentowanym bodźcem miał on zapisać efekt akustyczny najlepiej odpowiadający temu bodźcowi.

140 uzyskanych ogółem błędnych rozpoznań daje średnią liczbę błędów na jednego słuchacza równą 5,60 (7,27 %). W przypadku warunków naturalnych (100 błędnych rozpoznań, 20 respondentów) wynosi ona 5 błędów na słuchacza (6.49 %), a dla



Tabl. 3.1. Wykaz błędnie odebranych wyrazów syntetycznych.

Wyraz nadany	Ogólna liczba błędnych rozpoznań		Wyrazy odebrane	liczba wystąpień w warunkach	
		%		laboratoryjnych	naturalnych
1	2	3	4	5	6
aleja	4	16	alenia Amelia anemia	- - -	1 2 1
jawny	7	28	diabli diabły diabły jabły	- 1 1 1	3 - - 1
lawa	4	16	laba lama plama	1 - -	1 1 1
leń	6	24	lej	3	3
lewy	1	4	plewy	-	1
linia	7	28	lilia liwa	1 -	5 1
lniany	8	32	Beniamin dynamit lłany rumiany ymiany	- - 3 - -	1 2 - 1 1
ława	5	20	bulawa Łaba Mława plawa	- - - -	1 2 1 1
Mania	13	52	Maja malia Meria	2 2 -	7 - 2
mania	10	40	malja Maria	1 -	- 9
marny	2	8	malny mały	- -	1 1

1	2	3	4	5	6
miara	2	8	liara wiara	- 1	1 -
mina	1	4	mila	1	-
ława	4	16	ława nuawa	2 1	1 -
łyn	2	8	łyn ...łyn	- -	1 1
miania	7	28	jaja lialia miara	2 1 -	2 - 2
nowy	1	4	łody	-	1
oliwa	1	4	poliwa	1	-
rama	5	20	łama	-	5
rano	1	4	łamo	-	1
rower	2	8	rober Robert	1 -	- 1
równy	9	36	Lublin lubly rlubny rubli rubly	- 1 1 - 2	3 - - 2 -
ruina	5	20	łina łima oułwa ujma wina	- - - - -	1 1 1 1 1
rura	5	20	kura Uła unra	- - 1	1 3 -
rywal	3	12	ejbal ejwal ewal -	1 - -	- 1 1
wanna	3	12	mama manna	1 -	1 1
wiara	1	4	liara	-	1

1	2	3	4	5	6
wina	10	40	lina mina Nina wita	3 1 - -	1 3 1 1
wrona	4	16	brona droga norma	- - -	1 1 2
wuj	1	4	bój	-	1
wy	6	24	dym ly my ny	- 2 - 1	1 - 2 -
Łączna liczba błędów	140	7.27	-	40	100

warunków laboratoryjnych (40 błędnych rozpoznań, 5 respondentów) 8 błędów na słuchacza (10.39 %).

### 3.3. Omówienie wyników.

Wyrazy, w przypadku których liczba błędnych rozpoznań wynosiła 1 lub - najwyżej - 2 (łącznie : 11 wyrazów) można uznać za wystarczająco zrozumiałe, a przypisane im niepoprawne odpowiedzi potraktować jako błędy słuchaczy. Kwalifikacji takiej nie można natomiast przyjąć w odniesieniu do pozostałych 20 wyrazów. Biorąc pod uwagę stosunkowo wysoki procent błędnych identyfikacji, zrozumiałość tych wyrazów należy uznać za nie w pełni zadowalającą, a przyczyn tego upatrywać w niedoskonałości syntezy<sup>1)</sup>.

Najwięcej niepoprawnych odpowiedzi zanotowano w przypadku następujących wyrazów :

Mania (13), mania (10), wina (10), równy (9), lniany (8), linia (7), niania (7), jawny (7), leń (6), wy (6).

Charakterystyczne jest, iż w grupie tej występują wszystkie wyrazy syntetyczne zawierające głoskę /p/. Stosunkowo mała wyrazistość tej głoski jest wynikiem ograniczenia się przy jej syntezie (podobnie jak w przypadku dwu pozostałych głosek nosowych)

<sup>1)</sup>

Pojęciem "błędna identyfikacja" można się w związku z tym posługiwać jedynie w sposób umowny.



wyłącznie do toru formantowego. Uzyskany na tej drodze efekt nazalizacji był zdecydowanie niedostateczny. Świadczy o tym fakt, że w większości błędnych odpowiedzi głoska ta została zastąpiona przez zbliżone do niej pod względem brzmienia głoski ustne - /j/ (np. Maja zamiast Mania, lej zamiast leń itp.) lub /l/ (np. lilia zamiast linia).

Analogiczne błędy wystąpiły także w przypadku dwu pozostałych głosek nosowych, a zwłaszcza /n/. Ilustrację stanowią tu odpowiedzi typu diabli (zamiast jawny) czy rubly (zamiast równy). Zaznaczyła się ponadto typowa dla badań percepcyjnych tendencja do zastępowania głoski /n/ przez /m/ (np. malmy zamiast marny, ujma zamiast ruina). Niejasne są natomiast przyczyny, dla których głoski ustne /l/ i /j/ w wyrazie eleja zostały odebrane jako /m/ i /n/.

Istotnym źródłem błędnej identyfikacji była również mała wyrazistość głoski /v/, której syntezy dokonano bez użycia toru szumowego. W większości niepoprawnych odpowiedzi głoska ta - ze względu na zbliżone miejsce artykulacji - została zinterpretowana jako /b/ (np. Łaba zamiast ława, rober lub Robert zamiast power czy diabli zamiast jawny), w kilku innych przypadkach - jako /m/ (np. mina zamiast wina) lub jako /l/ (np. ly zamiast wy).

Pewna liczba błędnych odpowiedzi była efektem wyraźnych trudności w rozpoznawaniu głoski /r/, zwłaszcza gdy występowała ona w nagłosie. Respondenci interpretowali ją najczęściej jako /l/ (np. lama zamiast rama) lub jako samogłoskę (np. ouiwa zamiast ruina, ejwel zamiast rywal), a niekiedy w ogóle jej nie postrzegali (np. ujma zamiast ruina).

Oprócz wspomnianych wyżej przypadków substytucji głosek zanotowano także dwa inne typy błędnych rozpoznań. W czterech odpowiedziach na wyrazy ława, lewy, ława i oliwa nagłosowe /l/, /w/ i /o/ zostały poprzedzone przez spółgłoskę zwartą /p/. Było to najprawdopodobniej konsekwencją zbyt szybkiego wzrostu wartości amplitudy w obrębie kilku początkowych fram każdego z tych wyrazów. Ponadto, dwa wyrazy dwusylabowe (lniany i ława)<sup>1)</sup>

<sup>1)</sup> Liczby sylab w odpowiedziach nuawa (zamiast Mława) i luyn (zamiast młyn) nie da się określić jednoznacznie.

odebrano w kilku przypadkach jako trzysylabowe (liniany : czterokrotnie, ława : jednokrotnie). Błędy takie mogły powstać co najmniej z kilku powodów (np. w wyniku nieprawidłowego rozpoznania występującej w rzadkim w nagłosie połączeniu z /p/ głoski /l/ jako segmentu samogłoskowego), trudno byłoby jednak ustalić, który z nich miał znaczenie decydujące.

Ogół uzyskanych w toku badań odsłuchowych odpowiedzi dzieli się na dwa typy. Pierwszy z nich reprezentują wyrazy sensu stricto, drugi - asemantyczne ciągi liter, będące graficznym odzwierciedleniem tych z zaprezentowanych bodźców, których respondenci nie potrafili zidentyfikować jako wyrazy. O ile interpretacja zapisów nieznaczących nie nastrocza specjalnych trudności (błędna odpowiedź jest wiernym odbiciem usłyszanego bodźca), o tyle przy analizie niepoprawnych rozpoznań będących wyrazami języka polskiego powstają niejakie wątpliwości, dotyczące - najogólniej - rodzaju "strategii decyzyjnych" stosowanych przez poszczególnych respondentów. Niektórzy z nich mogli, przykładowo, dążyć do wpisywania odpowiedzi wyrazowych (tj. znaczących) we wszystkich przypadkach, nawet gdy usłyszany bodziec zaledwie kojarzył im się brzmieniowo z podaną ostatecznie odpowiedzią. O słuszności tego przypuszczenia świadczą odpowiedzi typu dynamit i Beniamin (zamiast liniany) czy norma (zamiast wrona), których brzmienie w niewielkim tylko stopniu przypomina brzmienie wyrazów nadanych. Nie należy w związku z tym wyciągać zbyt daleko idących wniosków co do rodzajów zamian głoskowych (np. zastąpienie wyrazu rura przez wyraz kura niekoniecznie oznacza, że /r/ było postrzegane jako /k/).

Prawdopodobieństwa błędnego rozpoznania danego wyrazu jest tym wyższe, im większa jest w języku liczba wyrazów zbliżonych do niego pod względem brzmienia. Przykładowo, niepoprawna identyfikacja wyrazu rura jest bardziej prawdopodobna (ze względu na występowanie w języku polskim takich wyrazów, jak : kura, góra, lura, fura, burra, jurra itp.) niż np. wyrazu manewry. Wybranie do zestawu doświadczalnego kilku wyrazów o podobnym brzmieniu (np. lina - mina - wina - ruina, ława - ława - Mława) spowodowało podwyższenie liczby niepoprawnych odpowiedzi.

Pewien wpływ na charakter tych odpowiedzi mogły wyrzecz



różnice w tzw. subiektywnym prawdopodobieństwie wyrazów. Stawiały one zapewne przyczynę, dla której np. wyraz mania został błędnie rozpoznany jako Marja (dziewięciokrotnie), a wyraz wina - jako lina (czterokrotnie) i jako mina (czterokrotnie). Powyższe wyrazy odebrane są bowiem subiektywnie bardziej prawdopodobne niż wyrazy nadane (por. [4]).

Porównanie wyników uzyskanych przez dwie grupy respondentów (warunki laboratoryjne i "naturalne") wykazało, że zrozumiałość wyrazów w warunkach naturalnych była zdecydowanie lepsza (por. rozdział 3.2.). Oznacza to, że "niesprzyjające" tło akustyczne wpłynęło na wzrost liczby poprawnych odpowiedzi. Ten pozorny paradoks wyjaśniają dwie następujące okoliczności :

- po pierwsze : bieżących odsłuchów przy syntezie wyrazów dokonywano, w warunkach akustycznych zbliżonych do tych, które w odniesieniu do eksperymentu percepcyjnego określono mianem "naturalnych"; zrozumiałość skonstruowanych bodźców była zatem optymalna właśnie dla tego typu warunków
- i
- po drugie : w warunkach laboratoryjnych dawały o sobie znać wszelkie niedoskonałości syntezy - zarówno te, będące efektem ograniczeń technicznych (np. niedostatecznie silna nazalizacja spółgłosek nosowych), jak i te, powstałe w wyniku niedociągnięć w doborze optymalnych wartości parametrów (bieżąca ocena odsłuchowa produktu syntezy była w znacznej mierze utrudniona ze względu na niemożliwe do wyeliminowania zakłócenia, których źródłem była wykorzystywana aparatura).

#### 3.4. Wnioski końcowe.

Zasadniczym powodem niskiej zrozumiałości części wyrazów syntetycznych były ograniczenia i niedoskonałość wykorzystywanego syntetyzatora. Z innych przyczyn wymienić należy zniekształcenia powstałe przy nagrywaniu materiału odsłuchowego (ze względu na nierównomierność przesuwu taśmy perforowanej sterującej syntezą) oraz niską jakość użytych przy odsłuchach kolumn głośnikowych. Duży rozrzut liczby błędów popełnianych przy rozpoznawaniu (od 1 do 10) wskazuje na konieczność odwołania

się w przyszłości do ocen większej liczby respondentów. Pozwoliłoby to na wyciągnięcie bardziej ogólnych wniosków w odniesieniu do błędnych rozpoznań.

Biorąc pod uwagę fakt ograniczenia sterowania syntezą do 4 parametrów, ogólny wynik oceny percepcyjnej należy uznać za zadowalający. Poprawę zrozumiałości i naturalności brzmienia wyrazów syntetycznych można będzie zapewne uzyskać poprzez właściwe kształtowanie przebiegów częstotliwości podstawowej.



DODATEK 4.

ZESTAWIENIE WARTOŚCI PARAMETRÓW DLA POSZCZEGÓLNYCH WYRAZÓW  
SYNTEZYCZNYCH (w porządku alfabetycznym).

W kolejnych rzędach, oznaczonych cyframi od "1" do "5", podane są następujące parametry :

- 1 - częstotliwość docelowa  $F_1$  (w Hz)
- 2 - częstotliwość docelowa  $F_2$  (w Hz)
- 3 - częstotliwość docelowa  $F_3$  (w Hz)
- 4 - amplituda (w dB); dla głoski uderzeniowej /r/ podane są dwie wartości : amplituda segmentu samogłoskowego i segmentu spółgłoskowego; przy wyrazach zawierających podwójne /n/ (np. "manna") została również podana wartość amplitudy na przejściu między /n/ poprzedzającym i następującym
- 5 - iloczyn (w ms); w nawiasach podano wartości czasu trwania ugięć formantowych występujących między poszczególnymi głoskami.

Aleja

	a	l	e	j	a
$F_1$ [Hz]	850	380	650	250	850
$F_2$ [Hz]	1300	1750	2100	3000	1400
$F_3$ [Hz]	2750	2800	3100	3250	2750
AV [dB]	17	18	24	18	14
IL [ms]	100(40)	70(30)	120(50)	30(90)	130

Areal

	a	r	e	a	w
1	850	500	650	850	350
2	1300	1900	2100	1300	720
3	2750	3200	3100	2750	2600
4	16	22/14	24	14	18
5	110(30)	70(30)	120(30)	90(120)	70

Armia

	a	r	m	j	a
1	850	500	200	250	850
2	1320	1300	1520	3100	1400
3	2670	2500	3200	3160	2670
4	18	22/14	22	18	14
5	150(30)	90(10)	90(10)	30(90)	140

Awaria

	a	v	a	r	j	a
1	850	200	850	500	250	850.
2	1300	950	1300	1900	3100	1400
3	2670	2900	2670	3200	3160	2670
4	18	14	18	22/15	16	16
5	110(30)	70(30)	140(30)	60(30)	50(90)	140

Emilia

	e	m	i	l	j	a
1	650	200	250	380	250	850
2	2100	1500	3100	2050	3100	1400
3	3100	3200	3160	2800	3160	2670
4	18	24	28	18	16	14
5	100(10)	100(10)	120(30)	90(30)	50(90)	140

Ile

	i	l	e
1	250	380	650
2	3100	1850	2100
3	3160	2800	3100
4	26	21	18
5	140(80)	20(30)	190

Iwona

	i	v	o	n	a
1	250	200	650	250	850
2	2550	950	950	2100	1300
3	3160	2900	2870	4700	2670
4	20	20	20	20	12
5	90(50)	50(30)	130(10)	80(10)	150



Ja		
	j	a
1	250	850
2	3100	1400
3	3160	2670
4	20	16
5	80(90)	230

Jawny					
	j	a	v	n	z
1	250	850	200	250	350
2	3000	1300	950	2000	2200
3	3160	2670	2900	4700	2850
4	16	16	21	22	22
5	70(90)	100(30)	80(10)	80(10)	150

Jem			
	j	e	n
1	250	650	200
2	3000	2100	1250
3	3160	3100	2200
4	19	22	24
5	80(90)	130(10)	180

Julia					
	j	u	l	j	a
1	250	350	400	250	850
2	2700	900	1750	2800	1400
3	3160	2670	2200	3160	2670
4	18	24	18	17	12
5	80(90)	80(40)	100(10)	70(90)	150

Lala				
	l	a	l	a
1	400	850	400	850
2	1600	1300	1600	1300
3	2800	2670	2800	2670
4	15	18	15	13
5	100(30)	130(30)	70(30)	170

Lawa

	l	a	v	a
1	400	850	200	850
2	1600	1300	950	1300
3	2800	2670	2900	2670
4	18	18	18	12
5	100(30)	150(30)	60(30)	160

Lenin

	l	e	ɲ	i	n
1	380	650	250	250	250
2	1750	2100	2050	3100	2100
3	2800	3100	4700	3160	4700
4	18	24	22	22	22
5	100(30)	140(30)	70(40)	80(10)	200

Leń

	l	e	ɲ
1	380	650	280
2	1750	2100	2800
3	2800	3100	4700
4	18	22	24
5	100(30)	80(140)	180

Lewy

	l	e	v	ɛ
1	380	650	250	350
2	1750	2100	1200	2170
3	2800	3100	2900	2850
4	18	22	14	22
5	100(30)	120(30)	40(30)	170

Lilia

	l	i	l	j	a
1	350	250	350	250	850
2	1900	3100	2100	3100	1400
3	2800	3160	2800	3160	2670
4	24	26	20	16	12
5	100(30)	110(30)	80(30)	50(90)	170



Lina

	l	i	n	a
1	350	250	200	850
2	2100	2600	2300	1300
3	2800	3160	4700	2670
4	18	26	22	12
5	100(30)	110(30)	70(70)	110

Linia

	l	i	ja	j	a
1	350	250	200	250	850
2	2100	2600	3200	2900	1400
3	2800	3160	4700	3160	2670
4	18	26	15	13	13
5	100(20)	30(110)	140(30)	80(90)	160

Lniany

	l	ja	a	n	i
1	400	230	850	250	350
2	1800	2250	1400	2100	2300
3	2800	4700	2670	4700	2850
4	19	24	16	22	20
5	100(50)	50(70)	130(10)	90(10)	170

Lojalny

	l	o	j	a	l	n	i
1	400	640	250	850	400	250	350
2	1500	950	2800	1300	1750	2000	2300
3	2500	2670	3160	2670	2800	4700	2850
4	16	20	18	18	24	24	22
5	100(30)	50(70)	30(90)	100(60)	60(10)	90(10)	160

Lawa

	w	a	v	a
1	300	850	200	850
2	800	1300	950	1300
3	2600	2670	2900	2670
4	20	17	21	12
5	100(50)	130(30)	60(30)	170

	Maj		
	m	a	j
1	200	850	250
2	1250	1400	3000
3	2200	2670	3160
4	24	17	19
5	120(10)	130(100)	120

	Majer				
	m	a	j	o	r
1	200	850	250	650	470
2	1250	1400	2800	1050	1100
3	2200	2670	3160	2670	1900
4	24	17	18	18	20/16
5	110(10)	120(90)	40(110)	70(30)	100

	Mała			
	m	a	w	a
1	200	850	350	850
2	1250	1400	800	1300
3	2200	2670	2600	2670
4	22	18	18	12
5	120(10)	120(50)	20(90)	150

	Mam		
	m	a	m
1	200	850	200
2	1250	1400	1250
3	2200	2670	2200
4	22	16	20
5	100(10)	150(10)	190

	Mama			
	m	a	m	a
1	200	850	200	850
2	1250	1400	1250	1400
3	2200	2670	2200	2670
4	24	17	21	12
5	120(10)	140(10)	90(10)	170

Manewry

	m	a	n	e
1	200	850	250	650
2	1250	1400	2100	2100
3	2200	2670	4700	3100
4	20	16	20	24
5	80(10)	100(10)	80(10)	130(30)

	v	r	z
1	250	500	350
2	1000	1900	2300
3	2900	3200	2850
4	18	20/14	22
5	70(30)	60(30)	150

Mania (imię)

	m	a	z	a
1	200	850	230	850
2	1250	1400	2100	1400
3	2200	2670	4700	2670
4	24	18	22	13
5	120(10)	150(40)	60(100)	150

Mania

	m	a	z	j	a
1	200	850	230	250	850
2	1250	1400	2500	3000	1400
3	2200	2670	4700	3160	2670
4	24	18	18	16	14
5	120(10)	100(90)	60(40)	80(90)	150

Manna

	m	a	n	n	a
1	200	850	250	250	850
2	1250	1400	2000	2000	1300
3	2200	2670	4700	4700	2670
4	22	18	18(24)	18	14
5	110(10)	170(10)	80(60)	80(10)	140



Marian

	m	a	r	j	a	n
1	200	850	500	250	850	250
2	1250	1340	1900	3100	1400	2000
3	2200	2670	2500	3160	2670	3200
4	24	18	22/16	16	16	18
5	100(10)	120(30)	70(30)	40(90)	100(10)	170

Marny

	m	a	r	n	±
1	200	850	500	250	350
2	1250	1340	1300	1800	2300
3	2200	2670	2500	4700	2850
4	24	16	24/15	23	19
5	120(10)	130(30)	90(10)	70(50)	120

Miał

	m	j	a	w
1	200	250	850	350
2	1250	3100	1300	750
3	2200	3160	2670	2600
4	24	16	16	16
5	100(10)	60(100)	80(140)	50

Miara

	m	j	a	r	a
1	200	250	850	500	850
2	1800	3100	1400	1900	1340
3	2600	3160	2670	3200	2670
4	20	16	18	22/14	12
5	100(10)	30(90)	110(30)	70(30)	140

Milion

	m	i	l	j	o	n
1	200	250	400	250	640	250
2	1500	3100	2000	2900	1000	2000
3	3200	3160	2800	3160	2870	4700
4	22	26	16	16	15	18
5	90(10)	120(30)	40(50)	10(80)	90(10)	170

Miły

	m	i	w	z
1	200	250	360	350
2	2000	3100	1000	2300
3	3200	3160	2600	2850
4	24	27	18	19
5	120(10)	140(30)	30(100)	120

Mina

	m	i	n	a
1	200	250	250	850
2	1250	3100	2300	1300
3	3200	3160	4700	2670
4	24	27	23	12
5	120(10)	120(30)	70(30)	160

Mława

	m	w	a	v	a
1	200	350	850	200	850
2	1250	800	1300	950	1300
3	2200	2600	2670	2900	2670
4	20	18	17	19	12
5	90(20)	60(90)	100(30)	60(30)	170

Młyn

	m	w	z	n
1	200	360	350	250
2	1250	800	2300	1800
3	2300	2700	3200	4700
4	24	16	24	21
5	120(10)	60(100)	100(10)	140

Mowa

	m	o	v	a
1	200	600	200	850
2	1100	900	900	1300
3	3300	2950	3000	2750
4	24	21	19	13
5	120(10)	150(30)	40(30)	150

	Mój		
	m	u	j
1	200	390	250
2	1100	800	2900
3	2300	2750	3250
4	24	26	21
5	140(10)	130(80)	80

	My	
	m	±
1	200	350
2	1250	2300
3	2300	3250
4	24	26
5	140(40)	230

	Mylny				
	m	±	l	n	±
1	200	350	350	250	350
2	1500	2300	2150	2200	2300
3	3300	3250	2900	4700	2950
4	22	26	22	20	22
5	120(10)	120(30)	110(30)	80(10)	140

	Naiwny					
	n	a	i	v	n	±
1	250	850	250	200	250	350
2	2100	1300	3100	1300	2000	2300
3	4700	2750	3250	3000	4700	2950
4	18	10	24	18	22	18
5	120(10)	90(30)	160(10)	60(10)	80(10)	120

	Nerwy				
	n	e	r	v	±
1	250	650	450	250	350
2	2100	2100	1800	1200	2150
3	4700	3200	2800	3000	2950
4	20	24	24/16	16	22
5	100(10)	130(30)	90(30)	70(30)	170



Niania

	ᵓ	a	ᵓ	a
1	230	850	230	850
2	2100	1300	2100	1300
3	4700	2750	4700	2750
4	24	18	22	13
5	130(100)	120(50)	50(100)	160

Niema

	ᵓ	e	m	a
1	250	650	200	850
2	2100	2100	1250	1400
3	4700	3200	2300	2750
4	18	22	22	14
5	60(140)	90(10)	70(10)	170

Norma

	n	o	r	m	a
1	200	650	500	200	850
2	2100	1050	1200	1250	1300
3	4700	2750	2000	2300	2750
4	19	22	22/15	20	12
5	100(60)	100(30)	90(10)	90(10)	150

Normalny

	n	o	r	m	a	l
1	200	650	500	200	850	400
2	1450	950	1200	750	1300	1750
3	4700	2750	2000	2300	2750	2900
4	19	18	20/14	20	18	22
5	100(10)	80(30)	50(10)	90(10)	120(60)	60(10)

	n	±
1	250	350
2	2000	2250
3	4700	2950
4	22	22
5	80(10)	160

Nowy

	n	o	v	z
1	200	650	250	350
2	1400	950	1200	2150
3	4700	2750	3000	2950
4	19	22	16	22
5	120(10)	150(30)	70(30)	150

Numer

	n	u	m	e	r
1	200	350	200	650	450
2	1700	900	1250	2100	1800
3	4700	3250	2300	3200	2800
4	14	26	20	20	22/15
5	100(80)	80(10)	100(10)	120(30)	110

Oliwa

	e	l	i	v	a
1	650	380	250	200	850
2	950	2050	3100	900	1300
3	2750	2900	3250	3000	2750
4	17	21	26	18	11
5	70(60)	60(30)	150(30)	50(30)	160

Oni

	o	o	i
1	650	200	250
2	950	2250	3100
3	2950	4700	3250
4	18	20	22
5	140(80)	40(80)	140

Raj

	r	a	j
1	500	850	250
2	1800	1400	2900
3	2800	2750	3250
4	22/15	18	19
5	100(30)	140(90)	120

Rema				
	r	a	m	a
1	500	850	200	850
2	1800	1400	1050	1400
3	2800	2750	2300	2750
4	22/15	18	22	13
5	100(30)	130(10)	100(10)	150

Ranny					
	r	a	n	n	±
1	500	850	250	250	350
2	1900	1400	2000	2000	2150
3	2800	2750	3300	3300	2950
4	22/15	17	17(23)	18	22
5	100(30)	130(10)	90(60)	80(10)	140

Rano				
	r	a	n	o
1	500	850	300	600
2	1800	1400	2000	900
3	2800	2750	3300	2950
4	22/15	17	22	16
5	100(30)	130(10)	100(10)	150

Rola				
	r	o	l	a
1	500	650	400	850
2	1400	950	1600	1300
3	3600	2750	2900	2750
4	22/15	20	16	12
5	100(30)	150(30)	60(30)	150

Rower					
	r	o	v	e	r
1	500	650	250	600	450
2	1400	950	1000	1750	1800
3	2000	2750	3000	2600	2800
4	20/15	20	20	16	22/15
5	100(30)	120(10)	80(30)	90(30)	110



	Rój		
	r	u	j
1	400	400	250
2	1000	800	2800
3	2000	2750	3250
4	22/15	26	16
5	100(30)	130(80)	90

	Równy				
	r	u	v	n	i
1	400	350	200	250	350
2	1000	900	900	2000	2300
3	2000	2750	3000	4700	2950
4	22/15	26	22	22	22
5	100(30)	130(10)	80(10)	80(10)	150

	Ruina				
	r	u	i	n	a
1	400	350	250	250	850
2	1000	900	3100	2100	1300
3	2000	2750	3250	4700	2750
4	20/14	26	26	22	12
5	100(30)	100(30)	140(10)	60(10)	160

	Rura			
	r	u	r	a
1	400	350	500	850
2	1000	900	1700	1400
3	2000	2750	2350	2750
4	22/15	26	20/14	13
5	100(30)	130(30)	60(30)	140

	Rywał				
	r	i	v	a	l
1	500	350	200	850	400
2	1900	2300	900	1300	1750
3	3300	2950	3000	2750	2900
4	22/15	24	18	13	16
5	100(60)	100(10)	70(30)	90(60)	160

Ulewa

	u	l	e	v	a
1	350	380	650	200	850
2	700	1750	2100	900	1300
3	2750	2900	3200	3000	2750
4	22	22	24	16	13
5	100(30)	70(30)	140(10)	70(30)	170

Umarz

	u	m	a	r	w
1	350	200	850	500	360
2	700	1250	1300	1300	800
3	2750	2300	2750	2600	2700
4	24	20	15	22/15	20
5	140(10)	90(10)	100(30)	70(10)	120

Wanna

	v	a	n	n	a
1	200	850	250	250	850
2	1400	1400	2000	2000	1300
3	3000	2750	4700	4700	2750
4	16	18	18(24)	18	14
5	100(30)	130(10)	80(60)	80(10)	140

Wawel

	v	a	v	e	l
1	200	850	250	650	350
2	900	1300	1000	2100	1750
3	3000	2750	3000	3250	2900
4	14	16	14	20	20
5	100(30)	130(30)	70(30)	90(60)	130

Weżna

	v	e	w	n	a
1	250	600	380	250	850
2	1000	1750	800	2100	1300
3	3000	3200	2700	4700	2750
4	22	22	20	22	12
5	100(30)	130(30)	100(10)	90(10)	160

Wiara

	v	j	a	r	a
1	200	250	850	500	850
2	1900	3000	1400	1900	1300
3	3000	3250	2750	3600	2750
4	22	16	18	22/14	12
5	100(30)	30(90)	110(30)	70(30)	140

Wina

	v	i	n	a
1	200	250	200	850
2	2100	2600	2300	1300
3	3200	3250	4700	2750
4	12	26	22	12
5	100(30)	110(30)	80(70)	110

Wojna

	v	o	j	n	a
1	200	650	250	250	850
2	900	950	3000	2100	1300
3	3000	2950	3250	4700	2750
4	18	20	21	21	12
5	100(30)	110(120)	50(10)	80(10)	170

Wolny

	v	o	l	n	i
1	200	650	400	250	350
2	900	950	1750	2000	2150
3	3000	2750	3600	4700	2950
4	18	21	22	22	22
5	100(30)	130(60)	80(10)	80(10)	150

Wrona

	v	r	o	n	a
1	200	500	650	250	850
2	1400	1400	950	2000	1300
3	3000	3600	2750	4700	2750
4	22	20/14	20	18	12
5	100(10)	60(30)	130(10)	80(10)	150



	Wuj		
	v	u	j
1	200	350	250
2	1050	700	2800
3	3000	2750	3250
4	14	26	18
5	100(30)	120(100)	100

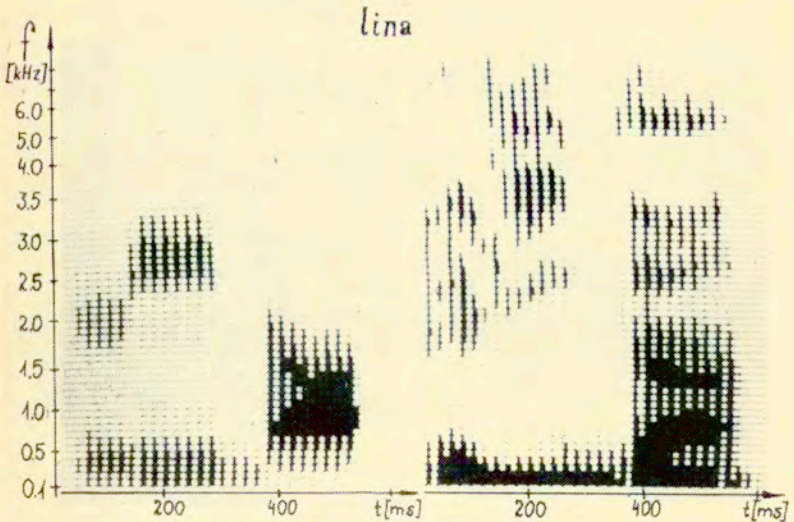
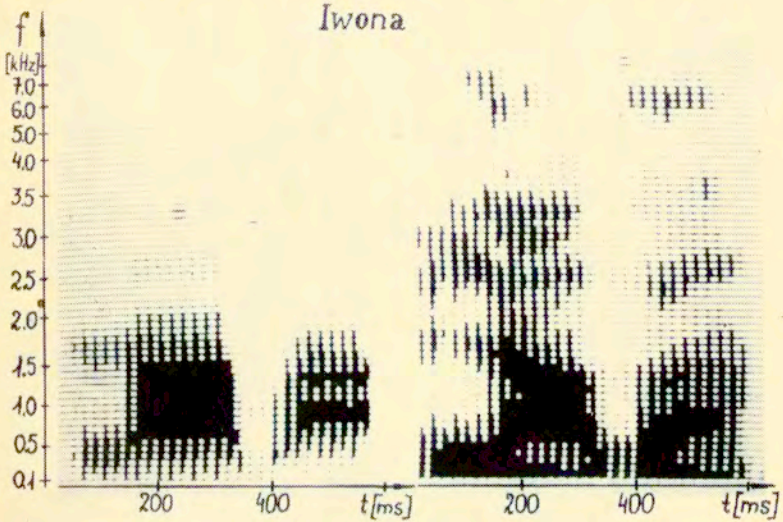
	Wy	
	v	z
1	200	3500
2	1900	2300
3	3000	2950
4	18	26
5	120(30)	240

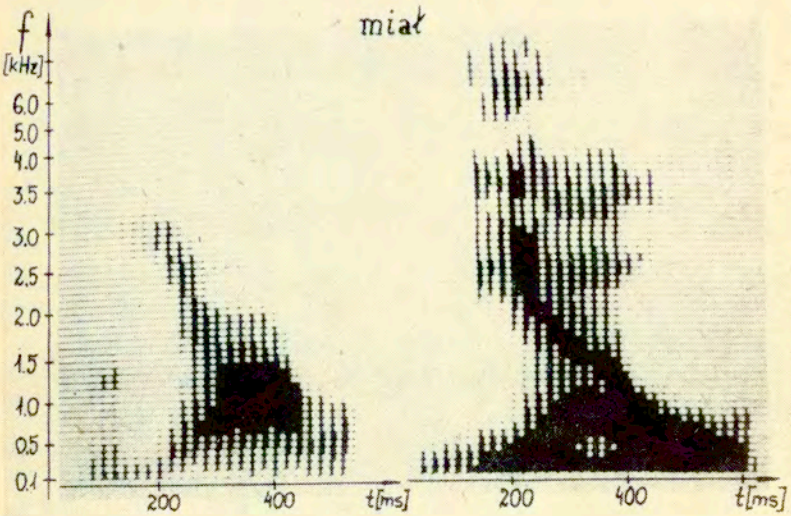
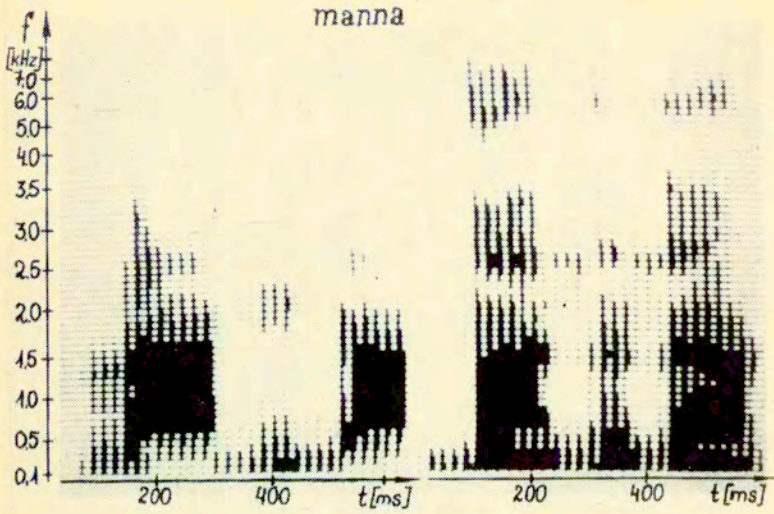
DODATEK B

SPEKTROGRAMY WYBRANYCH PAR WYRAZÓW SYNTETYCZNYCH I NATURALNYCH

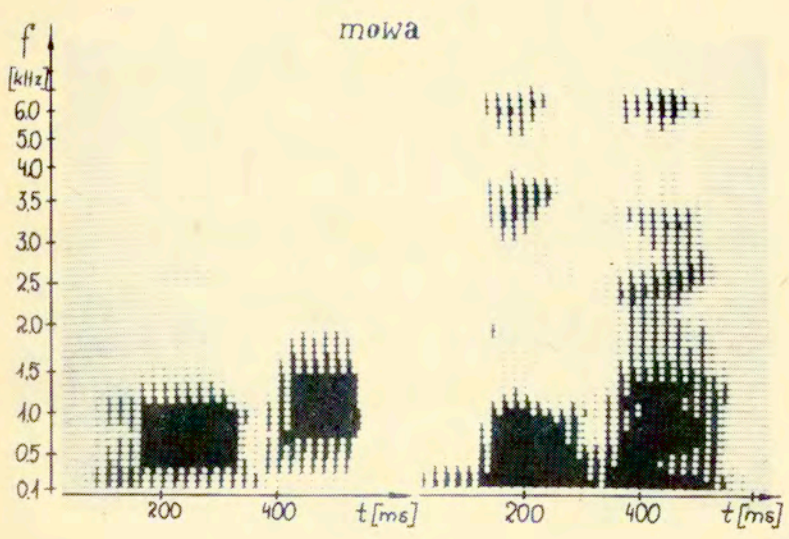
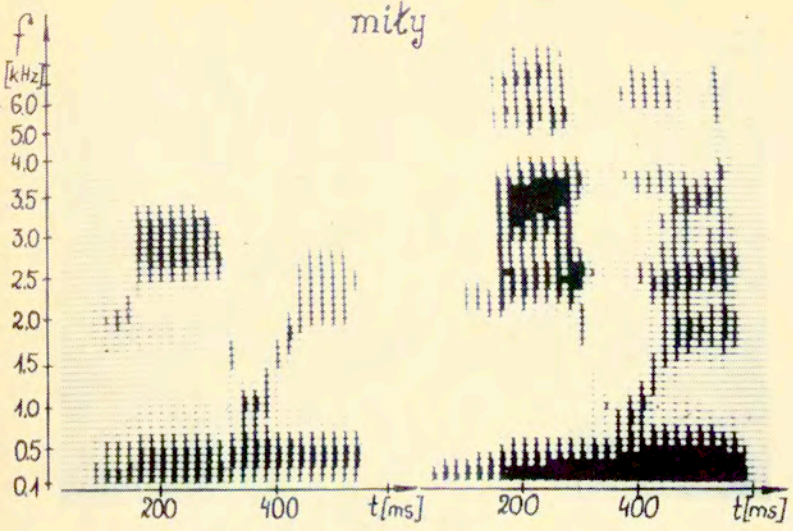
SPEKTROGRAM SYNTETYCZNY

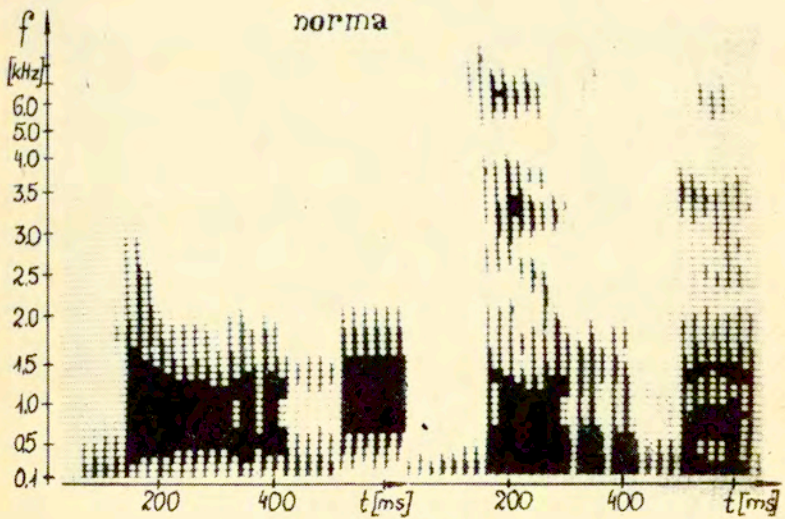
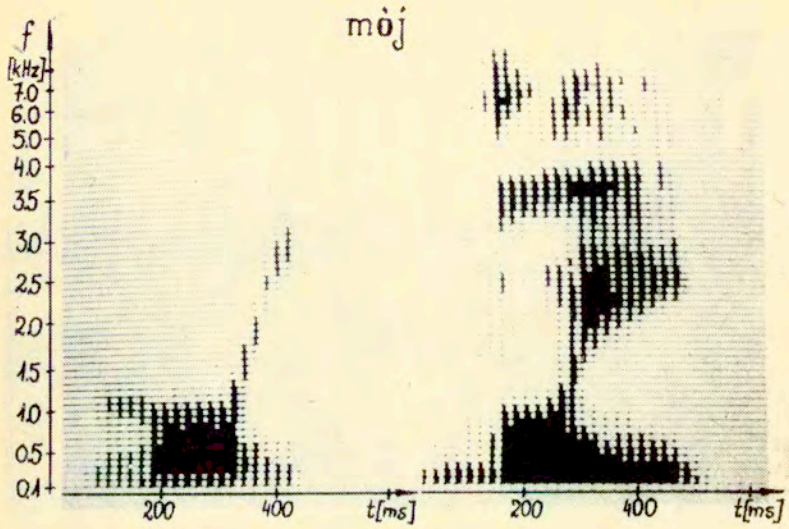
SPEKTROGRAM NATURALNY

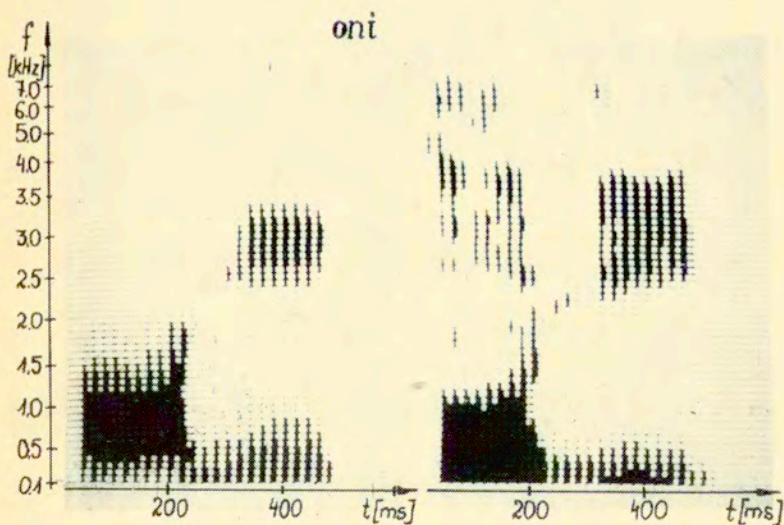
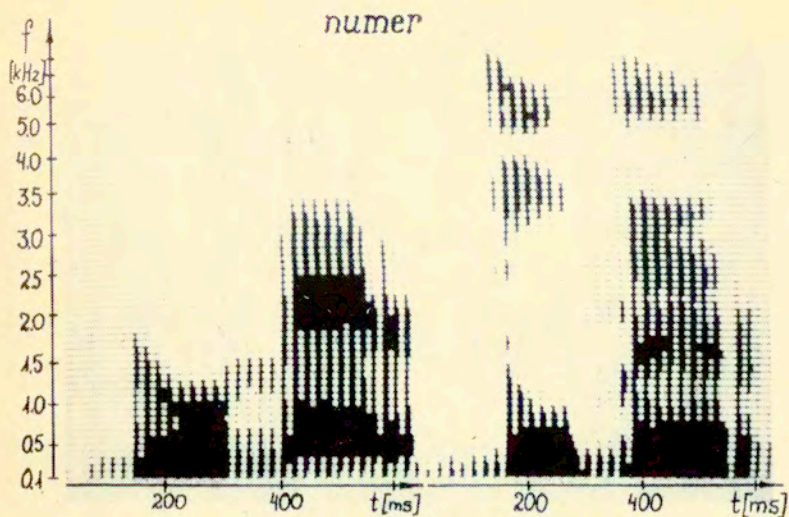




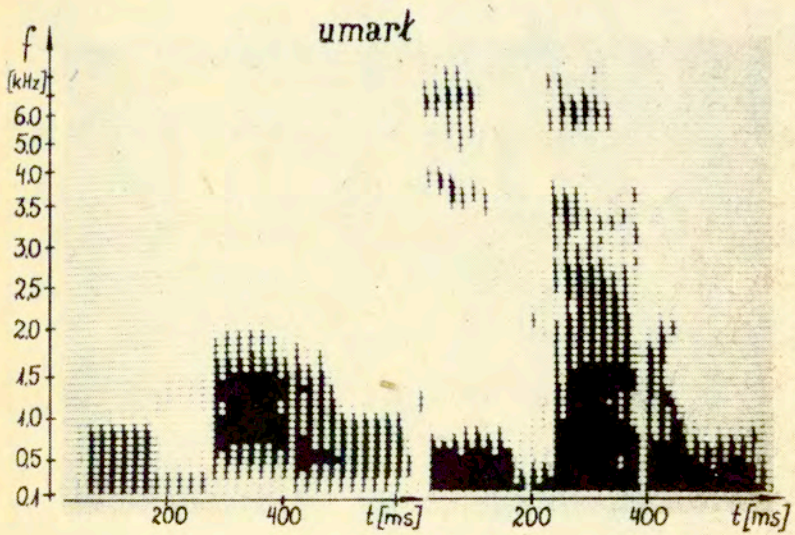
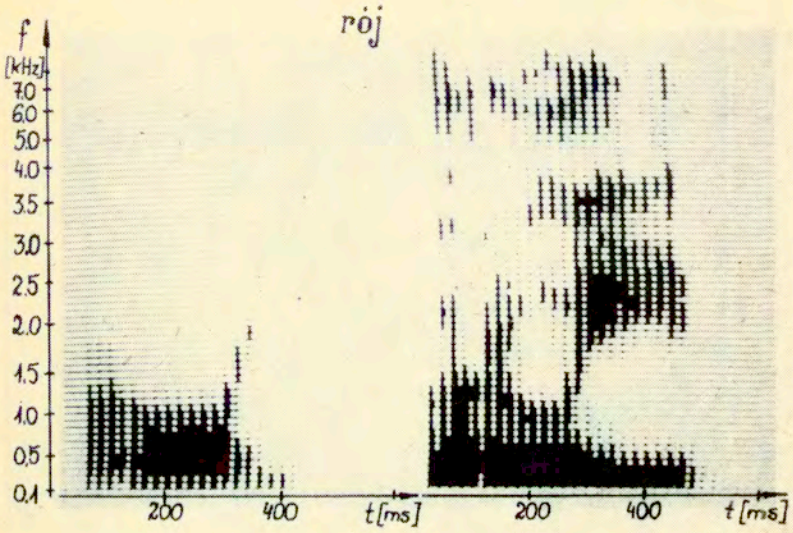












BIBLIOGRAFIA

- [1] CIARKOWSKI, R., Algorytmizacja i oprogramowanie działań operatorskich dla parametrycznej syntezy mowy za pomocą układu CT-1 sterowanego z minikomputera MERA 303, Prace IPPT 14/1983, Warszawa, 1983.
- [2] CIARKOWSKI, R., Sterowana z minikomputera MERA 303 synteza wybranych diał polskich i ich percepcja, Prace IPPT, (w druku).
- [3] DUKIEWICZ, L., Polskie głoski nosowe. Analiza akustyczna, PWN, Warszawa, 1967.
- [4] IMIOLCZYK, J., Subiektywne prawdopodobieństwo wyrazów. Podstawowy słownik frekwencyjny języka polskiego, PWN, (w druku).
- [5] JASSEM, W., Podstawy fonetyki akustycznej, PWN, Warszawa, 1973.
- [6] KOSIEL, U., Acoustic-phonetic parameters of the Polish frictionless consonants /l/ and /r/, w : Speech Analysis and Synthesis (ed. W.Jassem), vol. 2, PWN, Warszawa, 1970.
- [7] ŁOBACZ, P., An Acoustic analysis of the opposition /j:i/ and /w:u/ in present-day Polish, w Speech Analysis and Synthesis (ed. W.Jassem), vol. 2, PWN, Warszawa, 1970.
- [8] RICHTER, L., Analiza statystyczna rytmicznej struktury wypowiedzi w mowie polskiej, Prace IPPT 8/1984, Warszawa, 1984.