

**Grażyna Demenko**

**BAZA DANYCH TEKSTOWYCH  
DLA ANALIZY INTONACJI  
JĘZYKA POLSKIEGO**

**20/1995**

P. 269



**WARSZAWA 1995**

ISSN 0208-5658

Praca wpłynęła do Redakcji dnia 21 grudnia 1993 r.

56589



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>

Grażyna Demenko  
Zakład Fonetyki Akustycznej  
IPPT PAN

## BAZA DANYCH TEKSTOWYCH DLA ANALIZY INTONACJI JĘZYKA POLSKIEGO

### Streszczenie

Dla analizy intonacji w mowie ciągłej przygotowano materiał lingwistyczny składający się ze zróżnicowanych semantycznie, strukturalnie oraz gramatycznie fragmentów wiadomości telewizyjnych oraz reportaży. Do szczegółowego opisu przebiegów melodycznych w zdaniach jednofrazowych opracowano dwa zestawy zróżnicowanych strukturalnie tekstów.

Przygotowany materiał lingwistyczny (o długości około 10 min. mowy ciągłej) został trzykrotnie odczytany w studio nagrań przez dwóch profesjonalnych spikerów telewizyjnych, jednego fonetyka oraz jedną osobę bez wykształcenia fonetycznego. Przeprowadzono ekstrakcję częstotliwości podstawowej przy użyciu własnych programów pomiaru parametru  $F_0$  (opracowanych w ZFA) oraz spektrografu cyfrowego Kay 5500. Analizę spektrograficzną wykorzystano głównie w celu segmentacji materiału słownego na sylaby. Przygotowano zbiór danych cyfrowych, ilustrujących zmienność przebiegu parametru  $F_0$  na poszczególnych sylabach analizowanych wypowiedzi. Zbiór ten stanowić będzie podstawę dla klasyfikacji wzorców intonacyjnych w języku polskim.

### 1. Wstęp

Problem modelowania intonacji w mowie ciągłej nie jest w pełni rozwiązany. Opisy zmienności melodycznej nawet dla jednego języka różnią się znacznie między sobą i nie ma zgodności, który z nich jest właściwy. Zagadnienie sterowania częstotliwością podstawową (reprezentującą fizycznie intonację) w syntezie, wymaga rozwiązania złożonych problemów wynikających z konieczności:

- a) uwzględnienia wielu równoległe funkcjonujących źródeł zmienności,
- b) oceny subiektywnej (percepcyjnej) i obiektywnej (fizycznej analizy) intonacji,
- c) segmentacji sygnału na prozodyczne jednostki,
- d) klasyfikacji intonacyjnych wzorców.

Opracowanie reguł syntezy intonacji wymaga analizy sygnału na poziomie leksykalnym, syntaktycznym i semantycznym. W ostatnio opracowywanych opisach prozodii coraz większą uwagę zwraca się na odpowiednie przygotowanie bazy danych. Przeprowadzone porównania różnych metod klasyfikacyjnych wykazały możliwość poprawy wyników rozpoznawania wybranych jednostek w granicach 5-10% poprzez zmianę metody klasyfikacyjnej. Przygotowaniu reprezentatywnego materiału oraz wyborowi adekwatnych cech charakteryzujących poszczególne obiekty poświęcono jak dotąd mało uwagi.

Na zjawisko to zwrócili między innymi uwagę Fant, Kohler (por. [4]) - "speech technology has relied heavily on linguistic redundancies to ensure an acceptable performance of synthesis as well as recognition", brak jest natomiast - "fundamental knowledge by large documentary projects around our data banks".

Dla języka polskiego również więcej uwagi poświęcano metodom analizy instrumentalnej intonacji niż zagadnieniom związanym z reprezentatywnością i uniwersalnością modelowania. Z tego powodu dotychczasowe opracowania opisu częstotliwości podstawowej mowy polskiej (np. [2], [3]), obejmujące swym zakresem głównie wypowiedzi izolowane (wyrazy, krótkie zdania) dostarczają tylko fragmentarycznych informacji o cechach melodycznych języka. Najobszerniejszy opis struktur intonacyjnych języka polskiego zawiera praca z lat 60-tych M. Steffen-Batogowej (por. [1]).

Opracowana w ostatnich latach w ZFA synteza wymaga reguł sterowania intonacją w mowie ciągłej (por. [5], [6]). Niniejsze opracowanie poświęcono przygotowaniu bazy danych, stanowiącej podstawę dla utworzenia opisu intonacji w mowie ciągłej. Wstępnie, uwzględniono wpływ zmienności lingwistycznej oraz osobniczej na przebiegi parametru  $F_0$ . W następnym etapie pracy

przewidziana jest klasyfikacja typowych zmian intonacyjnych, między innymi w celu optymalizacji reguł syntezy mowy polskiej.

## 2. Materiał lingwistyczny

Dobór reprezentatywnego materiału słownego ilustrującego melodyczne struktury danego języka jest jednym z istotniejszych fragmentów badań prozodii mowy. Z uwagi na matematyczno - statystyczną analizę zbioru danych, dogodny jest taki dobór materiału lingwistycznego, który umożliwia kontrolę funkcjonowania poszczególnych źródeł zmienności sygnału.

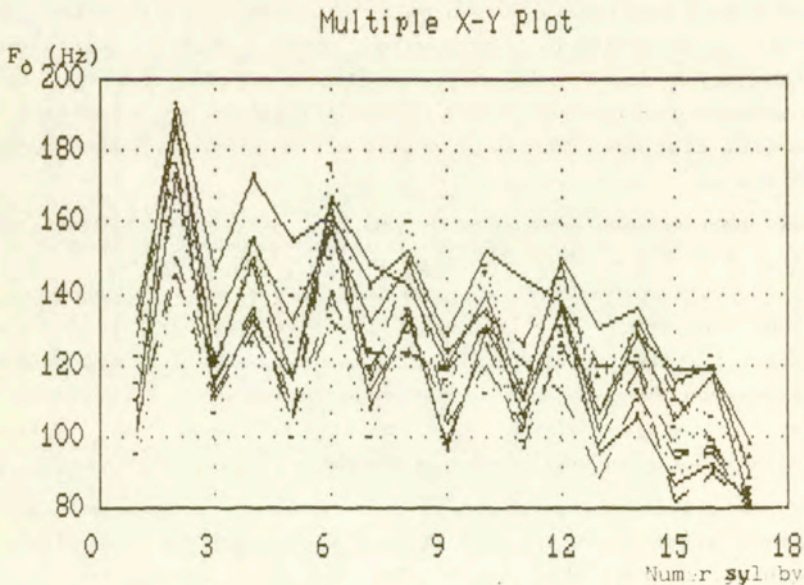
Systematyczne uwzględnianie wpływu takich czynników jak np.: realizacji tematycznej, leksykalnej, akcentu, struktury gramatycznej i fonetycznej frazy jest możliwe tylko w przypadku sztucznie ułożonych tekstów. Przygotowanie materiału o złożonej, sztucznej strukturze, szczególnie w przypadku analizy cech prozodycznych, negatywnie wpływa na wyniki badań. Dla modelowania struktur prozodycznych, optymalna jest analiza tekstów odpowiadających mowie potocznej, codziennej. Można założyć, że przygotowanie dostatecznie dużej liczby fragmentów np. reportaży, wiadomości, dialogów umożliwi opracowanie reprezentatywnej bazy danych dla analizy intonacji.

Na obecnym etapie badań przyjęto, że realizacja akustyczna wybranego materiału lingwistycznego powinna być wzorcowa. Z tego powodu, do czytania tekstów wybrano osoby piynnie posługujące się językiem polskim. Do analizy wyselekcjonowano fragmenty wiadomości telewizyjnych oraz dwa zestawy sztucznie przygotowanych dwunastu zdań o założonej strukturze leksykalnej. Zestawy zawierają zdania o różnych konfiguracjach sylab akcentowanych i nieakcentowanych o długości od 3 do 35 sylab.

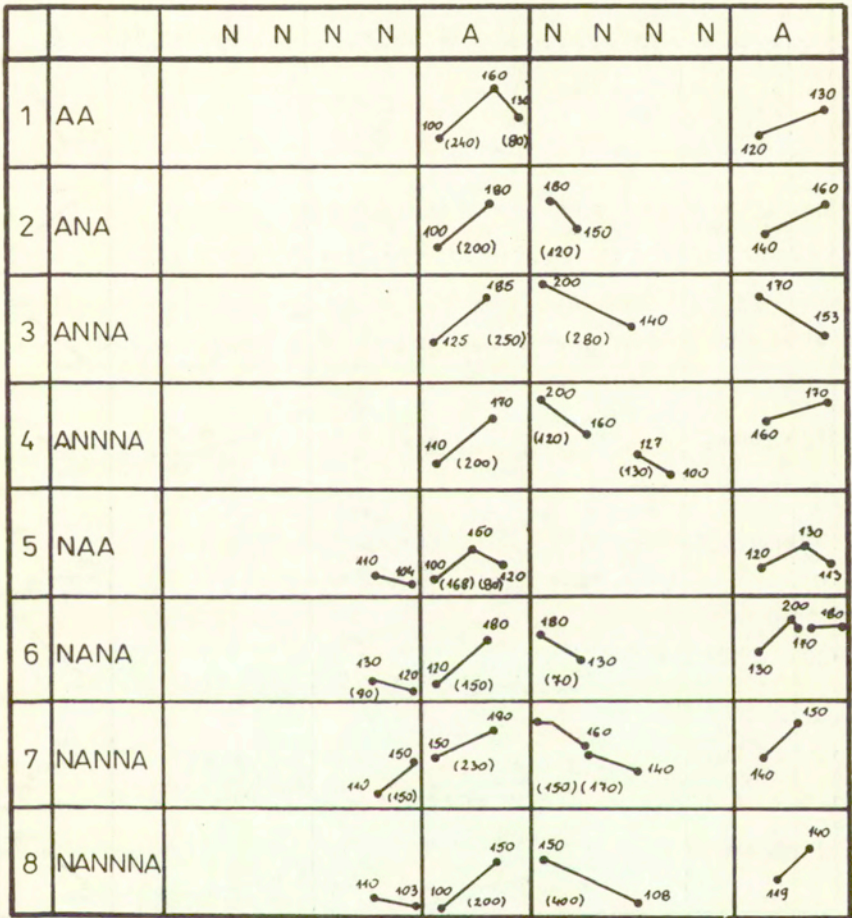
W materiale lingwistycznym w zestawie pierwszym (T1) samogłoski akcentowane znajdują się w sąsiedztwie spółgłosek dźwięcznych, w zestawie drugim (T2) - w otoczeniu spółgłosek bezdźwięcznych. W załączniku 1 podano struktury ułożonych zdań. Do szczegółowej analizy przyjęto materiał składający się z wypowiedzi 3 głosów męskich.

### 3. Analiza częstotliwości podstawowej w zdaniach jednofrazowych

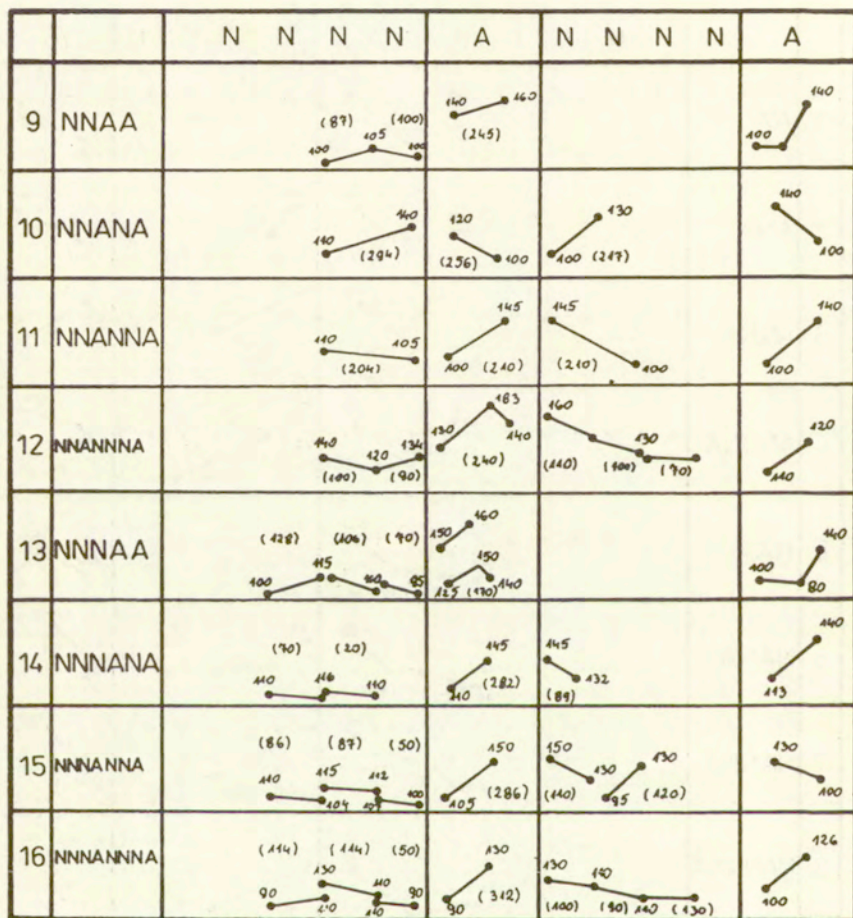
Przebiegi częstotliwości podstawowej w zdaniach jednofrazowych wykazały znaczne podobieństwa w replikacjach wypowiedzi przez poszczególne osoby, jak i w wypowiedziach różnych mówców. Rys.1 przedstawia przebiegi parametru  $F_0$  w replikacjach jednej z analizowanych wypowiedzi. Kontury częstotliwości podstawowej opisane zostały przez kolejne ekstrema: wartości maksymalne na sylabach akcentowanych i minimalne pomiędzy nimi. Przeprowadzona analiza korelacji (wartość współczynnika korelacji zmieniała się w granicach 0,87 - 0,98 - por. załącznik 2), wykazała istotne statystyczne podobieństwa przebiegów parametru  $F_0$  w analizowanych wypowiedziach.



Rys.1. Przebiegi częstotliwości podstawowej w 9 replikacjach wypowiedzi: "Londyn ma zostać jedyną stolicą europejską nie posiadającą ogrodu zoologicznego" (otrzymane dla 3 mówców).



Rys. 2. a. Przebiegi parametru  $F_0$  na początkach wypowiedzi w 16-stu różnych strukturach sylab akcentowanych i nieakcentowanych. Literą A oznaczono pozycję sylaby akcentowanej, literą N - nieakcentowanej



Rys.2.b. Przebiegi parametru  $F_0$  na początkach wypowiedzi. Na początkach przebiegów podano wartości parametru  $F_0$  (w Hz), w nawiasach czas trwania zmiany częstotliwości podstawowej (w ms)



Szczegółowej analizie poddano zmiany częstotliwości podstawowej na początkach wypowiedzi. Rys.2a i 2b ilustrują uśrednione przebiegi parametru  $F_0$  na pierwszych sylabach zdań jednofrazowych. Zależnie od struktury początku wypowiedzi obserwuje się różne zmiany intonacyjne na kolejnych sylabach.

Największy zakres zmienności częstotliwości podstawowej (ok. 100 Hz w głosie męskim) obserwowano w zdaniach, które rozpoczynały się od sylaby akcentowanej. Jeśli sylaba akcentowana rozpoczynała się od spółgłoski dźwięcznej, przebieg parametru  $F_0$  miał charakter rosnący z maksimum przypadającym na następną spółgłoskę. W przypadku spółgłoski bezdźwięcznej na początku sylaby akcentowanej otrzymywano najczęściej przebiegi parametru  $F_0$  równe lub lekko rosnące z wysoką wartością początkową (170-180 Hz). W zdaniach rozpoczynających się od sylab nieakcentowanych, obserwowano na początku wypowiedzi niewielką zmienność parametru  $F_0$  (w granicach 10-15 Hz), a na pierwszej sylabie akcentowanej, zakres zmienności częstotliwości podstawowej (ok. 60 Hz), znacznie mniejszy niż w wypowiedziach zaczynających się od sylaby akcentowanej.

Przebieg parametru  $F_0$  na drugiej sylabie akcentowanej, uzależniony jest od struktury i ilości poprzedzających sylab nieakcentowanych.

W tabelach 1 oraz 2 podano przykładowo znormalizowane wartości parametru  $F_0$  w wypowiedziach zawierających 3 sylaby akcentowane. Normalizację danych przeprowadzono zgodnie z zależnością

$$\ln F_{ni} = \ln F_i - \ln F_{min}$$

gdzie:  $F_{ni}$  - wartość znormalizowana,

$F_i$  - kolejna wartość normalizowana,

$F_{min}$  - wartość minimalna parametru  $F_0$ .

Przyjęto, że wartość  $F_{min}$  jest względnie stała dla poszczególnych mówców (por. np. [7]) i może stanowić punkt odniesienia dla porównywania przebiegów częstotliwości podstawowej w wypowiedziach tej samej osoby. Wartość  $F_{min}$  obliczono jako średnią arytmetyczną z minimalnych wartości tego

TABELA 1

Znormalizowane wartości parametru Fo na 3 kolejnych leksykalnie akcentowanych sylabach A1, A2 oraz A3 dla 2 mówców, trzech replikacji oraz tekstów T1 i T2. Zaznaczono wartości początkowe i końcowe parametru Fo na sylabach. Dodatkowo, strzałkami oznaczono wzrost lub spadek wartości. Przykładowo, zapis dla mówcy nr 2, w tekście T1, replikacji 1 dla pierwszej sylaby akcentowanej A1 należy odczytać jako wzrost wartości parametru Fo (0,08 do wartości 0,21), a następnie spadek (od wartości 0,21 do wartości 0,07).

Tekst	Mówca	Numer replikacji	Pozycja sylaby		
			A1	A2	A3
T1	1	1	0,33 0,07	0,01	0,01 0
		2	0,27 0,32 0,32 0,09	0,04 0,02	-
		3	0,33 0,07	0	-
	2	1	0,08 0,21 0,21 0,07	0,03	0,02
		2	0,22 0,11	0,05 0,02	0,01
		3	0,1 0,23 0,23 0,05	0,05 0,01	0
T2	1	1	0,32 0,37	0,15 0,1	0,1 0,5
		2	0,34 0,36	0,17 0,05	0,02
		3	0,32 0,28	0,15	0,01
	2	1	0,26 0,36	0,02	-
		2	0,28 0,32	0,17	0,1 0,02
		3	0,34 0,36	0,06	0,02

TABELA 2

Wartości parametru Fo w 3 replikacjach, dwóch osób w wypowiedzi zawierającej 3 sylaby akcentowane i 2 nieakcentowane

Tekst	Mówca	Numer replikacji	P o z y c j a   s y l a b y				
			A	N	A	N	A
T1	1	1	0,02 0,29	0,26 0,05	0,16 0,04	0,04 0,01	0,03
		2	0,05 0,34	0,28 0,05	0,13 0,08	0,08 0,01	-
		3	0,21 0,31	0,30 0,13	0,13 0,01	0,04	-
	2	1	0,03 0,17	0,15	0,15 0,03	0,02	0,12 0,01
		2	0,02 0,19	0,15 0,11	0,19 0,15	0,13 0,23	0,18 0,01
		3	0,03 0,21	0,02 0,1	0,11 0,06	0,06	-
T2	1	1	0,32 0,36	0,26	0,09	0,09	0,03 0
		2	0,39	0,26 0,18	0,09 0,07	0,07 0,01	0,01 0
		3	0,31 0,36	0,29 0,24	0,09 0,08	0,06	0,11 0,01
	2	1	0,19 0,26	0,01	0,15	0,08	0,01
		2	0,19 0,28	0,2 0,13	0,19 0,15	0,16 0,09	0,01
		3	0,23 0,3	0,14	0,14 0,05	0,05	-

parametru dla danego mówcy. W wypowiedziach rozpoczynających się od spółgłoski dźwięcznej (zestaw zdań T1) obserwowano na pierwszej sylabie akcentowanej przebiegi rosnące lub rosnąco opadające, z maksimum przypadającym na następującą po samogłosce akcentowanej spółgłoskę.

Zakres zmian parametru  $F_0$  mieści się w granicach 0,02-0,33, wartość początkowa parametru jest niska. W wypowiedziach (z zestawu T2) rozpoczynających się od spółgłoski bezdźwięcznej przebieg parametru  $F_0$  ma charakter lekko rosnący - lub równy (zakres zmian: 0,28 - 0,37) z wysoką wartością początkową. Na dwóch pozostałych sylabach akcentowanych, przebiegi parametru  $F_0$  są opadające lub równe i mają niewielki zakres zmienności. Przebieg na sylabach nieakcentowanych jest kontynuacją przebiegu parametru  $F_0$  na poprzedzającej sylabie akcentowanej. W niektórych przypadkach, na końcu wypowiedzi niemożliwy był pomiar parametru  $F_0$  z uwagi na zjawisko laryngalizacji - występowania niskich, nieregularnych drgań.

W tabeli 3 podano wartości parametru  $F_0$  w zdaniach o złożonej strukturze. Na sylabach akcentowanych zauważa się podobieństwo przebiegów częstotliwości podstawowej w poszczególnych replikacjach wypowiedzi. Na sylabach nieakcentowanych obserwuje się różne realizacje przebiegów parametru  $F_0$ , nawet dla jednego mówcy.

Tabela 4 ilustruje podobieństwa i różnice zmienności częstotliwości podstawowej w 3 replikacjach 26 sylabowej wypowiedzi tego samego mówcy. Stosunkowo konsekwentnie realizowane są przebiegi parametru  $F_0$  na początkach i końcach wypowiedzi. W środku wypowiedzi zauważa się zróżnicowanie w poszczególnych replikacjach.

W tabeli 5 podano średnie znormalizowane wartości parametru  $F_0$  w charakterystycznych punktach zdań. W wypowiedziach rozpoczynających się od sylaby akcentowanej, zawierającej na początku spółgłoskę bezdźwięczną (zdanie 1,5,6,7,8,10,12 w tekście T2) przebiegi częstotliwości podstawowej rozpoczynają się od wysokiej wartości (0,24-0,5). Jeżeli pierwszą samogłoskę akcentowaną poprzedza spółgłoska dźwięczna, to wartości parametru  $F_0$  na początku jest niska (0,06-0,15). Wartości maksymalne częstotliwości podstawowej w zdaniach krótkich występowały na pierwszej akcentowanej sylabie. W zdaniach długich (zdanie nr 9,10) wartości maksymalne parametru  $F_0$  występowały w dalszych fragmentach wypowiedzi. Realizacja przebiegów częstotliwości

podstawowej na końcach wypowiedzi trudna jest do wytlumaczenia wyłącznie przy uwzględnieniu struktury wypowiedzi.

TABELA 3

Wartości parametru Fo w 3 replikacjach dwóch osób w 2 zdaniach o złożonej strukturze

Mówca	Nr replik.	P o z y c j a s y l a b y											
		NN	A	NN	A	NN	A	NN					
Z D A N I E	1	1	0,08 0,21	0,3 0,34	0,29 0,05	0,09 0,32	0,33 0,05	0,08	0,01				
		2	0,11 0,26	0,26 0,36	0,36 0,11	0,10	0,07 0,05	0,8	-				
		3	0,08 0,29	0,3 0,35	0,35 0,05	0,05 0,32	0,33 0,01	0,09	0,04				
	2	1	0,12 0,08	0,08 0,28	0,28 0,04	0,01 0,21	0,21	0,08	0,02				
		2	0,13 0,09	0,09 0,32	0,32 0,23	0,23 0,5	0,05 0,14	0,11	-				
		3	0,15 0,07	0,11 0,35	0,35 0,1	0,18 0,1	0,09	0,14	0,03				
Z D A N I E	1		N	A	NN	A	NN	A	NN	A	NN	A	
		1	0,13 0,23	0,23 0,36	0,36 0,05	0,25 0,29	0,08 0,06	0,22 0,07	0,15 0,22	0,09 0,03	0,03	-	
		2	0,13 0,21	0,21 0,35	0,35 0,04	0,25 0,3	0,08 0,05	0,23 0,06	0,06 0,13	0,13 0,03	0,02	0,01	
	2	1	0,16 0,13	0,14 0,28	0,21 0,05	0,23	0,13 0,08	0,19 0,23	0,23 0,11	0,15 0,06	0,04 0,03	0,5	
		2	0,11 0,2	0,23 0,3	0,32 0,1	0,23 0,24	0,24 0,06	0,18 0,24	0,24 0,1	0,18 0,09	0,08 0,03	0,07	
		3	0,13 0,14	0,16 0,31	0,3 0,08	0,25	0,18 0,09	0,21 0,23	0,28 0,08	0,2 0,1	0,02	0,01	

TABELA 4

Wartości parametru Fo w 3 replikacjach wypowiedzi tego samego mówcy (26-sylabowe zdanie)

Replikacje		1		2		3	
P	NNN	0,08 0,04	0,04 0,12	0,08 0,07		0,09 0,05	
	A	0,12 0,28	0,28 0,32	0,09 0,29	0,39 0,36	0,35 0,09	0,28 0,28
Z	NNN	0,32 0,05		0,3 0,03	0,03 0,08	0,3 0,05	
	A	0,05 0,29	0,29 0,31	0,05 0,32	0,32 0,3	0,05 0,3	0,3 0,28
J	NNNNNN	0,21 0,13	0,7 0,13	0,23 0,3		0,2 0,4	
	A	0,05 0,22		0,03 0,21		0,12 0,21	0,21 0,1
S	N	0,4		0,21 0,01		0,16 0,04	
	A	0,4		0,6		0,13	
L	N	0,107		0,02		0,13 0,08	
	A	0,36 0,37		0,06		0,36 0,37	
B	N	0,18 0,11		0,18 0,05		0,21 0,11	
	A	0,11 0,1		0,05 0,13		0,11	
	NN	0,1		0,11 0,01		0,1	
	A	0,035		0,01		0,05 0,03	
	N	0,03		0		0,03	
	A	0,02		-		0,01 0	

TABELA 5

Wartości średnie parametru  $F_0$  dla 3 mówców w 12-stu zdaniach. Oznaczenia:  $F_{pd}$  - początkowa wartość parametru  $F_0$  w zdaniach rozpoczynających się od spójgłoski dźwięcznej,  $F_{pb}$  - wartość początkowa  $F_0$  w zdaniach rozpoczynających się od spójgłoski bezdźwięcznej,  $F_{A1}$  - wartość maksymalna na pierwszej sylabie akcentowanej,  $F_{max}$  - wartość maksymalna  $F_0$  w całej wypowiedzi,  $F_{Ak-1}$  - wartość maksymalna  $F_0$  na przedostatniej sylabie akcentowanej,  $F_{Ak}$  - na ostatniej sylabie akcentowanej

Nr zdania	Nr mówcy	$F_{pd}$	$F_{pb}$	$F_{A1}$	$F_{max}$	$F_{Ak-1}$	$F_{Ak}$
1	1	0,16	0,39	0,38	0,38	0,04	0,02
	2	0,11	0,29	0,32	0,32	0,02	-
	3	0,2	0,25	0,3	0,3	0,08	0,01
2	1	0,06	0,09	0,42	0,42	0,06	0,02
	2	0,15	0,17	0,35	0,35	0,15	0,07
	3	0,08	0,09	0,3	0,3	0,1	0
3	1	0,07	0,8	0,36	0,36	0,16	0,05
	2	0,12	0,13	0,32	0,32	0,21	0,08
	3	0,11	0,1	0,39	0,39	0,19	0,07
4	1	0,09	0,2	0,35	0,35	0,18	0,08
	2	0,13	0,14	0,31	0,31	0,21	0,1
	3	0,11	0,08	0,3	0,3	0,16	0,06
5	1	0,06	0,08	0,32	0,32	0,07	0,03
	2	0,17	0,2	0,25	0,25	0,1	0,04
	3	0,09	0,1	0,28	0,28	0,08	0
6	1	0,13	0,32	0,29	0,29	0,08	0,01
	2	0,09	0,25	0,2	0,2	0,18	0,16
	3	0,16	0,26	0,24	0,26	0,1	0,04
7	1	0,14	0,36	0,36	0,36	0,1	0,05
	2	0,16	0,26	0,18	0,18	0,16	0,04
	3	0,18	0,24	0,26	0,26	0,14	0,07
8	1	0,13	0,36	0,39	0,41	0,16	0,05
	2	0,17	0,28	0,32	0,3	0,19	0,08
	3	0,14	0,26	0,36	0,39	0,12	0,03
9	1	0,06	0,3	0,32	0,52	0,08	0,01
	2	0,12	0,18	0,3	0,38	0,14	-
	3	0,04	0,5	0,33	0,48	0,13	0,03
10	1	0,22	0,42	0,43	0,48	0,16	0,01
	2	0,16	0,33	0,38	0,39	0,13	0,08
	3	0,18	0,28	0,43	0,5	0,09	-

TABELA 5 cd.

Nr zdania	Nr mówcy	F <sub>pd</sub>	F <sub>pb</sub>	F <sub>A1</sub>	F <sub>max</sub>	F <sub>Ak-1</sub>	F <sub>Ak</sub>
11	1	0,08	0,08	0,38	0,38	0,12	0,01
	2	0,14	0,06	0,33	0,34	0,14	-
	3	0,15	0,18	0,28	0,3	0,16	0,03
12	1	0,2	0,34	0,35	0,35	0,16	0,01
	2	0,16	0,29	0,3	0,3	0,08	0,02
	3	0,11	0,22	0,28	0,3	0,06	-

#### 4. Analiza przebiegów parametru Fo w zdaniach złożonych

Dla analizy częstotliwości podstawowej w zdaniach wielofrazowych wybrano materiał lingwistyczny, zawierający fragmenty reportaży oraz wiadomości (zał. 3). Przygotowane zdania różniły się między sobą pod względem liczby, struktury i długości poszczególnych fraz. (W przyjętym znaczeniu fraza oznacza odcinek mowy, zawierający się między znakami interpunkcyjnymi). Poszczególne frazy składały się z różnej liczby grup akcentowych o odmiennych strukturach lingwistyczno-fonetycznych. Jako grupę akcentową zdefiniowano prozodyczną jedność składającą się z sylaby akcentowanej oraz sąsiednich sylab nieakcentowanych. Kompleksowa, zwłaszcza statystyczna, ocena czynników wpływających na kontur częstotliwości podstawowej w zdaniu złożonym wymaga przygotowania i przeanalizowania bardzo dużej liczby danych.

Techniczna analiza cech prozodycznych, ekstrakcja częstotliwości podstawowej, segmentacja sygnału wymaga znacznego nakładu pracy i manualnej korekty danych.

Efektywnym rozwiązaniem wydaje się analiza zróżnicowanego lingwistycznie materiału, wybór różnych konfiguracji wzorców intonacyjnych podobnie realizowanych przez poszczególnych mówców i stopniowe poszerzanie bazy danych przez analizę różnorodnych tekstów.

Przeprowadzono ekstrakcję parametru Fo oraz segmentację



poszczególnych wypowiedzi na sylaby, a wyniki analiz w postaci danych cyfrowych zapisano w pamięci komputera. Wstępnej ocenie poddano replikacje wypowiedzi trzech osób. Wiarygodny pomiar częstotliwości podstawowej z głosu czwartego mówcy był utrudniony, z uwagi na częste nieregularne obniżanie przez niego tonu podstawowego poniżej 65 Hz.

Analiza dziewięciu replikacji (3 mówców x 3 powtórzenia), wykazała możliwość oceny podobieństw i zróżnicowań konturów intonacyjnych w obrębie poszczególnych wypowiedzi.

W tabeli 6 podano przykładowo wartości parametru  $F_0$  w pewnych charakterystycznych punktach konturów intonacyjnych dla 9-ciu replikacji 4-frazowego zdania. Analizując dane dla poszczególnych mówców, zauważa się cechy indywidualne w realizacji konturów intonacyjnych. Na przykład mówca nr 3 konsekwentnie stosuje na ostatniej sylabie frazy wzrost kontynuacyjny. Dla porównania konsekwencji w realizacji różnych struktur intonacyjnych przez jednego mówcę podano w załączniku 4 przykład danych ilustrujących przebiegi parametru  $F_0$  (w 3 replikacjach) na kolejnych sylabach bardzo długiego (65-sylabowego) zdania.

TABELA 6

Wartości parametru  $F_0$  w kolejnych fazach wypowiedzi "Korzystając z okazji, jaką było spotkanie z grupą dziennikarzy włoskich, pułkownik Muamar Kaddafi oświadczył, że nie wyklucza swojego udziału w następnych wyborach we Włoszech".

Przez  $F_p$  oznaczono wartość początkową parametru  $F_0$ , przez  $F_{Ap}$ ,  $F_{Ak}$ ,  $F_{Nk}$  - odpowiednio: maksymalne wartości częstotliwości podstawowej na pierwszej sylabie akcentowanej, końcowej akcentowanej i ostatniej sylabie nieakcentowanej frazy

Mówca		1			2			3		
Nr replikacji		1	2	3	1	2	3	1	2	3
	$F_p$	121	126	131	120	121	116	98	101	98
	$F_{Ap}$	136	204	232	142*	170	160	162	176	179

TABELA 6 cd

Mówca		1			2			3		
Nr replikacji		1	2	3	1	2	3	1	2	3
W A R T O Ś Ć	F <sub>Ak</sub>	92	89	132	111	104	100	80	111	100
	F <sub>Nk</sub>	80	82	93	103	90	85	116	107	117
	F <sub>P</sub>	90	99	91	115	106	110	93	107	101
	F <sub>ap</sub>	112	129	121 <sup>*</sup>	151	142	148	150	131	126 <sup>*</sup>
P A R A M E T R U	F <sub>ak</sub>	88	102	150	90	89	90	88	87	91
	F <sub>Nk</sub>	152	160	129	167	160	150	91	113	116
	F <sub>P</sub>	106	120	122	123	121	119	101	100	98
	F <sub>ap</sub>	126	142	150	160	182	170	186	150	160
FO	F <sub>ak</sub>	86	126	102	111	130	120	113	89	88
	F <sub>Nk</sub>	116	132	128	96	169	135	99	113	116
	F <sub>P</sub>	101	96	115	124	104	110	98	100	107
	F <sub>ap</sub>	129	121 <sup>*</sup>	110 <sup>*</sup>	157	172	160	124	167	120
	F <sub>ak</sub>	88	98	112	105	100	111	92	85	101
	F <sub>Nk</sub>	81	84	89	98	89	85	80	83	82

Z A Ł A C Z N I K 1

Struktury 12-stu zdań.

Litera A oznaczono pozycję sylaby leksykalnie akcentowanej.

Litera N pozycję sylaby nieakcentowanej.

1. A N N A N N A N N A N N A N
2. N A N N A N N A N N A N N A N
3. N A N N A N N A N A N N A N N A N N A N N N A N
4. N N A N N A N N A N
5. A N A N A N A N A N
6. A N A N A
7. A N A N A N
8. A N A N N N N A N A N A N N N N N A N N N A N N A N
9. N N A N N N A N N A N N A N N A N A N A N A N A N
10. A N N N N A N N A N N A N N A N N A N N A N N A N A N A N A N N N A  
N A N
11. N A N A N A N
12. A A A

Z A Ł A C Z N I K 2

Wartości współczynników korelacji dla 9 replikacji wypowiedzi  
"Londyn ma zostać jedyną stolicą europejską nie posiadającą  
ogrodu zoologicznego"

Row	Var1	Var2	Var3	Var4	Var5	Var6	Var7	Var8	Var9
1	1	0,94	0,89	0,85	0,94	0,90	0,80	0,84	0,81
2	0,94	1	0,94	0,88	0,97	0,92	0,78	0,86	0,88
3	0,89	0,94	1	0,95	0,96	0,93	0,72	0,85	0,78
4	0,85	0,88	0,95	1	0,91	0,85	0,75	0,89	0,76
5	0,94	0,97	0,96	0,91	1	0,94	0,78	0,88	0,86
6	0,90	0,92	0,93	0,85	0,94	1	0,72	0,79	0,82
7	0,80	0,78	0,72	0,75	0,78	0,72	1	0,91	0,68
8	0,84	0,86	0,85	0,89	0,88	0,79	0,91	1	0,74
9	0,81	0,88	0,78	0,76	0,86	0,82	0,68	0,74	1

Z A Ł A C Z N I K 3

Fragmenty analizowanych tekstów:

1.

Przez jeden dzień pasażerowie podróżujący moskiewskim metrem nie musieli uiścić żadnej opłaty za przejazd. Stało się tak za sprawą umowy, jaką dyrekcja metra podpisała z pewną amerykańską firmą, która zgodziła się opłacić koszty jednego dnia eksploatacji, w zamian za ciągłe nadawanie programu reklamowego. Wysokość transakcji nie została ujawniona, jednak amerykańscy przedsiębiorcy - po otrzymaniu wyników badania skuteczności oddziaływania tej formy reklamy - zdecydowali, że do końca roku opłacą jeszcze dwa dni darmowej jazdy.

2.

Korzystając z okazji, jaką było spotkanie z grupą dziennikarzy włoskich, pułkownik Muamar Kadafi oświadczył, że nie wyklucza swojego udziału w następnych wyborach prezydenckich we Włoszech. Zdziwionym rozmówcom wyjaśnił, że urodził się jako obywatel włoski i nikt mu oficjalnie tego obywatelstwa nie odebrał. Kadafi przedstawił też zarys swojego programu. Najważniejszym punktem okazało się wypędzenie Amerykanów, których imperialistyczne zapędy ograniczają - jego zdaniem - wolność Włoch.

3.

Londyn ma zostać jedyną stolicą europejską nie posiadającą ogrodu zoologicznego. Mimo ciągnących się już od dwóch miesięcy dyskusji, nadal nie ma źródła, z którego można by finansować działalność tej liczącej sto pięćdziesiąt lat instytucji. Władze miasta nie mają brakującej sumy (około dwudziestu jeden milionów dolarów), a rząd odmawia subwencji. W ogrodzie przebywa osiem tysięcy zwierząt, z których większość - w razie likwidacji zoo - zostanie uśpiona, gdyż w związku z zastojem na międzynarodowym rynku zoologicznym nie ma na nie nabywców.

Z A Ł A C Z N I K 4

Przebiegi parametru Fo na kolejnych sylabach, w 3 replikacjach wypowiedzi. Przez Fp oznaczono wartość początkową parametru Fo na sylabie, przez Fk - wartość końcową.

z t e l e w i z y j n e g o k i n a n o c n e g o w c i a ż z s y m

F <sub>p</sub> F <sub>k</sub>	1	129	126	121	128	150	119	126	102	100	170*	93	152	115
		126	123	116	152	146	116	102	100	100	155	89	165	115
F <sub>p</sub> F <sub>k</sub>	2	124	120	115	115	134	126	131	93	81	156	167	150	140
		120	115	110	134	165	120	96	84	80	176*	148	144	123
F <sub>p</sub> F <sub>k</sub>	3	122	120	126	138	182*	162	148	105	92	148	165	124	134
		118	120	116	182	173	150	105	92	93	165	138	113	113

p a t i a w s p o m i n a m c z ł o w i e k a k t ó r y s i ę

F <sub>p</sub> F <sub>k</sub>	1	132	140	113	105	105	116	113	113	106*	95	102
		116	124	110	105	108	98	117	105	95	95	92
F <sub>p</sub> F <sub>k</sub>	2	144	117	108	117	123	121	155	142	126	112	119
		142	82	117	121	115	106	152	136	112	110	97
F <sub>p</sub> F <sub>k</sub>	3	140	140	116	119	119	115	132	117	115	99	106
		138	113	119	120	123	107	131	115	99	90	89

z m n i e j s z a ł f i l m z g a t u n k u f i k c j i n a u k o w e j

F <sub>p</sub> F <sub>k</sub>	1	110	80	117*	106	102	102	98	86	86	93	126	196
		81	78	97	102	104	100	86	86	93	93	196	169
F <sub>p</sub> F <sub>k</sub>	2	98	83	148*	105	106	100	80	95	83	85	132	152
		83	75	100	100	100	100	82	83	85	80	152	208
F <sub>p</sub> F <sub>k</sub>	3	111	87	126	94	94	97	107	93	80	80	123	140
		87	-	83	94	92	96	106	81	80	78	140	172

a więc raczej fantastyczny i raczej

$F^P$	1	152	92	92	91	97	89	126	129	128	128	129
$F_k$		100	92	90	86	93	87	117	121	129	131	112
$F^P$	2	117	128	117	101	112	88	134*	134	99	91	88
$F_k$		121	111	117	84	93	86	130	132	91	88	88
$F^P$	3	107	105	89	85	104	90	121	138	92	82	80
$F_k$		95	84	85	80	95	87	111	144	81	80	78

nie na ukoży ale za to podobaja

$F^P$	1	119	140	93	126	146	157	150	107	108	116	96	100	96
$F_k$		140	93	90	146	189	150	103	105	106	112	87	85	86
$F^P$	2	88	112	97	136	131	138	140	152	148	140	131	108	100
$F_k$		120	97	97	131	144	140	144	152	150	131	99	84	98
$F^P$	3	107	115	90	117	124	148	179	160	160	162	129	121	98
$F_k$		115	83	90	124	170	179	138	158	152	142	121	80	96

cy wyobraznie

$F^P$	1	86	96	87	83	83
$F_k$		83	87	86	83	75
$F^P$	2	86	87	83	87	84
$F_k$		87	80	84	87	-
$F^P$	3	88	101	92	80	74
$F_k$		84	91	85	75	-

Bibliografia

- [1] Steffen-Balogowa M.  
Analiza struktury przebiegu melodii polskiego języka ogólnego, rozprawa doktorska, Poznań, 1963.
- [2] Jassem W., Demenko G.  
On Extracting Linguistic Information from Fo Traces, w: Intonation in Discourse (c. Johns-Lewis ed.), Croom Helm, London I-18, 1986.
- [3] Demenko G., Jassem W., Krzyśko M.  
Classification of basic Fo patterns using discriminant functions, *Phonetica*, 41, I-12, 1988.
- [4] Kohler K. J.  
Prosody in speech synthesis: the interplay between basic research and TTS application, *Journal of Phonetics*, 19, 121-138, 1991.
- [5] Imiołczyk J., Nowak J., Demenko G.  
A Text-to-Speech System for Polish Eurospeech Proceedings, 1993, Vol. 2, ss. 885-889.
- [6] 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.
- [7] Nöth E.  
Prosodische Information in der automatischen Spracherkennung, Max Niemeyer Verlag, Tübingen 1991.