Mariusz Owsianny

BADANIE WPŁYWU WYSOKOŚCI GŁOSU NA PERCEPCJĘ SYNTETYCZNYCH SAMOGŁOSEK POLSKICH

10/1994

WARSZAWA 1994

ISSN 0208-5658

Praca wpłynęła do Redakcji dnia 21 grudnia 1993 r.

Instytut Podstawowych Problemów Techniki PAN
Nakład 100 egz. Ark.wyd.1,0 Ark.druk. 1,50
Oddano do drukarni w marcu 1994 r.

Wydawnictwo Spółdzielcze sp. z o.o.
Warszawa, ul.Jasna 1

Mariusz Owsianny
Zakład Fonetyki Akustycznej IPPT PAN
Poznań

BADANIE WPŁYWU WYSOKOŚCI GŁOSU NA PERCEPCJĘ SYNTETYCZNYCH SAMOGŁOSEK POLSKICH

Streszczenie

Przeprowadzono badanie oddziaływania współzależności między wysokością głosu a częstotliwościami formantowymi na percepcję samogłosek syntetycznych.

Wykorzystując software'owy, formantowy syntezator mowy SMOK pracujący w układzie szeregowo-równoległym wg. systemu D.H. Klatta, umożliwiający generację sygnału mowy o dobrej naturalności brzmienia resyntetyzowano samogłoski polskie: /i/, /ɨ/, /e/, /a/, /o/, /u/ wypowiedziane przez kobiecych i męskich spikerów. Te prototypowe głoski poddano modyfikacji. Zmieniano częstotliwość podstawową *F0* w zakresie ±1 oktawa w stosunku do zarejestrowanego uprzednio przebiegu tego parametru.

W przeprowadzonych badaniach odsłuchowych brało udział 15-stu słuchaczy w tym 5 kobiet i 10-ciu mężczyzn. Słuchacze mieli za zadanie zidentyfikować prezentowane w kolejności losowej samogłoski.

Wyniki wskazują, że w percepcyjnej identyfikacji samogłosek istotne znaczenie ma nie tylko samo położenie formantów na skali częstotliwości, lecz również ich relacja do *F0*. Istnienie niedopasowania pomiędzy częstotliwościami formantowymi i wysokością głosu prowadzi do zmiany kategorii fonetycznej.

1. Wstęp.

Barwa dźwięku samogłoskowego w tradycyjnym opisie akustycznym zależy od obwiedni widma energetycznego i jest niezależna od częstotliwości podstawowej *F0*. Tymczasem jednak badania percepcyjne dowodzą wpływu zarówno częstotliwości właściwej samogłosek (ang. intrinsic pitch), jak i zdecydowanego wpływu częstotliwości podstawowej *F0* na ich jakość. I tak Carlson, Fant i Granström [4] pokazali, że taka sama struktura formantów może być postrzegana jako dwie, różne samogłoski w zależności od częstotliwości źródła

pobudzającego. Traunmüller [26] dowiódł wpływu odległości *F1-F0* wyrażonej w jednostkach subiektywnej skali wysokości dźwięku - barkach, na percepcję stopnia otwartości samogłosek. W swoich eksperymentach wykazał, że jednoczesne przesuwanie częstotliwości podstawowej *F0* i częstotliwości pierwszego formantu *F1* wzdłuż skali tonalnej zasadniczo nie zmienia wrażenia otwartości, podczas gdy przesuwanie tylko *F1* powoduje zmianę kategorii fonetycznej. Postrzegana otwartość samogłoski zmienia się w zależności od *F0* nawet, jeżeli częstotliwości formantowe, które powszechnie uważa się za odpowiedzialne za percepcyjną tożsamość samogłosek, pozostają te same. Niewielkie zmiany położenia formantów przy stałej częstotliwości *F0*, prowadzą do zmiany brzmienia samogłoski, większe różnicują kategorię fonetyczną. Analogicznie, gdy zmiany *F0* przekroczą pewną wartość krytyczną, może dojść do zmiany kategorii fonetycznej. Efekt taki jest funkcją podobieństwa fonemów samogłoskowych danego języka, jest tym większy im percepcyjna przestrzeń samogłoskowa jest gęściej wypełniona.

Podobne sugestie, znaleźć można w wielu pracach [25, 26, 27, 28, 29], jednak w zakresie języka polskiego najbardziej wnikliwą analizę przedstawił Imiołczyk [10, 11], który wykorzystując syntetyczne samogłoski stacjonarne należące do różnych kategorii głosowych, a będące odpowiednio przeskalowanymi męskimi samogłoskami (posłużono się czynnikami skalującymi podanymi przez Fanta [9]) przeprowadził szereg interesujących eksperymentów odsłuchowych. Analiza wyników doprowadziła go do wniosku, że częstotliwość podstawowa jest zasadniczym czynnikiem decydującym o uznaniu wypowiedzi za męską, kobiecą bądź dziecięcą. Odgrywa ona w związku z tym kluczową rolę w percepcyjnej normalizacji toru głosowego. Częstotliwości formantowe mogą być właściwie zinterpretowane dopiero na podstawie informacji, której nośnikiem jest *F0*. Formanty charakteryzujące daną samogłoskę muszą łączyć się z odpowiednią częstotliwością podstawową. Jeśli *F0* jest w stosunku do nich zbyt niska, postrzegana samogłoska staje się bardziej otwarta. Odwrotnie, jeśli *F0* jest zbyt wysoka samogłoska postrzegana jest jako bardziej przymknięta. Prace te są w opozycji do znacznie starszej (1973 r.) publikacji Kosiel [19], w której autorka bezskutecznie poszukiwała korelacji między *F0* a

częstotliwościami formantowymi. Wyniki przeprowadzonego przez nią doświadczenia nie dały podstaw do odrzucenia ogólnie wtedy przyjętej tezy, jakoby częstotliwość podstawowa była sterowana przez mówiącego niezależnie od sterowania narządami formującymi ponadkrtaniowy tor głosowy w zakresie wytwarzania segmentów wokalicznych.

Celem niniejszej pracy było potwierdzenie wyników i sugestii zaprezentowanych przez Imiołczyka, w odniesieniu jednak do syntetycznych samogłosek polskich będących wiernymi kopiami głosek naturalnych oraz próba znalezienia lub choćby przybliżonego oszacowania granicznych wartości parametru *F0*, przy których zachodzi zmiana kategorii fonetycznej.

Przekonanie o bezpośredniej zależności między położeniem formantów w głosach kobiecych, męskich i dziecięcych a *F0* było przyczyną, dla której Maurer i inni [21] badali rzeczywistą wokalizację samogłosek. Dowiedli oni, że różnice w położeniu formantów częściowo zanikają, gdy wartość *F0* jest identyczna dla różnych grup mówców. Dotyczy to formantu pierwszego oraz drugiego dla samogłosek tylnych.

Zagadnienia wpływu wysokości głosu na częstotliwości formantowe wiążą się w sposób ścisły i nierozerwalny z próbami wyjaśnienia mechanizmu, wspomnianej już wcześniej, percepcyjnej normalizacji toru głosowego nadawcy, a więc eliminacji zmienności osobniczej. Normalizacja powoduje, że słuchacze bezbłędnie potrafią rozpoznać i zaklasyfikować tą samą głoskę wypowiedzianą przez głos męski, kobiecy czy dziecięcy, choć pod względem struktury akustycznej dźwięki te różnią się znacznie. Wyjaśnienie mechanizmu normalizacyjnego jest podstawową kwestią w procesie automatycznego rozpoznawania mowy. Wydaje się, że synteza mowy może w dużej mierze przyczynić się do rozwiązania tego zagadnienia. Obecnie, bodaj najbardziej popularne są dwie hipotezy. Według pierwszej z nich normalizacja *F0* jest rezultatem audytywnej integracji częstotliwości podstawowej *F0* i harmonicznych w okolicy pierwszego formantu *F1* na spektrogramie w "środek ciężkości" (ang. SGC tj. spectral center of gravity - efekt zrelacjonowany po raz pierwszy przez Chistovich i innych [6]) [25, 26, 29]. Johnson w swoich pracach z 1988 roku [14, 15] próbuje dowieść prawdziwości drugiej

hipotezy. Dotyczy ona procesu dostosowania się do słuchacza (ang. adjustment to talker), w którym częstotliwość podstawowa *F0* służy, jako wskaźnik do identyfikacji mówcy, a normalizacja samogłosek jest rezultatem dostrojenia wewnętrznej przestrzeni samogłoskowej, z którą porównywane są napływające dźwięki. Wzrost czasów reakcji słuchowej podczas identyfikacji bodźców o bardzo zróżnicowanych *F0* w stosunku do identyfikacji bodźców o jednakowej częstotliwości podstawowej jest jednym z argumentów przemawiających za prawdziwością tej hipotezy.

2. Opis doświadczenia.

2.1. Przygotowanie wzorców.

W pracy posłużono się przygotowanymi wcześniej wzorcami sześciu samogłosek polskich: /i/, /ɨ/, /e/, /a/, /o/, /u/ wypowiedzianymi jednym ciągiem w sposób izolowany. Przeciętna długość każdej z samogłosek wynosiła ok. 200 ms. Nagrano je zwracając szczególną uwagę, aby poziom sygnału był możliwie najwyższy oraz by przebieg częstotliwości podstawowej *F0* nie wykazywał znacznych wahań. Wypowiedzi, poddane filtracji dolnoprzepustowej oraz próbkowaniu z częstotliwością 10 kHz w przetworniku A/C zapisywano bezpośrednio na dysku komputera. Następnie, przy pomocy skonstruowanego przez P. Domagałę software'owego analizatora sygnału mowy wykorzystującego metodę LPC [7], dokonywano ekstrakcji parametrów niezbędnych do resyntezy samogłosek. Dla każdej framy, której długość ustalono na 128 próbek, program wyliczał następujące parametry: częstotliwość podstawową *F0*, częstotliwości i szerokości formantów oraz wzmocnienie, traktowane jako amplituda tonu krtaniowego. Powyższe parametry, jak również parametry stałe dla danej wypowiedzi: czas trwania, liczba fram i częstotliwość próbkowania transformowano przy pomocy dodatkowego programu do formatu dostosowanego do syntezatora formantowego SMOK [22, 23]. Po wprowadzeniu niezbędnych korekt do przebiegu wyekstrahowanych parametrów oraz uzupełnieniu ich o nowe, używane do wysterowania wspomnianego syntezatora, wygenerowano syntetyczne kopie samogłosek (sterowanie syntezatorem

SMOK odbywa się przy pomocy 38 parametrów tj. 13-stu stałych i 25-ciu zmiennych w czasie, które pozwalają wybrać optymalną konfigurację układu, opisują kształt przebiegu tonu krtaniowego oraz funkcję transmitancji kanału głosowego). Ogółem wysyntetyzowano opisaną metodą sześć głosów: cztery kobiece i dwa męskie wypowiadające samogłoski polskie. Ich naturalność zweryfikowali słuchacze w przeprowadzonych badaniach odsłuchowych zaprezentowanych w publikacji [22]. Słuchacze mieli za zadanie zidentyfikować prezentowane w kolejności losowej samogłoski naturalne i syntetyczne, zaznaczyć gdy brzmią sztucznie, podać płeć nadawcy oraz w kolejnym teście wskazać jego imię, po uprzednim nauczeniu się rozpoznawania głosów wypowiadających wzorcowe samogłoski. Otrzymane rezultaty wskazują na wysoką naturalność głosów uzyskanych przy pomocy software'owego formantowego syntezatora mowy, uwzględniającą cechy osobnicze oraz na przydatność metody LPC do ekstrakcji parametrów do syntezy.

Tak zachęcające rezultaty oraz łatwość ingerencji w parametry użyte do syntezy, co zapewnia wytwarzanie powtarzalnych i jednoznacznych elementów mowy, skłaniają do wykorzystania ich w badaniach percepcyjnej zależności między wysokością głosu a częstotliwościami formantowymi występującymi w samogłoskach polskich.

Na rysunkach 1 i 2 przedstawiono w sposób graficzny odpowiednie przebiegi częstotliwości podstawowej *F0* i częstotliwości formantowych samogłosek polskich użytych do wysterowania syntezatora formantowego. Wyliczone na ich podstawie średnie wartości wspomnianych wyżej parametrów zaprezentowano w tabeli 1. Częstotliwości *F0* i *F1* posłużyły do wyliczenia przedstawionej w ostatniej kolumnie różnicy *F1-F0* wyrażonej w barkach - jednostkach subiektywnej skali wysokości dźwięku. Wykorzystano przy tym powszechnie znaną zależność wysokości tonu w paśmie krytycznym od częstotliwości:

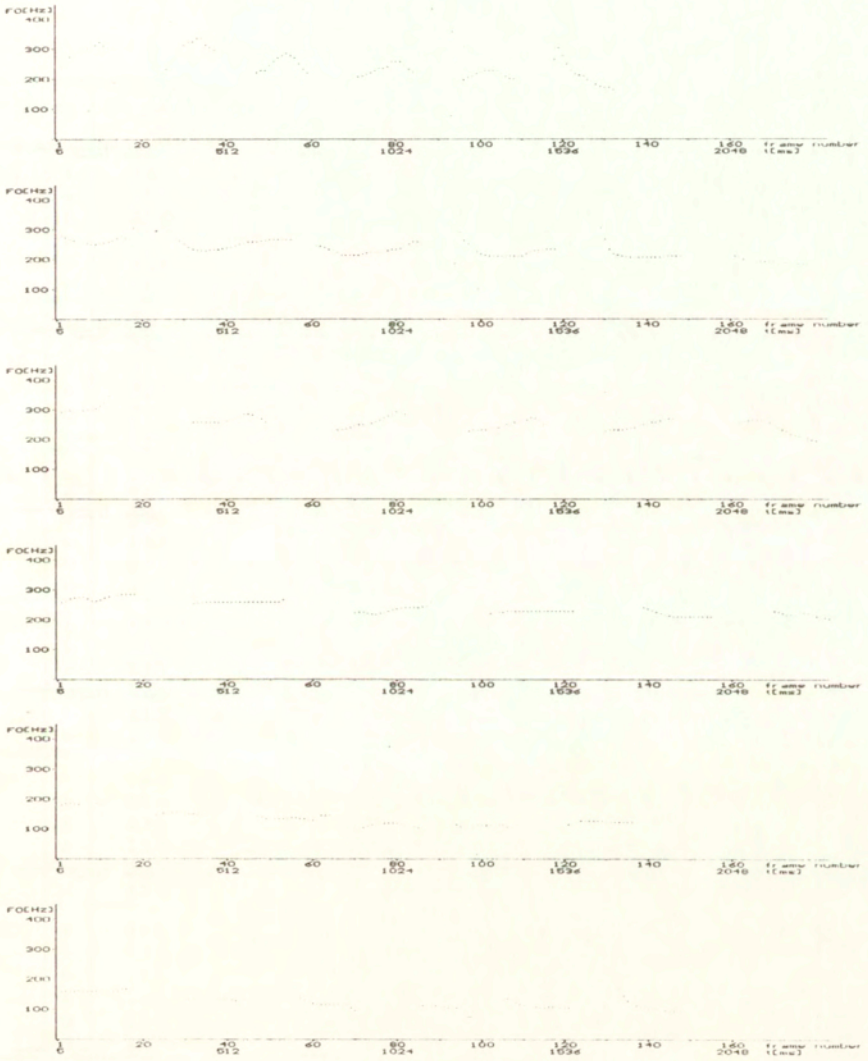
$$BARK = 13\arctan(0.76x) + 3.5(\arctan(\frac{x}{7.5}))^2, \quad (1)$$

gdzie x oznacza częstotliwość w kHz. Oczywiście, najpierw transformowano wartości *F1* i *F0* na skalę barkową, a następnie obliczano różnicę. Uzyskane wartości, zgodnie z przedstawioną

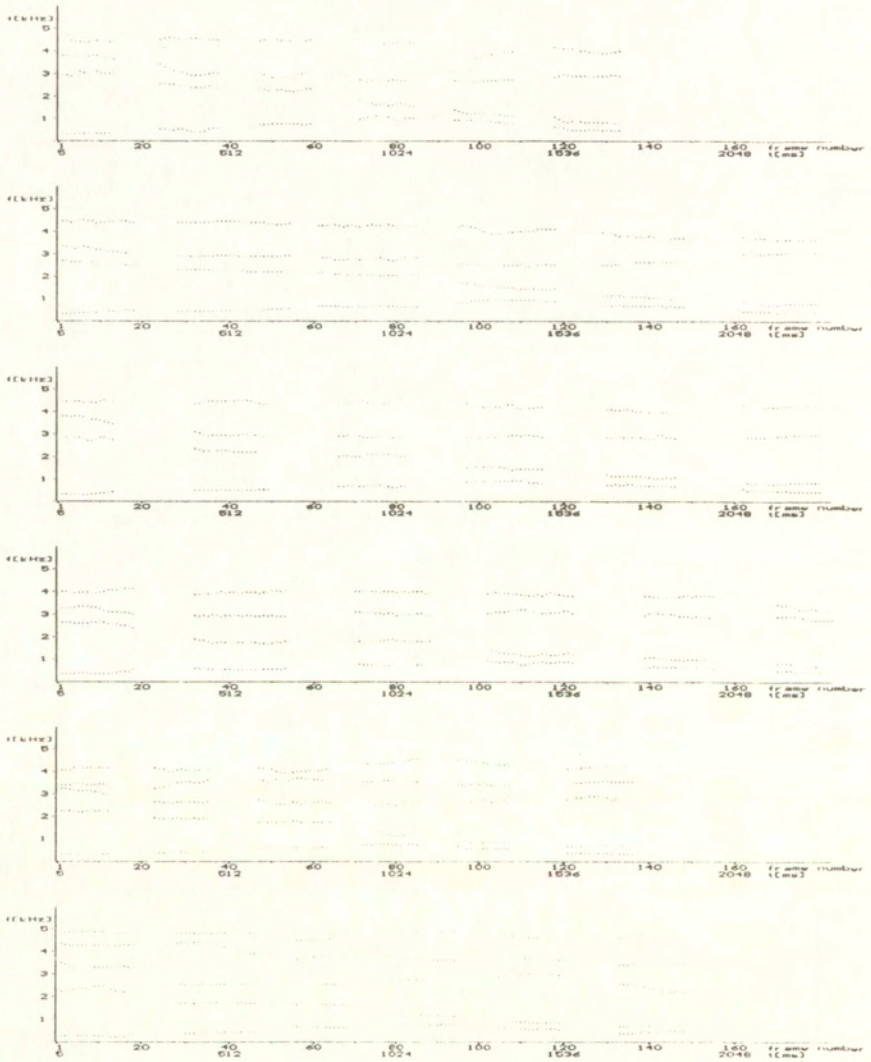
wcześniej teorią Traunmüllera [26], są miarą otwartości wysyntetyzowanych samogłosek. Nietrudno zauważyć, że tylko w przypadku samogłosek wypowiedzianych przez spikerów oznaczonych GD, EN, BS szereg samogłosek polskich uporządkowanych od najbardziej przymkniętej do najbardziej otwartej jest identyczny. W innych przypadkach istnieją nieznaczne różnice. Szereg rozpoczyna samogłoska /i/, następnie mamy /ɨ/ lub /u/, potem /e/ lub /o/, na końcu zdecydowanie najbardziej otwartą samogłoskę /a/. Gdyby wyliczyć średnie wartości różnicy *F1-F0* dla poszczególnych głosek szereg przyjąłby postać: /i/, /ɨ/, /u/, /e/, /o/, /a/.

Opisane, wzorcowe samogłoski będące kopią samogłosek naturalnych zostały zmodyfikowane. Ograniczono się do zmiany tylko jednego parametru a mianowicie częstotliwości podstawowej *F0*, chociaż w naturalnym głosie taka modyfikacja pociąga za sobą zmianę częstotliwości formantowych oraz parametrów opisujących ton krtaniowy: *OQ[%]* określającego procentowy udział fazy otwartej w okresie pobudzenia krtaniowego i modelującego widmo w zakresie niskich częstotliwości oraz *TL[dB]* określającego spadek obwiedni widma dla częstotliwości powyżej 3 kHz w stosunku do obwiedni wykazującej typowy spadek wynoszący -12 dB/oktawę (por. [16, 17, 18, 24]). Przyjęty zakres zmian *F0* wynosił ±1 oktawa ze skokiem wynoszącym 1/4 oktawy w stosunku do zarejestrowanego dla poszczególnych głosów przebiegu. Modyfikacji dokonano przy użyciu edytora parametrów wbudowanego do syntezatora SMOK, przesuwając równolegle zaprezentowane na rys. 1 przebiegi *F0* w górę i w dół skali częstotliwości, a następnie dokonując syntezy i zapisu otrzymanych przebiegów na dysku komputera. W ten sposób dotychczasowy inwentarz bodźców tj. 6 głosów * 6 samogłosek powiększono dziewięciokrotnie, co daje ogółem 324 bodźce, z czego 216 pochodziło od wzorców kobiecych a 108 od męskich. Każda samogłoska miała więc 54 realizacje pochodzące od 6-ciu różnych wzorców z 9-cioma różnymi przebiegami *F0*.

Cały ten materiał testowy poddano randomizacji za pomocą specjalnie w tym celu napisanego programu i nagrano na taśmę magnetofonową. Odstęp czasu pomiędzy poszczególnymi bodźcami wynosił 3 s. Po każdych 45-ciu sygnałach następował krótki muzyczny przerywnik trwający ok. 7s. Taki podział materiału testowego



Rys. 1. Wykres przebiegu częstotliwości podstawowej *F0* w funkcji numeru framy (czasu) dla samogłosek polskich: /i/, /ɨ/, /e/, /a/, /o/, /u/ wypowiedzianych przez czterech kobiecych (AS, GD, EN, IP) i dwóch męskich (BS, MO) mówców.



Rys. 2. Wykres przebiegu formantów w funkcji numeru framy (czasu) dla samogłosek polskich: /i/, /ɨ/, /e/, /a/, /o/, /u/ wypowiedzianych przez czterech kobiecych (AS, GD, EN, IP) i dwóch męskich (BS, MO) mówców.

wzorzec		F0[Hz]	F1[Hz]	F2[Hz]	F3[Hz]	F4[Hz]	F5[Hz]	F1-F0 [Bark]
AS	i	292	311	3015	3741	4454	-	0.18
	ɨ	285	493	2457	3076	4545	-	1.90
	e	250	753	2277	2930	4477	-	4.35
	a	232	1033	1654	2690	4347	-	6.45
	o	214	871	1200	2693	4019	-	5.55
	u	211	500	860	2902	4010	-	2.67
GD	i	264	383	2581	3192	4440	-	1.11
	ɨ	248	471	2253	2907	4402	-	2.06
	e	229	630	2078	2772	4258	-	3.59
	a	224	897	1508	2502	4054	-	5.63
	o	222	654	1043	2593	3781	-	3.85
	u	192	407	762	3008	3641	-	2.02
EN	i	308	321	2768	3665	4441	-	0.12
	ɨ	266	494	2209	2940	4433	-	2.09
	e	260	669	2045	2862	4427	-	3.60
	a	247	871	1474	2858	4223	-	5.23
	o	247	703	1089	2821	4019	-	3.99
	u	235	458	809	2870	4228	-	2.07
IP	i	273	351	2568	3192	4008	-	0.73
	ɨ	256	513	1738	2902	3938	-	2.35
	e	231	699	1797	3002	3976	-	4.12
	a	226	824	1224	3093	3878	-	5.10
	o	215	598	970	2918	3769	-	3.46
	u	215	406	703	2806	3251	-	1.79
BS	i	184	300	2207	3118	3402	4122	1.11
	ɨ	150	388	1887	2607	3451	4053	2.26
	e	135	589	1745	2593	3618	4040	4.14
	a	113	731	1172	2516	3558	4426	5.51
	o	110	569	889	2677	3412	4400	4.24
	u	123	342	672	2805	3525	4201	2.10
MO	i	161	258	2289	3314	4279	4853	0.94
	ɨ	133	383	1716	2519	4230	4807	2.38
	e	118	627	1619	2491	3753	4586	4.64
	a	114	752	1164	2695	3658	4853	5.66
	o	112	560	867	2945	3654	4617	4.15
	u	109	424	694	2360	3450	4710	2.99

Tabela 1. Średnie wartości częstotliwości podstawowej *F0* i częstotliwości formantowych obliczone na podstawie parametrów określających samogłoski polskie wypowiedziane przez czterech kobiecych (AS, GD, EN, IP) i dwóch męskich (BS, MO) mówców. W ostatniej kolumnie przedstawiono wyrażoną w barkach różnicę *F1-F0*.

odpowiadał rozkładowi miejsc przeznaczonych na odpowiedzi w ankiecie, którą wypełniali słuchacze, ułatwiając im kontrolę numeru sygnału, który aktualnie identyfikują.

2.2. Badania odsłuchowe.

W badaniach odsłuchowych przeprowadzonych w warunkach studyjnych w bezechowej komorze uczestniczyło 15-stu słuchaczy w tym 5 kobiet i 10-ciu mężczyzn. 5-ciu słuchaczy odsłuchiwało test trzykrotnie. Stwierdzono jednak wysoką zgodność otrzymywanych przezposzczególnych słuchaczy w kolejnych seriach rezultatów i dlatego zaniechano dalszych wielokrotnych przesłuchań. Od tej pory każdy słuchacz jednokrotnie wypełniał dostarczoną ankietę wysłuchując 324-ech testowych samogłosek. Ogółem uzyskano 25 wypełnionych ankiet. Słuchacze mieli za zadanie zidentyfikować podawane w kolejności losowej samogłoski.

3. Wyniki i wnioski.

Liczba "fałszywie ocenionych" (tzn. zakwalifikowanych do innych kategorii fonetycznych, niżby to sugerował układ formantów) samogłosek na ogólną liczbę 324-ech prezentowanych w czasie jednego testu wahała się w granicach od 21 do 79. Średnio jednak, wartość ta wynosiła 44.3 błędy, co daje 13.7 % nieprawidłowo zidentyfikowanych głosek. Oczywiście na liczbę tę składały się błędy przypadkowe oraz wynikłe z pewnej nienaturalności podawanych bodźców (samogłoski syntetyczne), sposobu prezentacji, występującej koartykulacji itd., a także będące przedmiotem niniejszego badania, "błędy" pochodzące z celowej zmiany częstotliwości podstawowej *F0*.

Wyniki badań odsłuchowych przedstawiono w tabeli 2. Podzielona jest ona na dwie części. W górnej, podano liczbę i rodzaj błędnych identyfikacji syntetycznych samogłosek polskich będących zmodyfikowanymi w zakresie częstotliwości podstawowej *F0* kopiami naturalnych głosek pochodzących od męskich mówców.Wartości podane w dolnej części tabeli dotyczą analogicznego zbioru samogłosek pochodzących jednak od kobiecych prototypów. Na ogólną liczbę 324-ech bodźców prezentowanych w sposób losowy w każdym teście, 108

	ΔF0 w oktawach								
	-1	-0.75	-0.5	-0.25	0	+0.25	+0.5	+0.75	+1
i				1*u*				14*ɨ* 10*u*	1*ɨ*
ɨ			1*e*				1*u*	2*u*	2*i* 1*u*
e				1*o*					
a				1*o*		1*o*	2*o*	15*o*	34*o*
o									1*u*
u	14*o*	14*o*	17*o*	14*o*	6*o*	6*o*	1*o*	1*o*	
i	3*e* 2*ɨ*	4*ɨ* 2*e*	5*e* 2*ɨ*	4*ɨ*	3*ɨ* 1*e*	3*ɨ* 1*e*	2*ɨ*		
ɨ	60*e* 1*i*	44*e*	47*e*	27*e*	18*e* 1*i*	6*e* 6*i* 1*u*	7*i* 4*e*	14*i* 1*e* 1*u*	28*i* 3*u*
e						5*ɨ* 2*o*	2*ɨ*	10*ɨ*	53*ɨ*
a				7*o*	1*o*	5*o*	6*o*	21*o* 2*e*	15*o*
o	54*a*	48*a*	29*a*	28*a*	20*a* 1*u*	11*a* 2*u*	20*a* 5*u*	19*a* 9*u*	20*a* 19*u*
u	64*o*	44*o*	40*o*	28*o*	19*o*	11*o*	3*o*	3*o*	

Tabela 2. Liczba i rodzaj błędnych identyfikacji syntetycznych samogłosek polskich będących zmodyfikowanymi kopiami wypowiedzi męskich (górna część tabeli) i kobiecych (dolna część tabeli) mówców. Częstotliwość podstawowa *F0* podlegała skokowej zmianie co 1/4 oktawy w granicach ±1 oktawa, w stosunku do przebiegów intonacji wyekstrahowanych z prototypów.

pochodziło od męskich, a 216 od kobiecych mówców. Suma błędnych identyfikacji dotycząca głosów męskich wyniosła 161 na ogólną liczbę 2700 bodźców ocenionych przez słuchaczy, kobiecych natomiast 927 na ogólną liczbę 5400, co daje odpowiednio 6.0 % i 17.2 %. Ten prawie trzykrotnie wyższy procent fałszywie rozpoznawanych samogłosek w przypadku głosów kobiecych o zmienionej w szerokich granicach częstotliwości *F0* dowodzi, iż głosy te są bardziej podatne na zmiany parametru *F0* i łatwiej zmieniają przynależność do danego fonemu pod jego wpływem. Prawdopodobnie niebagatelne

znaczenie ma położenie tych głosów na skali *F0*. Zakres możliwych głosów kobiecych jest ograniczony z jednej strony przez głosy męskie, a z drugiej dziecięce. To implikuje bardziej krytyczną ocenę słuchaczy w odniesieniu do głosów kobiecych niż męskich czy dziecięcych i jest jedną z istotnych przyczyn trudności w uzyskaniu wysokiej jakości syntetycznej mowy kobiecej.

W czasie przygotowywania materiału-testowego zwrócono uwagę, iż identyfikacja samogłosek o zmienionej częstotliwości podstawowej nie nastręczała żadnych trudności i była jednoznaczna, gdy były one podawane w bloku o zbliżonej wartości *F0* (należały do jednego głosu). Przypadek taki miał miejsce nawet wtedy, kiedy formanty kobiecego /o/ połączono z "męską" (zmienioną o -1 oktawę) częstotliwością podstawową. Wystarczyło jednak ocenić słuchowo opisaną realizację samogłoski /o/ w izolacji, by zdecydowana większość słuchających zaklasyfikowała ją do fonemu /a/. Obserwacja taka jest zgodna z efektem kontrastu występującym podczas normalizacji częstotliwości podstawowej *F0*, zgłoszonym przez Johnsona [15: str.255].

Wysokość głosu wywiera istotny wpływ na percepcję stopnia otwarcia samogłosek: im, dla danego zestawu formantów, jest ona wyższa, tym samogłoska wydaje się być bardziej przymknięta; i na odwrót: im jest niższa, tym bardziej otwarta staje się samogłoska. Gdy zmiany *F0* przekroczą pewną wartość krytyczną, może dojść do zmiany kategorii fonetycznej. Efekt taki jest w dużym stopniu zależny od podobieństwa fonemów samogłoskowych danego języka. W przypadku języka polskiego najbardziej podobne, i to zarówno pod względem artykulacyjnym, akustycznym i percepcyjnym, są głoski /a/ oraz /o/. Biorąc pod uwagę powyższe fakty oraz zaprezentowany w paragrafie 2.1, a wyliczony na podstawie wyrażonej w barkach różnicy *F1-F0*, szereg samogłosek polskich uporządkowanych od najbardziej przymkniętej do najbardziej otwartej, łatwiej właściwie ocenić i zrozumieć wyniki zaprezentowane w tabeli 2. Należy jednocześnie pamiętać, że maksymalna liczba błędnych odpowiedzi dla poszczególnych samogłosek i odpowiednio zmienionych częstotliwości *F0* (dotyczy jednej kratki w tabeli) w przypadku głosów męskich wynosi 50 (2*liczba słuchaczy) a dla głosów kobiecych 100 (4*liczba słuchaczy). Teraz już nietrudno zauważyć, że dla głosów męskich

zmiana kategorii fonetycznej może nastąpić, gdy formanty charakterystyczne dla /a/ połączyć z podwojoną częstotliwością podstawową. Aż w 34-ech przypadkach na 50 możliwych, co stanowi 68% wszystkich odpowiedzi w tej kategorii, słuchacze wskazali na samogłoskę /o/. Charakterystyczny jest stopniowy wzrost liczby odpowiedzi /o/, wraz ze wzrostem parametru *F0*, a tym samym zwiększenie niedopasowania między wysokością głosu i częstotliwościami formantowymi.

Jak już wspomniano, samogłoski wypowiedziane przez kobiecych spikerów są "precyzyjniej zdefiniowane" od męskich, co sprzyja zmianie kategorii fonetycznej. I tak, zdecydowane obniżenie wartości *F0* powoduje, że /i/ jest identyfikowane jako /e/, /o/ jako /a/, a /u/ jako /o/. Natomiast podwyższenie wartości częstotliwości podstawowej dla tej kategorii głosów wywoła percepcyjną zamianę /e/ w /i/. O ile, percepcyjne przejście męskiego /a/ w /o/ pod wpływem zwiększenia parametru *F0*, czy też odwrotne, kobiecego /o/ w /a/ przy obniżeniu *F0*, łatwo można wyjaśnić podobieństwem częstotliwości formantowych męskiego /a/ i kobiecego /o/, to jednak aby zrozumieć np. zidentyfikowanie kobiecego /e/ przez ponad połowę słuchaczy jako /i/ przy zwiększeniu *F0*, należy posłużyć się teorią przedstawioną przez Traunmüllera [26]. Średnia różnica *F1-F0* dla kobiecego /e/ w rozpatrywanych głosach wynosi 3.9 barka. W miarę zwiększenia *F0* zmniejsza się ona, aż osiągnie wartość mniejszą od krytycznej, wynoszącej 3 barki. Teraz mogą już zaistnieć słuchowe procesy integracji *F0* i harmonicznych w pobliżu *F1* w jeden "środek ciężkości". Gdy obniży się on i osiągnie wartość zbliżoną do pierwszego formantu dziecięcego /i/, tak właśnie zostanie oceniony przez słuchaczy. Teoria audytywnej integracji dobrze tłumaczy rodzaj poszczególnych przejść i określa płynną granicę, przy której ten efekt ma miejsce.

Liczby i wskazania opisujące reakcję słuchaczy na prezentację poszczególnych samogłosek o niezmodyfikowanej wartości *F0* przedstawiono w środkowej, wyróżnionej kolumnie tabeli. Stanowią one rzeczywisty błąd systematyczny, który powinno się uwzględnić w rozważaniach. Przyczyny jego wystąpienia należy szukać przede wszystkim w zjawisku koartykulacji zachodzącym pomiędzy sąsiednimi głoskami w szeregu artykulacyjnym, według którego wzorcowe,

naturalne samogłoski były wypowiadane. Samogłoski występujące w testach są wiernymi kopiami tych głosek. Na rysunkach 1 i 2, prezenujących przebiegi intonacji i częstotliwości formantowych łatwo zaobserwować to pozornie niekorzystne zjawisko utrudniające właściwą identyfikację. Z drugiej jednak strony, koartykulacja jest naturalną cechą języka mówionego, nierozerwalnie z nim związaną. Dlatego właśnie pozwolono, by "zakłócała" ona przebieg doświadczenia.

Modyfikowanie położenia formantów (szczególnie tych najniższych), przy stałej wysokości głosu, powoduje zmianę barwy samogłoski. Większe zmiany, różnicują kategorię fonetyczną. Analogicznie, niewielkie manipulacje częstotliwością podstawową *F0* zmieniają brzmienie samogłoski poprzez wpływ na percepcję stopnia jej otwarcia. Skutkiem większych zmian *F0*, gdy przekroczona zostanie pewna wartość krytyczna, może być, jak dowiedziono powyżej, również zmiana kategorii fonetycznej. Daje się więc zauważyć pewne rekompensacyjne oddziaływanie wysokości głosu względem częstotliwości formantowych w percepcji samogłosek. W identyfikacji samogłosek istotne znaczenie ma nie tylko samo położenie formantów na skali częstotliwości, lecz także ich relacja do *F0*. Istnienie niedopasowania pomiędzy częstotliwościami formantowymi i wysokością głosu prowadzi do percepcyjnej zmiany kategorii fonetycznej.

Bibliografia

[1] R.A.W. Bladon, B. Lindblom, Modeling the judgment of vowel quality differences, J. Acoust. Soc. Am., **69**, 5, 1414-1422, (1981).

[2] A. Bladon, Arguments against formants in the auditory representation of speech. In R. Carlson and B. Granström (editors), *The Representation of Speech in the Peripheral Auditory System*: New York, Elsevier Biomedical Press 1982, 95-102.

[3] G. Bloothooft, R. Plomb, Spectral analysis of sung vowels. II. The effect of fundamental frequency on vowel spectra, J. Acoust. Soc. Am., **77**, 4, 1580-1588, (1985).

[4] R. Carlson, G. Fant, B. Granström, Two-formant models, pitch, and vowel perception. In G. Fant and M.A.A. Tatham (editors), *Auditory Analysis and Perception of Speech*: Academic Press, London, 1975, 55-82.

[5] R. Carlson, J. Glass, Vowel clasification based on analysis-by-synthesis, Speech Transmission Laboratory, Quarterly Progress and Status Report, **4**, (1992).

[6] L.A. Chistovich, R.L. Sheikin, V.V. Lublinskaya, 'Centers of gravity' and spectral peaks as the determinants of vowel quality. In B. Lindblom and S. Öhman (editors), *Frontiers of Speech Communication Research*, Academic Press, London, 1979, 143-157.

[7] P. Domagała, Wielowymiarowa statystyczna analiza parametrów LPC segmentów fonetycznych w typowych połączeniach, Prace IPPT, **21**, (1988).

[8] C.G.M. Fant, *Acoustic theory of speech production*, Mouton, The Hague, 1960.

[9] G. Fant, Non-uniform vowel normalization, Speech Transmission Laboratory, Quarterly Progress and Status Report, **2-3**, (1975).

[10] J. Imiołczyk, Determination of perceptual boundaries between the male female and child's voices in isolated synthetic polish vowels, Archives of Acoustics, vol.**16**, 2, 305-323, (1991).

[11] J. Imiołczyk, Wpływ relacji między częstotliwościami

formantowymi i częstotliwością podstawową na percepcję samogłosek polskich, Księga pamiątkowa ofiarowana Pani Profesor Marii Steffen-Batogowej oraz Panu Profesorowi Tadeuszowi Batogowi, red. J. Pogonowski, Wydawnictwo Naukowe UAM, Poznań 1993, (przyjęte do druku).

[12] W. Jassem, *Podstawy fonetyki akustycznej*, PWN, Warszawa 1973.

[13] W. Jassem, Acoustic-phonetic variability of polish vowels, Archives of Acoustisc, vol.**17**, 2, 217-233, (1992).

[14] K. Johnson, Intonational context and F0 normalization, Research on speech perception, Progress Report No. **14**, 81-108, (1988).

[15] K. Johnson, F0 normalization and adjusting to talker, Research on speech perception, Progress Report No. **14**, 237-258, (1988).

[16] I. Karlsson, Evaluations of acoustic differences between male and female voices; a pilot study, Speech Transmission Laboratory, Quarterly Progress and Status Report, **1**, (1992).

[17] D.H. Klatt, Software for a cascade/parallel formant synthesizer, J. Acoust. Soc. Am., **67**, 971-995, (1980).

[18] D.H. Klatt & L.C. Klatt, Analysis, synthesis, and perception of voice quality variations among female and male talkers, J. Acoust. Soc. Am., **87**, 2, 820-857, (1990).

[19] U. Kosiel, Correlations between fundamental frequency and formant frequencies in Polish vowels, Speech analysis and synthesis, vol.**3**, 117-120, PWN, Warszawa 1973.

[20] H. Kubzdela, M. Owsianny, Wpływ poziomu formantów na percepcję syntetycznych dźwięków samogłoskowych, Prace IPPT, **41**, (1991).

[21] D. Maurer, N. Cook, T. Landis, Ch. d'Heureuse, Are measured differences between the formants of men, women and children due to F0 differences?, J. Internat. Phonetic Assoc., **21**, 2, (1992).

[22] M. Owsianny, The synthesis of female voices using a software synthesizer, Archives of Acoustics, (przyjęte do druku).

[23] M. Owsianny, Software'owa realizacja formantowego syntezatora mowy, Księga pamiątkowa ofiarowana Pani Profesor Marii Steffen-Batogowej oraz Panu Profesorowi Tadeuszowi Batogowi, red. J. Pogonowski, Wydawnictwo Naukowe UAM, Poznań 1993, (przyjęte do druku).

[24] K.N. Stevens, The contribution of speech synthesis to phonetics: Dennis Klatt's legacy, Proceedings of the 12th International Congress of Phonetic Sciences, Aix, vol.**1**, 28-37, (1991).

[25] A.K. Syrdal, Aspects of a model of the auditory representation of American English vowels, Speech Commun., **4**, 121-135, (1985).

[26] H. Traunmüller, Perceptual dimension of openness in vowels, J. Acoust. Soc. Am., **69**, 5, 1465-1475, (1981).

[27] H. Traunmüller, Articulatory and perceptual factors controlling the age- and sex-conditioned variability in formant frequencies of vowels, Speech Commun., **4**, 49-61, (1984).

[28] H. Traunmüller, Paralinguistic variation and invariance in the characteristic frequencies of vowels, Phonetica, **45**, 1, (1988).

[29] S.A. Zahorian & A.J. Jagharghi, Speaker normalization of static and dynamic vowel spectral features, J. Acoust. Soc. Am., **90**, 1, 67-75, (1991).