

Praca wstępna do Rozprawy doktorskiej z dnia 2 marca 1995 r.

**Grażyna Demenko**

**SYNTEZA PODSTAWOWYCH TYPÓW  
PRZEBIEGÓW INTONACYJNYCH**

**19/1995**

P. 269

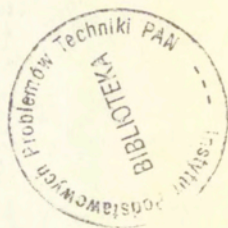


**WARSZAWA 1995**

<http://rcin.org.pl>

ISSN 0208-5658

Praca wpłynęła do Redakcji dnia 2 marca 1995 r.



56590



Na prawach rękopisu

---

Instytut Podstawowych Problemów Techniki PAN  
Nakład 100 egz. Ark. wyd. 1,5 Ark. druk. 2,0  
Oddano do drukarni w maju 1995 r.

---

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

<http://rcin.org.pl>

## SYNTEZA PODSTAWOWYCH TYPÓW PRZEBIEGÓW INTONACYJNYCH

### I. Streszczenie

Korzystając z pakietu programów analizy i syntezy mowy metodą predykcji liniowej, przygotowano kilkadziesiąt wypowiedzi syntetycznych różniących się przebiegami częstotliwości podstawowej. Przetestowano kilka hipotez dotyczących wpływu różnych akustycznych realizacji wypowiedzi na percepcję akcentu. Badano następujące czynniki: a) istotność zmian częstotliwości podstawowej oraz iloczasu b) dynamikę, charakter oraz umiejscowienie w obrębie wypowiedzi zmian parametru  $F_0$ . Doświadczenie odsłuchowe przeprowadzono z grupą 30 studentów. Eksperyment wykazał, że istotny wpływ na percepcję akcentu w języku polskim ma przebieg częstotliwości podstawowej, iloczas i amplituda mogą być parametrami dodatkowymi ułatwiającymi rozpoznanie akcentu. Poprzez modelowanie przebiegu parametru  $F_0$  można uzyskać zmianę miejsca akcentu realnego i zmianę znaczenia wypowiedzi. Słuchacze oceniają intonację wypowiedzi globalnie. Dla percepcji akcentu ważna jest szybkość i dynamika oraz miejsce wystąpienia zmiany parametru  $F_0$ .

## II. Wstęp.

Pomimo szeregu prac poświęconych syntezie intonacji, problem sterowania częstotliwością podstawową w mowie ciągłej nie jest nadal wystarczająco dobrze rozwiązany. Zwiększenie naturalności mowy wymaga weryfikacji i uzupełnienia dotychczas wykorzystywanych zasad generowania parametru  $F_0$ . Opracowana w ostatnich latach w Zakładzie Fonetyki Akustycznej synteza języka polskiego wymaga reguł sterowania informacją prozodyczną w mowie ciągłej (por. 1,3,4). Dla sformułowania precyzyjnych reguł syntezy intonacji w mowie ciągłej niezbędna jest kompleksowa analiza sygnału. Złożoność problemu wynika z konieczności: a) uwzględnienia wielu równoległych działających źródeł zmienności, b) subiektywnej i obiektywnej oceny intonacji, c) segmentacji sygnału na prozodyczne jednostki, d) klasyfikacji intonacyjnych wzorców, e) reguł modelowania grup akcentowych (por.np.[6]). Istota akcentu określona w terminach akustycznych pozostaje w dalszym ciągu kontrowersyjna. Uwydatnienie (zaakcentowanie) sylaby może być w różny sposób akustycznie oraz fonetycznie zrealizowane. W niektórych językach określa się akcent poprzez długość sylaby, przebieg intensywności, częstotliwości podstawowej (por. np. 9,10). W większości opracowań podkreśla się jako najistotniejszą cechę zmienności parametru  $F_0$ . Jako parametry drugorzędne przyjmuje się zmiany intensywności i długości sylaby (np. 5,7). Dla języka polskiego analiza akustyczna (por.np. [8],[11]), oraz ostatnio przeprowadzone przez Jassema, Dommelena doświadczenia percepcyjne [7] potwierdziły istotność zmian częstotliwości podstawowej w akcentuacji języka polskiego.

Niniejsze zadanie badawcze poświęcone jest przygotowaniu syntetycznych przebiegów parametru  $F_0$  na względnie prostych, kontrastywnych próbkach mowy oraz percepcyjnej weryfikacji utworzonych wzorców. Uznano za konieczne przetestowanie kilku hipotez dotyczących wpływu różnych akustycznych realizacji wypowiedzi na percepcję akcentu. Badano następujące czynniki: a) udział częstotliwości podstawowej oraz iloczasu w realizacji akcentu, b) charakter i dynamikę zmian parametru  $F_0$ , c) wpływ intonacji całej wypowiedzi na percepcję sylaby akcentowanej. W dalszym etapie pracy zakłada się implementację utworzonych wzorców intonacyjnych w mowie ciągłej w systemie text-to-speech.

### III. Techniczna realizacja syntezy.

Analizę i syntezę LPC przeprowadzono w systemie komputera PC486 i spektrografu cyfrowego Kay 5500. Wyekstrahowane parametry LPC przedstawione graficznie i numerycznie, zawierają informacje o wartościach formantów, szerokościach pasm i względnych poziomach formantów oraz o wartości częstotliwości podstawowej. W tabeli 1 (w załączniku) umieszczono przykładowy wydruk parametrów LPC. Wyekstrahowane parametry mogą być zmienione poprzez edytor i weryfikowane w procesie resyntezy. Z uwagi na ograniczenia pamięciowe komputera, analizę LPC przeprowadzono dla próbek mowy o maksymalnej długości 5 sekund i ograniczonych do pasma 4000Hz. W rezultacie otrzymane próbki mowy syntetycznej charakteryzowały się gorszą jakością niż w przypadkach analizy sygnału do 8000Hz. Ponieważ w eksperymencie badane są cechy przodocyczne sygnału założono, że lepszym rozwiązaniem jest przyjęcie do analizy wypowiedzi dłuższych o nieco gorszej jakości niż bardzo krótkich (do 2 sekund) nieograniczonych częstotliwościowo. Program umożliwia analizę LPC pojedynczych fram sygnału lub całej wypowiedzi metodą kowariancji lub korelacji. Po procesie analizy LPC, numeryczny edytor umożliwia ingerencje użytkownika w parametry LPC i wybór rodzaju syntezy. Standardowa konfiguracja analizy pozwala na korektę wartości minimalnej i maksymalnej dla parametru  $F_0$ , liczby współczynników predykcji, długości framy LPC, współczynników emfazy lub preemfazy sygnału. Dla modelowania przebiegów parametru  $F_0$  wykorzystano opcję wpisywania wartości tego parametru do kolejnych 10 milisekundowych fram sygnału. Programy umożliwiają natychmiastowy odsłuch uzyskanej wypowiedzi syntetycznej i naturalnej, analizę spektrograficzną oraz poprzez numeryczny edytor korektę ich wartości. Ponieważ proces analizy nie jest pozbawiony błędów, uzyskanie syntetycznej wypowiedzi dobrej jakości jest pracochłonne i wymaga na bieżąco korekty parametrów w procesie analizy-resyntezy. Jakość wszystkich wykorzystanych w pracy syntetycznych wypowiedzi oceniano percepcyjnie i na bieżąco optymalizowano. Analiza na spektrografie Kay 555 umożliwiła segmentację materiału oraz weryfikację pomiaru. Materiał zapisany na arkuszach kalkulacyjnych będzie stanowić podstawę do matematycznej klasyfikacji przebiegów intonacyjnych.

#### IV. Przygotowanie materiału lingwistycznego.

Na taśmie magnetofonowej zapisano 2 pary wypowiedzi

- 1 a) *mama ma kawę?*
- b) *ma mama kawę?*
- 2 a) *będziemy jeść, bambino?*
- b) *będziemy jeść bambino?*

wymówione 3-krotnie przez dwóch fonetyków. W wypowiedzi pierwszej akcent prymarny pada na wyraz *kawę*. W przypadku wypowiedzi 1a zakres zmiany parametru  $F_0$  na tym wyrazie jest znaczny (około trzykrotny wzrost wartości parametru), w przypadku wypowiedzi 1b dynamika zmiany częstotliwości parametru  $F_0$  jest znacznie mniejsza (około dwukrotny wzrost wartości). Główna różnica w wypowiedziach 1a i 1b zawarta jest w interwale parametru  $F_0$  na wyrazie *kawę* - w wypowiedzi 1a reprezentowana jest intonacja pełna rosnąca, w wypowiedzi 1b - intonacja rosnąca. Zmienne miejsce akcentu w wypowiedzi 2 (w wypowiedzi 2a na wyrazie *jeść*, w wypowiedzi 2b na wyrazie *bambino*) decyduje o zmianie znaczenia zdania. Wypowiedzi 1 i 2 były w dalszej części doświadczenia traktowane jako wzorcowe. Z 24 zdań (2 mówców x 4 wypowiedzi x 3 replikacje) wybrano zestaw 4 wzorcowych wypowiedzi. Na podstawie analizy spektrograficznej przyjęto do syntezy te pary wypowiedzi, między którymi zaistniały najmniejsze różnice w wahaniach poziomu sygnału oraz w długościach segmentów fonetycznych.

W zdaniach 1a i 1b (a - *mama ma kawę*, b - *ma mama kawę*) różnice w poziomach amplitud na samogłoskach nie przekraczały 3-4 db, również różnice w długościach kolejnych sylab w zdaniach (a) i (b) były stosunkowo niewielkie ( rzędu 20-30ms). W wypowiedziach 2a i 2b (a - *będziemy jeść, bambino* i b - *będziemy jeść bambino*) główna różnica polegała na udźwięcznieniu lub ubezdźwięcznieniu zbitki spółgłoskowej *śc* w wyrazie *jeść*. Na rys.1 zilustrowano przebiegi parametru  $F_0$  w wypowiedzi 1a i 1b. W wypowiedzi 1a obserwuje się wzrost parametru  $F_0$  na sylabach *mama* (ok. 35Hz), spadek częstotliwości podstawowej na sylabach *maka* (ok.60Hz) i wzrost na sylabie *we* (do 260Hz). W wypowiedzi 1b wzrost parametru  $F_0$  występuje na sylabie *ma* (ok.70Hz), spadek na sylabach *mamaka* (ok. 80Hz) i wzrost do 140 Hz na sylabie *we*. Ryc.2 ilustruje przebiegi częstotliwości podstawowej w wypowiedzi 2a i 2b. Na sylabach *będziemy* zauważa się małe różnice w

zmienności parametru  $F_0$ , główne różnice obserwuje się na sylabach *bambino*. W przypadku wypowiedzi *lb* na wyrazie *jeść* zauważa się spadek parametru  $F_0$  do wartości minimalnej (84Hz), po którym następuje wzrost parametru  $F_0$  na sylabach *bambi*. W wypowiedzi *lb* wartość minimalną przebiegu obserwuje się na sylabie *bi* (75Hz), po której następuje na sylabie *no* wzrost parametru  $F_0$  do wartości 227Hz.

## V. Synteza wybranych wypowiedzi.

Opracowano kilka testów mających na celu ustalenie wpływu różnych realizacji akustycznych wypowiedzi na percepcję akcentu.

Testy dotyczyły:

- a) sprawdzenia jakości mowy syntetycznej
- b) ustalenia wpływu długości sylaby
- c) wpływu intonacji w obrębie całej wypowiedzi
- d) wpływu miejsca ekstremum, dynamiki i szybkości zmian parametru  $F_0$ .

W doświadczeniu odsłuchowym uczestniczyło 30 studentów Wydziału Fizyki.

### 5.1. test 0

Sprawdzenie jakości mowy syntetycznej do oceny intonacji.

Do odsłuchu podano wypowiedzi naturalne i syntetyczne. Słuchacz decydował którą wersję wypowiedzi usłyszał oraz odpowiadał na pytanie czy melodie w wypowiedzi naturalnej i syntetycznej są do siebie podobne. Dla określenia stopnia podobieństwa słuchacze wykorzystywali skalę ocen od 0-4, według następującej zasady : ocenę 0 otrzymywały pary wypowiedzi usłyszane jako zupełnie inne, ocenę 1- niepodobne, 2- podobne ,3- bardzo podobne 4 - takie same. Na arkuszu ocen słuchacze zakreślali odpowiedzi według następującego schematu

	wersja	wersja
	a	b
wypow. naturalna	.....	....
wypow. syntetyczna	.....	....
	podobieństwo wypowiedzi	
wypow. naturalna		
i syntetyczna	0	1 2 3 4

Przyjęto długie przerwy między wypowiedziami przy ocenie wersji wypowiedzi (6s) , i krótkie przy ocenie podobieństwa między parami wypowiedzi (1s). Otrzymano, średnio w 94% zgodność odpowiedzi słuchaczy przy rozróżnianiu wersji a i b w poszczególnych parach wypowiedzi. W ocenie podobieństwa wypowiedzi w 90% stwierdzono, że melodie są takie same, w 10%, że są bardzo podobne. W dalszych testach wykorzystano wypowiedzi syntetyczne. Ryc. 3 i 4 ilustrują spektrogramy wzorcowych wypowiedzi 1 i 2 naturalnych oraz syntetycznych.

## 5.2 test 1.

Analiza wpływu długości sylaby na akcentuację.

Zsyntezowano 3 realizacje wypowiedzi *mamama kawę* z różnymi długościami sylab *ma*. W realizacji 1 wypowiedzi *mamama kawę* - przyjęto odpowiednio na kolejnych sylabach *ma* wartości - 160ms, 80ms, 80ms. W realizacji 2 wypowiedzi *mamama kawę* przyjęto wartości 80ms, 160ms, 80ms, w realizacji 3 - *mamama kawę* - długości sylab wynosiły odpowiednio 80ms, 80ms, 160ms. Przyjęto stałą wartość parametru  $F_0$  w obrębie całej wypowiedzi równą 110 Hz. Słuchacze mieli za zadanie rozróżnić wersję a i b w wypowiedzi 1 oraz 2. Odpowiedzi słuchaczy podano w tabeli 2.



Tabela 2

Oceny (w%) rozpoznania przez słuchaczy wersji a i b w trzech realizacjach wypowiedzi *mamama kawę*.

	wersja a	wersja b
realizacja 1	52%	48%
realizacja 2	54%	46%
realizacja 3	45%	55%

Wyniki testu nie wykazują istotności zróżnicowania w rozpoznawaniu wersji a i b wypowiedzi. Około połowa słuchaczy określała tą samą wypowiedź jako posiadającą znaczenie (a) połowa jako posiadającą znaczenie (b). Długość sylaby, jako samodzielny parametr nie odgrywa więc w akcentuacji decydującej roli.

### 5.3. Test 2.

Ocena istotności przebiegu parametru F0 w rozpoznaniu akcentu.

Do odsłuchu podano wypowiedzi z intonacją naturalną i wypowiedzi ze stałą wartością parametru F0- równą 110 Hz. Dla przetestowania hipotezy o priorytecie intonacji w akcentowaniu, w zdaniu o znaczeniu (a) podłożono przebieg ze zdania (b), w zdaniu o znaczeniu (b), podłożono przebieg ze zdania a. Dla wypowiedzi 1 - *mamama kawę* otrzymano następujące wersje wypowiedzi:  
wypowiedź 1.

a.

- a1 - *mama ma kawę?* - intonacja naturalna
- a2 - *mama ma kawę?* stała wartość parametru Fo=110Hz
- a3 - *mama ma kawę?* - intonacja z wypowiedzi b1

b.

- b1- *ma mama kawę?* - intonacja naturalna
- b2- *ma mama kawę?* - stała wartość parametru Fo=110Hz
- b3 *ma mama kawę?* - intonacja z wypowiedzi a1

wypowiedź 2.

a.

- a1- będziemy jesć, bambino? - intonacja naturalna
- a2- będziemy jesć, bambino? - stała wartość parametru  $F_0=110\text{Hz}$
- a3- będziemy jesć, bambino? - intonacja z wypowiedzi b1

b.

- b1- będziemy jesć bambino? - intonacja naturalna
- b2- będziemy jesć bambino? - stała wartość parametru  $F_0=110\text{Hz}$
- b3- będziemy jesć bambino? - intonacja z wypowiedzi a1

W tabelach 3 i 4 przedstawiono (w procentach) rozpoznane przez słuchaczy wersje wypowiedzi 1 i 2.

Tabela 3.

Oceny (w %) rozpoznania przez słuchaczy wersji a i b wypowiedzi 1.

Wypowiedź	Wersja	
	a	b
a1	90	10
a2	61	39
a3	19	81
b1	29	71
b2	42	58
b3	66	34

Wypowiedzi z intonacją naturalną zostały rozpoznane w 90% (wersja a) i 71% (wersja b). Nie rozróżniono wersji a i b w wypowiedziach ze stałą wartością parametru  $F_0$  (61% i 42%). Wypowiedź (a) z intonacją z wersji (b) została rozpoznana w 81%, wypowiedź (b) z intonacją z wersji (a) rozpoznano poprawnie w 66%.

Tabela 4.

Oceny (w %) rozpoznania przez słuchaczy wersji (a) i (b) wypowiedzi 2.

wypowiedź	znaczenie	
	a	b
a1	91	9
a2	56	44
a3	23	77
b1	84	16
b2	60	40
b3	64	36

Wypowiedzi z intonacją naturalną (a1 i b1) zostały rozpoznane odpowiednio w 91% i 84%. Wypowiedzi zawierające stałą intonację zostały rozpoznane poprawnie w 56% i 60%. Wypowiedź z intonacją (b) rozpoznano w 77%, wypowiedź (b) z intonacją (a) w 68%. Można przypuszczać, że ten stosunkowo niski procent rozpoznania wynika z konieczności uwzględnienia cechy udźwięcznienia (w wersji (b) wypowiedzi) lub ubezdźwięcznienia (w wersji (a) wypowiedzi) zbitki spółgłoskowej *ść*.

#### 5.4. Test 3.

Wpływ intonacji wypowiedzi na percepcję sylaby akcentowanej.

##### Test3.1.

Założono, że w percepcji zmienności częstotliwości podstawowej na poszczególnych sylabach odgrywa istotną rolę intonacja całej wypowiedzi. Zsyntezowano 6 wypowiedzi *mama ma kawę* z intonacją na *mama ma* naturalną i przebiegiem parametru F0 na sylabach *kawę* zmienianym zgodnie z poniższym schematem.

1. Kontur  $F_0$  przesunięty o 20Hz.  $F_{min}$  na wyrazie *kawę* =  $F_{min}$  na wyrazie *mama* (90Hz). Wzrost parametru  $F_0$  na sylabach *kawę* w zakresie 90-280Hz.
2. Do wartości parametru  $F_0$  na sylabie *ka* dodano 50 Hz.  $F_{min}$  na sylabach *kawę* - równy  $F_{max}$  na sylabach *mama* (130 Hz). Wzrost na wyrazie *kawę* w zakresie 130-280 Hz.
3. Na sylabie *ka* przyjęto wartość  $F_0 = 130$  Hz, wzrost na *we* do 130Hz.
4. Na sylabie *ka* przyjęto wartość  $F_{min} = 70$ Hz, na *we* wzrost  $F_0$  do 130Hz.
5. Na sylabie *ka* przyjęto wartość  $F_{min} = 90$ Hz, na sylabie *we* - wzrost  $F_0$  do 130 Hz.
6. Na sylabie *ka* przyjęto wartość  $F_0 = 70$  Hz, na sylabie *we* -  $F_0 = 90$ Hz.

Na ryc. 5a zilustrowano podane przebiegi. Jako wypowiedzi o znaczeniu (a) zostały rozpoznane przez słuchaczy realizacje 1,2 i 6 (dla pierwszej realizacji otrzymano 79% poprawnego rozpoznania, dla drugiej 77%, dla szóstej 76%). W wypowiedziach tych (1,2,6), zauważa się obecność wyraźnego akcentu na wyrazie *kawę* - intonacja pełna rosnąca - (trzykrotny wzrost wartości parametru  $F_0$ ). Znaczna zmiana parametru  $F_0$  na końcu zdania decyduje o percepcji poszczególnych akcentów. Słuchacz percypuje globalnie intonację, mniejszą uwagę zwraca na lokalne akcenty. Małoprawdopodobny byłby wybór *ma mama kawę* (zaakcentownie dwóch kolejnych wyrazów).

### Test 3.2.

Analogiczne doświadczenie przeprowadzono dla wypowiedzi 1b. - *ma mama kawę*. Dla przebiegu parametru  $F_0$  stosowano różne wzrosty (zilustrowane na ryc.5b). Wypowiedzi z dużym wzrostem parametru  $F_0$  (230 Hz) na sylabach *kawę* zostały niejednoznacznie rozpoznane. Dobrze (tzn. powyżej 70%) zostały rozpoznane te, na których wzrost na sylabach *kawę* nie był większy niż 180Hz. Mamy tutaj do czynienia z sytuacją odwrotną niż w poprzednim przypadku. Słuchacze rozpoznali poprawnie wersję (b) w tych wypowiedziach, w których na wyrazie *kawę* stosowano intonację rosnącą a nie pełną rosnącą.

## 5.5 Test 4.

Zmiana miejsca akcentu.

### 1. Zmiana parametru F0 na pierwszych dwóch sylabach

ma mama kawę?

Analizowano możliwość uwydatnienia pierwszych dwóch sylab przez podwyższenie częstotliwości podstawowej (2 ekstrema przebiegu parametru F<sub>0</sub>, na kolejnych dwóch sylabach, z dynamiką zmiany parametru F<sub>0</sub> na sylabie 50Hz). Słuchacze nie rozróżniali wersji (a) i (b) wypowiedzi. Otrzymano niejednoznaczne wyniki. Przy obniżeniu częstotliwości na drugiej sylabie (względem sylaby pierwszej o ok. 50 Hz) rozpoznawano wersję (a) wypowiedzi.

### 2. Zmiana parametru F0 na pierwszej i trzeciej sylabie

mama ma kawę?

Na rys. 6 zilustrowano przebieg parametru F<sub>0</sub> z uwydatnieniem 1 i 3 sylaby. Przy dynamice zmiany 40Hz (zmniejszenie o 40Hz wartości parametru F<sub>0</sub> na drugiej sylabie względem sylaby pierwszej i trzeciej), 85% słuchaczy rozpoznało znaczenie (a) wypowiedzi. Przy mniejszej dynamice zmiany otrzymano rozpoznanie niejednoznaczne (mniej niż 60% poprawnych odpowiedzi). Na rys. 7 zilustrowano inną możliwość uwydatnienia pierwszej i trzeciej sylaby. Na dwóch pierwszych sylabach *mama* przyjęto następujące zmiany parametru F<sub>0</sub>:

- na dwóch pierwszych sylabach *mama* spadek F<sub>0</sub> (130Hz-100Hz), na trzeciej sylabie *ma* wzrost F<sub>0</sub> (100Hz-150Hz);
- analogicznie jak w punkcie (a): spadek F<sub>0</sub> (120Hz -100Hz) i wzrost F<sub>0</sub> (100Hz-150 Hz) oraz
- wartość stałą parametru F<sub>0</sub> na dwóch pierwszych sylabach (100-100) Hz i wzrost F<sub>0</sub> (100Hz-150Hz).

Wszystkie 3 wersje wypowiedzi a,b,c zostały rozpoznane jako wypowiedź *mama ma kawę*.

Nie rozpoznano poprawnie znaczenia wypowiedzi (mniej niż 65%) w

przypadku zastosowania: a) spadku  $F_0$  na trzech sylabach *mamama* (spadek zmieniano w zakresie 140-100Hz) oraz b) stosowania równego spadku na dwóch pierwszych sylabach *mama* (150-100Hz) i równego wzrostu na trzeciej sylabie *ma* (100-150Hz).

### 3. Zmiany parametru $F_0$ na na pierwszej sylabie

a) ma mama kawę?

b) mama ma kawę?

Przetestowano różne możliwości akcentuacji pierwszej sylaby. Na rys.8 zilustrowano najprostszy przebieg parametru  $F_0$  decydujący o akcencie na 1 sylabie. Na pierwszej sylabie zastosowano wzrost  $F_0$  w zakresie 100-160 Hz, na drugiej sylabie założono raptowny spadek wartości parametru  $F_0$  (60 Hz w 20 ms). Słuchacze w 85% określili tę wypowiedź jako wersję (b). Na rys. 10,11,12,13,14 zilustrowano przykłady modelowania akcentu dla pierwszej sylaby. Na ryc.10 na pierwszej sylabie *ma* przyjęto wartości parametru  $F_0$  w zakresie 80-100Hz, na drugiej sylabie *ma* założono wartość  $F_0=180$ Hz, na trzeciej sylabie przyjęto spadek wartości parametru w zakresie 180-100Hz. Na ryc.11 przedstawiono analogiczny przebieg, ale z większą dynamiką zmian. Ryc. 12 i 13 ilustrują przebieg częstotliwości podstawowej z wolniejszymi niż w przypadkach zilustrowanych na ryc.10 i 11 zmianami parametru  $F_0$ . Wersje 10 i 11 zostały ocenione przez słuchaczy (68-81%) jako posiadające znaczenie (b), wersje 12 i 13 jako znaczenie (a). Na ryc.14,14a,14b zilustrowano przebiegi parametru  $F_0$  różniące się dynamiką zmian na pierwszej sylabie *ma*. Wypowiedzi zawierające te przebiegi rozpoznano (w71%) jako wersję (a) wypowiedzi.

### 4. Zmiany parametru $F_0$ na drugiej sylabie

*ma* mama kawę.

Na drugiej sylabie *ma* zastosowano szybki wzrost parametru (100Hz-160Hz) i taki sam symetryczny szybki spadek. Ryc.9 ilustruje syntezowany przebieg. Większość słuchaczy (76%) rozpoznała wypowiedź zawierającą ten przebieg jako wersję (b).

#### 5.6. Test 5.

Analiza szybkości zmiany parametru  $F_0$ .

Na ryc.15 podano przykład zmiany szybkości spadku parametru  $F_0$  na trzeciej sylabie *ma*. Na dwóch pierwszych sylabach *mama* zastosowano wzrost parametru  $F_0$  w zakresie 100-160 Hz. Na trzeciej sylabie *ma* zastosowano trzy możliwości spadku wartości parametru 1) bardzo szybki spadek wartości parametru  $F_0$  (60Hz w ciągu 20ms), 2) wolniejszy spadek (60Hz w ciągu 50ms), 3) wolny spadek wartości (60Hz w ciągu 80ms). Wypowiedzi zawierające przebieg parametru  $F_0$  z szybkim spadkiem wartości parametru na trzeciej sylabie, sugerowały słuchaczom (w 73%) wersję (b) wypowiedzi. Wypowiedzi z wolnym spadkiem wartości nie rozpoznano jednoznacznie. Podobne zjawisko przesuwania akcentu opisano dla jęz. angielskiego (por.np.10). Zwrócić należy tutaj uwagę na obecność intonacji rosnącej na wyrazie *kawe* na końcu wypowiedzi.

#### 5.7. Test 6.

Analiza akcentu realnego w wypowiedzi *będziemy jeść bambino*.

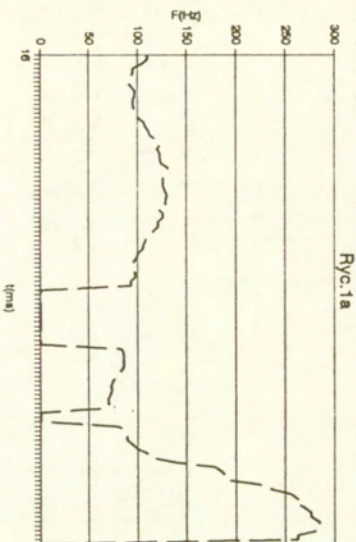
W celu ustalenia wpływu miejsca akcentu na znaczenie (a) lub (b) wypowiedzi, przeprowadzono analogiczne jak w poprzednim przypadku testy. Dodatkowo, zastosowano aproksymację prostoliniową zmian parametru na sylabach *będziemy*, *jeść*, *bambino*. Stosowano zmianę parametru  $F_0$  w zakresie 90Hz-180Hz. Ryc.15a i 15b ilustrują syntezywane przebiegi. Analizowano wpływ przesuwania konturu parametru  $F_0$  na sylabach *bambino* (ryc.16a i 16b). Istotną rolę w zmianie znaczenia wypowiedzi odgrywa miejsce wystąpienia minimum wartości parametru  $F_0$  (na sylabach *jeść* oraz *bambi*). Dla dobrego rozpoznania (powyżej 85%) znaczenia (b) wypowiedzi, wymagane jest na sylabie *bi* w wyrazie *bambino* lokalne minimum (około 15Hz-20Hz). Aproksymacja prostoliniowa wypowiedzi nie powoduje trudności percepcyjnych w lokalizacji akcentu realnego.

## VI. Wnioski.

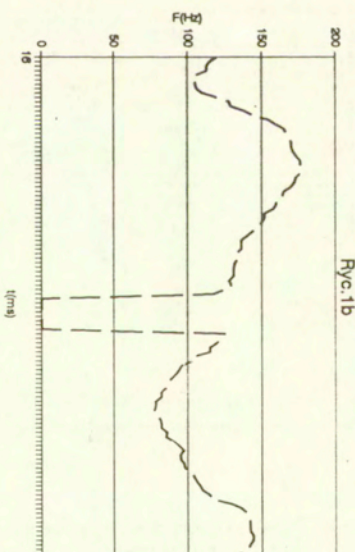
1. Istotna dla akcentuacji w języku polskim jest zmiana częstotliwości podstawowej, zmiany iloczasu i amplitudy nie odgrywają pierwszorzędnej roli.
2. Poprzez zmianę przebiegu częstotliwości podstawowej uzyskuje się zmianę miejsca akcentu realnego - można uzyskać zmianę znaczenia wypowiedzi.
4. Istotny dla percepcji akcentu na poszczególnych sylabach jest przebieg parametru  $F_0$  na całej wypowiedzi. Słuchacz rozróżnia intonację pełną rosnącą i niską rosnącą.
5. Dla percepcji akcentu ważna jest szybkość zmiany, dynamika oraz miejsce wystąpienia zmiany wartości parametru  $F_0$ .
6. Wyniki klasyfikacji percepcyjnej pozwalają tworzyć podstawę do klasyfikacji automatycznej.



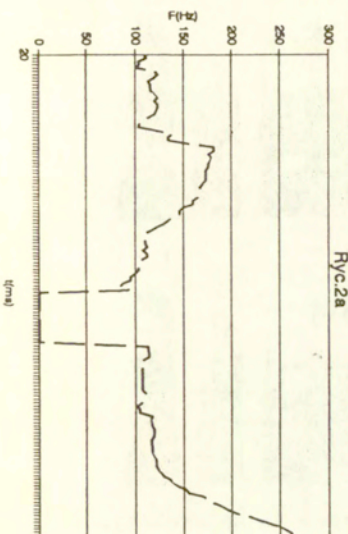
### Przebieg parametru F0



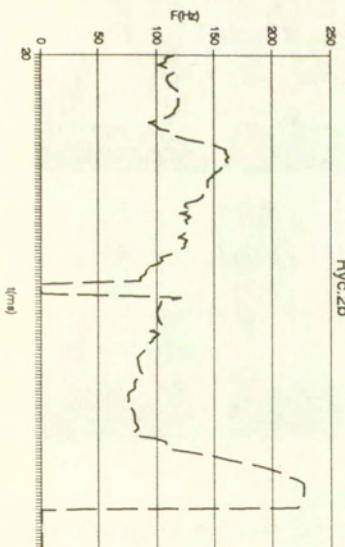
### Przebieg parametru F0

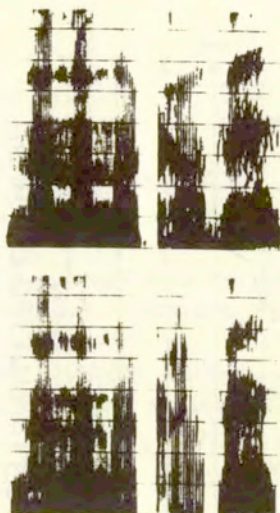


### Przebieg parametru F0

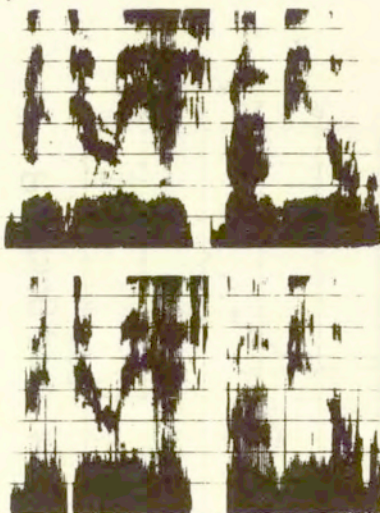


### Przebieg parametru F0





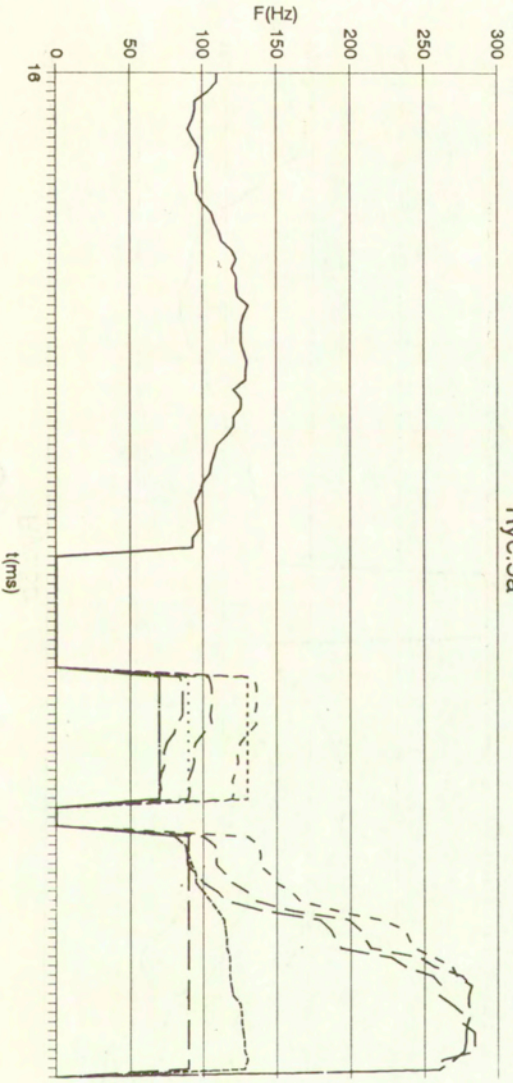
Rys. 3. Spektrogram wypowiedzi *mama ma kawę* a) naturalnej, b) syntetycznej



Rys 4. Spektrogram wypowiedzi *będziemy jeść bambino* a) naturalnej b) syntetycznej

# Przebieg parametru F0

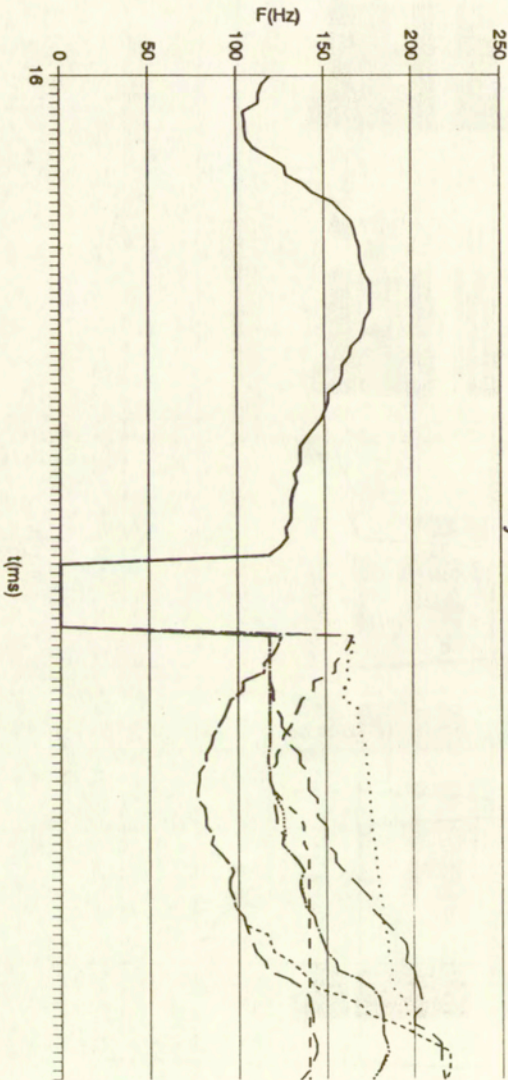
Ryc.5a



Przebieg parametru F0

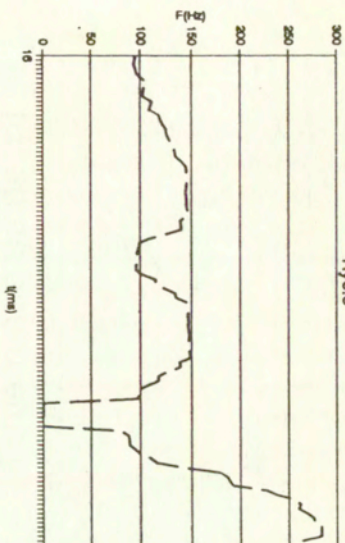
# Przebieg parametru F0

Ryc.5b



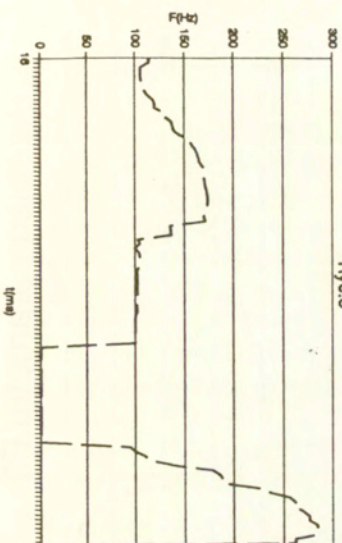
### Przebieg parametru F0

Ryc. 6



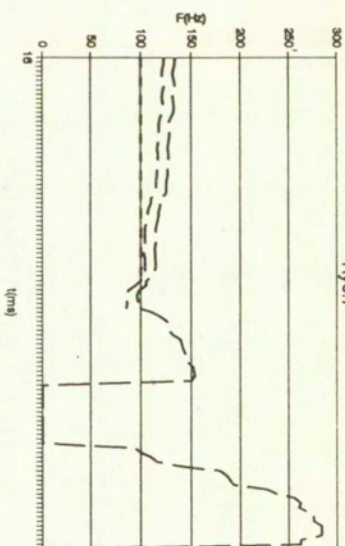
### Przebieg parametru F0

Ryc. 8



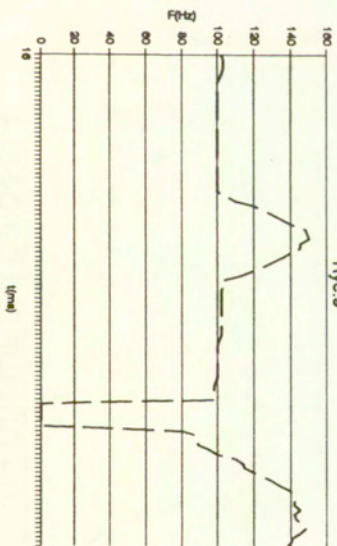
### Przebieg parametru F0

Ryc. 7



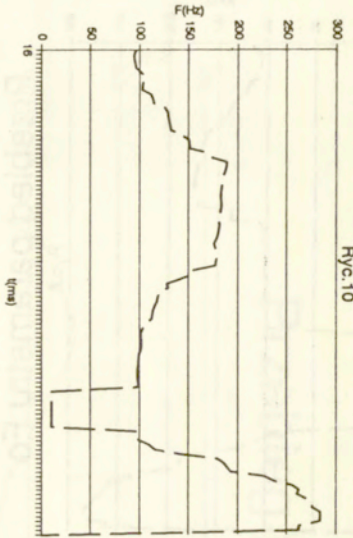
### Przebieg parametru F0

Ryc. 9



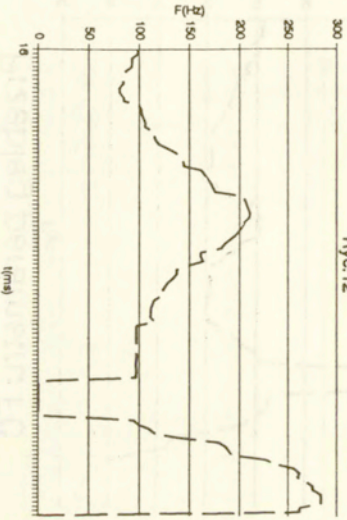
### Przebieg parametru F0

Ryc.10



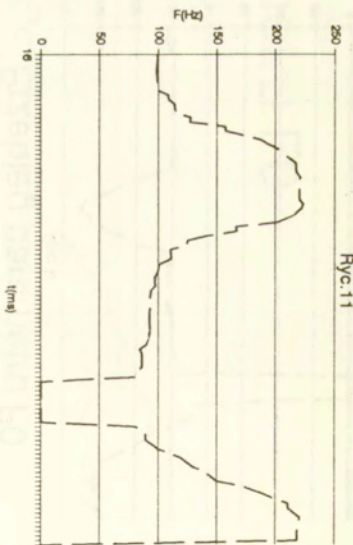
### Przebieg parametru F)

Ryc.12



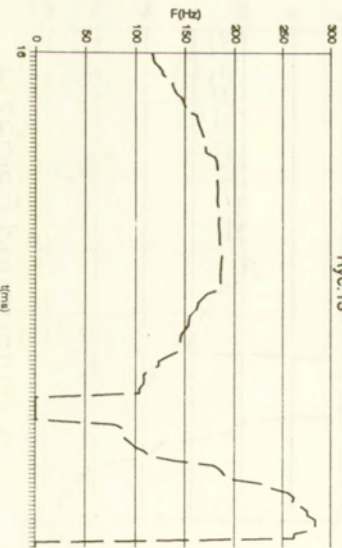
### Przebieg parametru F0

Ryc.11



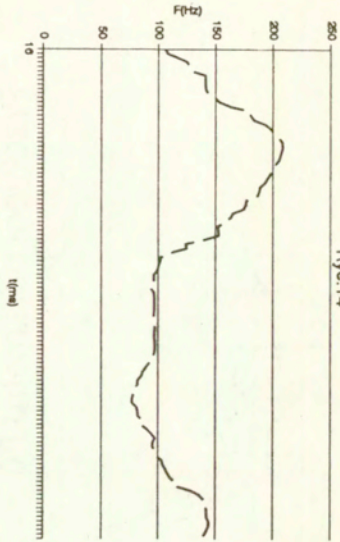
### Przebieg parametru F0

Ryc.13



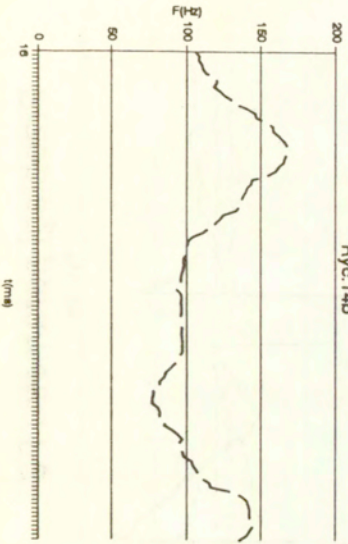
### Przebieg parametru F0

Ryc. 14



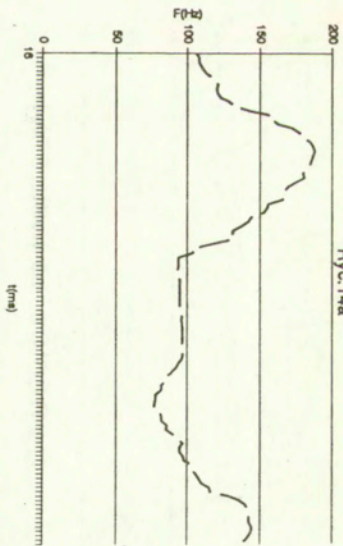
### Przebieg parametru F0

Ryc. 14b



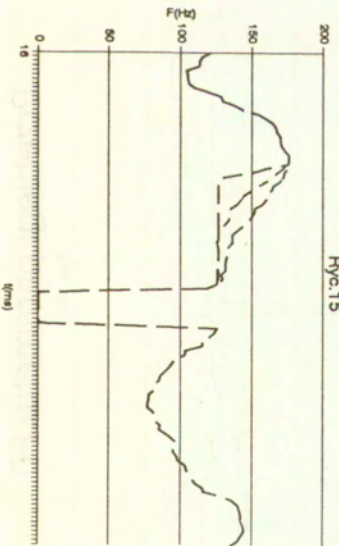
### Przebieg parametru F0

Ryc. 14a

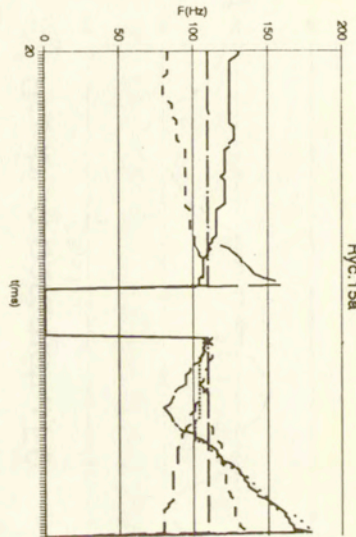


### Przebieg parametru F0

Ryc. 15

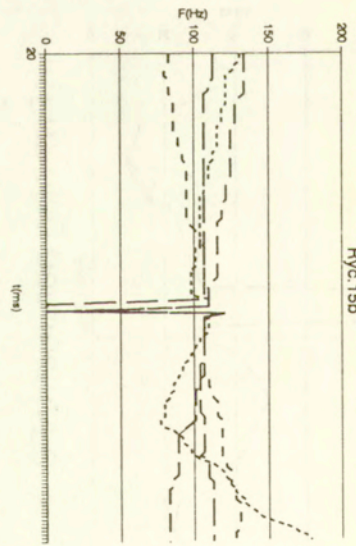


### Przebiegi parametru F0



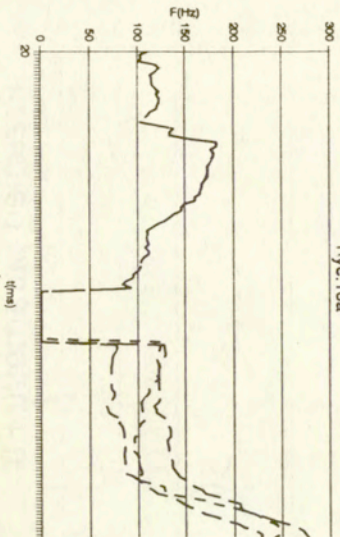
Ryc. 15a

### Przebiegi parametru F0



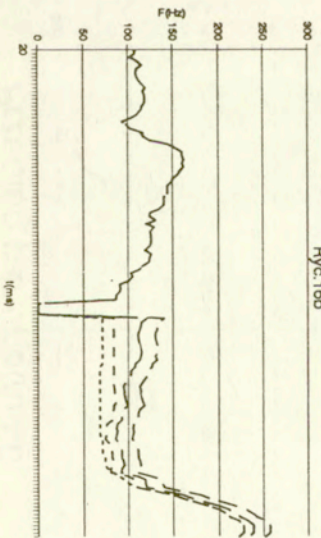
Ryc. 15b

### Przebiegi parametru F0



Ryc. 16a

### Przebiegi parametru F0



Ryc. 16b



Załącznik

#	M	dB	F0	F1	B1	F2	B2	F3	B3	F4	B4
10	0	35	0	351	90	1321	284	1724	260	2581	288
11	0	46	110	220	90	1397	188	1909	100	2695	200
12	0	51	110	220	66	1335	162	1800	100	2597	100
13	0	54	104	228	166	1396	106	1985	124	2702	78
14	0	55	95	220	166	1340	140	1953	206	2875	192
15	0	55	95	319	134	1525	120	1993	542	2702	58
16	0	55	93	330	142	1449	244	1884	364	2770	218
17	0	55	90	292	124	1351	230	1821	198	2727	96
18	0	56	93	245	122	1452	194	1805	472	2747	104
19	0	58	97	243	82	1299	76	1894	174	2033	100
20	0	60	97	294	100	1300	62	1347	52	2100	100
21	0	60	95	582	46	1357	52	1687	210	2200	66
22	0	60	95	590	54	1200	44	1764	254	2250	96
23	0	61	96	500	90	1350	80	1397	48	2300	168
24	0	61	96	580	86	1100	100	1392	62	2300	252
25	0	61	100	574	78	1300	100	1387	92	2250	300
26	0	61	106	580	80	1350	100	1360	92	2250	270
27	0	61	108	582	78	1300	90	1335	94	2400	244
28	0	61	111	478	100	1100	62	1277	62	2600	198
29	0	60	113	500	150	1000	56	1267	98	2450	232

Tabela 1. Wydruk przykładowych parametrów umieszczonych w edytorze LPC. Kolejno w poszczególnych kolumnach podano wartość poziomu sygnału, parametru F0, częstotliwości i szerokości pasm 4 pierwszych formantów.

## BIBLIGRAFIA

1. Demenko G., Nowak J., Imiołczyk J.: Analysis and synthesis of Pitch Movements in a read Polish Text, Eurospeech Proceedings, 1993, Vol.2, ss.793-797.
2. Hasegawa, Y., Kazuehata : Fundamental frequency as an acoustic cue to accent perception , Language and Speech, vol.31 part 1-2, 1992.
3. Imiołczyk J., Nowak I., Demenko G. : High-Intelligibility Text-to-speech Synthesis for Polish, 1994, Arch. of Acoustics, vol.19, No2, pp.161-172.
4. Imiołczyk J., Nowak I., Demenko G.: A Text-to Speech System for Polish Eurospeech Proceedings, 1993, vol.2, pp.885-889.
5. Jassem W.: Akcent języka polskiego, wyd.PAN, Wrocław 1962.
6. Jassem W., Demenko G.: On Extracting Linguistic Information from Fo traces, w Intonation in Discourse (c. Johns-Lewis ed), Croom Helm, London 1-18, 1986.
7. Jassem W., Dommelen W.: Perception of Polish Accent In A Re-synthesized Speech Signal, Archives Of Acoustics 15, 3-4, 325-348 (1990).
8. Jassem W., Morton J., Steffen-Batóg M.: The perception of stress in syntfetic speech-like stimuli by Polish listeners, Speech Analysis and Synthesis, vol.1, 1968.
9. Kohler K.J.: Prosody in speech synthesis: the interplay between basic research and TTS application, Journal of Phonetics, 19, 1991. 121-138.
10. Mobius B.: Ein quantitatives Modell der deutschen Intonation, Niemeyer, Tübingen, 1993.
11. Nowakowska W.: Rola częstotliwosci podstawowej i poziomu intensywnosci w percepcji akcentu w mowie polskiej, Prace IPPT 74/1977, Warszawa 1977.
12. Steffen-Batogowa M.: Analiza struktury przebiegu melodii polskiego języka ogólnego, rozprawa doktorska, Poznań 1963.