

**R. Ciarkowski, J. Imiolczyk**

**SYNTEZA MOWY POLSKIEJ:  
DIADY I WYRAZY  
ZE SPÓLGŁOSKAMI ZWARTYMI**

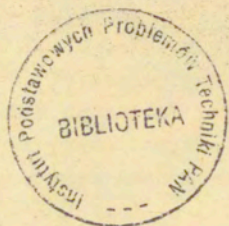
**42/1985**

P. 269.



**WARSZAWA 1985**

Praca wpłynęła do Redakcji dnia 5 września 1985 r.



56907



Na prawach rękopisu

---

Instytut Podstawowych Problemów Techniki PAN  
Nakład 160 egz. Ark.wyd. 1,1 Ark.druk. 1,75  
Oddano do drukarni w listopadzie 1985 r.  
Nr zamówienia 690/85

---

Warszawska Drukarnia Reukowa, Warszawa,  
ul. Śniadeckich 8

Ryszard Ciarkowski  
Janusz Imiołczyk  
Pracownia Fonetyki Akustycznej  
IPPT PAN

SYNTEZA MOWY POLSKIEJ :  
DIADY I WYRAZY ZE SPÓŁGŁOSKAMI ZWARTYMI <sup>1)</sup>

Streszczenie

Cel eksperymentu stanowiła synteza polskich spółgłosek zwartych w pozycji nagłosowej i wygłosowej. Jego realizacji dokonano przy wykorzystaniu zestawu minikomputerowego MERA 303 (wraz z wyspecjalizowanymi urządzeniami peryferyjnymi), pracującego pod kontrolą systemu programowego SPOS1 ([1], [2]).

Uzyskany zbiór 24 diad (typu : spółgłoska zwarta + samogłoska) oraz 24 wyrazów syntetycznych (zawierających spółgłoskę zwartą w wygłosie) poddano ocenie percepcyjnej przez 20 słuchaczy. Obok prezentacji otrzymanego bioru diad i wyrazów syntetycznych praca zawiera również wnioski dotyczące ogólnych zasad stosowanych w tym eksperymencie przy syntezie spółgłosek zwartych. Wyniki testu percepcyjnego pozwalają stwierdzić, że dla syntezy polskich spółgłosek zwartych wystarczający byłby zestaw pięciu lub - najwyżej - sześciu parametrów sterujących.

1. Techniczne aspekty syntezy spółgłosek zwartych.

Synteza była realizowana przy wykorzystaniu zestawu minikomputerowego MERA 303, obejmującego jednostkę centralną MOMIE 8B o pamięci operacyjnej 8 kbajtów wraz z typowymi urządzeniami peryferyjnymi ([1]):

- zestawem : drukarka znakowo-mozaikowa DZM 180 i klawiatura operatora (alfanumeryczna i numeryczna),
  - modułem pamięci na dyskach elastycznych MDE-300 (4x250 kbajtów),
  - perforatorem (DT-105) i czytnikiem taśmy (CT 1001A)
- oraz wyspecjalizowanymi urządzeniami peryferyjnymi ([1], [2]):

1)

Praca wykonana w ramach planu C-1

- 63 - kanałowym analizatorem widmowym (zakres 80-8310 Hz),
- kanałem funkcji analogowych KF-01 (sawierającym m.in. 8-bitowy przetwornik A/C),
- formantowym syntetyzátorem mowy COMPUTALKER CT-1 oraz
- monitorem graficznym MEMOSKOP (wielkość obrazu 64x256 punktów).

Sterujący zestawem system programowy SPOS1 umożliwiał syntezę fragmentu mowy o czasie trwania do 910 ms oraz bieżącą analizę mowy syntetycznej i naturalnej, która prowadziła do uzyskania spektrogramu wyświetlanego na ekranie monitora graficznego ([2]).

Zasadniczym ograniczeniem w realizacji syntezy spółgłosek zwartych były następujące wady układu syntetyzującego CT-1 (szerzej omówione w pracy [3]):

- nieefektywne sterowanie formantem szumowym (ze względu na niewłaściwe widmo szumu i niską jego energię oraz nieodpowiednią filtrację szumu),
- stała szerokość pasma dla poszczególnych formantów,
- zbyt niski poziom trzeciego formantu (w praktyce oznaczało to ograniczenie sterowania parametrami toru formantowego do dwóch pierwszych formantów).

## 2. Synteza spółgłosek zwartych.

Celem eksperymentu było poszerzenie zbioru pozostałych do dyspozycji głosek i diad syntetycznych, będących efektem poprzedniego doświadczenia ([2]), o spółgłoski zwarte /b/, /p/, /d/, /t/, /g/, /k/, /c/ oraz /ʃ/.

Układ parametrów syntezy, którego efektem miała być określona spółgłoska zwarta, zależał od pozycji, jaką spółgłoska ta zajmowała w wybranej „wypowiedzi” syntetycznej ([6]). Przykładowo głoskę /t/ występującą w nagłosie cechował odmienny układ parametrów niż tę samą głoskę /t/ występującą w wygłosie (różnice te dotyczyły zarówno iloczasu i przebiegu amplitudy jak i częstotliwości formantowych). Fakt ten był - w połączeniu ze stosunkowo małą wyrazistością zwartych wygłosowych w diadach VC - przyczyną przeprowadzenia doświadczenia w dwóch etapach: pierwszy z nich obejmował syntezę spółgłosek zwartych w diadach typu CV, drugi - syntezę wyrazów, w których spółgłoski zwarte występowały w pozycji wygłosowej (nie wyłącznie).

Parametrami sterującymi wykorzystywanymi przy syntezie spółgłosek zwartych były: częstotliwość podstawowa  $F_0$ , częstotliwość pierwszego, drugiego i trzeciego formantu ( $F_1$ ,  $F_2$ ,  $F_3$ ), amplituda tonu krtaniowego  $A_T$  oraz amplituda szumu aspiracji  $A_H$ . Pierwsza synteza polskich głosek zwartych przy użyciu syntetyzatora formantowego została opisana w pracy [7].

Wychodząc z założenia, że intonacja jest w zasadzie cechą mowy ciągłej, zrezygnowano z różnicowania przebiegów  $F_0$ , przyjmując (poza absolutnym początkiem i końcem syntezowanych fragmentów oraz poza obrębem głosek zwartych) stałą wartość tego parametru równą 122 Hz.

### 2.1. Zasady syntezy diad typu CV.

Ze względu na zbyt niski poziom  $F_3$  niemożliwe okazało się uzyskanie w nagłosie spółgłosek tylnojęzykowych /k/, /g/, /c/ i /j/. Dokonano w tej sytuacji syntezy pozostałych spółgłosek zwartych (/p/, /b/, /t/, /d/) w połączeniach z samogłoskami /i/, /ɛ/, /e/, /a/, /o/ i /u/.

Przyjęto, że średni czas trwania stanu ustalonego samogłoski powinien wynosić ok. 200 - 220 ms. Optymalny czas trwania spółgłosek kształtował się w granicach :

100 - 110 ms - dla głosek dźwięcznych /b/ i /d/ oraz

30 - 40 ms - w przypadku głosek bezdźwięcznych /p/ i /t/.

Podany czas jest sumą czasów trwania segmentów składowych głoski zwartej : zwarcia, płozi i aspiracji (w przypadku głosek bezdźwięcznych - tylko płozi i aspiracji). Długość przejścia między spółgłoską swartą a samogłoską wynosiła z reguły ok. 20 - 30 ms.

Przebieg amplitudy tonu krtaniowego  $A_T$  kształtowano dla obu spółgłosek dźwięcznych /b/ i /g/ w identyczny sposób. Po stopniowym wzroście wartości amplitudy w ciągu 50 ms (do 8 dB<sup>1)</sup>), następował dość gwałtowny skok amplitudy (do 20 - 22 dB) odpowiadający 30-milisekundowej płozi i dalej - niemal równie gwałtowny jej spadek. Podobnie kształtowano przebieg  $A_T$  w przypadku spółgłosek /p/ i /t/, z tym jednak, że okres początkowego narastania amplitudy był bardzo krótki (rzędu 10 - 20 ms), a jej

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

wzrost w obrębie 10-milisekundowej plosji - znacznie mniejszy : dla /p/ - 10 dB, dla /t/ - 14 dB.

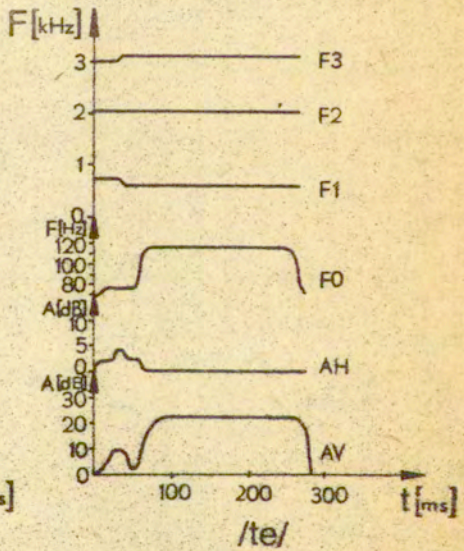
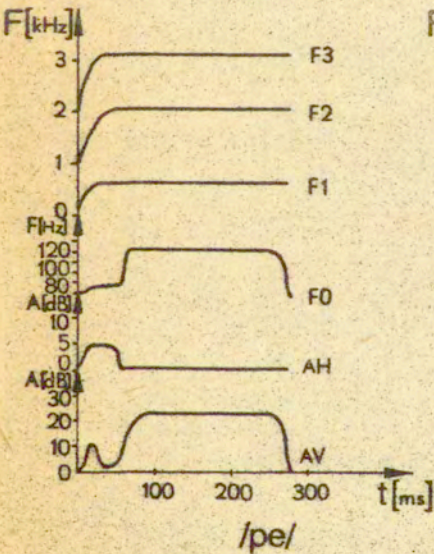
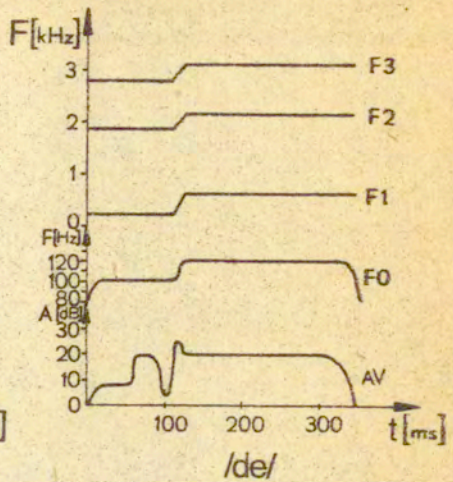
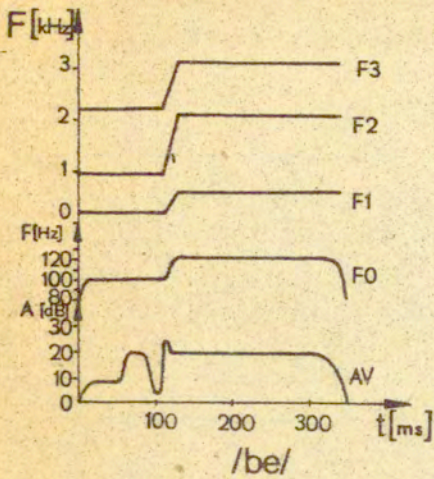
Parametrem, którego wartość decydowała o sposobie artykulacji spółgłosek (t.j. o ich „szwarcności”) była częstotliwość  $F_1$ . W obrębie spółgłosek /b/, /d/ i /p/ parametrowi temu nadano najniższą możliwą wartość równą ok. 200 Hz (w obrębie spółgłoski /t/ - równą 670 Hz). Wrażenie odpowiedniego miejsca artykulacji uzyskano poprzez właściwe wysterowanie  $F_2$ . Dla spółgłosek /p/ i /b/ wartość  $F_2$  wynosiła 900 Hz, dla /t/ przyjęte natomiast częstotliwość docelową w granicach od 1300 do 2500 Hz (w zależności od częstotliwości docelowej następującej samogłoski) i - na tej samej zasadzie - dla /d/ 1700 - 2400 Hz.

Jak wspomniano, znaczenie  $F_3$  było przy syntezie minimalne ze względu na jego niski poziom. Pewien wpływ na jakość uzyskiwanych efektów akustycznych miał natomiast parametr  $F_0$ . W obrębie spółgłosek dźwięcznych przyjęto częstotliwość podstawową równą 100 Hz (wartość  $F_0$  w segmentach samogłoskowych była stała i wynosiła ok. 120 Hz).

W celu poprawy naturalności brzmienia spółgłosek bezdźwięcznych posłużono się parametrem  $A_H$  (szumem aspiracji), wprowadzając szum o stosunkowo niskiej intensywności w obrębie plosji i aspiracji spółgłoski szwartej.

Przy syntezie przejść międzysłgłoskowych łączono  $F_1$ ,  $F_2$  i  $F_3$  głoski szwartej odpowiednio z  $F_1$ ,  $F_2$  i  $F_3$  następującej po niej samogłoski. We wszystkich przypadkach wzrost (lub spadek) częstotliwości formantowych w obrębie przejścia był równomierny. Jeżeli przykładowo, częstotliwość docelowa  $F_2$  spółgłoski wynosiła 1600 Hz, a samogłoski - 1900 Hz, to przy wymaganej długości ugięcia formantowego równej 20 ms, przyrost częstotliwości w kolejnych 10-milisekundowych przedziałach czasowych miał wartość 100 Hz (... 1600 | 1700 1800 | 1900 ..Hz ):

W pierwszej części dodatku niniejszej pracy zamieszczone zestawienie wartości poszczególnych parametrów sterujących syntezą ( $F_1$ ,  $F_2$ ,  $F_3$ ,  $F_0$ ,  $A_V$ ,  $A_H$ ) oraz dane dotyczące iloczesów głoskowych dla wszystkich 24 diad CV zawierających spółgłoski szwarte. Przykładowe przebiegi parametrów dla wybranych diad przedstawiane zostały na rys. 1.



Rys.1. Wykresy przebiegów parametrów akustycznych dla diad **CV** zawierających głoski zwarte /b/, /d/, /p/ i /t/.

## 2.2. Synteza wyrazów z głoskami zwartymi w wygłosie.

Ze względu na stosunkowo niską wyrazistość spółgłosek zwartych występujących w wygłosie (co ma miejsce również w mowie naturalnej) uznano za bardziej celowe dokonanie ich syntezy raczej w wyrazach niż w asemantycznych diadach typu VC. Dla celów syntezy wybrano zestaw następujących 24 wyrazów jedno- i dwusylabowych (19 i 5) :

<u>byk</u>	<u>polip</u>	<u>bit</u>
<u>bek</u>	<u>typ</u>	<u>mit</u>
<u>lek</u>	<u>lep</u>	<u>byt</u>
<u>brak</u>	<u>pułap</u>	<u>bilet</u>
<u>mak</u>	<u>trop</u>	<u>tupet</u>
<u>bok</u>	<u>bób</u>	<u>beret</u>
<u>buk</u>	<u>drób</u>	<u>brat</u>
	<u>lub</u>	<u>ażot</u>
		<u>but</u>

Stopień dowolności w wyborze był ograniczony ze względu na niekompletność zbioru pozostających do dyspozycji (oprócz spółgłosek zwartych) głosek i diad syntetycznych, będących efektem poprzednich doświadczeń ([2]). Ostatecznie selekcji należało dokonać wśród wyrazów, w skład struktury fonetycznej których wchodziły głoski /i, t, e, a, o, u, m, n, p, l, r, j, w, v, b, d, p, t/ oraz wygłosowa głoska /k/. Preferowano wyrazy zawierające spółgłoski zwarte nie tylko w pozycji wygłosowej. Wzięto również pod uwagę frekwencję wyrazów. Uwzględniając ewentualną praktyczną przydatność zsyntezowanych wyrazów, jak też konieczność przeprowadzenia badań odsłuchowych w celu uzyskania ich oceny percepcyjnej zdecydowano, że zestaw powinien zawierać słowa powszechnie znane i dość często używane. Ze względu na ograniczenia fonotaktyczne całkowite spełnienie tego założenia okazało się niemożliwe. Stąd też w zestawie występują również wyrazy o zdecydowanie mniejszym subiektywnym prawdopodobieństwie, jak np. polip, bit, pułap (por. [5]).

Wyrazy syntetyczne uzyskiwano w oparciu o bibliotekę diad syntetycznych (por.[2]) i zbiór syntetycznych diad CV zawierających głoski zwarte w sposób opisany w pracach [1] i [3]. Zasadniczy problem tej części doświadczenia polegał na doborze



odpowiednich konfiguracji parametrów w obrębie wygłosowych spółgłosek zwartych /p/, /t/ (/b/ i /d/ nie występują w wygłosie).

Optymalna długość głosek zwartych w wygłosie (rozumiana jako suma zwarcia, płożji i aspiracji) była znacznie większa niż w nagłosie i wahała się w granicach od 110 do 170 ms w zależności od długości (liczby sylab) wyrazu. Odmienny niż w nagłosie był także przebieg amplitudy  $A_T$ . W obrębie przejścia między głoską poprzedzającą a głoską zwartą jej wartość znacznie malała (do 1 - 3 dB) i pozostawała na tym poziomie przez okres ok. 60 - 90 ms (czas zwarcia spółgłoski). Następnie gwałtownie wzrastała do 14 - 18 dB dla głosek /p/ i /t/ lub do 24 - 28 dB dla głoski /k/ (ok. 20 - 30 milisekundowa płożja), by wreszcie w czasie 40 - 50 ms zmaleć do wartości zerowej. W celu poprawy naturalności brzmienia spółgłosek zwartych posłużono się (podobnie jak w nagłosie) szumem aspiracji, którego amplitudę  $A_H$  kertałtowano na podobnej zasadzie, jak amplitudę tonu krtaniowego (lecz przy założeniu stosunkowo niskiej intensywności rzędu kilku dB). Zależnie od poprzedzającej głoski parametry formantowe dla spółgłosek zwartych wynoszą :

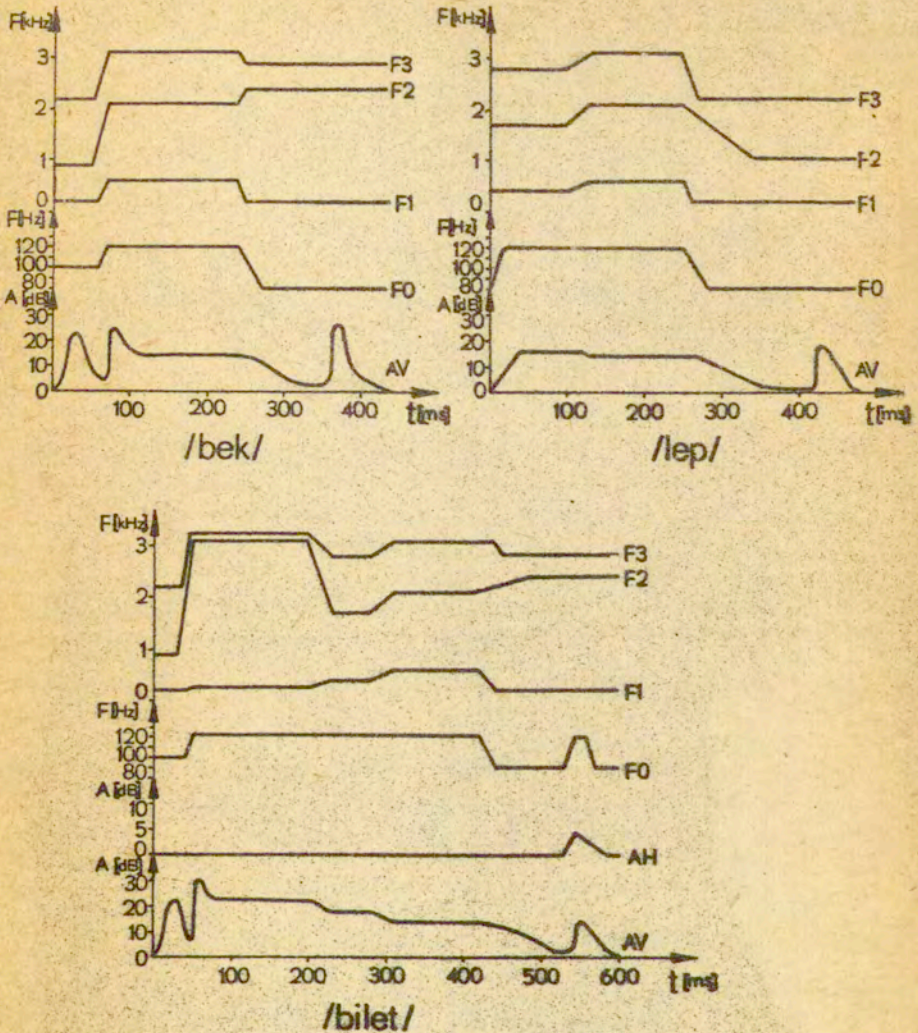
dla /k/	$F_1=200$ Hz	$F_2=1300 - 2400$ Hz	$F_3=2850$ Hz
/p/	$F_1=200$ Hz	$F_2=900 - 1300$ Hz	$F_3=2200$ Hz
/t/	$F_1=200$ Hz	$F_2=1900 - 2850$ Hz	$F_3=2850$ Hz

(por. np. dane w pracy [4], str. 100).

Przejście od częstotliwości formantowych samogłosek do częstotliwości formantowych spółgłosek zwartych realizowano w zależności od kontekstu głoskowego na przestrzeni :

dla $F_1$ i $F_3$	-	10 - 30 ms
dla $F_2$	-	40 - 110 ms

Wykresy przebiegów parametrów syntezy dla spółgłosek zwartych występujących w pozycji wygłosowej przedstawia rysunek 2 (dla wybranych wyrazów z zeszytowanego zestawu). Dodatek niniejszej pracy zawiera zestawienie wartości poszczególnych parametrów sterujących syntezą ( $F_1$ ,  $F_2$ ,  $F_3$ ,  $F_0$ ,  $A_T$ ,  $A_H$ ) oraz dane dotyczące iloczynów głoskowych dla wszystkich 24 wyrazów zeszytowanego zestawu.



Rys.2. Wykresy przebiegów parametrów akustycznych dla wybranych wyrazów zawierających głoski zwarte /k/, /p/, /t/ w wygłosie.

### 3. Ocena percepcyjna diad i wyrazów.

Podstawę do zobiiektywizowanej oceny produktu syntezy stanowiły wyniki badań odsłuchowych, w których zadaniem respondentów (20 osób) było zapisanie prezentowanych kolejno bodźców w wersji ortograficznej. Uzyskane w ten sposób listy diad i wyrazów odebranych przez respondentów dały bezpośrednią i jednoznaczna odpowiedź na temat wyrazistości poszczególnych bodźców.

#### 3.1. Metoda prezentacji.

Zestaw diad i wyrazów uzyskanych w eksperymencie nagrano na taśmie magnetofonowej w dwóch wersjach. Wersja pierwsza - będąca częścią adaptacyjną - zawierała pełny zestaw diad i pełny zestaw wyrazów występujących w porządku losowym w 3-sekundowych odstępach. W wersji drugiej, stanowiącej właściwy materiał testowy, w obu zestawach (diad i wyrazów) po dwukrotnej prezentacji każdego bodźca (odstęp pomiędzy wypowiedziami 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 oswojenie respondentów z brzmieniem mowy syntetycznej, odbiegającym od brzmienia mowy naturalnej. Badania odsłuchowe poproszono każdorazowo instrukcją, mającą wyjaśnić respondentom istotę eksperymentu i ich w nim zadanie.

#### 3.2. Wyniki oceny percepcyjnej diad.

Wyniki oceny percepcyjnej diad syntetycznych zawierających głoski swarte podane zostały w tabeli 1. Uzyskanych ogółem 190 błędnych rozpoznań<sup>1)</sup> daje średnią liczbę błędów przypadających na jednego słuchacza równą 8,5 (35,4 %).

Wśród błędnie odebranych diad można wyróżnić cztery grupy :

- diady ze spółgłoską /d/ odbieraną w większości jako /l/ lub rzadziej jako /b/,
- diady ze spółgłoską /t/ myloną często z /k/,
- diady ze spółgłoską /p/ odbieraną jako /t/,
- diady z samogłoską /i/ interpretowaną jako samogłoska /i/.

<sup>1)</sup> Pojęcia „niepoprawna odpowiedź”, „błędne rozpoznanie” należy traktować w sposób umowny.

Tabl. 1. Wyniki percepcyjnej oceny diad syntetycznych typu CV zawierających głoski zwarte.

Diada nada- na	Liczba popraw- nych rozpo- znań	Ogólna liczba błędnych rozpo- znań		Diady odebrane i liczba ich wystąpień					
			%						
1	2	3	4	5					
bi	20	-	-						
bá	8	12	60	bi	11	má	1		
be	19	1	5			me	1		
ba	20	-	-						
bo	14	6	30	be	1	mo	1	vo	4
bu	16	4	20			mu	2	vu	1
								vgu	1
dí	0	20	100	bi	8	li	11	ri	1
dá	0	20	100	bá	5	bi	4	lá	2
de	6	14	70	be	1	le	12	li	8
da	8	12	60	ba	1	ga	1	la	9
do	2	18	90	bo	7	lo	11	dla	1
du	2	18	90			lu	18		
pi	19	1	5			ti	1		
pá	11	9	45	pi	3	tá	5	ká	1
pe	20	-	-						
pa	19	1	5			ta	1		
po	15	5	25			to	5		
pu	17	3	15			tu	3		
ti	4	16	80			di	2	ki	13
tá	6	14	70			ká	5	gi	1
te	13	7	35	ti	5	ke	2	ki	3
ta	12	8	40	ti	1	tá	1	pe	1
to	19	1	5			ka	8	e	1
tu	20	-	-			ko	1	a	1
Łącz- na liczba	290	190	39.6						

Pierwsza grupa błędów (zamiana spółgłoski zwartej /d/ na /l/) świadczy o zbyt słabej płozji - skok amplitudy w obrębie płozji nie był dostatecznie gwałtowny. Interpretacja spółgłoski /t/ jako spółgłoski /k/ mająca miejsce w drugiej grupie błędnie odebranych dźwięków, wynika najprawdopodobniej ze zbyt ostrego wzrostu amplitudy w krótkim czasie (10 ms) oraz z analogicznego kształtowania przebiegu amplitudy szumu aspiracji ( $A_H$ ) na zbyt wysokim poziomie. W efekcie uzyskano wrażenie silnej aspiracji charakterystycznej dla spółgłoski zwartej /k/. Przyczyny wyleśnienia spółgłoski zwartej /p/ ze spółgłoską /t/ należy upatrywać w ich dużym subiektywnym podobieństwie.

Błędny odbiór samogłoski /a/ (zamienianej na /i/) wynika ze struktury formantowej samogłosek /a/ oraz /i/. Obie te samogłoski charakteryzuje niska częstotliwość pierwszego formantu i wysoka - drugiego. Jest to przyczyną subiektywnego podobieństwa tych samogłosek [8].

### 3.3. Wyniki oceny percepcyjnej wyrazów.

Wyniki oceny percepcyjnej wyrazów zawierających spółgłoski zwarte w pozycji wygłosowej przedstawia tabela 2. 226 uzyskanych ogółem błędnych rozpoznań daje średnią liczbę błędów przypadających na jednego słuchacza równą 9,4 (39.1%).

Ze względu na typową również dla mowy naturalnej stosunkowo niską wyrazistość spółgłosek zwartych, mały stopień zrozumiałości zsintezowanych wyrazów nie stanowi zaskoczenia. Trzeba przy tym uwzględnić fakt, że niektóre wyrazy cechowało niebyst wysokie subiektywne prawdopodobieństwo występowania.

Najwięcej niepoprawnych odpowiedzi uzyskano dla następujących wyrazów :

drób (20), brat (18), byt (18), lub (17), tupet (17), mit (17), bób (16), bit (15), beret (11).

Charakterystyczne jest, iż w grupie tej zdecydowana większość wyrazów zawiera w pozycji wygłosowej spółgłoskę /t/. Jej identyfikacja jako /k/, charakteryzującego się silną płozją i znacznym poziomem szumu aspiracji, ma swoją przyczynę w zbyt dużej amplitudzie przyjętej dla parametru  $A_H$  w ramach głoski /t/, jak również w kształtowaniu przebiegu tego parametru w sposób wypaklający płozję. Stąd też otrzymane odpowiedzi :

Tabl. 2. Wyniki percepcyjnej oceny wyrazów syntetycznych zawierających głoski zwarte w pozycji wygłosowej.

Wyraz nadany	Liczba poprawnych rozpoznań	Ogólna liczba błędnych rozpoznań		Wyrazy odebrane i liczba ich wystąpień	
		3	4		
1	2	3	4	5	
byk	15	5	25	bić 1 bip 1 byp 1	cyrk 1 tyk 1
bek	12	8	40	beg 1 bejt 1 berg 1	berek 3 beret 2
lek	12	8	40	lep 8	
brak	14	6	30	bab 1 blat 2	brat 2 byłam 1
mak	19	1	5	mag 1	
bok	17	3	15	bob 2	bop 1
buk	13	7	35	bób 2 bóg 2 bu 1	pug 1 purk 1
polip	12	8	40	poić 1 poiw 2 poji 1	poli 1 tolip 1 tonik 2
typ	12	8	40	byk 1 cyrk 3 kyk 1	kyrk 1 pip 1 tyk 1
lep	17	3	15	chleb 2	lek 1
pułap	19	1	5	tuła 1	
trop	14	6	30	klop 1 krok 1	tlók 1 top 1 trok 2
bób	4	16	80	bóg 2 bu 5 buk 5	but 3 hu 1

Wyraz nadany	Liczba popraw- nych rozpoz- nań	Ogólna liczba błędnych rozpoznań		Wyrazy odebrane i liczba ich wystąpień										
		3	4	5										
1	2	3	4	5										
drób	0	20	100	dół 1	lum 2	lód 5	lup 2	lub 2	lut 1	luk 7				
lub	3	17	85	lu 7	luk 3	lód 6	bób 1							
bit	5	15	75	bi 2	bilet 1	bicz 1	birk 1	bić 2	byk 1	big 1	mik 1	bik 3	mike 1	bil 1
mit	3	17	85	mi 2	mil 2	mig 8	kip 1	mik 4						
byt	2	18	90	bić 1	byk 12	bicz 1	byrt 1	by 2	kły 1					
bilet	14	6	30	bile 1	miłek 1	biret 1	piłe 1	byłem 1	piłek 1					
tupet	3	17	85	czubek 1	tubek 6	kubek 8	tuber 1	numer 1						
beret	9	11	55	bele 1	belek 4		berek 6							
brat	2	18	90	blat 3	brak 15									
młot	16	4	20	chłop 1	tłok 1	młot 2								
but	17	3	15	bób 1	buk 2									
Łącz- na liczba	254	226	47.1											

brak zamiast brat, byk zamiast byt, kubek i tubek zamiast tupet, mig i mik zamiast mit, big i bik zamiast bit, belek i berek zamiast beret.

Brak poprawnej odpowiedzi dla słowa „drób” wiąże się ze źle zrealizowanym połączeniem głosek /d/ i /r/ i - co za tym idzie - błędną interpretacją całego wyrazu.

W przypadku wyrazów „bób” (odbieranego jako bóg i buk) oraz wyrazu „lub” (odbieranego jako luk i lód) przyczyną błędnych rozpoznań należy upatrywać w zbyt dużym poziomie szumu aspiracji i kształtowaniu przebiegu jego amplitudy w sposób uwypuklający płoszę.

#### 4. Wnioski końcowe.

Zasadniczymi powodami niskiej wyrazistości diad i zrozumiałości wyrazów zawierających spółgłoski zwarte były :

- ograniczenia i niedoskonałość wykorzystywanego syntetyzatora,
- błędy popełniane w trakcie syntezy, dotyczące doboru wartości amplitudy dla parametrów  $A_V$  i  $A_H$  oraz kształtowania ich przebiegu (szczególnie w obrębie płoszi).

Mimo że ocena wyrazistości bodźców syntetycznych okazała się znacznie niższa od oczekiwanej, wyniki doświadczenia należy uznać za zachęcające. Regularność wśród „nieprawidłowych” rozpoznań wskazuje wyraźnie, jakiego rodzaju korekty powinny przynieść pożądane rezultaty. W przypadku kontynuowania prac nad syntezą przy użyciu dotychczasowego układu syntetyzującego konieczne byłoby dokonanie eksperymentu stanowiącego kolejny etap niniejszej pracy i dającego szczegółową odpowiedź na temat zasad syntezy spółgłosek zwartych.

Istotnym rezultatem eksperymentu jest dokonanie syntezy spółgłosek zwartych przy użyciu minimalnej liczby parametrów sterujących.



D O D A T E K

ZESTAWIENIE WARTOŚCI PARAMETRÓW DLA POSZCZEGÓLNYCH  
DIAŁ I WYRAZÓW SYNTETYCZNYCH.

W kolejnych rzędach oznaczonych cyframi od "1" do "7",  
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 - częstotliwość docelowa  $F_0$  (w Hz)
- 5 - amplituda tonu krtaniowego podawanego na tor formantowy (w dB), dla głosek zwartych podane są trzy wartości :  
wartość amplitudy segmentu poprzedzającego płoźję,  
szczytowa wartość amplitudy płoźji, wartość amplitudy  
segmentu następującego po płoźji, dla głoski uderzeniowej  
/r/ podane są dwie wartości : amplituda segmentu samo-  
głoskowego i spółgłoskowego
- 6 - amplituda szumu podawanego na tor formantowy aspiracji,  
(w dB)
- 7 - iloczas (w ms), w nawiasach podano wartości czasu trwania  
ugięć formantowych występujących między poszczególnymi  
głoskami, dla głosek zwartych podane są trzy wartości  
iloczasu : dla segmentu poprzedzającego płoźję, płoźji  
i segmentu następującego po płoźji.

	b	i	by	b	í
$F_1$ [Hz]	200	250	1	200	350
$F_2$ [Hz]	900	3100	2	900	2300
$F_3$ [Hz]	2200	3200	3	2200	2850
$F_0$ [Hz]	90	122	4	100	122
$A_v$ [dB]	8-20- 8	26	5	8-20- 8	22
$A_H$ [dB]	0	0	6	0	0
IL [ms]	50-30-20(20)	220	7	50-30-20(20)	220

	b	e	ba	b	a
1	200	.600	1	200	850
2	900	2100	2	900	1300

3	2200	3100	3	2200	2670
4	100	122	4	100	122
5	8-20- 8	20	5	8-22- 8	18
6	0	0	6	0	0
7	50-30-20(20)	220	7	50-30-20(20)	220
bo	b	o	bu	b	u
1	200	600	1	200	350
2	900	1000	2	900	850
3	2200	2870	3	2870	2670
4	100	122	4	100	122
5	8-22- 8	20	5	8-20- 8	26
6	0	0	6	0	0
7	50-30-20(20)	220	7	50-30-20(20)	220
di	d	i	di	d	i
1	200	250	1	200	350
2	2400	2780	2	2100	2300
3	2870	3160	3	2850	2850
4	100	122	4	100	122
5	8-22- 8	24	5	8-22- 8	24
6	0	0	6	0	0
7	50-30-10(50)	200	7	50-30-20(30)	210
de	d	e	da	d	a
1	200	600	1	200	850
2	1900	2100	2	2100	1300
3	2870	3100	3	2910	2670
4	100	122	4	100	122
5	8-20- 8	20	5	8-20- 8	18
6	0	0	6	0	0
7	50-30-30(10)	220	7	50-30-30(10)	220

do	d	o	du	d	u
1	200	600	1	200	350
2	1700	1000	2	1900	850
3	2870	2870	3	2870	2670
4	100	122	4	100	122
5	8-20- 8	20	5	8-20- 8	26
6	0	0	6	0	0
7	50-30-30(10)	220	7	50-30-30(40)	190

pi	p	i	py	p	i
1	200	250	1	200	350
2	900	3100	2	900	2300
3	2200	3200	3	2200	2850
4	90	122	4	90	122
5	3-10- 6	26	5	3-10- 6	22
6	4	0	6	4	0
7	10-10-10(30)	210	7	10-10-10(20)	220

pe	p	e	pe	p	e
1	200	600	1	200	850
2	900	2100	2	900	1300
3	2200	3100	3	2200	2670
4	90	122	4	90	122
5	3-10- 6	20	5	3-10- 6	18
6	4	0	6	4	0
7	10-10-10(20)	220	7	10-10-10(20)	220

po	p	o	pu	p	u
1	200	600	1	200	350
2	900	1000	2	900	850
3	2200	2870	3	2200	2670
4	90	122	4	90	122
5	3-10- 6	26	5	3-10- 6	26
6	4	0	6	4	0
7	10-10-10(20)	220	7	10-10-10(20)	220

ti	t	i	ty	t	±
1	670	250	1	670	350
2	3300	3100	2	2500	2300
3	3600	3200	3	3100	2850
4	90	122	4	90	122
5	7-14-10	26	5	7-14-10	26
6	1-3-3	0	6	2-4-2	0
7	20-10-10(30)	210	7	20-10-10(20)	220

te	t	e	ta	t	a
1	670	600	1	670	850
2	2100	2100	2	1750	1300
3	3000	3100	3	2670	2670
4	90	122	4	90	122
5	6-16- 8	18	5	6-14- 8	16
6	2-4-2	0	6	2-4-2	0
7	20-10-10(30)	210	7	20-10-10(20)	220

to	t	o	tu	t	u
1	670	600	1	670	350
2	1300	1000	2	1500	850
3	2870	2870	3	2670	2670
4	90	122	4	90	122
5	6-14- 8	16	5	6-14-10	22
6	2-4-2	0	6	2-4-2	0
7	20-10-10(20)	220	7	20-10-10(30)	200

bek	b	e	k
1	200	600	200
2	900	2100	2400
3	2200	3100	2870
4	100	122	80
5	3-22- 8	14	2-26- 8
6	0	0	0
7	10-30-10(20)	170(80)	40-20-50

beret	b	e	r	e	t
1	200	600	500	600	200
2	900	2100	1850	2100	2300
3	2200	3100	3200	3100	3700
4	100	122	122	122	110-130-90
5	3-22-6	16	22/10	12	3-18-10
6	0	0	0	0	2-4-2
7	10-30-10(10)	140(30)	90(30)	50(90)	40-20-50

bilet	b	i	i	e	t
1	200	250	400	600	200
2	900	3100	1700	2100	2400
3	2200	3200	2800	3100	2850
4	100	122	122	122	90-122-90
5	3-22-6	22	18	14	3-14-8
6	0	0	0	0	2-4-2
7	10-30-10	150(30)	50(30)	100(80)	50-20-40

bit	b	i	t
1	200	250	200
2	900	3100	900
3	2200	3200	2850
4	100	122	90-122-90
5	3-22-6	22	3-16-8
6	0	0	2-3-2
7	10-30-10(10)	160(90)	60-20-50

bok	b	o	k
1	200	600	200
2	900	1000	1400
3	2200	2870	2850
4	100	122	80
5	3-22-8	14	2-26-8
6	0	0	0
7	10-30-20(10)	160(20)	100-20-50

brak	b	r	a	k
1	200	500	850	250
2	900	1400	1300	1700
3	2200	3500	2670	3200
4	122	122	122	80
5	2-22- 8	24/10	10	2-24-8
6	0	0	0	0
7	10-30-20(20)	80(30)	110(50)	90-20-40

brat	b	r	a	t
1	200	500	850	200
2	900	1400	1300	2100
3	2200	3500	2670	2850
4	100	122	122	90-122-90
5	3-20-8	24/12	10	3-16-10
6	0	0	0	3-4-3
7	10-20-20(20)	80(30)	120(90)	40-20-40

bób	b	u	p
1	200	350	200
2	900	800	900
3	2870	2670	2850
4	100	122	90
5	8-20- 8	20	1-14- 4
6	0	0	0-4-2
7	50-30-30(10)	170(40)	100-30-20

buk	b	u	k
1	200	350	200
2	900	800	1300
3	2200	2670	2850
4	100	122	80
5	3-22-8	20	2-26-8
6	0	0	0
7	10-30-20(10)	160(40)	100-20-50

but	b	u	t
1	200	350	200
2	900	800	1900
3	2670	2670	2650
4	100	122	90-122-90
5	8-20-8	20	3-16-8
6	0	0	0
7	50-30-30(10)	150(40)	100-20-50

byk	b	z	k
1	200	350	200
2	900	2300	1900
3	2200	2850	2650
4	100	122	80
5	3-22-8	20	2-26-8
6	0	0	0
7	10-30-10(20)	120(40)	90-20-50

byt	b	z	t
1	200	350	200
2	900	2250	2400
3	2200	2850	2650
4	100	122	90-122-90
5	3-22-6	20	3-16-6
6	0	0	1-3-1
7	10-30-10(10)	160(70)	90-30-40

drób	d	r	u	p
1	200	400	300	200
2	1700	1050	800	900
3	2670	3500	2670	2650
4	100	122	122	90-120-90
5	10-16-12	22/15	22	1-15-6
6	0	0	0	0
7	20-20-20(30)	80(30)	100(10)	150-30-30

lck	l	e	k
1	400	600	200
2	1700	2100	2400
3	2800	3100	2850
4	122	122	80
5	16	14	2-24- 8
6	0	0	0
7	100(30)	130(10)	110-20-50

lep	l	e	p
1	400	600	200
2	1700	2100	1050
3	2800	3100	2200
4	122	122	80
5	16	14	1-18- 4
6	0	0	0
7	100(30)	110(90)	90-30-20

lub	l	u	p
1	400	300	200
2	1650	750	900
3	2800	2670	2850
4	122	122	90
5	16	20	1-12- 8
6	0	0	0-4-2
7	100(70)	90(70)	80-30-20

mak	m	a	k
1	200	850	200
2	1250	1400	1800
3	2200	2670	2850
4	122	122	80
5	22	10	2-26- 8
6	0	0	0
7	120(10)	110(40)	100-20-50



mit	m	i	t
1	200	250	200
2	1250	3100	2400
3	2200	3200	2850
4	122	122	90-122-90
5	20	22	3-16-8
6	0	0	2-3-2
7	120(10)	130(80)	60-20-50

młot	m	w	o	t
1	200	350	600	200
2	1050	800	1000	1950
3	2200	2600	2870	2850
4	122	122	122	90-122-90
5	20	16	12	3-18-8
6	0	0	0	2-3-2
7	100(10)	30(90)	100(80)	70-20-40

polip	p	o	l	i	p
1	200	600	350	250	200
2	950	1000	2400	3100	1050
3	2200	2870	2800	3200	2200
4	90	122	122	122	90
5	3-10-6	16	16	20	1-14-4
6	2-4-4	0	0	0	1-2-1
7	10-10-10(10)	130(30)	100(30)	90(90)	70-30-20

pułap	p	u	w	a	p
1	200	350	350	850	200
2	950	800	750	1300	900
3	2200	2670	2600	2670	2200
4	90	122	122	122	80
5	3-16-6	22	14	8	2-18-8
6	2-4-4	0	0	0	0
7	10-10-10(20)	40(80)	30(80)	90(110)	60-20-30

trop	t	r	o	p
1	350	500	600	200
2	1700	1400	1000	900
3	2670	3500	2670	2200
4	80-90-100	122	122	90
5	7-16-7	22/14	12	1-14-8
6	2-4-0	0	0	0-4-2
7	20-20-10(10)	80(30)	120(90)	90-20-30

tupet	t	u	p	e	t
1	250	350	200	600	250
2	1500	900	900	2100	2300
3	2200	2670	2200	3100	2850
4	100	122	90	122	90-122-90
5	3-20-12	22	1-20-14	16	3-16-10
6	0	0	0	0	1-3-2
7	10-10-20(10)	110(30)	80-10-30(10)	150(60)	40-20-50

typ	t	±	p
1	200	350	200
2	1900	2250	1300
3	2850	2850	2200
4	110-130-110	122	90
5	3-16-6	22	1-18-4
6	0	0	0
7	10-20-10(10)	160(110)	80-30-20

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 diad polskich i ich percepcja, Prace IPPT 7/1984, Warszawa, 1984.
- [3] CIARKOWSKI, R., IMIOŁCZYK, J., Synteza wybranych wyrazów polskich i ich percepcja, Prace IPPT 1/1985, Warszawa, 1985.
- [4] FRĄKOWIAK-RICHTER, L., Vowel - formant transitions at stop-consonant boundaries in Polish, Speech Analysis and Synthesis, Vol. II, PWN, Warszawa, 1970.
- [5] IMIOŁCZYK, J., Subiektywne prawdopodobieństwo wyrazów. Podstawowy słownik frekwencyjny języka polskiego. PWN, (w druku).
- [6] JASSEM, W., Podstawy fonetyki akustycznej, PWN, Warszawa, 1973.
- [7] KACPROWSKI, J., MIKIEL, W., Simplified rules for parametric synthesis of nasal and stop consonants in C-V syllables by means of the "terminal-analog" speech synthesizer, Acustica, vol. 16, zeszyt 6, s. 356-364, S.Hirzel Verlag - Stuttgart 1965/66.
- [8] ŁOBACZ, P., DEMENKO, G., Zależność percepcji segmentalnych cech samogłosek polskich od struktury długotrwałej pamięci leksykalno - fonematycznej, Prace IPPT 40/1983, Warszawa, 1983.