

2.23 — akustyka mowy



Janusz Imiolczyk

**OKREŚLENIE GRANIC PERCEPCYJNYCH
MIĘDZY GŁOSEM MĘSKIM, KOBIECYM
I DZIECIĘCYM
W IZOLOWANYCH SYNTETYCZNYCH
SAMOGŁOSKACH POLSKICH**

5/1990

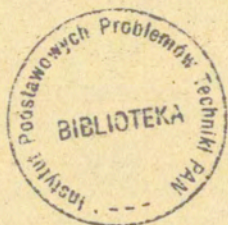
P. 269



WARSZAWA 1990

<http://rcin.org.pl>

Praca wpłynęła do Redakcji dnia 20 listopada 1989 r.



56783



Na prawach rękopisu

Instytut Podstawowych Problemów Techniki PAN

Nakład 140 egz. Ark.wyd. 1,85 Ark.druk. 2,5

Oddano do drukarni w lutym 1990 r.

Nr zamówienia 64/90

Warszawska Drukarnia Naukowa, Warszawa,
ul.Śniadeckich 8

<http://rcin.org.pl>

OKREŚLENIE GRANIC PERCEPCYJNYCH MIĘDZY GŁOSEM MĘSKIM,
KOBIECYM I DZIECIĘCYM W IZOLOWANYCH SYNTETYCZNYCH
SAMOGŁOSKACH POLSKICH¹⁾.

Streszczenie.

W pracy podjęto problem granic percepcyjnych między głosem męskim i kobiecym oraz kobiecym i dziecięcym. Wykorzystany materiał badawczy obejmował 777 wypowiedzi syntetycznych, stanowiących realizacje sześciu ustnych samogłosek polskich: /i/, /ɨ/, /e/, /a/, /o/ i /u/. Wartości częstotliwości formantowych i F0 były kształtowane w tych wypowiedziach w taki sposób, aby dla każdej samogłoski uzyskać realizacje reprezentujące typowy głos męski i typowy głos dziecięcy, oraz szereg realizacji pośrednich.

W badaniach odsłuchowych wzięło udział 20 osób, których zadaniem było rozpoznanie każdego z bodźców jako określonej samogłoski polskiej oraz zaklasyfikowanie go jako wypowiedzi mężczyzny, kobiety lub dziecka.

Analiza uzyskanych odpowiedzi wykazała, że częstotliwość podstawowa jest zasadniczym czynnikiem decydującym o uznaniu wypowiedzi za męską, kobiecą bądź dziecięcą i że odgrywa ona w związku z tym kluczową rolę w percepcyjnej normalizacji toru głosowego: częstotliwości formantowe, powszechnie uważane za określające fonematyczną przynależność samogłoski, mogą być właściwie zinterpretowane dopiero na podstawie informacji, której nośnikiem jest F0.

Wstęp.

W wydanej w 1986 r. pracy zbiorowej pt. "Invariance and Variability in Speech Processes" [29] kilkudziesięciu autorów na kilkuset stronach koncentruje swoją uwagę na różnych aspektach problemu zmienności w sygnale mowy. Mimo wielu lat badań i setek eksperymentów nie udało się wyjaśnić, w jaki sposób zmienność ta, która przysparza tak wiele trudności np. w

1) Praca wykonana w ramach CPBP 02.03 .

automatycznym rozpoznawaniu mowy, eliminowana jest z niezwykłą skutecznością w procesie percepcji.

Poszukując odpowiedzi na to pytanie niektórzy autorzy (zwłaszcza K. N. Stevens i S. Blumstein) starali się wykazać, że w zmiennym z pozoru sygnale występują pewne cechy inwariantne, które stanowią "punkt zaczepienia" dla percepcji. Inni badacze uznali ten kierunek poszukiwań za jałowy, wychodząc z założenia, że należy się raczej skoncentrować na próbach ustalenia, w jaki sposób słuchacze radzą sobie z charakteryzującą sygnał mowy zmiennością (np. [11]).

Jedno z najistotniejszych źródeł zmienności, z jakimi ma do czynienia słuchacz w procesie percepcji stanowi zróżnicowanie osobnicze głosów. Jak zauważa Holmes [20], różnice w zakresie struktury akustycznej między realizacjami tej samej głoski przez głos męski, kobiecy i dziecięcy są tak znaczne, że fonetyczna ekwiwalentność tych wypowiedzi musi mieć podłoże percepcyjne. Stwierdza on dalej, że między cechami akustycznymi występują zapewne systematyczne relacje, które umożliwiają interpretację zupełnie różnych sygnałów akustycznych jako tożsamyh językowo (str. 347). Wynikałoby stąd, iż (1) niezbędny warunek skutecznej percepcyjnej eliminacji zmienności typu osobniczego stanowi zdolność rozróżniania co najmniej trzech głównych rodzajów głosu, a mianowicie głosu męskiego, kobiecego i dziecięcego oraz (2) informacje o rodzaju głosu zawarte są w sygnale mowy pod postacią odpowiednich relacji cech akustycznych.

Możliwość percepcyjnej identyfikacji głosu jako męskiego, kobiecego czy dziecięcego implikuje istnienie mniej lub bardziej wyraźnych granic między tymi trzema typami głosów. Jak można przypuszczać, określenie tych granic oraz ustalenie, jakie parametry akustyczne determinują ich przebieg powinno przyczynić się do wyjaśnienia istotnych kwestii dotyczących mechanizmu percepcyjnej normalizacji toru głosowego nadawcy, tj. eliminacji z sygnału mowy zmienności o charakterze osobniczym.

Za najbardziej odpowiedni z punktu widzenia przyjętego celu pracy, uznano materiał obejmujący izolowane samogłoski. Posłużenie się wypowiedziami dłuższymi (np. sylabami typu CV)

wprowadziłyby dodatkowe źródła zmienności i mogłyby w konsekwencji znacznie utrudnić interpretację wyników.

1. Nagranie, analiza i resynteza samogłosek naturalnych.

Etap wstępny pracy obejmował nagranie w warunkach studyjnych sześciu ustnych samogłosek polskich: [i], [ɨ], [e], [a], [o], [u], wymówionych przez jeden głos męski. Wykorzystano w tym celu mikrofon austriacki RS 249-946 oraz magnetofon kasetowy Revox B710. Zwrócono uwagę, aby poziom nagrania oraz przebieg F0 w poszczególnych wypowiedziach (rosnąco-opadający) były możliwie zbliżone.

Wypowiedzi zostały następnie poddane filtracji dolno-przepustowej, spróbkowane z częstotliwością 10 kHz i zapisane na dysku komputera MASSCOMP MC5400.

Kolejny krok stanowiła wszechstronna analiza wypowiedzi przy użyciu pakietu wyspecjalizowanych programów do analizy sygnału mowy o nazwie AUDLAB [2]. Wykorzystano następujące możliwości tego pakietu:

- ekstrakcja F0
- wyznaczanie widm chwilowych i średnich (FFT, 256 i 512 punktów)
- wyznaczanie spektrogramów 2- i 3-wymiarowych,
- pomiar długości segmentów.

Dane pochodzące z analizy zostały wykorzystane do (re-)syntezy samogłosek przy użyciu software'owego syntetyzatora formantowego Klatta¹⁾. Posłużono się częścią równoległą syntetyzatora, co stworzyło możliwość indywidualnego sterowania amplitudami poszczególnych formantów. Dążąc do osiągnięcia maksymalnie naturalnego brzmienia samogłosek, wybrano drugie z dwu dostępnych źródeł pobudzania ("ss" = 2). Przyjęto częstotliwość konwersji C/A równą 10 000 próbek/sek²⁾.

Oprócz "standardowych" parametrów występujących we wszelkiego rodzaju syntetyzatorach formantowych (np. częstotliwości formantowe, F0 itp.), posłużono się dwoma

1) Wykorzystano jedną z późniejszych wersji programu opisanego w [22]; zob. też [1].

2) Filtr antyaliasingowy miał nominalną częstotliwość odcinania równą 4,8 kHz.

dodatkowymi: "no" i "tl". Pierwszy z nich odpowiada liczbie próbek w fazie otwartej okresu tonu pobudzania, natomiast drugi pozwala regulować kąt nachylenia obwiedni widma tego tonu. Zasady posługiwania się tymi dwoma parametrami zostaną podane w dalszej części pracy (punkty 2.2 i 2.3).

Samogłoskę zsyntetyzowaną w oparciu o dane z analizy samogłoski naturalnej analizowano ponownie przy pomocy AUDLAB-u, porównując jej cechy widmowe z cechami widmowymi jej wzorca naturalnego. Zabieg ten powtarzano do czasu uzyskania niemal idealnej (możliwej do uzyskania) zgodności w obrazie widmowym obu głosek (brano pod uwagę wszystkie widma chwilowe składające się na całą wypowiedź).

Aby zminimalizować pozajakościowe różnice między poszczególnymi samogłoskami syntetycznymi, zuniformalizowano ich długość (do 440 ms) oraz nadano im identyczny przebieg amplitudy oraz częstotliwości podstawowej, wyekstrahowane z naturalnej samogłoski [e]¹⁾. Także wartości parametrów "no" i "tl" były dla wszystkich samogłosek jednakowe.

W efekcie, różnice między samogłoskami syntetycznymi ograniczone były do: (1) częstotliwości, (2) amplitud i (3) szerokości wstęg formantów oraz (4) ogólnego poziomu wzmocnienia, którego modyfikacja miała na celu osiągnięcie w przybliżeniu jednakowej głośności wszystkich samogłosek.

Wymienione wyżej parametry miały w każdej samogłosce syntetycznej stałą, typową dla niej wartość. Zmienne w czasie były natomiast - co oczywiste - przebiegi amplitudy tonu pobudzania oraz częstotliwości podstawowej, których wartości uaktualniano co 5 ms. Czas narastania amplitudy do wartości docelowej na początku samogłoski wynosił 35 ms, zaś czas jej wygasania - 50 ms. Kształt konturu F0 był następujący:

t [ms]	0	100	245	390	435
F0 [Hz]	139,0	144,0	119,0	112,0	119,5

1) Ściśle rzecz biorąc, oba przebiegi stanowiły aproksymacje przebiegów charakteryzujących naturalną samogłoskę [e].

Zmiany F0 pomiędzy kolejnymi momentami "zwrotnymi" ($t = 0$ ms, $t = 100$ ms itd.) miały charakter liniowy.

Poniżej (Tablica 1) przedstawiono wartości częstotliwości formantowych dla poszczególnych samogłosek.

Tablica 1

Formant [Hz]	Samogłoska					
	i	ɨ	e	a	o	u
F1	250	365	600	800	620	340
F2	2170	1960	1720	1220	950	760
F3	3100	2530	2620	2760	2730	2250
F4	3770	3230	3050	3760	3920	3030
F5	4500	4100	4350	4850	4560	4140

2. Synteza samogłosek kobiecych.

2.1. Częstotliwości formantowe.

Samogłoski syntetyczne o wartościach częstotliwości formantowych podanych w Tablicy 1 przyjęto za reprezentatywne dla głosu męskiego. Wartości te zostały następnie przeskalowane w taki sposób, aby dla każdej samogłoski pierwotnej (męskiej) uzyskać jej fonematycznie ekwiwalentny odpowiednik o częstotliwościach formantowych typowych dla wypowiedzi głosu kobiecego. Posłużono się w tym celu tzw. czynnikami skalującymi k (*k scaling factors*) podanymi przez Fanta ([14], Tablica 1, str. 87).

Jak wiadomo, długość ponadkrtaniowego toru głosowego jest u kobiet o kilkanaście procent mniejsza niż u mężczyzn (np. [5], [14]). Jednakże głównie ze względu na niejednakowe proporcje długości jamy gardłowej i ustnej u obu płci częstotliwości odpowiadających sobie formantów męskich i kobiecych nie da się ze sobą powiązać przy użyciu jednego tylko czynnika

skalującego¹⁾. Jak wykazał Fant ([14], [15]), związane z pićią różnice w zakresie częstotliwości (co najmniej) trzech najniższych formantów można wyrazić jedynie poprzez użycie odrębnych czynników skalujących (odpowiednio: k1, k2 i k3), których wartość zależy ponadto od typu samogłoski.

W Tabelicy 2.1 uwidocznione są wartości czynników skalujących przyjęte w niniejszej pracy. Są one wynikiem pewnej modyfikacji danych Fanta ([14], tab. 1), uwzględniającej artykulacyjno-akustyczną odmienność samogłosek polskich od ich szwedzkich i amerykańskich "odpowiedników". Czynnikiem k4 i k5, nie występującym w cytowanej pracy, nadano wartość 17 %, równą w przybliżeniu średniej arytmetycznej ze wszystkich wartości k3 zestawionych przez Fanta. W Tabelicy 2.1 podano także kobiece częstotliwości formantowe poszczególnych samogłosek, uzyskane poprzez pomnożenie formantów męskich przez odpowiednie czynniki k.

Tabelica 2.1

Samogł	k1 [%]	F1 [Hz]	k2 [%]	F2 [Hz]	k3 [%]	F3 [Hz]	k4 [%]	F4 [Hz]	k5 [%]	F5 [Hz]
i	8	270	22	2647	16	3596	17	4411	17	5265
ɨ	10	402	24	2430	20	3036	17	3779	17	4797
e	24	744	20	2064	19	3118	17	3569	17	5090
a	16	928	16	1415	16	3202	17	4399	17	5675
o	15	713	14	1083	15	3140	17	4586	17	5335
u	10	374	5	798	19	2678	17	3545	17	4844

Ze względu na zróżnicowanie procentowych przyrostów wartości poszczególnych formantów przy przejściu od głosu męskiego do kobiecego dla poszczególnych samogłosek dobór pośrednich zestawów formantowych wymagał potraktowania każdej z samogłosek

1) Zależności między długościami poszczególnych odcinków toru artykulacyjnego i częstotliwościami formantów samogłoskowych opisane są w klasycznej pracy Fanta [13] (zob. np. tablice na str. 120).

w odrębny sposób. Przyjęto założenie, że różnica między kolejnymi zestawami wartości formantów odpowiadającymi skróceniu toru głosowego nie będzie dla żadnego z trzech pierwszych formantów większa od podanych przez Flanagan [16] progów rozróżnialności równych: ± 20 Hz dla F1, ± 50 Hz dla F2 i ± 75 Hz dla F3. W tej sytuacji otrzymano dla każdej z samogłosek następującą liczbę zestawów formantów: [i] - 11, [ɨ] - 11, [e] - 9, [a] - 8, [o] - 7 i [u] - 7. Zestawy te przedstawione są w Tablicach 2.2 + 2.7.

Tablica 2.2

Zestaw formant.	F1 [Hz]	F2 [Hz]	F3 [Hz]	F4 [Hz]	F5 [Hz]
1	250	2170	3100	3770	4500
2	252	2218	3150	3834	4577
3	254	2265	3199	3898	4653
4	256	2313	3249	3962	4730
5	258	2361	3298	4026	4806
6	260	2409	3348	4091	4883
7	262	2456	3398	4155	4959
8	264	2504	3447	4219	5036
9	266	2552	3497	4283	5112
10	268	2599	3546	4347	5189
11	270	2647	3596	4411	5265

Samogłoska [i]: $\Delta k_1 = 2$ Hz, $\Delta k_2 = 47,7$ Hz, $\Delta k_3 = 49,6$ Hz,
 $\Delta k_4 = 64,1$ Hz, $\Delta k_5 = 76,5$ Hz

Tablica 2.3

Zestaw formant.	F1 [Hz]	F2 [Hz]	F3 [Hz]	F4 [Hz]	F5 [Hz]
1	365	1960	2530	3230	4100
2	369	2007	2581	3285	4170
3	372	2054	2631	3340	4239
4	376	2101	2682	3395	4309
5	380	2148	2732	3450	4379
6	383	2195	2783	3505	4449
7	387	2242	2834	3559	4518
8	391	2289	2884	3614	4588
9	394	2336	2935	3669	4658
10	398	2383	2985	3724	4727
11	402	2430	3036	3779	4797

Samogłoska [ɛ]: $\Delta k_1 = 3,7$ Hz, $\Delta k_2 = 47$ Hz, $\Delta k_3 = 50,6$ Hz,
 $\Delta k_4 = 55$ Hz, $\Delta k_5 = 69,7$ Hz

Tablica 2.4

Zestaw formant.	F1 [Hz]	F2 [Hz]	F3 [Hz]	F4 [Hz]	F5 [Hz]
1	600	1720	2620	3050	4350
2	618	1763	2682	3115	4443
3	636	1806	2744	3180	4535
4	654	1849	2807	3245	4628
5	672	1892	2869	3310	4720
6	690	1935	2931	3375	4813
7	708	1978	2993	3439	4905
8	726	2021	3055	3504	4998
9	744	2064	3118	3569	5090

Samogłoska [e]: $\Delta k_1 = 18$ Hz, $\Delta k_2 = 43$ Hz, $\Delta k_3 = 62,2$ Hz,
 $\Delta k_4 = 64,9$ Hz, $\Delta k_5 = 92,5$ Hz

Tablica 2.5

Zestaw formant.	F1 [Hz]	F2 [Hz]	F3 [Hz]	F4 [Hz]	F5 [Hz]
1	800	1220	2760	3760	4850
2	818	1248	2823	3851	4968
3	837	1276	2886	3943	5086
4	855	1304	2950	4034	5204
5	873	1332	3013	4125	5322
6	892	1360	3076	4217	5440
7	910	1387	3139	4308	5557
8	928	1415	3202	4399	5675

Samogłoska [a]: $\Delta k_1 = 18,3$ Hz, $\Delta k_2 = 27,9$ Hz, $\Delta k_3 = 63,2$ Hz,
 $\Delta k_4 = 91,3$ Hz, $\Delta k_5 = 117,9$ Hz

Tablica 2.6

Zestaw formant.	F1 [Hz]	F2 [Hz]	F3 [Hz]	F4 [Hz]	F5 [Hz]
1	620	950	2730	3920	4560
2	636	972	2798	4031	4689
3	651	994	2867	4142	4818
4	667	1017	2935	4253	4947
5	682	1039	3003	4364	5076
6	698	1061	3072	4475	5206
7	713	1083	3140	4586	5335

Samogłoska [o]: $\Delta k_1 = 15,5$ Hz, $\Delta k_2 = 22,2$ Hz, $\Delta k_3 = 68,3$ Hz,
 $\Delta k_4 = 111$ Hz, $\Delta k_5 = 129,1$ Hz

Tablica 2.7

Zestaw formant.	F1 [Hz]	F2 [Hz]	F3 [Hz]	F4 [Hz]	F5 [Hz]
1	340	760	2250	3030	4140
2	346	766	2321	3116	4257
3	351	773	2393	3202	4375
4	357	779	2464	3287	4492
5	363	785	2535	3373	4609
6	368	792	2607	3459	4727
7	374	798	2678	3545	4844

Samogłoska [u]: $\Delta k_1 = 5,6$ Hz, $\Delta k_2 = 6,3$ Hz, $\Delta k_3 = 71,3$ Hz,
 $\Delta k_4 = 85,8$ Hz, $\Delta k_5 = 117,3$ Hz

Przy syntezie wszystkich wypowiedzi reprezentujących tę samą samogłoskę wartości amplitud i szerokości wstęg poszczególnych formantów były stałe i równe wartościom charakteryzującym zestaw 1. Częstotliwości F5 wyższe od 5 kHz nie były uwzględniane.

2.2. Częstotliwość podstawowa.

Na ogół stwierdza się, iż częstotliwość podstawowa jest średnio w głosie kobiecym o ok. oktawę wyższa niż w męskim (np. [13], str.242) i wynosi ok. 220 Hz, choć oczywiście różnicowania osobnicze są dość znaczne. W niniejszej pracy przyjęto (za Klatt'em i Klatt [23]) za typową dla głosu kobiecego częstotliwość podstawową równą 1,7 męskiej. Kobięcy przebieg F0 oraz 6 przebiegów pośrednich uzyskano przez pomnożenie wartości F0 w każdym z punktów nieciągłości przebiegu męskiego przez odpowiedni czynnik skalujący (1,1, 1,2...1,7). Średnia wartość F0 wynosiła w przebiegu męskim 128 Hz, a w kobiecym - 218 Hz (Tablica 2.18).

Jak wykazały wyniki szeregu prac (np. [19], [23], [28]), proporcja długości pulsu krtaniowego i długości fazy zamkniętej okresu jest w głosie żeńskim wyższa niż w męskim, a

sam kształt pulsu - bardziej symetryczny (sinusoidalny). Konsekwencją tego jest m. in. większa stromość obwiedni widma tonu krtaniowego w głosie kobiecym. W porównaniu z głosem męskim, w którym obniżenie poziomu widma wynosi ok. -12 dB/oktawę (np. [12], [13], [17]), głos kobiety charakteryzuje bardziej stromy spadek wynoszący ok. -15 + -18 dB/oktawę ([23], [28]). Monsen i Engebretson [28] stwierdzają jednakże, iż w przypadku fonacji o względnie wysokiej (podwyższonej w stosunku do normalnej)¹⁾ częstotliwości podstawowej również w głosie męzczyzny spadek poziomu obwiedni widma może się zbliżać do -18 dB/oktawę.

Tablica 2.8

czyn. skal.	F0 [Hz]					średn. F0 [Hz]	no	% dług. okr.	t1
	0ms	100ms	245ms	390ms	435ms				
1,0	139,0	144,0	119,0	112,0	119,5	128	33	42%	0
1,1	152,9	158,4	130,9	123,2	131,5	141	33	47%	3
1,2	166,8	172,8	142,8	134,4	143,4	154	33	51%	6
1,3	180,7	187,2	154,7	145,6	155,4	166	33	55%	9
1,4	194,6	201,6	166,6	156,8	167,3	179	33	59%	12
1,5	208,5	216,0	178,5	168,0	179,3	192	33	63%	15
1,6	222,4	230,4	190,4	179,2	191,2	205	33	68%	18
1,7	236,6	244,8	202,3	190,4	203,2	218	33	72%	21

(Opis w tekście)

Uwzględniając powyższe obserwacje, przyjęto w niniejszej pracy dwie zasady dotyczące charakterystyki tonu podstawowego (zob. Tablica 2.8):

- (1) wraz ze wzrostem F0 zwiększana była (za pomocą parametru "t1") stromość spadku obwiedni widma; wartości "t1" = 0 odpowiadał spadek równy ok. -12 dB/oktawę, zaś dla "t1" = 21

1) Według Fanta ([13], str. 242), zakres zmienności F0 w typowym głosie męskim wynosi od 60 Hz do 240 Hz.

wynosił on ok. -17 dB/oktawę

(2) długość fazy otwartej okresu wynosiła we wszystkich przypadkach 3,3 ms ("no"= 33), co oznacza, że stanowiła ona od 42 % do 72 % całkowitej długości okresu.

Każdy z ośmiu przebiegów F0 połączono na drodze syntezy z każdym z zestawów formantowych, wyodrębnionych dla poszczególnych samogłosek (Tablice 2.2 + 2.7)¹⁾. Ponadto zsyntetyzowano: (1) dla wszystkich samogłosek - wypowiedzi będące kombinacjami (a) pierwszego oraz (b) ostatniego zestawu formantów z ostatnim (żeńskim) przebiegiem F0, w którym jednakże wartość "t1" była równa 0 oraz (2) dla samogłosek [e] i [a] - po kilkanaście dodatkowych bodźców, w których "t1" również było równe 0.

Ogółem dokonano na tym etapie syntezy 471 bodźców, w tym: dla [i] - 90, dla [ɨ] - 90, dla [e] - 93, dla [a] - 82, dla [o] - 58 i dla [u] - także 58 bodźców.

3. Synteza samogłosek dziecięcych.

3.1. Częstotliwości formantowe.

Syntezy samogłosek dziecięcych dokonano przyjmując jako punkt wyjścia obserwację Fanta [14], iż różnice w zakresie częstotliwości formantowych samogłosek kobiecych i dziecięcych

Tablica 3.1

Samogł.	F1	F2	F3	F4
i	319	3123	4243	5205
ɨ	474	2867	3582	4459
e	878	2436	3679	4211
a	1095	1670	3778	5191
o	841	1278	3705	5411
u	441	942	3160	4183

1) Jak zauważa Fant [12], obwiednia widma tonu krtaniowego nie zależy w istotny sposób od rodzaju samogłoski.

można wyrazić za pomocą prostego czynnika, niezależnego od typu samogłoski. Wszystkie wartości "kobięcych" częstotliwości formantowych (tj. wartości występujące w ostatnim zestawie każdej z Tablic 2.2 + 2.7) pomnożono w związku z tym przez czynnik $k = 18\%$ (zob. Tablica 3.1 powyżej).

Oprócz "docelowych" samogłosek dziecięcych zsyntetyzowano także samogłoski o pośrednich (między kobiecymi a dziecięcymi) częstotliwościach formantowych, kierując się przy tym przyjętą wcześniej zasadą (por. pkt. 2.1), iż $\Delta k_1 < 20$ Hz, $\Delta k_2 < 50$ Hz, $\Delta k_3 < 75$ Hz. Otrzymano następującą liczbę podziałów dla poszczególnych samogłosek: [i] - 10, [ɨ] - 9, [e] - 8, [a] - 9, [o] - 8 i [u] - 7. W Tablicach 3.2 + 3.7 przedstawiono zestawy formantowe odpowiadające tym podziałom. Numeracja tych zestawów stanowi kontynuację numeracji z Tablic 2.2 + 2.7.

Tablica 3.2.

Zestaw formant.	F1 [Hz]	F2 [Hz]	F3 [Hz]	F4 [Hz]
12	275	2695	3661	4490
13	280	2742	3725	4570
14	285	2790	3790	4649
15	290	2837	3855	4729
16	295	2885	3920	4808
17	299	2933	3984	4887
18	304	2980	4049	4967
19	309	3028	4114	5046
20	314	3075	4178	5126
21	319	3123	4243	5205

Samogłoska [i]: $k_n = 18\%$; $\Delta k_1 = 4,9$ Hz, $\Delta k_2 = 47,6$ Hz, $\Delta k_3 = 64,7$ Hz, $\Delta k_4 = 79,4$ Hz

Tablica 3.3

Zestaw formant.	F1 [Hz]	F2 [Hz]	F3 [Hz]	F4 [Hz]
12	410	2479	3097	3855
13	418	2527	3157	3930
14	426	2576	3218	4006
15	434	2624	3279	4081
16	442	2673	3340	4157
17	450	2722	3400	4233
18	458	2770	3461	4308
19	466	2819	3522	4384
20	474	2867	3582	4459

Samogłoska [i]: $k_n = 18\%$; $\Delta k_1 = 8$ Hz, $\Delta k_2 = 48,6$ Hz, $\Delta k_3 = 60,7$ Hz, $\Delta k_4 = 75,6$ Hz

Tablica 3.4

Zestaw formant.	F1 [Hz]	F2 [Hz]	F3 [Hz]	F4 [Hz]
10	761	2111	3188	3649
11	778	2157	3258	3730
12	794	2204	3328	3810
13	811	2250	3398	3890
14	828	2297	3469	3971
15	845	2343	3539	4051
16	861	2390	3609	4131
17	878	2436	3679	4211

Samogłoska [e]: $\Delta k_n = 18\%$; $\Delta k_1 = 16,8$ Hz, $\Delta k_2 = 46,5$ Hz, $\Delta k_3 = 70,1$ Hz, $\Delta k_4 = 80,3$ Hz

Tablica 3.5

Zestaw formant.	F1 [Hz]	F2 [Hz]	F3 [Hz]	F4 [Hz]
9	947	1443	3266	4487
10	965	1472	3330	4575
11	984	1500	3394	4663
12	1003	1528	3458	4751
13	1021	1557	3522	4839
14	1040	1585	3586	4927
15	1058	1613	3650	5015
16	1077	1641	3714	5103
17	1095	1670	3778	5191

Samogłoska [a]: $\Delta k_n = 18 \%$; $\Delta k_1 = 18,6$ Hz, $\Delta k_2 = 28,3$ Hz,
 $\Delta k_3 = 64,0$ Hz, $\Delta k_4 = 88,0$ Hz

Tablica 3.6

Zestaw formant.	F1 [Hz]	F2 [Hz]	F3 [Hz]	F4 [Hz]
8	729	1107	3211	4689
9	745	1132	3281	4796
10	761	1156	3352	4895
11	777	1181	3422	4998
12	793	1205	3493	5102
13	809	1229	3564	5205
14	825	1254	3634	5308
15	841	1278	3705	5411

Samogłoska [o]: $\Delta k_n = 18 \%$; $\Delta k_1 = 16,0$ Hz, $\Delta k_2 = 24,4$ Hz,
 $\Delta k_3 = 70,3$ Hz, $\Delta k_4 = 103,1$ Hz

Tablica 3.7

Zestaw formant.	F1 [Hz]	F2 [Hz]	F3 [Hz]	F4 [Hz]
8	384	819	2747	3636
9	393	839	2816	3727
10	403	860	2885	3818
11	412	880	2954	3909
12	422	901	3023	4001
13	432	922	3091	4092
14	441	942	3160	4183

Samogłoska [u]: $\Delta k_n = 18\%$; $\Delta k_1 = 9,8$ Hz, $\Delta k_2 = 20,5$ Hz,
 $\Delta k_3 = 68,9$ Hz, $\Delta k_4 = 91,1$ Hz

Podobnie jak na poprzednim etapie syntezy, amplitudy i szerokości wstęg formantów nie były zmieniane. Wszystkie F5 oraz te F4, których częstotliwość przekraczała 5 kHz pominięto.

3.2 Częstotliwość podstawowa.

Jak stwierdza Fant ([13], str. 242), częstotliwość podstawowa głosu dziecka w wieku mniej więcej 10 lat wynosi średnio ok. 300 Hz, przy czym zróżnicowania międzysobnicze są duże. Jest to zasadniczo zgodne ze spostrzeżeniami Hasek i Singha [18], choć w pracy tej dwójki autorów przeciętne wartości F0 u 5..10-letnich chłopców i dziewcząt są nieco niższe.

Biorąc pod uwagę przytoczone obserwacje, zsyntetyzowano w sposób opisany w punkcie 2.2 5 przebiegów F0, których czynnik skalujący wynosił od 1,8 do 2,2 (por. Tablica 2.8). Przy braku danych na temat charakterystyki tonu krtaniowego dzieci przyjęto dla wszystkich 5 przebiegów wartość "t1" = 21, tj. taką samą, jak w "docelowym" przebiegu żeńskim (Tablica 2.8), oraz wartość "no" = 26, co odpowiadało procentowemu udziałowi fazy otwartej w całkowitej długości okresu w granicach 61 % do 72 %.

Odpowiednie wartości przedstawia Tablica 3.8.

Tablica 3.8

czyn. skal.	F0 [Hz]					średn. F0 [Hz]	no	% dług. okr.	tl
	0ms	100ms	245ms	390ms	435ms				
1,8	250,2	259,2	214,2	201,6	215,1	230	26	60%	21
1,9	264,1	273,6	226,1	212,8	227,1	243	26	63%	21
2,0	278,0	288,0	238,0	224,0	239,0	256	26	67%	21
2,1	291,9	302,4	249,9	235,2	251,0	269	26	70%	21
2,2	305,8	316,8	261,8	246,4	262,9	282	26	72%	21

Każdy z 5 przebiegów F0 z Tablicy 3.8 oraz, dodatkowo, przebieg o czynniku skalującym 1.7 (zob. Tablica 2.8) połączono na drodze syntezy z każdym z zestawów formantowych występujących w Tablicach 3.2 + 3.7. W efekcie otrzymano 306 bodźców, w tym: dla [i] - 60, dla [ɛ] - 54, dla [e] - 48, dla [a] - 54, dla [o] - 48 i dla [u] - 42 bodźce.

4. Badania odsłuchowe.

4.1. Materiał testowy. Sposób prezentacji.

Łączna liczba bodźców wygenerowanych na pierwszym i drugim etapie syntezy wynosiła 777 (471 + 306), Składało się na nią:

- 150 bodźców typu [i]
- 144 bodźce typu [ɛ]
- 141 bodźców typu [e]
- 136 bodźców typu [a]
- 106 bodźców typu [o]
- i 100 bodźców typu [u].

Cały ten materiał poddano randomizacji przy użyciu specjalnej procedury wchodzącej w skład wykorzystywanego do syntezy pakietu programów autorstwa Klatta [22], a następnie nagrano na magnetofonie kasetowym Revox B710. Odstęp między kolejnymi bodźcami wynosił na taśmie testowej 3 s. Przy jego doborze oparto się na rezultatach pracy Crowdera [7], według których

taki właśnie jest czas wygaszania pamięci audytywnej w teście dyskryminacji samogłosek i czas zaniku wpływu kontekstu na wyniki rozpoznawania samogłosek.

Odstęp między grupami 10 bodźców wynosił 4,5 sek.

W badaniach odsłuchowych wzięło udział 20 osób, których zadaniem było rozpoznanie każdego bodźca jako określonej samogłoski polskiej oraz zaklasyfikowanie go jako wypowiedzi mężczyzny, kobiety lub dziecka. Odbywało się to poprzez wpisanie odpowiednich oznaczeń dwuliterowych (np. "YM", "AK", "UD" itp.) w dostarczonej słuchaczom ankiecie.

Każda z osób odsłuchiwała materiał w trakcie dwóch sesji odległych od siebie o kilka dni. Na początku obu sesji instruowano słuchaczy co do ich zadania i prezentowano im (nie podlegający ocenie) zestaw osiemnastu "docelowych" samogłosek męskich, kobiecych i dziecięcych.

4.2. Wyniki.

4.2.1. Identyfikacja samogłosek.

Na ogólną liczbę 15 540 odpowiedzi odnotowano łącznie 557 błędów (3,6 %), przy czym zdecydowanie najwięcej wystąpiło ich w przypadku bodźców typu [o]. Tablica 4.1 prezentuje odnośne dane liczbowe.

Tablica 4.1

Samogł. nadana	liczba bodźców	liczba błędów	% błędów	rozpoznania błędne *
u	2000	2	0,1	-(2)
e	2820	4	0,1	-(4)
i	3000	18	0,6	± (14), u(1), -(3)
a	2720	28	1,0	o(24), -(4)
±	2880	190	6,6	i(145), e(38), -(7)
o	2120	315	14,9	a(312), -(3)

* "-" w tej kolumnie oznacza brak odpowiedzi w ankiecie

Procentowy udział błędów w identyfikacji samogłosek [u],

[e], [i] oraz [a] można uznać za mieszczący się w granicach normy. Przemawia za tym fakt, iż wśród odpowiedzi odnoszących się do wymienionych samogłosek błędy rozłożyły się w zasadzie w sposób losowy. Znaczna liczba błędnych identyfikacji uzyskanych dla [ɛ] i, zwłaszcza, dla [o] wskazywałaby natomiast na pewne nieprawidłowości w strukturze (niektórych) bodźców reprezentujących te samogłoski. Prawdopodobne przyczyny tego faktu omówione zostaną w punkcie 5.1.

4.2.2. Rozpoznawanie kategorii głosu.

Obliczając sumę wystąpień kwalifikatorów "M" (mężczyzna), "K" (kobieta) i "D" (dziecko) dla poszczególnych bodźców, nie wzięto pod uwagę tych przypadków, w których identyfikacja samogłoski nie była poprawna. W odniesieniu do niektórych bodźców - głównie typu [o] - spowodowało to znaczne ograniczenie liczby "ważnych" rozpoznań kategorii głosu.

W Tablicach 4.2 + 4.7 podane są liczby wystąpień odpowiedzi typu "M", "K" i "D" dla kombinacji poszczególnych zestawów formantowych samogłosek i przebiegów F0. W każdej komórce tych Tablic uwidoczniona jest liczba odpowiedzi reprezentujących dominującą liczebnie kategorię głosu oraz, ewentualnie, liczba błędnych identyfikacji samogłoski ("B") w tych przypadkach, gdy przewyższa ona 2.

Tablica 4.2

Zest. form.	Srednia F0 [Hz]							
	128	141	154	166	179	192	205	218
1	20M	18M	18M	19M	16M	11M	10K	11K
2	19M	19M	17M	16M	12M	15K	14K	18K
3	20M	18M	16M	15M	15M	13K	15K	18K
4	19M	18M	17M	18M	10M	14K	16K	19K
5	20M	18M	16M	15M	14M	16K	12K	17K
6	19M	15M	16M	18M	14M	15K	18K	19K
7	18M	17M	16M	12M	10M	10M	19K	18K
8	20M	19M	16M	13M	11M	17K	19K	20K
9	19M	15M	16M	12M	11M	12K	16K	20K
10	19M	16M	16M	13M	16M	12K	18K	19K
11	16M	16M	19M	13M	12K	16K	18K	18K
Zest. form.	Srednia F0 [Hz]							
	218	230	243	256	269	282		
12	18K	20K	14K	18K	11K	12D		
13	18K	16K	20K	11D	14K	11D		
14	17K	16K	14K	11D	13K	11D		
15	18K	15K	14K	11K	15D	18D		
16	17K	15K	11D	16D	16D	19D		
17	15K	12K	11K	19D	14D	18D		
18	15K	14K	12K	12D	20D	20D		
19	11D	13D	14D	14D	18D	18D		
20	14K	10D	13D	15D	19D	18D		
21	11K	12K	12D	19D	16D	18D		

Tablica 4.3

Zest. form.	Średnia F0 [Hz]							
	128	141	154	166	179	192	205	218
1	20M	19M	20M	17M	17M	11M	11K	10K
2	20M	19M	16M	16M	16M	12M	11K	12K
3	19M	19M	18M	16M	15M	10M	12K	14K
4	18M	17M	19M	19M	15M	10K	12K	16K
5	19M	19M	18M	16M	15M	11M	13K	13K
6	19M	18M	17M	14M	12M	11K	15K	12K, 3B
7	17M	19M	16M	18M	13M	12M	12K	15K
8	19M	18M	16M	13M	12M	10K	16K	13K
9	18M	19M	14M, 4B	12M	10M	14K	13K	14K, 3B
10	17M	16M	16M	12M	10M	14K	10M	15K
11	18M	16M	15M, 4B	16M	14K	15K	16K, 3B	19K
Zest. form.	Średnia F0 [Hz]							
	218	230	243	256	269	282		
12	18K	17K	15K	8K, 5B	9K, 4B	8K, 5B		
13	18K	17K	12K, 3B	12K	13D, 3B	10D, 3B		
14	19K	15K	14K	10D	11D	10D, 3B		
15	16K	15K	10K	16D	15D	11D		
16	19K	15K	11D	9D	13D, 3B	15D, 4B		
17	15K	10D	10D	15D	9K	13D, 5B		
18	11K, 3B	13D	12D	17D	13D, 5B	16D, 4B		
19	13K	11K, 3B	12D	16D	15D, 4B	17D, 3B		
20	10K, 3B	10K	14D	13D	13D, 3B	14D, 5B		

Tablica 4.4

Zest. form.	Średnia F0 [Hz]							
	128	141	154	166	179	192	205	218
1	20M	20M	19M	16M	14M	14M	12K	11M
2	20M	19M	19M	18M	12M	11K	12M	13K
3	20M	19M	20M	18M	16M	12M	13M	13K
4	20M	19M	18M	17M	13M	11K	14K	20K
5	20M	19M	19M	17M	13M	12K	15K	16K
6	19M	18M	19M	16M	11K	13K	17K	17K
7	20M	18M	17M	18M	12M	15K	14K	16K
8	19M	18M	18M	14M	11M	17K	15K	16K
9	20M	20M	18M	14M	12K	14K	16K	19K
Zest. form.	Średnia F0 [Hz]							
	218	230	243	256	269	282		
10	13K	13K	10K	12K	14D	16D		
11	16K	13K	12D	16D	13D	17D		
12	13K	11K	11K	12K	16D	19D		
13	16K	11D	17D	19D	18D	17D		
14	12K	14D	16D	14D	19D	20D		
15	10K	11K	17D	20D	19D	20D		
16	13D	11D	20D	17D	20D	20D		
17	12D	16D	19D	20D	19D	20D		

Tablica 4.5

Zest. form.	Średnia F0 [Hz]							
	128	141	154	166	179	192	205	218
1	20M	20M	18M	16M	13M	11M	15K	13K, 3B
2	19M	19M	19M	14M	12M	10K	17K	17K
3	19M	19M	15M	12M	11K	13K	18K	13K
4	20M	18M	19M	11M	12K	11M	15K	16K
5	20M	18M	18M	16M	14K	15K	18K	17K
6	19M	19M	18M	11M	11K	12K	14K	18K
7	19M	17M	14M	12M	10K	12K	18K	17K
8	19M	18M	17M	10M	10K	18K	18K	19K
Zest. form.	Średnia F0 [Hz]							
	218	230	243	256	269	282		
9	17K	11D	12K	16D	19D	17D		
10	18K	12K	12D	15D	14D	20D		
11	14K	14D	11D	12D	17D	19D		
12	18K	14K	12D	16D	20D	18D		
13	12K	13D	16D	18D	19D	19D		
14	10K	17D	11D	19D	18D	17D		
15	20D	16D	13D	16D	19D	18D		
16	16D	13D	17D	18D	19D	20D		
17	15D	17D	19D	19D	20D	20D		

Samogłoska [a]

Tablica 4.6

Zest. form.	Srednia F0 [Hz]							
	128	141	154	166	179	192	205	218
1	17M	17M	19M	15M	11K	17K	17K	17K
2	18M	14M	15M	11M	11K	15K	16K	19K
3	19M	17M	16M	9K, 3B	10K	16K	18K	16K
4	15M, 3B	13M, 3B	11M, 3B	10M, 3B	16K	15K	17K	17K
5	12M, 6B	12M, 4B	10M, 3B	9K, 3B	13K	16K	16K	17K
6	5M, 15B	14M, 4B	16M	9M	9K, 5B	9K, 6B	12K	15K
7	5M, 15B	15M, 4B	8M, 11B	7M, 9B	9K, 5B	12K, 4B	14K, 3B	16K
Zest. form.	Srednia F0 [Hz]							
	218	230	243	256	269	282		
8	14K	14K	12K	9D, 3B	11D	18D		
9	15K	12K	10D	12D	18D	17D		
10	8K, 6B	10D	8K, 8B	10D	16D, 3B	19D		
11	12K, 3B	9K, 3B	15D	17D	17D	20D		
12	6D, 9B	10D, 5B	11D, 6B	17D	16D	18D		
13	6D, 9B	12D, 3B	9D, 5B	15D, 5B	14D, 4B	16D, 4B		
14	3D, 16B	11D, 3B	11D, 9B	17D	14D, 5B	18D		
15	5D, 14B	8D, 6B	11D, 7B	14D, 5B	17D, 3B	15D, 3B		

Samogłoska [o]

Tablica 4.7

Zest. form.	Srednia F0 [Hz]							
	128	141	154	166	179	192	205	218
1	20M	19M	20M	18M	13M	15M	12M	12K
2	19M	20M	17M	19M	15M	13M	13K	14K
3	20M	18M	19M	18M	16M	12M	11K	12K
4	19M	17M	19M	16M	10K	12K	14K	10K
5	20M	18M	16M	18M	13M	11K	15K	18K
6	20M	18M	20M	18M	11M	11M	16K	19K
7	18M	19M	18M	18M	11M	17K	14K	18K
Zest. form.	Srednia F0 [Hz]							
	218	230	243	256	269	282		
8	16K	17K	19K	16K	15K	11K		
9	17K	19K	17K	11K	13D	10D		
10	20K	18K	17K	19K	11K	11D		
11	19K	15K	16K	15K	15D	11D		
12	16K	16K	14K	10K	12D	13D		
13	18K	14K	13K	13D	15D	15D		
14	19K	15K	11D	12K	13D	14D		

Samogłoska [u]

5.1. Analiza przyczyn błędnych identyfikacji samogłosek.

Przyczyny, dla których zdecydowaną większość wśród błędnych rozpoznań samogłosek [o] i [i] stanowiły odpowiednio samogłoski [a] oraz [i] są oczywiste i wiążą się z artykulacyjno-akustyczno-percepcyjnym podobieństwem tych par samogłosek (zob. np. [26]). Zastanawiający jest natomiast brak symetrii w błędnych reakcjach: o ile np. [o] zostało rozpoznane jako [a] aż 312 razy, substytucja odwrotna wystąpiła zaledwie w 24 przypadkach; analogicznie, [i] odebrano jako [i] 145 razy.

natomiast [i] jako [ɨ] - tylko 14 razy. Fakty te wskazywałyby na to, że wartości czynników skalujących przyjęte dla samogłosek [o] i [ɨ] nie były właściwe, a ściślej - były zbyt wysokie, w wyniku czego nastąpiło nadmierne podwyższenie F2 i F3 w kobiecych i dziecięcych realizacjach tych samogłosek, dające w efekcie ich akustyczne (i percepcyjne) zbliżenie do samogłosek [a] i [i]. Podczas gdy wyjaśnienia takiego nie można - zwłaszcza w przypadku [o] - kategorycznie odrzucić, należy zwrócić także uwagę na kilka dodatkowych faktów mogących mieć w tym kontekście istotne znaczenie.

W myśl teorii Stevensa [33], samogłoski typu [i], [a] i [u], zwane przez niego samogłoskami "kwantalnymi" (ang. *quantal vowels*), mają w języku specyficzny status, który wyraża się m. in. tym, iż nie są one identyfikowane jako punkty na continuum, lecz tworzą kategorie percepcyjne dyskretne. Konsekwencją tego jest m. in. mniejsze uwrażliwienie (większa tolerancja) percepcyjnych mechanizmów klasyfikujących na występujące w konkretnych realizacjach tych samogłosek odchylenia (typu akustycznego) od ich hipotetycznych prototypów, co w praktyce oznaczałoby ich łatwiejszą identyfikację.

Zarówno dla [o], jak i dla [ɨ] zanotowano pewną, co prawda niewielką, liczbę błędnych odpowiedzi typu (odpowiednio) [a] oraz [i] także w przypadku tych bodźców, w których zestawy formantowe były "zdecydowanie męskie", tj. albo odpowiadały ściśle wartościom formantów wyekstrahowanych z wypowiedzi naturalnych samogłosek [o] oraz [ɨ], albo też były do tych wartości bardzo zbliżone. Błędy tego rodzaju, jak i - z czym należy się liczyć - pewna część ogólnej liczby błędów miały charakter przypadkowy.

Analiza odpowiedzi udzielonych przez poszczególnych słuchaczy wykazuje, że fonetyczne prototypy (zob. np. [31], [32]) samogłosek nie mają charakteru absolutnego. U niektórych słuchaczy zaobserwowano wyraźne tendencje do preferowania określonego typu odpowiedzi (*response bias*). I tak, na ogólną liczbę 145 substytucji typu [ɨ] → [i] aż 116 (tj. ponad 80 %) wystąpiło w odpowiedziach zaledwie czterech osób; analogiczne, z 312 odnotowanych przypadków rozpoznania [o] jako [a] aż 127 (tj. prawie 42 %) "było dziełem" zaledwie dwóch (innych) osób.

Wypowiedzi, które większość uznała za reprezentujące fonemy /ɪ/ i /o/ plasowały się więc w percepcyjnej przestrzeni samogłoskowej niektórych słuchaczy poza granicami tych kategorii fonetycznych - bliżej hipotetycznych prototypów /i/ oraz /a/. Gdyby, przykładowo, pominąć błędne oceny osób, które preferowały odpowiedź typu [i], to wówczas liczba niepoprawnych identyfikacji żadnego z bodźców reprezentujących fonem /ɪ/ nie przekraczałaby 2.

Obok 145 przypadków zidentyfikowania [ɪ] jako [i] odnotowano także 38 odpowiedzi klasyfikujących tę samogłoskę jako /e/. Fakt ten przeczyłby hipotezie o zbyt wysokich wartościach czynników skalujących dla [ɪ]. Jeśli bowiem rzeczywiście F2 i F3 w (niektórych) bodźcach reprezentujących /ɪ/ miałyby wartości zawyżone, reakcje typu [e] nie powinny były wystąpić, zwłaszcza w związku z dość znaczną odległością percepcyjną tych samogłosek (por. np. [26]).

Analiza przypadków identyfikacji [ɪ] jako [e] wykazuje, że w przeważającej mierze miały one miejsce, gdy częstotliwość podstawowa wypowiedzi nie była "dopasowana" do jej częstotliwości formantowych, a ściślej: gdy była w zestawieniu z nimi za niska. Typowymi przykładami są tu bodźce stanowiące kombinacje "kobięcych" formantów z męską F0 lub "dziecięcych" formantów z kobiecą F0.

Tego samego rodzaju niedopasowanie F0 i formantów było także - przynajmniej w części - przyczyną znacznej liczby błędnych identyfikacji niektórych bodźców typu [o]: świadczy o tym fakt, iż najczęściej identyfikowane jako [a] były te wypowiedzi, które charakteryzował największy rozróżnienie między częstotliwościami formantowymi i F0. Przykładowo: bodźce stanowiące kombinacje "kobięcych" formantów i męskiej F0 oraz "dziecięcych" formantów i kobiecej F0 zostały zaklasyfikowane jako /a/ przez aż kilkunastu (14 + 16) słuchaczy.

Nie ulega wątpliwości, że częstotliwość podstawowa oraz odległość F1-F0 odgrywają pewną rolę w percepcji otwartości samogłosek (np. [3], [9], [27], [35], [36], [37]). Trau Müller [36] wykazał na przykład, że postrzegana otwartość samogłoski zmienia się w zależności od F0 nawet jeżeli częstotliwości formantowe pozostają te same. Carlson i in. [3] stwierdzają, że aby dana samogłoska zachowała tożsamość fonetyczną, wzrostowi

F0 musi towarzyszyć odpowiednie podwyższenie formantów.

Należy w tym miejscu zwrócić uwagę, że charakteryzujący część bodźców brak zgodności między F0 i częstotliwościami formantów wynika już z samej koncepcji niniejszego eksperymentu. Celem jego nie było bowiem wygenerowanie wzorcowych (a więc łatwo identyfikowalnych) realizacji poszczególnych fonemów lecz bodźców, w których różne "konfiguracje" formantów łączyłyby się z różnorodnymi przebiegami F0. Przypadki "niedopasowania" omawianych parametrów stanowią tego nieuchronną konsekwencję.

Ze względu na zapewne dość znaczną odległość od odpowiednich prototypów samogłoskowych, jak też na fakt, iż niektóre wypowiedzi były być może wręcz "nierealistyczne" (np. kobiece formanty w połączeniu z męską F0), identyfikacja części bodźców musiała nastąpić sporo trudności, zwłaszcza w przypadku samogłosek nie-kwantalnych i zbliżonych percepcyjnie do innych (a więc - [i], [o]). Dodatkowe utrudnienie stanowiła niewątpliwie różnorodność "głosów" (fizycznie, żadne dwa bodźce w całym zestawie nie były identyczne), występujących ponadto w kolejności losowej (zob. np. [21]). Ten rodzaj prezentacji zmuszał słuchaczy do traktowania każdego bodźca w sposób odrębny - w oderwaniu od poprzednich.

Z akustycznego punktu widzenia, realizacje tej samej samogłoski przez głos męski i dziecięcy znacznie się między sobą różnią ([20]). Mimo to słuchacze potrafią oczywiście rozpoznać obie wypowiedzi jako reprezentujące jeden fonem [38]. Ustalenie tej percepcyjnej ekwiwalentności dokonuje się na drodze normalizacji głosu, której mechanizm nie został jeszcze w pełni poznany. Dowodem występowania tego procesu w percepcji jest fakt, iż czas reakcji przy identyfikacji samogłosek pochodzących od kilku różnych głosów jest dłuższy niż w przypadku jednego tylko głosu ([34]). Oznacza to, że konieczność dokonywania normalizacji (tak częsta w niniejszym eksperymencie) utrudnia w pewnym stopniu identyfikację wypowiedzi i może spowodować spadek jej efektywności.

Zapewne ze względu na losową kolejność prezentacji bodźców nie zaznaczył się w odpowiedziach tzw. efekt kontrastu, polegający na preferowaniu reakcji reprezentujących inna kategorię fonetyczną niż ta, do której został zaklasyfikowany

bodziec poprzedni (zob. np. [4], [30], [31])¹⁾.

Interesujący przyczynek do kwestii braku zgodności między częstotliwościami formantowymi a F0 stanowią rezultaty prac [8], [24] i [38]. Jak się okazuje, sprawne "wmontowanie" w nagranie kilkuwyrazowej wypowiedzi dziecka jednego wyrazu wymówionego przez mężczyznę imitującego wysoką F0 głosu dziecięcego w bardzo znaczny sposób obniża poprawność percepcyjnej identyfikacji samogłoski zawartej w tym wyrazie. Dowodzi to w pierwszym rzędzie istotnej roli szerszego kontekstu w procesie rozpoznawania głosek: kontekst poprzedzający dostarcza "ram odniesieniowych", które w warunkach naturalnej komunikacji słownej ułatwiają (optymalizują) identyfikację kolejnych samogłosek ([38]). W przypadku "mystyfikacji" jaką posłużono się w cytowanych doświadczeniach, brak odpowiedniości między strukturą akustyczną wypowiedzi męskiej a percepcyjnymi ramami odniesieniowymi musiał spowodować znaczną liczbę błędów (w pracy [8] - ok. 54 %). Należy zwrócić uwagę, że podczas gdy F0 w wypowiedziach mężczyzny była bardzo bliska typowej F0 dziecka (stąd uznanie wypowiedzi za dziecięcą), różnice między częstotliwościami formantowymi samogłoski występującej w tej wypowiedzi i tej samej samogłoski wymówionej przez dziecko były zasadnicze. Przykładowo, w pracy [38] męskie [ɛ] upodobniło się do dziecięcego [æ], natomiast męskie [æ] - do dziecięcego [ɪ]. Można założyć, że nawet gdyby bodźce tego typu zostały zaprezentowane w izolacji, ów brak zgodności między formantami i F0 spowodowałby niepoprawne rozpoznanie (zamierzonych) samogłosek. Zjawisko takie wystąpiło w niniejszej pracy w przypadku niektórych realizacji samogłoski /o/, a jego przyczyną był fakt, iż kobiece częstotliwości formantowe [o] (zwłaszcza F1 i F2) w niewielkim tylko stopniu różniły się od męskich częstotliwości formantowych [a], które były niemal identyczne z częstotliwościami dziecięcego [o] (zob. Tablice 2.5, 2.6, 3.5, 3.6). Zbyt niska - w stosunku do "właściwej" -

1) Według cytowanego już Crowdera [7], efekt taki nie powinien w zasadzie wystąpić w identyfikacji samogłosek przy zachowaniu 3-milisekundowego odstępu między kolejno prezentowanymi bodźcami.

częstotliwość podstawowa powodowała w tej sytuacji zmianę kategorii fonetycznej bodźca.

Niemal powszechnie uważa się, że percepcyjną tożsamość samogłoski określa położenie jej formantów, zwłaszcza F1 i F2. Równie powszechnie zwraca się także uwagę, że percepcja samogłosek ma charakter stosunkowo (w porównaniu z percepcją spółgłosek) nie-kategorialny, czego dowodem jest fakt, iż dyskryminacja samogłosek jest bardziej efektywna od ich identyfikacji (np. [6], [30], [25]). Modyfikowanie położenia formantów w pewnym zakresie przy stałej częstotliwości podstawowej prowadzi więc do zmiany brzmienia samogłoski, ale niekoniecznie - do zmiany jej kategorii fonetycznej. Analogiczne zjawisko występuje również przy manipulowaniu F0 w wypowiedziach o stałej konfiguracji formantów ([36]). Zmiana kategorii fonetycznej zachodzi wówczas, gdy zakres modyfikacji formantów lub F0 przekroczy pewną wartość krytyczną (jak należy przypuszczać, przesunięcia formantów odgrywają tu bardziej decydującą rolę)¹⁾.

Z powyższych rozważań wynika, że w percepcyjnej identyfikacji samogłosek istotne znaczenie ma nie tylko samo położenie formantów w skali częstotliwości, ale także ich relacja do F0. Słuszności tego wniosku wydają się przeczyć rezultaty uzyskane przez Summerfielda i Haggarda [34], którzy stwierdzają w konkluzji, że o ile znaczenie F0 w percepcyjnym rozpoznawaniu głosek nie ulega wątpliwości, jej wykorzystanie jako czynnika normalizującego w identyfikacji samogłosek jest jedynie percepcyjną opcją. Przyczynę tej konkluzji stanowi jednakże zapewne fakt, iż różnica w zakresie F0 między głosem I i pozostałymi wynosiła w tej pracy zaledwie 20 Hz, była więc najprawdopodobniej zbyt mała, by mogło się ujawnić normalizacyjne działanie tego parametru (wszystkie cztery głosy reprezentowały te same klasę głosu, a mianowicie głos męski).

Interpretując uzyskane przez siebie wyniki, cytowani już van Bergem i in. [38] wysuwają hipotezę o istnieniu w pamięci

1) Pewne znaczenie ma tu z pewnością liczba fonemów samogłoskowych w danym języku: "ciasne" wypełnienie percepcyjnej przestrzeni samogłoskowej i łączące się z tym znaczne podobieństwo sąsiadujących w niej ze sobą samogłosek sprzyja zapewne występowaniu tego zjawiska.

słuchaczy usrednionych wzorców (ang. *templates*) samogłosek męskich, kobiecych i dziecięcych, wytworzonych w oparciu o przeszłe doświadczenie językowe. Błędy, które wystąpiły w ich eksperymencie autorzy ci tłumaczą posłużeniem się przez słuchaczy w identyfikacji samogłosek wzorcami reprezentującymi niewłaściwy rodzaj głosu.

Bez względu na to, czy powyższą hipotezę uznamy za słuszną, czy nie, uzasadnione wydaje się przypuszczenie, że FO jest istotnym nośnikiem informacji na temat klasy głosu, a tym samym spełnia niewątpliwie ważną funkcję normalizacyjną, umożliwiając w ten sposób bardziej efektywną komunikację. Słuszność tego przypuszczenia potwierdzają omówione niżej rezultaty percepcyjnej kategoryzacji głosów.

5.2. Kategoryzacja głosów.

W celu zapewnienia maksymalnie obiektywnych warunków eksperymentu i uniknięcia możliwości jakiegokolwiek wpływu na kryteria decyzyjne słuchaczy nie informowano ich o celu pracy, a zwłaszcza nie uprzedzano ich, że w prezentowanym zestawie mogą wystąpić np. wysokie głosy męskie czy niskie głosy kobiece. Podejmowane przez słuchaczy decyzje miały więc we wszystkich przypadkach charakter w pełni niezależny.

Już nawet dość pobieżny, przegląd uzyskanych odpowiedzi wykazuje, że - podobnie jak w przypadku identyfikacji samogłosek - u części słuchaczy dała się zaobserwować tendencja do preferowania określonego typu reakcji, świadcząca o względności "prototypów" kategorii głosu. Dotyczyło to zwłaszcza trzech osób, z których dwie znacznie częściej od pozostałych wybierały odpowiedź "kobieta" jako reakcję na bodziec uznany przez większość za wypowiedź męską, trzecia zaś wybierała tę alternatywę (tj. "K") w przypadku wypowiedzi zaklasyfikowanych w przeważającej mierze jako dziecięce. Można domniemywać, iż pewien wpływ na kształtowanie się prototypów kategorii głosu, ma rodzaj głosów dominujących w najbliższym otoczeniu danej osoby.

Jak wynika z analizy odpowiedzi przedstawionych w Tablicach 4.2 + 4.7, dokonując kategoryzacji głosów słuchacze pozostawali w jakimś stopniu pod wpływem kontekstu poprzedzającego, który okazał się być nieistotny przy

identyfikacji samogłosek. Świadczy o tym fakt, iż niektóre z bodźców w sposób wyraźny różnią się pod względem przypisanej im liczby i kategorii kwalifikatorów głosu od swojego bezpośredniego otoczenia, charakteryzującego się zbliżonymi wartościami F0 i częstotliwości formantowych. W odpowiedziach uwidocznił się zarówno efekt kontrastu, polegający na preferowaniu reakcji odmiennej od reakcji na bodziec poprzedni, jak też efekt atrakcji (przyciągania). Jeżeli, przykładowo, różnica między dwoma kolejnymi bodźcami była dość znaczna, a pierwszy z nich został przez słuchaczy uznany za wypowiedź zdecydowanie męską, zaznaczała się większa niż przeciętnie zgodność słuchaczy w zaklasyfikowaniu drugiego bodźca jako wypowiedzi kobiety (efekt kontrastu). Jeżeli natomiast bodziec poprzedzający został, przykładowo, uznany za wypowiedź męską, a różnica między nim i bodźcem następnym nie wydała się słuchaczom zasadnicza, zaznaczała się tendencja do klasyfikowania drugiego z bodźców jako "również wypowiedzi męskiej (efekt atrakcji).

5.2.1. Głos męski a głos kobiecy.

Zasadniczym kryterium, jakim kierowali się słuchacze przy kwalifikacji głosu jako męskiego bądź kobiecego była niewątpliwie wysokość F0. Świadczy o tym wyraźnie fakt, iż (1) wysokie położenie formantów samo w sobie nie stanowiło gwarancji wystąpienia reakcji typu "kobieta" oraz (2) nawet wypowiedzi o stosunkowo niskich (a więc potencjalnie męskich) formantach lecz o wysokiej F0 uznawane były za kobiece. Jak wynika z Tablic 4.2 + 4.7, odpowiedzi typu "K" pojawiają się przy średniej F0 wypowiedzi równej 179 Hz, a przy F0 wynoszącej 192 Hz zaczynają dominować liczebnie. W tym zakresie częstotliwości przebiega zapewne percepcyjna granica między męską i kobiecą F0.

Należy zauważyć, że nawet w przypadku tych bodźców, które charakteryzował istotny brak zgodności między formantami i F0, i które nie zostały w związku z tym poprawnie zidentyfikowane przez pewną liczbę słuchaczy, kwalifikator głosu w rozpoznaniach błędnych był w przeważającej mierze taki sam, jak w większości rozpoznań poprawnych. Fakt ten potwierdza zarówno dominującą rolę częstotliwości podstawowej w określaniu

kategorii głosu, jak i jej wpływ na identyfikację samogłosek.

Porównanie par bodźców, które różniły się jedynie wartością "t1" ("t1" = 0 lub "t1" = 21) wykazuje, iż wypowiedzi o bardziej stromym spadku obwiedni widma tonu pobudzania (tj. o "t1" = 21) były uznawane za kobiece zdecydowanie częściej niż te, w których "t1" równe było 0.

5.2.2. Głos kobiecy a głos dziecięcy.

Percepcyjne granice między głosem kobiecym i dziecięcym wydają się mniej wyraziste niż między głosem kobiecym i męskim. Może to m. in. wynikać z pewnej niedookreśloności kategorii głosu dziecięcego czy pojęcia "dziecko" w ogóle. Podczas gdy np. głos dziecka 5...6-letniego jest stosunkowo łatwo identyfikowalny, różnice w strukturze akustycznej wypowiedzi kobiety i 12...13-letniego chłopca mogą być niewielkie. Trudno jest oczywiście ocenić, jaką definicję dziecka przyjęli na użytek niniejszego doświadczenia poszczególni słuchacze, można jednak przypuszczać, że ich definicje nie były całkowicie zbieżne.

Owa niedookreśloność głosu dziecięcego spowodowała m. in. większy niż w przypadku granicy mężczyzna-kobieta efekt kontrastu i atrakcji (por. "Tablice 4.2 + 4.7). Wydaje się także, iż przy wyborze przez słuchaczy odpowiedzi typu "dziecko"/"kobieta" pewną (nieco większą) rolę odgrywały częstotliwości formantowe wypowiedzi, choć - bez względu na położenie formantów - wysoka F0 przesądzała o uznaniu wypowiedzi za dziecięcą. Dla połączonych danych samogłoskowych percepcyjna granica między głosem kobiecym i dziecięcym przebiega w zakresie wartości F0 od 230 Hz (przewaga odpowiedzi typu "K") do 243 Hz (niewielka przewaga odpowiedzi typu "D"). Ze schematu tego "wyłamuje się" w pewnym stopniu samogłoska /u/, w przypadku której granica "K"/"D" jest nieco przesunięta ku wyższym wartościom F0.

-
- 1) Z funkcjonalnego punktu widzenia, odróżnienie głosu kobiety od głosu starszego dziecka może nie być istotne ze względu na zbliżone rozmiary ponadkrtaniowego toru artykulacyjnego i wiążące się z tym niewielkie różnice w zakresie typowych częstotliwości formantowych.

Podsumowanie.

Wbrew rozpowszechnionej opinii, o percepcyjnej tożsamości fonetycznej samogłoski nie decydują wyłącznie jej częstotliwości formantowe: mogą one być zinterpretowane dopiero na podstawie informacji, której nośnikiem jest częstotliwość podstawowa. Aby wypowiedź samogłoskowa była wyrazista (łatwa do zidentyfikowania), charakteryzujące ją formanty muszą więc łączyć się z odpowiednią częstotliwością podstawową. Jeśli F0 jest w stosunku do tych formantów za niska, postrzegana samogłoska staje się bardziej otwarta; jeśli natomiast F0 jest zbyt wysoka, samogłoska postrzegana jest jako bardziej przymknięta.

Ustalenie kategorii głosu wydaje się niezbędnym warunkiem poprawnej identyfikacji treści wypowiedzi. Innymi słowy, aby móc zrozumieć co zostało powiedziane, trzeba wpiery wiedzieć jaka kategorie głosu reprezentuje autor wypowiedzi. Jak wykazały uzyskane rezultaty, zasadniczym czynnikiem decydującym o percepcyjnej kwalifikacji głosu jako męskiego, kobiecego lub dziecięcego jest częstotliwość podstawowa. Oznacza to, iż F0 w istotny sposób ukierunkowuje proces percepcyjnej normalizacji toru głosowego, a tym samym wpływa na efektywność komunikacji słownej.

Fakt istnienia zależności między F0 i postrzeganym stopniem otwarcia samogłoski, decydujący w konsekwencji o wyrazistości samogłoski, ma istotne znaczenie m. in. w syntezie mowy: brak zgodności między F0 i przyjętymi dla poszczególnych samogłosek częstotliwościami formantów może prowadzić do obniżenia naturalności, a może nawet i zrozumiałości generowanego sygnału.

Nie jest wykluczone, że F0 mogłaby być także wykorzystana jako czynnik normalizacji głosu w automatycznym rozpoznawaniu mowy: konieczność zgodności F0 i formantów implikuje bowiem możliwość prognozowania lokalizacji formantów poszczególnych głosek w skali częstotliwości na podstawie średniej (typowej) wartości F0 głosu operatora. Weryfikacja tej hipotezy wymagałaby jednak przeprowadzenia odrębnych badań.

BIBLIOGRAFIA

1. ALLEN, J., M.S. HUNNICUTT, D.H. KLATT, *From Text to Speech: The MITalk System*, 1987, Cambridge University Press, Cambridge.
2. *AUDLAB User's Guide*, Centre for Speech Technology Research, 1987, University of Edinburgh.
3. CARLSON, R., G. FANT, B. GRANSTRÖM, *Two-formant models, pitch, and vowel perception*, w: G. Fant and M.A.A. Tatham (eds.): *Auditory Analysis and Perception of Speech*, 1975, Academic Press, London, str. 55 - 82.
4. CENTMAYER, K., *Interrelations of vowel perception and linguistic context*, w: G. Fant and M.A.A. Tatham (eds.): *Auditory Analysis and Perception of Speech*, 1975, Academic Press, London, str. 143 - 152.
5. CHIBA, T., M. KAJIYAMA, *The Vowel - Its Nature and Structure*, 1941, Tokyo.
6. CHISTOVICH, L., *Central auditory processing of peripheral vowel spectra*, 1985, J. Acoust. Soc. Amer., vol. 77, str. 789 - 805.
7. CROWDER, R.G., *Decay of Auditory Memory in Vowel Discrimination*, 1982, J. Exptl. Psychol.: Learning, Memory and Cognition, vol. 8, No 2, str. 153 - 162.
8. DECHOVITZ, D., *Information conveyed by vowels: a confirmation*, 1977, Haskins Status Report, SR - 51/52, str. 213 - 219.
9. DI BENEDETTO, M.-G., *On vowel height: acoustic and perceptual representation by the fundamental and the first formant frequency*, 1987, Proceedings of the 11th International Congress of Phonetic Sciences, Tallin, vol. 5, str. 198 - 201.
10. EIMAS, P.D., J.L. MILLER, P.W. JUSCZYK, *On infant speech perception and the acquisition of language*, w: S. Harnad (ed.): *Categorical Perception: The Groundwork of Cognition*, 1987, Academic Press, Cambridge, str. 161 - 195.
11. ELMAN, J.L., J.L. MC CLELLAND, *Exploiting Lawful Variability in the Speech Wave*, w: J.S. Perkell, D.H. Klatt (eds.): *Invariance and Variability in Speech Processes*, 1986, Lawrence Erlbaum Assoc., Hillsdale, str. 360 - 380.
12. FANT, G., *On the predictability of formant levels and spectrum envelopes from formant frequencies*, w: M. Halle, H.G. Lunt, M. McLean (eds.): *For Roman Jakobson*, 1956, Mouton, The Hague, str. 109 - 120.
13. FANT, G., *Acoustic Theory of Speech Production*, 1970, Mouton, The Hague (2 wyd.).

14. FANT, G., *A Note on Vocal Tract Size Factors and Nonuniform F-Pattern Scalings*, w: G. Fant: *Speech Sounds and Features*, 1973, The MIT Press, Cambridge, Massachusetts, str. 84-93.
15. FANT, G., *Non-uniform Vowel Normalization*, 1975, *Speech Transmission Laboratory QPSR* 2-3, str. 1 - 19.
16. FLANAGAN, J. L., *Estimates of the Maximum Precision Necessary in Quantizing Certain "Dimensions" of Vowel Sounds*, 1957, *J. Acoust. Soc. Amer.*, vol. 29, str. 533 - 534.
17. FLANAGAN, J.L., *Speech Analysis, Synthesis and Perception*, 1965, Springer-Verlag, Berlin.
18. HASEK, C.S., S. SINGH, *Acoustic attributes of preadolescent voices*, 1980, *J. Acoust. Soc. Amer.*, vol. 68, str. 1262 - 1265.
19. HOLMBERG, E.B., R.E. HILLMAN, J.S. PERKELL, *Glottal airflow and pressure measurements for female and male speakers in soft, normal and loud voice*, referat wygłoszony na: 114 Meeting of Acoust. Soc. Amer., 1987, Florida.
20. HOLMES, J. N., *Normalisation in Vowel Perception*, w: J.S. Perkell, D.H. Klatt (eds.): *Invariance and Variability in Speech Processes*, 1986, Lawrence Erlbaum Assoc., Hillsdale, str. 346 - 357.
21. KAHN, D., *On the identifiability of isolated vowels*, 1978, *UCLA Working Papers in Phonetics*, No 41, str. 26 - 31.
22. KLATT, D. H., *Software for a Cascade/Parallel Formant Synthesizer*, 1980, *J. Acoust. Soc. Amer.*, vol. 67, str. 971 - 995.
23. KLATT, D.H., L.C. KLATT, *Voice quality variations within and across female and male talkers: implications for speech analysis, synthesis and perception*, referat wygłoszony na: 114 Meeting of Acoust. Soc. Amer., 1987, Florida.
24. LADEFOGED, P., P. BROADBENT, *Information Conveyed by Vowels*, 1957, *J. Acoust. Soc. Amer.*, vol. 29, str. 98-104.
25. LIBERMAN, A. M., F. S. COOPER, D.P. SHANKWEILER, M. STUDDERT-KENNEDY, *Perception of the Speech Code*, w: E.E. David, Jr., P.B. Denes (eds.): *Human Communication: A Unified View*, 1972, McGraw-Hill, New York, str. 13 - 50.
26. ŁOBACZ, P., G. DEMENKO, *Zależność percepcji segmentalnych cech samogłosek polskich od struktury długotrwałej pamięci leksykalno-fonematycznej*, *Prace IPPT* 40/1983, Warszawa.
27. MAJEWSKI, W., J. ZALEWSKI, *Rola częstotliwości podstawowej w procesie percepcji syntetycznych sygnałów dźwiękowych mowy*, *Prace Naukowe Instytutu Telekomunikacji i Akustyki Politechniki Wrocławskiej*, 1973, Nr 13, str. 37 - 50.

28. MONSEN, R.B., A.M. ENGBRETSON, *Study of variations in the male and female glottal wave*, 1977, J. Acoust. Soc. Amer., vol. 62, str. 981 - 993.
29. PERKELL, J. S., D.H. KLATT (eds.), *Invariance and Variability in Speech Processes*, 1986, Lawrence Erlbaum Assoc., Hillsdale.
30. REPP, B.H., A.F. HEALY, R.G. CROWDER, *Categories and Context in the Perception of Isolated Steady-State Vowels*, 1979, J. Exp. Psychol.: Human Perception and Performance, vol. 5 (1), str. 129 - 145.
31. REPP, B.H., A.M. LIBERMAN, *Phonetic category boundaries are flexible*, w: S. Harnad (ed.): *Categorical Perception: The Groundwork of Cognition*, 1987, Academic Press, Cambridge, str. 89 - 112.
32. SAMUEL, A.G., *Phonetic Prototypes*, 1982, Perception and Psychophysics, vol. 31 (4), str. 307 - 314.
33. STEVENS, K.N., *The Quantal Nature of Speech: Evidence from Articulatory-Acoustic Data*, w: E.E. David, Jr., P.B. Denes (eds.): *Human Communication: A Unified View*, 1972, McGraw-Hill, New York, str. 51 - 66.
34. SUMMERFIELD, A.Q., M.P. HAGGARD, *Vocal tract normalisation as demonstrated by reaction times*, w: G. Fant, M.A.A. Tatham (eds.): *Auditory Analysis and Perception of Speech*, 1975, Academic Press, London, str. 115 - 141.
35. SYRDAL, A. K., *Aspects of a model of the auditory representation of American English vowels*, 1985, Speech Communication, vol. 4, str. 121 - 135.
36. TRAUNMOLLER, H., *Perceptual dimension of openness in vowels*, 1981, J. Acoust. Soc. Amer., vol. 69, str. 1465 - 1475.
37. TRAUNMOLLER, H., F. LACERDA, *Perceptual relativity in identification of two-formant vowels*, 1987, Speech Communication, vol. 6, str. 143 - 157.
38. VAN BERGEM, D.R., L.C.W. POLS, F.J. KOOPMANS-VAN-BEINUM, *Perceptual Normalization of the Vowels of a Man and a Child in Various Contexts*, 1988, Speech Communication, vol. 7, str. 1 - 20.