

561891

**Grażyna Demenko**

**STATYSTYCZNE WŁASNOŚCI ROZKŁADÓW  
CHWILOWYCH WARTOŚCI PARAMETRU  $F_0$   
W MOWIE CIĄGŁEJ**

**WARSZAWA 1980**

Praca wpłynęła do Redakcji dnia 11 sierpnia 1980 r.

Zarejestrowana pod nr 31/1980



027911

Na prawach rękopisu

---

Instytut Podstawowych Problemów Techniki PAN

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

Oddano do drukarni w sierpniu 1980r.

Nr zamówienia 566/0/80

---

Warszawska Drukarnia Naukowa, Warszawa,  
ul. Śniadeckich 8

E. 159/80

Grażyna Demenko  
Pracownia Fonetyki Akustycznej IPPT PAN

STATYSTYCZNE WŁASNOŚCI ROZKŁADÓW CHWILOWYCH WARTOŚCI  
PARAMETRU  $F_0$  W MOWIE CIĄGŁEJ

Streszczenie

Przeprowadzono próbę określenia długoterminowych matematycznych własności częstotliwości podstawowej poprzez statystyczną analizę rozkładów chwilowych wartości  $F_0$ . Dla ustalenia wpływu częstotliwości odczytu obwiedni zapisu tonograficznego na kształtowanie się rozkładów wykonano je w kilku wersjach przy zmiennych warunkach analizy. W celu scharakteryzowania poszczególnych prób posłużono się testami statystycznymi badającymi równość parametrów analizowanych rozkładów. Uzyskano 90 rozkładów, na podstawie których sugeruje się przyjęcie częstotliwości próbkowania synchronicznej z okresem podstawowym oraz taki dobór czasu trwania sygnału aby można uważać go było za stacjonarny.

1. Wstęp.

Parametr  $F_0$  jest jednym z podstawowych nośników informacji osobniczej i lingwistycznej. Zasadniczą trudność w określeniu stopnia przydatności tego parametru w rozmaitych problemach fonetyki, polega na braku wypracowania odpowiedniego modelu umożliwiającego opis czasowej zmienności częstotliwości podstawowej. Jedną z przyczyn tej trudności jest niemożliwość przeanalizowania wszystkich zjawisk, które należałoby zbadać, aby uzyskać jednoznaczną interpretację wyników. Dlatego coraz częściej w tego typu zagadnieniach stosuje się analizę statystyczną, której zadaniem jest wskazanie właściwych metod przygotowania i interpretowania danych, będących podstawą sformułowania ostatecznych wniosków.

Zadanie niniejszej pracy polegało na :

- (1) opracowaniu algorytmów umożliwiających wczytywanie danych z tonometru TM-3, do pamięci MERY 303 oraz ich analizę,
- (2) analizie statystycznej parametru  $F_0$ ,
- (3) próbie określenia badanych głosów za pomocą parametrów charakteryzujących rozkłady częstotliwości podstawowej w celu określenia najprostszego algorytmu umożliwiającego adekwatny opis statystycznych własności  $F_0$ .

Wykaz oznaczeń :

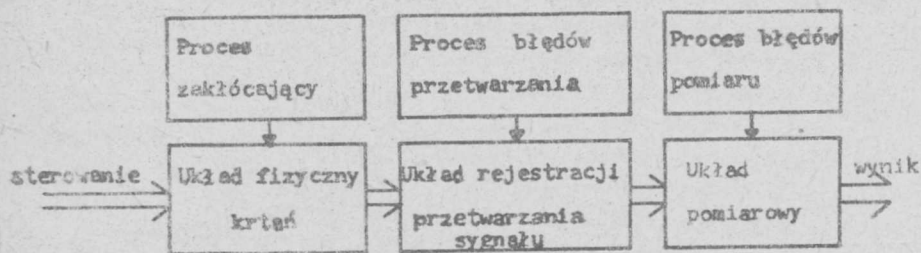
- $N$  - liczebność próby,
  - $\bar{x}$  - średnia arytmetyczna próby,
  - $s^2$  - wariancja próby,
  - $s$  - odchylenie standardowe próby,
  - $G$  - odchylenie standardowe populacji,
  - $x_i$  - wartość zmiennej losowej,
  - $d_3$  - współczynnik asymetrii,
  - $d_4$  - współczynnik ekscesu,
  - $f_{oi}$  - liczebności eksperymentalne,
  - $f_{ei}$  - liczebności teoretyczne,
  - $j$  - liczba przedziałów częstotliwościowych,
  - $d$  - szerokość przedziału częstotliwościowego,
  - $r_g$  - górna granica rozkładu liczebności,
  - $r_l$  - dolna granica rozkładu liczebności,
  - $r_d$  - dolna mierzona częstotliwość,
  - $m$  - ilość niezależnych prób,
  - $k_{rl}$  - zawartość komórki pamięciowej,
  - $V$  - liczba stopni swobody,
  - $d$  - poziom istotności,
  - $V$  - współczynnik zmienności.
2. Charakterystyka badanego sygnału.

Przed przystąpieniem do jakiegokolwiek analizy należy ustalić podstawowe cechy sygnału, rozważyć jego stopień zdeterminowania, stacjonarność i zbadać zgodność rozkładów charakteryzujących sygnał z rozkładem normalnym.

W wielu zagadnieniach trudno rozstrzygnąć, czy sygnał ma charakter zdeterminowany, czy też nie. Teoretycznie, jeżeli sygnał można opisać zależnościami matematycznymi, należy go zaliczyć do zdeterminowanych. Można twierdzić jednak, że żaden fizyczny



sygnał nie jest ściśle zdeterminowany (zawsze może zaistnieć nieprzewidziana sytuacja, która zmieni jego charakter), z drugiej jednak strony, żaden sygnał nie jest czysto losowy, bo przy dostatecznej znajomości mechanizmu jego wytwarzania, można opisać go mniej lub więcej dokładnie. W praktyce, jeśli wielokrotne powtarzanie doświadczenia daje te same rezultaty, sygnał możemy uważać za zdeterminowany. Częstotliwość podstawowa sygnału mowy nie może być opisana ścisłymi zależnościami matematycznymi, bo wynik każdej obserwacji jest niepowtarzalny. Warunki stacjonarności procesu losowego są ściśle określone, w przypadku jednak analizy pojedynczej funkcji losowej stosuje się nieco inne pojęcie (por. [1]). Gdy mówi się o stacjonarności jednej realizacji oznacza to zazwyczaj, że charakterystyki wyznaczone dla krótkich, niezależnych przedziałów czasowych, nie zmieniają się znacznie dla kolejnych przedziałów (więcej niż można by tego oczekiwać biorąc pod uwagę typową losową zmienność). W przypadku pomiarów częstotliwości podstawowej należy rozważyć model przedstawiony na ryc. 1.



Ryc.1. Schemat blokowy układu analizy częstotliwości podstawowej.

Jest oczywiste, że zbiór przyczyn determinujący określony skutek (tutaj wynik pomiaru  $F_0$ ) jest zbiorem o dużej liczebności. Możemy wyróżnić w nim jedną przyczynę główną, a o pozostałych zakładamy, że ich jednostkowe oddziaływanie jest

porównywalne oraz wzajemnie różnokierunkowe.

Rozważmy relację przyczynowo-skutkową wiążącą zbiór wzajemnie niezależnych zmiennych  $(G, \varphi_1 \dots \varphi_i)$  ze zmienną zależną  $S$  :

$$S = f(G, \varphi_1 \dots \varphi_i) = f(G) + f(\varphi_i)$$

Częstotliwość podstawowa jest więc zmienną losową określoną deterministycznie w zakresie stałych uwarunkowań strukturalnych, anatomicznych, genetycznych czynnika  $G$  oraz probabilistycznie w zakresie efektów różnorodnych (niezależnych w rozsądnym przybliżeniu) wielokrotnych oddziaływań czynników  $\varphi_i$  w skład których wchodzi :

- zmienne określające warunki drgania strun głosowych,
- czynniki psychiczne i emocjonalne,
- rodzaj realizacji dźwięku, intonacji, artykulacji.

Podstawowym pytaniem jest, jak dalece charakterystyka zmiennej  $S$  tzn. skutku jest zależna od charakterystyk zmiennych losowych  $\varphi_i$ . Twierdzenia Lindberga i Lapunowa dają odpowiedź na to pytanie. Pod warunkiem ograniczonego zakresu zmienności zmiennych losowych (w praktyce zawsze w przyrodzie spełnionego) rozkład zmiennej zależnej  $S$  będzie niemal zawsze zbliżony do normalnego, jeśli tylko poczynione są założenia co do mechanizmu oddziaływania wzajemnie niezależnych i losowych przyczyn pobocznych. Biorąc to pod uwagę nasuwa się pytanie, czy można zmienne losowe wpływające na probabilistyczny charakter  $F_0$  uważać za niezależne. Intuicyjnie można poczynić takie założenie, należy jednak dokładnie zbadać mechanizm oddziaływania tych zmiennych. Symetria i niezależność, aczkolwiek bardzo często w praktyce w przyrodzie spełnione, nie są absolutnymi prawami natury. Koniecznym jest więc fakt weryfikacji hipotezy stacjonarności i normalności charakterystyk  $F_0$  przy uwzględnieniu stopnia ich zdeterminowania w określonych warunkach.

### 3. Techniczna realizacja układu analizy $F_0$ .

Przetwarzanie sygnałów biologicznych ze względu na ich specyfikę stwarza perspektywy dla zastosowania hybrydowej techniki obliczeniowej. Schemat układu pomiarowego częstotliwości podstawowej przedstawia ryc. 2. Z nagranych na taśmie magnetofonowej materiału językowego dokonywana jest ekstrakcja

parametru  $F_0$  przy pomocy tonometru TM3. Sygnał wyjściowy ekstraktora obserwowano na ekranie oscyloskopu z długą poświatą lub rejestrowano na papierze świetliczołym przy pomocy oscylografu pętlicowego. Tonometr posiada 24 warianty pomiarowe które umożliwiają właściwy dobór warunków ekstrakcji. Pomiar okresu polega na generacji napięcia rosnącego liniowo w czasie równym długości mierzonego okresu. Wynikiem pomiaru jest ciąg napięciowych impulsów trójkątnych o długościach wprost proporcjonalnych do długości poszczególnych okresów. Dokładne omówienie sposobu ekstrakcji parametru  $F_0$  przedstawione jest w pracy [9]. Wynik ekstrakcji w postaci ciągu impulsów przesyłany jest do jednego z bloków Kanału Funkcji Analogowych, do czasomierza, gdzie dokonuje się cyfrowy pomiar okresu. KFA jest jednostką organizującą transmisję informacji między rejestrem akumulatora minikomputera i pewną klasą urządzeń peryferyjnych (np. magnetofon, mikrofon, tonometr). KFA posiada 4 niezależne kanały przepływu informacji : kanał konwertera, czasomierza, dwa kanały wyprowadzania informacji na wyjścia analogowe. Dokładny sposób współpracy KFA z jednostką sterującą minikomputera podany jest w pracy [10]. Każdy ze składników układu wprowadza swój stały błąd pomiarowy związany głównie z charakterystyką przenoszenia sygnału oraz błędy losowe. Maksymalny błąd ekstrakcji  $F_0$  wynosi 3 %, natomiast błąd wprowadzany przez układ przetwornika analogowo-cyfrowego około 2,5 %. Dokładniejsza analiza błędów przedstawiona jest w pracy [9] oraz [10].

#### 4. Realizacja podstawowych algorytmów.

##### 4.1. Algorytm wczytywania danych z tonometru do pamięci minikomputera.

Ze względu na specyfikę analizowanego sygnału - częstotliwości podstawowej algorytm ten musiał być skonstruowany elastycznie i umożliwiać wprowadzanie danych :

- a/ ze zmienną częstotliwością od 10 do 50 ms,
- b/ synchronicznie z częstotliwością podstawową,
- c/ z dowolnego miejsca taśmy magnetofonowej.

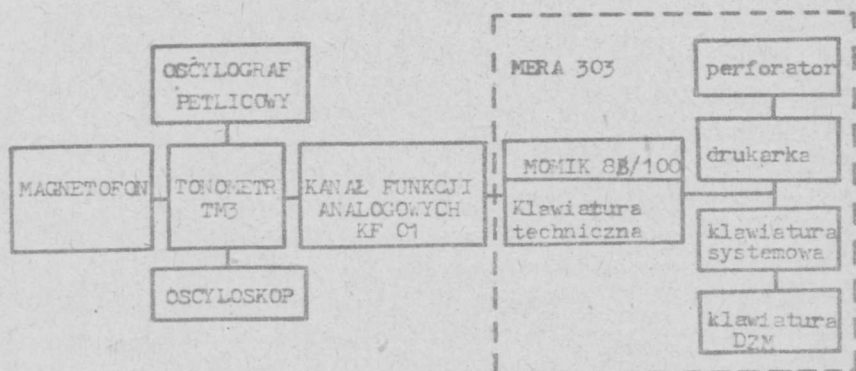
Tablica 1 przedstawia częstotliwość próbkowania wyników pomiaru czasu określoną programowo opóźnieniem czasowym sterowanym z klawiatury. Zależnie od nastawianej klawiszami wartości następuje różne opóźnienie czasowe i w efekcie różna

częstotliwość próbkowania, z możliwością regulacji w dość szerokich granicach.

Tablica 1.

Regulacja częstotliwości próbkowania za pomocą opóźnienia czasowego określonego układem klawiatury.

Częstotliwość próbkowania Hz	Numer wciśniętego klawisza	
	klawiatura techniczna	klawiatura systemowa
25	0,1,2,3	3,4,6,7
50	0,1,2,3,4	3,4,6,7
100	0,1,2,3,4,5	3,4,5



Ryc.2. Schemat układu pomiaru i analizy  $F_0$ .

Wprowadzanie danych synchronicznie z okresem  $F_0$  można uzyskać bezpośrednio poprzez analizę przerw (w konstrukcji czasomierza przewidziano sygnał służący do wygenerowania przerwy bezpośrednio po zakończeniu pomiaru okresu) lub przez analizę

danych po ich wprowadzeniu.

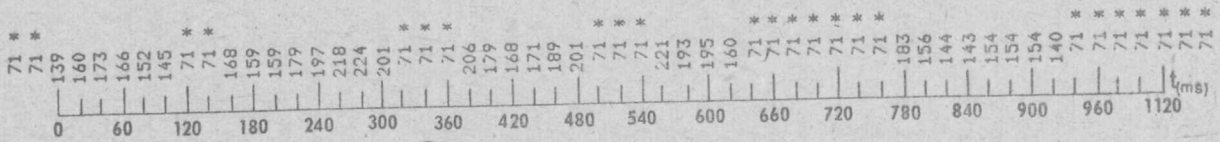
W zależności od przyjętej częstotliwości próbkowania sygnału w pamięci minikomputera można zmieścić więcej lub mniej danych. Ponieważ do analizy przewidziane zostały odcinki 50 s sygnału mowy, istniała konieczność podzielenia ich na odcinki krótsze i sekwencyjne wprowadzanie ich do pamięci. Istotną trudnością było wczytanie danych dokładnie z tego samego miejsca, w którym zakończono wczytanie uprzednio analizowanej części materiału. Metoda ręcznego zatrzymywania i uruchamiania przesuwu taśmy magnetofonowej z oczywistych powodów była zbyt niedoskonała, dlatego opracowany został automatyczny sposób wyszukiwania żądanego ciągu danych. Z chwilą wczytania pierwszej danej uruchamiany zostaje licznik i w zależności od jego ustawienia, do pamięci wczytywany jest odpowiedni fragment sygnału.

Licznik ten zrealizowano programowo. Od dowolnie wybranego miejsca na taśmie magnetofonowej mierzony jest czas jaki upływa podczas wprowadzania kolejnych danych do pamięci minikomputera. Po upływie odmierzanego zawsze od tego samego miejsca na taśmie magnetofonowej czasu, wprowadzamy do pamięci następną partię danych. W celu natychmiastowego uwidocznienia wyników ekstrakcji parametru  $F_0$  opracowany został algorytm umożliwiający wyprowadzanie bloków danych, w postaci cyfrowych tonogramów, z pamięci na oscyloskop. Dysponując oscyloskopem z pamięcią można prześledzić przebiegi intonacyjne, oraz wybrać najbardziej poprawny wariant ekstrakcji  $F_0$ .

Ryc. 3 przedstawia wydruki  $F_0$  z początkowego fragmentu jednej z analizowanych wypowiedzi. Program wydruku polega na tym, że jednocześnie drukowany jest ciąg wartości rzędnych - częstotliwości podstawowej oraz orientacyjny wykres w postaci punktów, których odległość od osi poziomej jest wprost proporcjonalna do wartości danej rzędnej. Poniżej wydruków podano skalę czasową. W punkcie (c) ryc. 3 przebieg  $F_0$  podano w formie wydruku kolejnych wartości tego parametru bez zaznaczenia skali czasowej, która jest nieliniowa. W segmentach bezdźwięcznych określono uawnie  $F_0 = 71$  Hz.



(a)

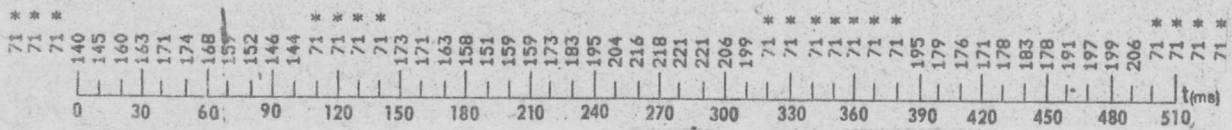


z a s a d n i t j e t a p

Ryc. 3. Wydruki częstotliwości podstawowej. Przebieg intonacji przy częstotliwości próbkowania (a) 20 ms

- 10 -

(b)



(c)

z a s a d n i t j e t

- 71
- 138
- 139
- 144
- 160
- 162
- 162
- 165
- 168
- 171
- 171
- 169
- 168
- 166
- 168
- 166
- 159
- 158
- 154
- 143
- 150
- 179
- 174
- 169
- 171
- 168
- 166
- 163
- 163
- 159
- 156
- 150
- 155
- 159
- 159
- 160
- 165
- 171
- 174
- 179
- 183
- 187
- 191
- 195
- 197

- 11 -

Ryc. 3. Przebieg intonacji przy częstotliwości próbkowania (b) 10 ms, (c) wydruk wartości  $F_0$  przy częstotliwości próbkowania równej okresowi podstawowemu.



#### 4.2. Algorytm grupowania danych.

Zbrane dla każdej próby dane uporządkowano w szereg rozdzielczy i sporządzono histogramy. Zaklasyfikowanie danej do odpowiedniego przedziału częstotliwościowego o szerokości 5 Hz poprzedzało określenie dolnej i górnej granicy rozkładów liczebności, zależne od dolnej mierzonej częstotliwości. W Tabelicy 2 podano początkowe wartości zakresów częstotliwości dla głosów męskich i żeńskich przy ich podziale na niskie i wysokie.

Tablica 2

Dolna mierzona częstotliwość dla poszczególnych zakresów w Hz.

M		Z	
N	W	N	W
70	100	140	200

Algorytm przedstawia ryc. 4.

#### 4.3. Algorytm podstawowej analizy statystycznej.

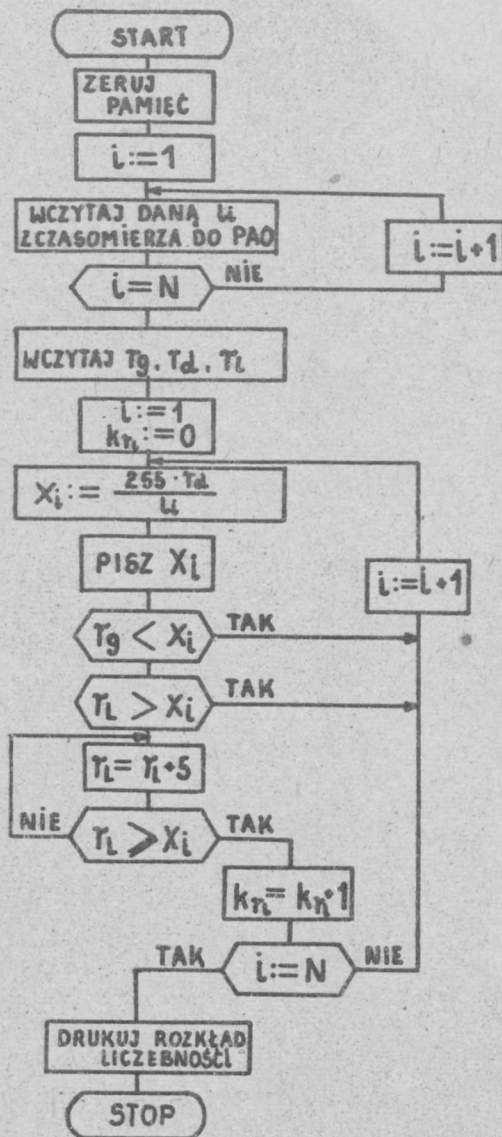
Wartości średnie oraz wariancje analizowanych zbiorów danych obliczono z rozkładów liczebności na podstawie zależności

$$\bar{x} = \frac{\sum_{i=1}^j f_{oi} x_i}{\sum_{i=1}^j f_{oi}} \quad (1)$$

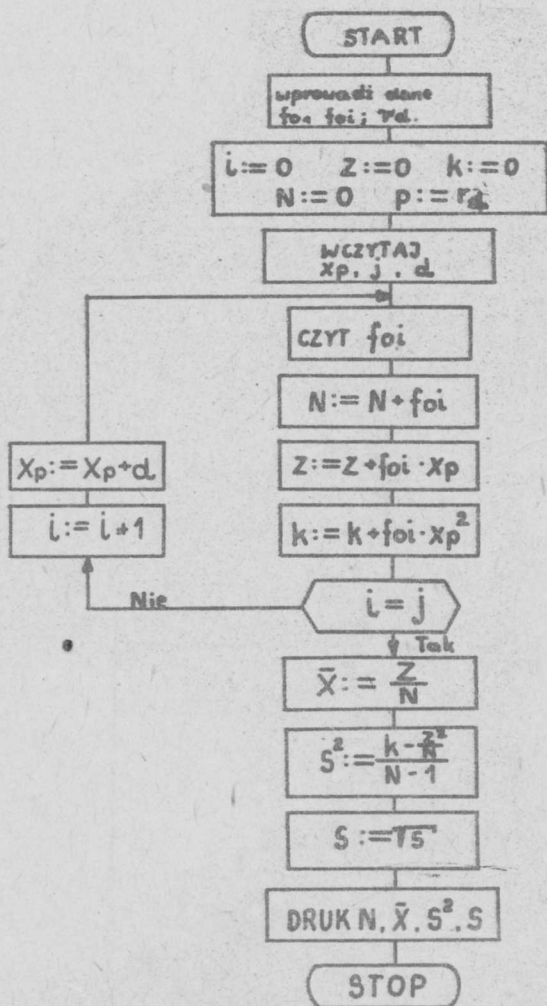
$$S^2 = \frac{\sum_{i=1}^j f_{oi} x_i^2 - \frac{\left(\sum_{i=1}^j f_{oi} x_i\right)^2}{N}}{N - 1} \quad (2)$$

Algorytm przedstawia ryc: 5

W celu określenia odchylenia rozkładu empirycznego od teoretycznego (normalnego) liczebności teoretyczne wyliczono na podstawie równania funkcji charakterystycznej rozkładu normalnego (por. np. [3], [5]).



Ryc.4. Algorytm czytania danych do PAO i wykonywania rozkładu liczebności.



Ryc.5. Algorytm obliczania podstawowych parametrów statystycznych : wartości średniej, wariancji oraz liczebności analizowanych zbiorów danych.

$$f(x) = \frac{N \cdot d}{\sqrt{2\pi} \cdot \sigma} e^{-\frac{1}{2} \left( \frac{x - \bar{x}}{\sigma} \right)^2} \quad (3)$$

Zaprogramowano również obliczenie asymetrii oraz ekscesu jako miar stopnia odchylenia danego rozkładu od normalnego. Asymetrię rozkładu empirycznego określa wzór

$$d_3 = \frac{\mu_3}{s^3} \quad (4)$$

gdzie  $\mu_3$  jest empirycznym momentem centralnym rzędu trzeciego

$$\mu_3 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^3 \quad (5)$$

Eksces rozkładu określa wzór

$$d_4 = \frac{\mu_4}{s^4} \quad (6)$$

gdzie  $\mu_4$  jest momentem centralnym rzędu czwartego

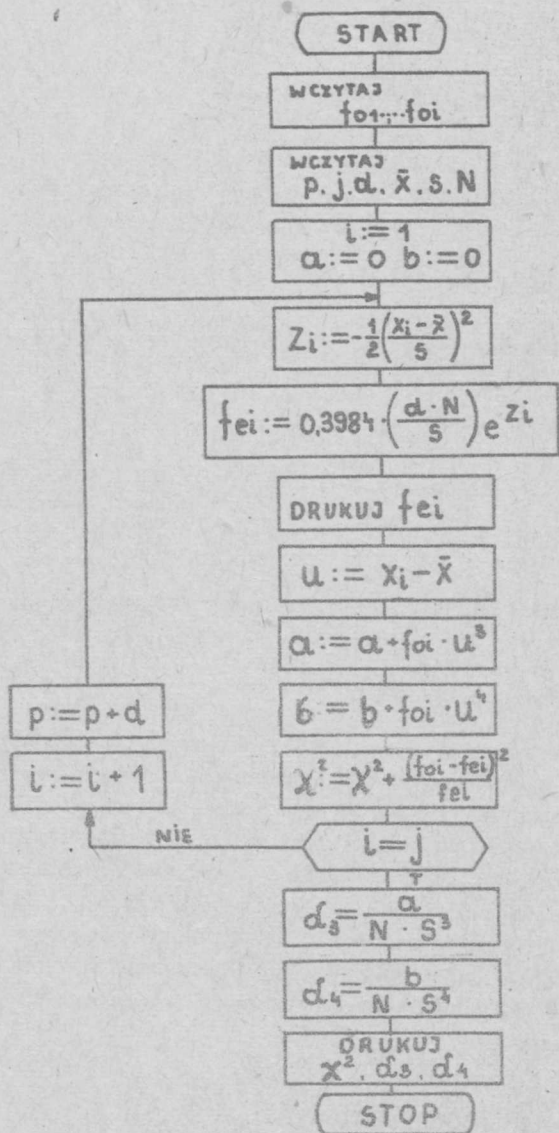
$$\mu_4 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^4 \quad (7)$$

Ryc. 6 przedstawia algorytm obliczania powyższych zależności.

Przewidziana została możliwość wykonania następujących testów statystycznych: testu  $\chi^2$  sprawdzającego hipotezę normalności rozkładu, testu Bartletta weryfikującego hipotezę równości wariancji, testu t określającego równość średnich w przypadku równych i różnych wariancji. Test zgodności  $\chi^2$  porównujący empiryczne częstotliwości z teoretycznymi nie dowodzi prawdziwości hipotezy normalności rozkładu, ale na zadanym poziomie istotności ustala zgodność lub niezgodność z danymi empirycznymi.

$$\chi^2 = \sum_{i=1}^j \frac{(f_{oi} - f_{ei})^2}{f_{ei}} \quad (8)$$

Test Bartletta daje możliwość weryfikacji różnic między miarami rozproszenia i odpowiedź na pytanie, czy sytuacja eksperymentalna powoduje większe czy mniejsze zróżnicowanie wyników.



Ryc.6. Algorytm obliczania stopnia odchylenia danego rozkładu od normalnego - współczynników asymetrii i ekscesu oraz wartości funkcji testu  $\chi^2$ .

Wartość funkcji testu Bartletta określa wzór (por. [5]).

$$\chi^2 = \frac{\nu \ln s^2 - \sum_{r=1}^m \nu_r \ln s_r^2}{c} \quad (9)$$

gdzie oznaczono

$$c = 1 + \frac{1}{3(m-1)} \left[ \sum_{r=1}^m \frac{1}{\nu_r} - \frac{1}{\nu} \right]$$

$$s^2 = \frac{\sum_{r=1}^m (N_r - 1) s_r^2}{N - m}$$

$$\nu = N - m$$

Tak określoną wartość  $\chi^2$  porównuje się z wartością krytyczną  $\chi^2(\alpha; \nu = m - 1)$ .

Test t Studenta jest testem istotności różnic między dwiema średnimi. Zastosowano dwie formy testu dla równych i różnych wariancji (por. [5]). Wartość funkcji testu Studenta wyznacza się z zależności :

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{S \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}}$$

gdzie

$$S \sqrt{\frac{1}{N_1} + \frac{1}{N_2}} = \sqrt{\frac{(N_1 - 1) s_1^2 + (N_2 - 1) s_2^2}{N_1 + N_2 - 2} \cdot \frac{N_1 + N_2}{N_1 N_2}}$$

i porównuje z wartością krytyczną  $t(\alpha; \nu = N_1 + N_2 - 2)$ .

Gdy nie można zakładać równości wariancji wartość  $S \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}$  oznaczamy

$$S \sqrt{\frac{1}{N_1} + \frac{1}{N_2}} = \sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}$$

a wartość  $\nu$  określamy z zależności



$$V = (N_1 + N_2 - 2) \left[ \frac{1}{2} + \frac{s_1^2 s_2^2}{s_1^4 s_2^4} \right]$$

#### 4.4. Oprogramowanie algorytmów i ich rozmieszczenie w pamięci MERY 303.

Organizacja maszyny cyfrowej decyduje o charakterystycznych cechach oprogramowania. Ograniczone możliwości software'owe narzuciły konieczność opracowania algorytmów w dwóch językach MOTIS oraz BASIC. Język MOTIS jest prostym językiem symbolicznym, który umożliwił oprogramowanie algorytmu wprowadzenia danych do PAO w czasie rzeczywistym oraz algorytmu wykonywania rozkładów liczebności. Po wydruku rozkładów wartości  $F_0$  z całego analizowanego materiału dokonano wymiany oprogramowania wczytując do pamięci minikomputera BASIC. BASIC jest uniwersalnym językiem konwersacyjnym umożliwiającym bezpośrednią komunikację operatora z systemem oraz przemiennie wykonywanie pod kontrolą systemu różnych działań. Język ten posiada szereg standardowych funkcji arytmetycznych, które pozwalają na stosunkowo proste oprogramowanie złożonych działań. W BASIC-u oprogramowano algorytmy obliczenia parametrów statystycznych badanych rozkładów oraz algorytmy wykonujące testy  $\chi^2$ , Bartletta i Studenta.

Poniżej podano rozmieszczenie poszczególnych programów w PAO, zaznaczono adresy początku każdego z nich w kodzie ósemkowym.

Minisystem - 70

Assembler - 1730

Program organizacyjny - 4740

Podprogramy operacji arytmetycznych programowanych

1/ DZ2 - dzielenie - 5400

2/ WP2 - wprowadzanie danych - 5650

3/ WY2 - wyprowadzanie danych - 6014

4/ DQ2 - dodawanie - 6353

Program wczytywania danych - 6453

Program wydruku danych - 6556

Program wydruku rozkładu liczebności - 7511

Program wypisu danych na oscyloskop - 7663

Dane - 7740 - 17777

## 5. Materiał doświadczalny i sposób przeprowadzenia eksperymentu.

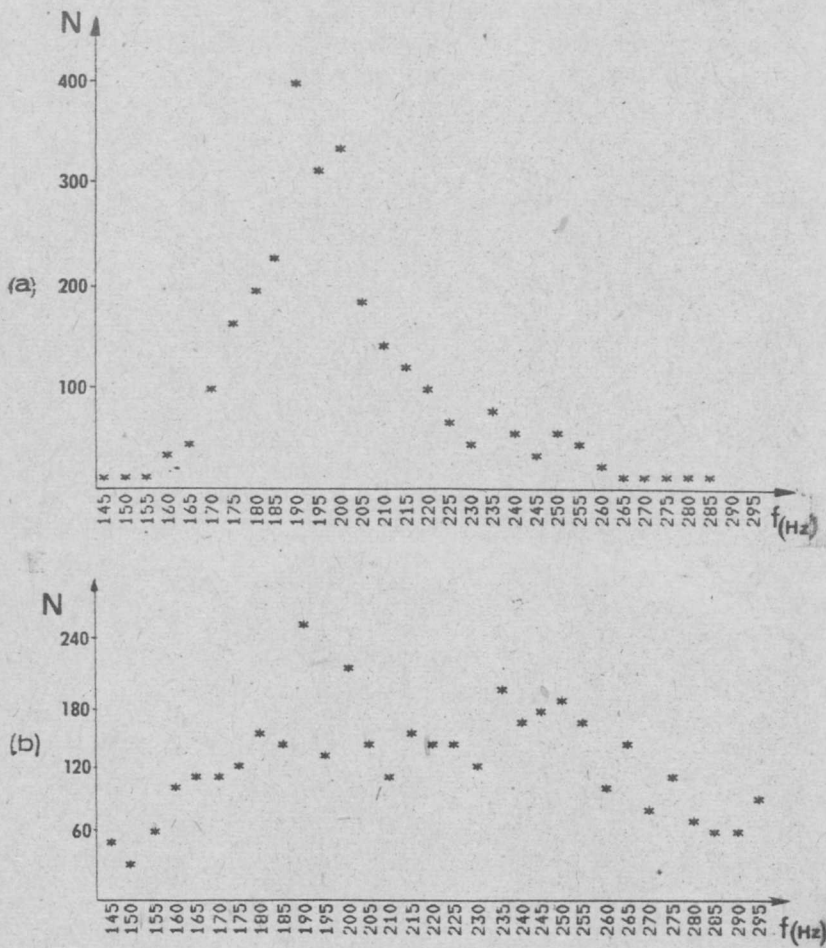
### 5.1. Wybór materiału fonetycznego.

W sposób ogólny rozkłady statystyczne były już uprzednio badane. Opierając się na pracach [7], [8] sposób przeprowadzenia eksperymentu ustalono następująco :

Każda badana osoba przeczytała ten sam tekst gazetowy o rozciągłości 50 s w różnych odstępach czasu, po godzinie pierwszego czytania, w dniu następnym oraz po tygodniu. Dodatkowo każda z badanych osób opowiadała improwizowany przez siebie 50-sekundowy tekst. W ten sposób otrzymano dla każdej osoby po 4 czytania tego samego tekstu (dla 3 osób po 3 próby z powodu nieuczestniczenia we wszystkich eksperymentach) oraz po jednej próbie swobodnego tekstu mówionego. Materiał został zapisany na taśmie i stanowił podstawę do dalszych analiz. Wobec oczywistej niemożliwości przeprowadzenia badania prostego, realizując badanie reprezentacyjne należy zapewnić reprezentatywność próby. Nie zwróciwszy dostatecznej uwagi na zagadnienie doboru obiektów próby nie można dokonywać precyzyjnego określenia zasięgu wniosków.

W doświadczeniu badanych było 9 osób, 5 mężczyzn i 4 kobiety (dobór osób był losowy - różne możliwości głosowe, różny wiek, wykształcenie) . Ponieważ czytany tekst gazetowy był dla wszystkich osób nieciekawym, można założyć, że zaangażowanie emocjonalne było niewielkie. Natomiast 50-sekundowy tekst mówiony odzwierciedlał predyspozycje intonacyjne badanych osób. W celu zbadania wpływu intonacji na kształtowanie się rozkładów wartości  $F_0$  wykonano dodatkowo następujący eksperyment. Dwie osoby miały za zadanie przeczytać ten sam 50-sekundowy tekst gaztowy naturalnie, lecz raz w sposób zupełnie obojętny, a drugi raz z zainteresowaniem. Dla jednej z osób podano wyniki analizy parametrów statystycznych w Tabl. 6, a rozkłady wartości  $F_0$  przedstawiono na ryc. 7.

Przy tak specyficznym zagadnieniu jakim jest analiza częstotliwości podstawowej sygnału mowy, aby osiągnąć dostateczną reprezentatywność badania, należałoby uwzględnić zarówno większą liczbę mówców i dokładniejsze zbadanie wszystkich czynników mających wpływ na określoną realizację próby. Przyjęta



Ryc.7. Rozkłady wartości  $F_0$ . Głos GD.

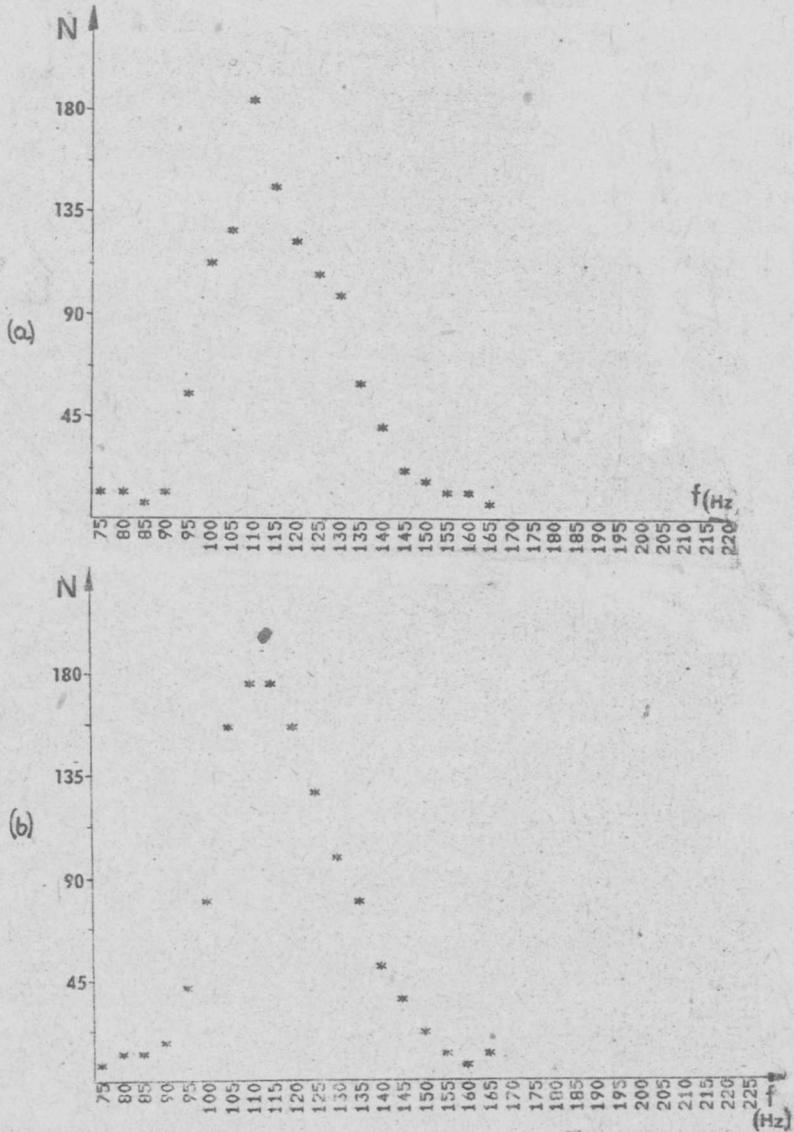
(a) czytanie tekstu w sposób obojętny

(b) czytanie tekstu ze zróżnicowaną intonacją.

ilość i rodzaj materiału językowego służyła głównie dla celu weryfikacji przyjętej techniki pomiarowej i obliczeniowej. Zwiększenie ilości materiału do analizy nie stanowi trudności a czas analizy i przetwarzania danych jest uwarunkowany głównie obsługą operatorską, ponieważ dysponując nawet małą maszyną cyfrową realizację algorytmów otrzymuje się prawie natychmiastowo.

## 5.2. Przygotowanie oraz analiza sygnału.

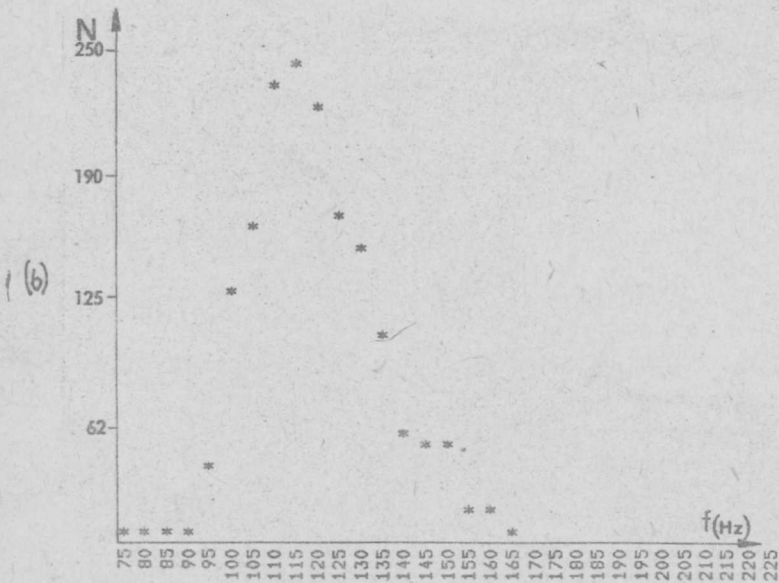
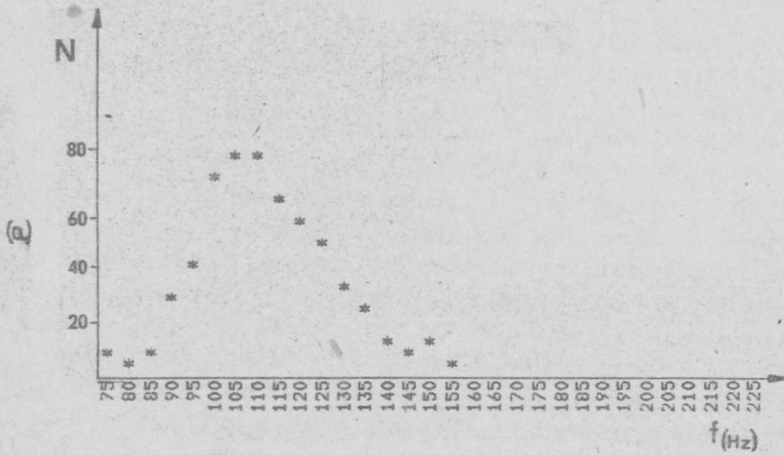
Właściwa metoda przetwarzania sygnału jest głównie zależna od charakteru badanego zjawiska oraz od technicznych celów przetwarzania. Przygotowanie sygnału odnosi się do procesu ekstrakcji tonu podstawowego sygnału mowy. Dokładność wyniku ekstrakcji parametru  $F_0$  zależy od wielu czynników wynikających zarówno z cech akustycznych sygnału, jak i przyjętej metody ekstrakcji. W celu wyboru najlepszego wariantu pomiarowego oraz określenia stopnia dokładności ekstrakcji zostały wykonane tonogramy z wybranego materiału. Dodatkowo jako kontrolę wprowadzono przy każdej analizie tonogramy cyfrowe. Następnym etapem jest proces dyskretyzacji sygnału. Próbkowanie dla cyfrowej analizy jest dokonywane zwykle w jednakowych odstępach czasu. Problemem jest przyjęcie odpowiedniego odstępu próbkowania. Przyjęto następujące wartości : 10, 20, 40 ms oraz częstotliwość próbkowania synchroniczną z częstotliwością podstawową. Po wykonaniu wstępnych analiz okazało się, że próbkując sygnał synchronicznie z  $F_0$  otrzymany z 25 s tekstu można uzyskać podobne wyniki jak przy częstotliwości próbkowania 20 ms sygnału otrzymanego z 50 s tekstu. Z tego powodu oraz kierując się zasadą, że każdy okres jest nośnikiem o równej wartości informacyjnej przyjęto zmienną częstotliwość próbkowania, synchroniczną z  $F_0$ . Po przeprowadzeniu dyskretyzacji sygnału lecz przed przystąpieniem do praktycznej analizy należy na otrzymanych danych dokonać pewnych operacji. Pierwszym zadaniem jest określenie zależności między danymi cyfrowymi a wartościami odpowiedniej wielkości fizycznej w jej jednostkach miary, a następnie należy dane pogrupować. Podział danych na przedziały klasowe umożliwia znaczną redukcję danych, przy zachowaniu podstawowych własności statystycznych zbioru obserwacji. Dla każdego analizowanego tekstu dokonano rozkładów liczebności  $F_0$  i wyprowadzono w formie graficznej na drukarce wierszowej.



Ryc. 8. Rozkłady wartości  $F_0$  przy częstotliwości próbkowania równej 20 ms. Głos WJ.

(a) z 25 sekundowego tekstu

(b) z 50 sekundowego tekstu.



Ryc.9. Rozkłady wartości  $F_0$  przy częstotliwości próbkowania synchronicznej z okresem podstawowym. Głos WJ.

(a) z pierwszych 25 sekund tekstu

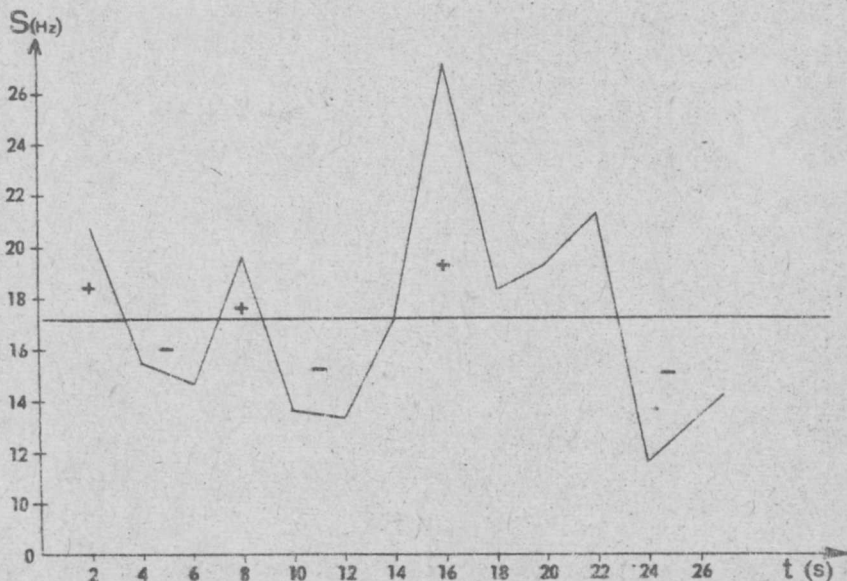
(b) z drugich 25 sekund tekstu.



Ryc. 8 przedstawia rozkłady liczebności  $F_0$  przy częstotliwości próbkowania równej 20 ms, a ryc. 9 przy częstotliwości próbkowania synchronicznej z  $F_0$ .

W celu zbadania stacjonarności analizowanego sygnału, 25 s tekstu przypadkowo wybranego podzielono na równe czasowo odcinki (dokonano tego przy pomocy tonogramów), w których zostały określone podstawowe parametry. Zbadano ciąg kolejnych wartości pod względem obecności podstawowego trendu i zmian innych niż te, które wynikają z oczekiwanej zmienności losowej. Przeprowadzono nieparametryczny test oparty na teorii serii. Każda z obserwacji zaliczona została do dwóch wyłączających się kategorii  $\oplus$  lub  $\ominus$ . Liczba serii występujących w ciągu daje wskazówkę czy wyniki są niezależnymi obserwacjami tej samej zmiennej.

Wyniki testu przedstawia ryc. 10.



Ryc. 10. Zmienność odchylenia standardowego wyznaczonego w 2-sekundowych przedziałach 26 sekundowego analizowanego odcinka.

Tablica 4

Wyniki analizy statystycznej dla 9 osób. W kolumnie 1 podano inicjały badanych osób, umieszczone przy nich numery odnoszą się do poszczególnych powtórzeń czytania tekstu, indeks t oznacza tekst mówiony. Numery 1 i 2 oznaczają odpowiednio pierwsze i drugie 25 sekund badanego 50-sekundowego fragmentu.

1	2	3	4	5	6	7	8	9	10	11
	N	$\bar{x}$	$s^2$	s	$d_3$	$d_4$	v%	$\chi^2$	$\chi^2$	t
HK1	1	2289	158.86	360.81	18.99	0.33	0.36	11.95	113	2.04
	2	2294	159.96	300.44	17.33	0.3	0.22	10.8	150	19.7 ⊖
HK2	1	2466	169.62	362.81	19.04	0.07	0.1	11.23	151	1.31
	2	1865	168.89	305.71	17.48	0.26	-0.41	10.35	171	15.4 ⊖
HK3	1	2641	174.49	518.36	22.76	-0.21	0.29	13.04	248	2.4
	2	2646	175.94	400.24	20.01	-0.12	-0.05	11.37	170	44.29 ⊖
HK4	1	2553	176.13	447.11	21.14	-0.05	0.38	12.01	258	1.42
	2	2481	176.95	383.49	19.58	-0.03	0.6	11.07	143	14.16 ⊖
HKt	1	2224	152.78	288.05	16.97	0.63	0.65	11.11	227	1.04
	2	2236	152.27	243.09	15.59	0.09	-0.12	10.23	134	16.03 ⊖
WJ1	1	1693	114.12	250.82	15.84	0.38	0.42	13.88	178	6.16
	2	1844	117.36	237.89	15.42	0.62	0.81	13.14	169	1.2 ⊖
WJ2	1	1922	121.81	360.7	18.99	0.21	-0.19	15.8	93	2.7
	2	1968	120.13	398.7	19.96	0.45	0.43	16.38	196	4.8
WJ3	1	2032	111.39	243.31	15.6	0.5	0.13	14.0	200	2.7
	2	2271	110.09	237.9	15.42	0.49	0.15	14.01	153	0.26
WJt	1	2393	121.22	246.8	15.7	0.31	0.54	12.96	99	27.5
	2	2267	137.57	565.2	23.77	-0.27	-0.32	17.27	131	390 ⊖ ⊖
PJ1	1	2009	142.16	281.21	16.77	-0.04	0.82	11.79	97	4.38
	2	1623	144.89	402.5	20.06	0.21	-0.28	13.84	72	58.09 ⊖
PJ2	1	2116	141.36	290.98	17.06	0.65	0.18	12.06	133	2.0
	2	1917	142.43	282.87	16.82	0.64	0.4	11.81	106	0.4
PJ3	1	2572	155.07	478.68	21.88	-0.18	1.07	14.11	290	7.36
	2	2193	150.5	428.88	20.71	0.17	1.03	13.76	395	7.1 ⊖
PJ4	1	2120	141.52	253.06	15.91	0.14	1.25	11.24	236	0.31
	2	2023	141.37	242.76	15.58	0.05	1.4	11.02	252	0.89
PJt	1	1990	152.35	386.35	19.65	0.14	0.36	12.89	188	12.17
	2	1858	145.38	248.87	15.77	0.4	1.95	10.85	395	91 ⊖ ⊖

1	2	3	4	5	6	7	8	9	10	11
	N	$\bar{x}$	$s^2$	s	$d_3$	$d_4$	v%	$\chi^2$	$\chi^2$	t
AM1	1 1724	127.3	475.06	21.79	0.56	0.28	17.11	167		6.92
	2 1638	132.2	388.76	19.72	0.05	-0.33	14.9	88	16.5	⊖
AM2	1 1600	116.09	325.39	18.04	0.84	0.86	15.54	302		1.8
	2 1605	117.2	290.94	17.06	0.58	0.49	14.5	224	5	
AM3	1 1678	112.16	254.44	15.95	0.53	0.08	14.2	209		10.8
	2 1654	118.2	258.86	16.09	0.28	-0.03	13.6	196	0.12	⊖
AM4	1 1737	119.55	356.43	18.88	0.47	-0.52	15.79	142		4.41
	2 1565	122.4	338.97	18.41	0.41	-0.28	15.04	186	1.04	⊖
AMt	1 1745	113.48	181.11	13.46	0.78	1.4	11.86	203		0.94
	2 1715	113.05	174.83	13.22	1.2	4.6	11.69	600	0.54	
KM1	1 2324	147.15	502.15	22.41	0.19	0.22	15.23	184		1.9
	2 2624	145.91	552.92	23.51	0.27	-0.52	16.11	247	5.7	
KM2	1 2157	144.31	527.18	22.96	0.26	-0.19	15.91	164		3.9
	2 2127	147.34	715.9	26.76	0.35	-0.6	18.16	101	49.9	⊖
KM3	1 2235	132.77	393.78	19.8	0.09	0.7	14.91	276		6.74
	2 2177	136.88	426.29	20.65	0.23	-0.2	15.09	253	3.46	⊖
KMt	1 2460	143.63	386.7	19.66	0.15	0.7	13.69	289		4.69
	2 2303	141.17	271.09	16.46	-0.36	1.9	11.66	1270	74.3	⊖
DK1	1 2448	213.43	623.11	24.96	0.65	0.45	11.69	435		0.96
	2 2421	212.73	655.74	25.6	0.59	0.62	12.04	495	1.6	
DK2	1 2564	214.29	597.44	24.4	0.58	0.39	11.41	572		0.24
	2 2612	214.12	610.75	24.7	0.63	0.22	11.54	505	0.31	
DK3	1 2720	214.93	740.54	27.2	0.36	-0.12	12.66	352		3.3
	2 2542	212.35	691.88	26.3	0.54	0.04	12.39	260	3.01	
DKt	1 2522	214.18	754	27.46	0.56	-0.34	12.82	441		10.3
	2 2217	222.23	687.2	26.22	0.35	-0.14	11.79	499	5.04	
ES1	1 2726	212.92	456.11	21.36	0.37	0.37	10.03	284	0.81	5.04
	2 2459	209.84	439.7	20.97	0.67	1.16	9.99	460		⊖
ES2	1 2682	210.17	462.77	21.51	0.78	1.05	10.23	735		1.28
	2 2652	210.94	497.65	22.31	0.71	0.63	10.57	640	3.52	
ES3	1 2626	208.86	540.28	23.24	0.53	0.65	11.13	470		1.48
	2 2987	207.96	524.66	22.91	0.49	0.21	11.02	420	0.62	

1	2	3	4	5	6	7	8	9	10	11
	N	$\bar{x}$	$s^2$	s	$d_3$	$d_4$	$v\%$	$\chi^2$	$\chi^2$	t
ES4	1 2665	203.93	427.24	20.67	0.69	0.67	10.13	240	11.7	1.7
	2 2829	202.95	486.99	22.07	0.7	0.01	10.87	160		
ES1	1 2986	209.29	541.6	23.27	0.82	0.69	11.12	300	166	0.57
	2 2994	209.66	730.01	27.02	0.61	0.27	12.88	570		
GD1	1 2661	217.39	467.4	21.61	0.18	0.72	9.94	236	25.6	4.85
	2 2945	220.33	566.33	23.79	0.41	0.13	10.79	374		
GD2	1 2913	223.52	636.1	25.22	0.78	0.55	11.28	563	0.16	4.58
	2 2643	226.64	645.95	25.42	0.39	0.16	11.22	213		
GD3	1 2857	220.29	586.65	24.22	0.75	0.71	10.99	473	18.2	4.35
	2 2790	217.36	689.05	26.25	0.89	0.67	12.07	592		
GD4	1 2764	210.32	526.44	22.94	0.76	0.79	10.91	319	0.66	0.03
	2 2850	210.34	510.56	22.59	0.94	1.3	10.74	414		
GDt	1 3512	231.18	588.47	24.58	0.08	0.44	10.63	200	4.11	29.17
	2 3011	213.3	631.83	25.14	0.9	0.98	11.79	280		
DG1	1 2926	232.93	765.84	27.67	0.28	-0.25	11.88	205	9.1	3.11
	2 3087	230.64	855.26	29.24	0.32	-0.5	12.68	317		
DG2	1 2990	226.09	800.38	28.29	-0.19	0.07	12.51	417	9.51	3
	2 2967	228.23	714.74	26.73	0.06	0.02	11.71	402		
DG3	1 2977	227.72	822.02	28.67	-0.26	-0.49	12.59	406	0.46	2.25
	2 2995	226.06	801.88	28.32	-0.05	-0.47	12.53	223		
DG4	1 3013	219.83	608.98	24.68	0.23	0.05	11.23	190	0.05	2.95
	2 3110	221.69	604.20	24.58	0.15	-0.2	11.08	196		
DGt	1 3029	242.81	867.64	29.46	-0.63	-0.63	12.13	188	45	-
	2 1967	208.52	581.84	24.12	-0.3	-0.3	11.57	144		

### 6. Opracowanie i interpretacja wyników.

Wyniki analizy statystycznej przedstawiono w Tabelicy 4.

W większości przypadków daje się zauważyć, że rozkłady liczebności z pierwszych i drugich 25 sekund tego samego 50-sekundowego analizowanego odcinka tekstu są do siebie podobne. Równoważnych par rozkładów posiadających podobne średnie, wariacje, asymetrię i eksces jest 16, całkowicie różnych par pod względem tych samych parametrów jest tylko 5, pozostałe różnią się między

sobą albo średnią, albo wariancją, asymetrią, ekscesem. Ogólna suma badanych par wynosi 42. Należy zaznaczyć, że w 4 przypadkach pary nie były ekwiwalentne fonetycznie - nagrano zbyt krótki tekst, tych par nie należałoby brać pod uwagę. Mimo to wątpliwe wydaje się oszacowanie 25 s jako stałej długości tekstu dla każdej osoby i dla każdego rodzaju mowy. Dla niektórych wypowiedzi stałość parametrów statystycznych daje się zauważyć już po 15-20 s, dla innych po 30-40, a nawet po 50 sekundach. Wykonano test stacjonarności, dla jednej dowolnie wybranej 25-sekundowej wypowiedzi. W Tablicy 5 umieszczono wartości parametrów statystycznych koniecznych dla wykonania testu stacjonarności.

Otrzymujemy następujący szereg wartości odchyień standardowych: 11.5, 13.3, 13.6, 13.9, 14.9, 15.4, 17.1, 18.3, 19.06, 19.8, 20.8, 21.3, 26.9

Medianą jest podkreślona wartość. Ryc. 10 przedstawia wahania odchylenia standardowego wokół mediany. Na poziomie istotności 0.05 odczytano wartości krytyczne dla  $r_{n, \frac{\alpha}{2}} < r < r_{n, \frac{\alpha}{2}}$

czyli dla  $r_{6, 0.975} < r < r_{6, 0.025}$  jako  $3 < r < 10$ .

Ponieważ  $r$  (liczba serii w badanym ciągu) wynosi 6, to hipotezę o jego stacjonarności można przyjąć z dużym stopniem pewności. Na ryc. 11 i 12 przedstawiono dystrybuanty badanych rozkładów na siatce rozkładu normalnego. Jeżeli badany rozkład jest normalny, to jego dystrybuenta ma kształt linii prostej.

W większości przypadków dla każdej osoby analizowano po 10 rozkładów, których dystrybuanty nakładały się jedna na drugą lub przecinały się wzajemnie. Dla uzyskania czytelności charakterystyk zaznaczono jedynie dla każdej z osób skrajne dystrybuanty.

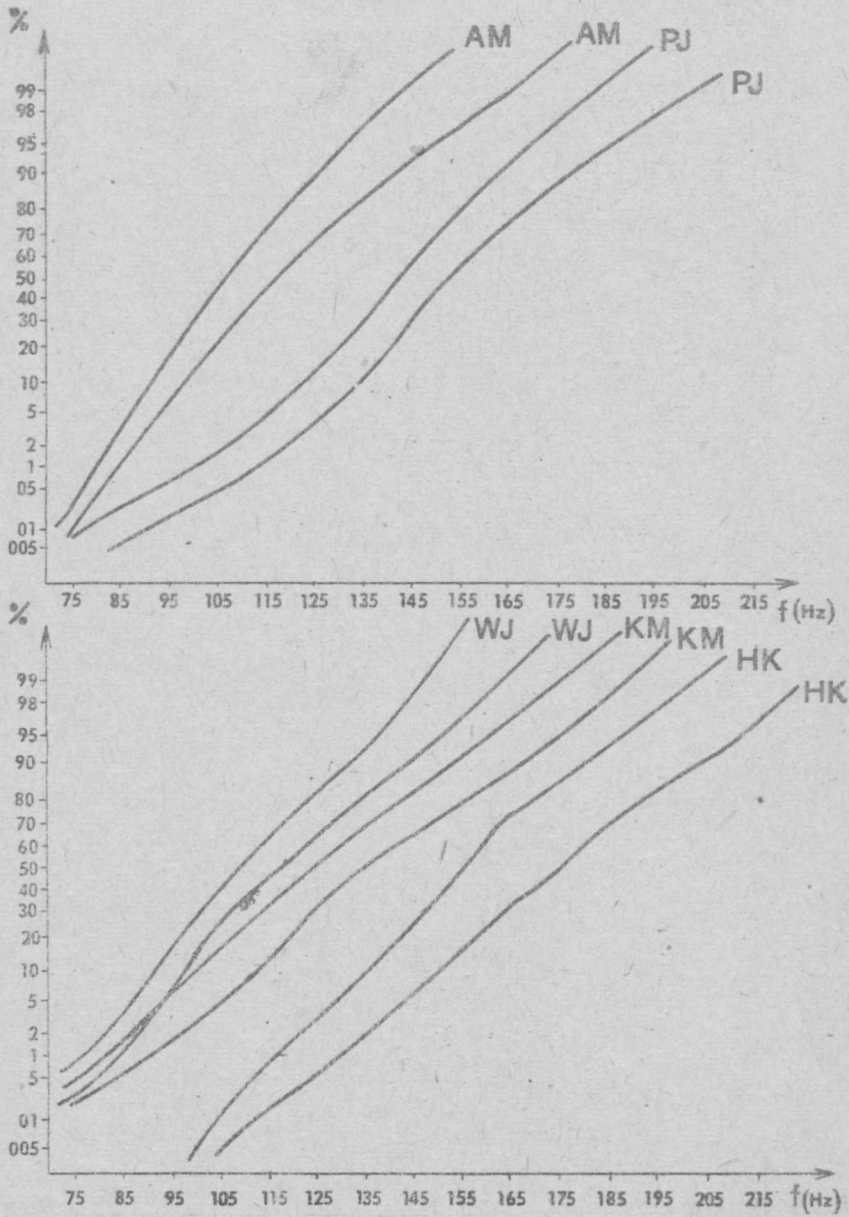
Naszkicowane krzywe w niektórych przypadkach mogą być aproksymowane liniami prostymi, w innych natomiast odchylenia krzywych są dość wyraźne i należy przeprowadzić bardziej dokładny test. Wykonany test  $\chi^2$  dał ujemne wyniki, przedstawione w kolumnie 9 Tablicy 4, nawet dla poziomu istotności 0.001 przy liczbie stopni swobody 25. Na podstawie tego, co powiedziano wcześniej, istnieją jednak przesłanki do traktowania tych rozkładów jako zbliżonych do normalnych. Mamy tutaj do czynienia z populacją



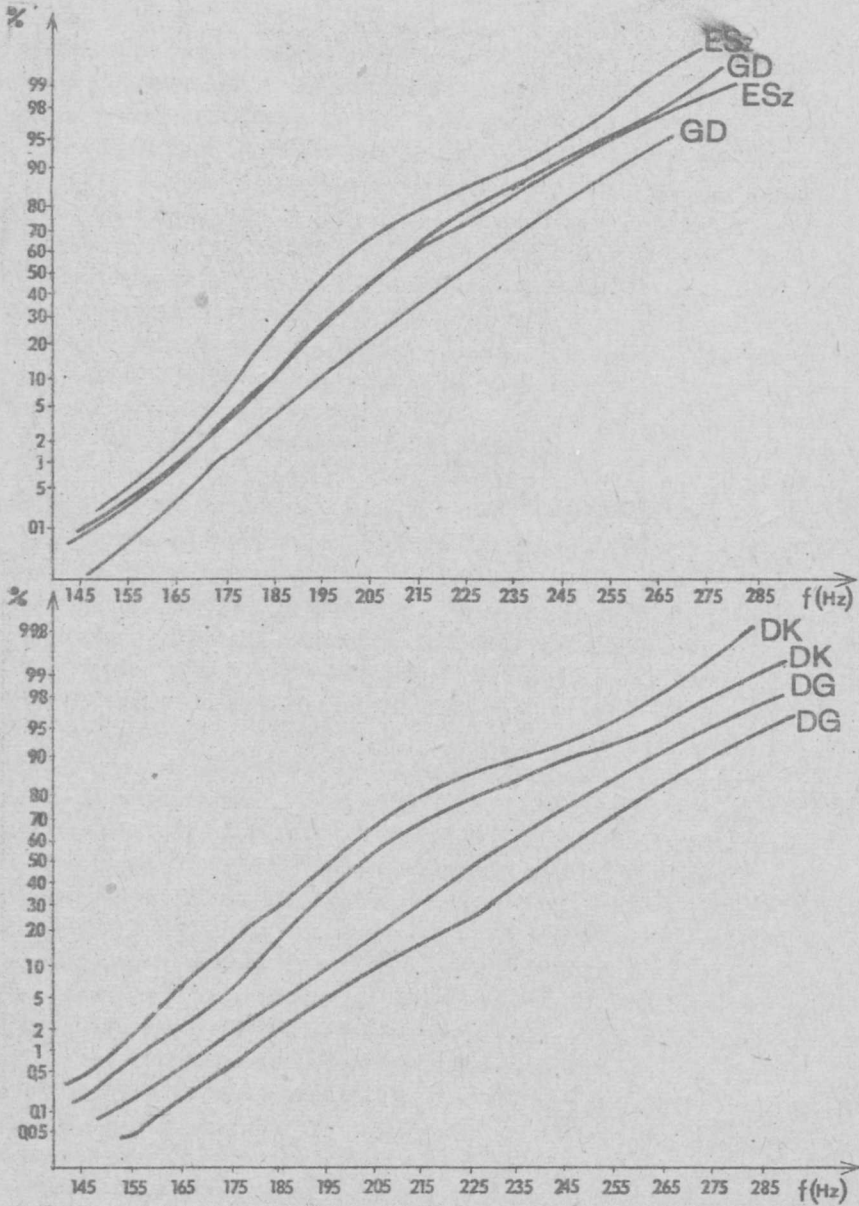
Tablica 5.  
 W wierszu 1 Tablicy zaznaczono podział 26-sekundowego odcinka czasowego na  
 trzynaście 2-sekundowych podziałów. W następnych wierszach, odpowiednio dla  
 każdego przedziału umieszczono wyznaczone parametry statystyczne.

	1	2	3	4	5	6	7	8	9	10	11	12	13
czas w sek.	0-2	2-4	4-6	6-8	8-10	10-12	12-14	14-16	16-18	18-20	20-22	22-24	24-26
2 N	158	122	111	135	144	121	120	63	118	118	109	112	126
3 $\bar{x}$	131.3	118.9	109.9	136.5	122.4	120.7	115.8	151.1	127.9	114.6	128.8	128.2	116.1
4 $S^2$	432.6	237.1	222.2	364.1	194.8	185.3	294.2	724	334.6	353.4	452.2	133.1	179.5
5 S	20.8	15.4	14.9	19.8	13.9	13.6	17.1	26.9	18.3	19.0	21.3	11.5	13.3





Ryc. 11. Dystrybuanty badanych rozkładów. Głosy męskie.



Ryc.12. Dystrybuanty badanych rozkładów. Głosy żeńskie.

składającą się z części opisanej normalnym rozkładem i z części zawierającej inne rozkłady. Pewne założenia o rozkładach populacji są konieczne dla większości statystycznych procedur. Często zdarza się, że zmiana skali, np. taka jak użycie pierwiastka kwadratowego lub logarytmu obserwacji, może spełnić więcej założeń. Zmiana skali jest trudna do określenia. W pracach W. Jassema i G. Ramiszwiliego sugeruje się transformację log normalną. Przeprowadzono dalsze obliczenia statystyczne (zakładające normalność charakterystyk), uwzględniając przy tym asymetrię i eksces jako miary odchylenia rozkładów empirycznych od normalnych oraz analizując wyniki z dużą uwagą. Porównując ze sobą wartości asymetrii daje się zauważyć dość ogólną tendencję do asymetrii prawostronnej. W głosie DG widoczna jest wyraźna skłonność do lewostronnej, co może być w tym przypadku cechą indywidualizującą głos. Porównano wariancję dla pierwszych i drugich 25 sekund tego samego 50-sekundowego odcinka, obliczając wartość funkcji Bartletta. Jako punkt krytyczny przyjęto wartość  $\chi^2(0.01; 1) = 6.6$ . Zaznaczono znakiem (-) te próby, które nie spełniają założenia równości wariancji. Analizując podane wartości w kolumnie 10 Tablicy 4 zauważa się, że w większości przypadków wartości testu leżą poza punktem krytycznym.

Przeprowadzono porównanie średnich wartości z pierwszych i drugich 25 sekund dla każdej próby za pomocą testu t. Jako punkty krytyczne przyjęto wartości  $t(0.01; \infty) = 2.57$  oraz  $t(0.001; \infty) = 3.29$ . Dla poziomu istotności 0.01 zaznaczono (-) w kolumnie 11 Tablicy 4 te próby, które nie spełniają założenia równości średnich.

Porównując ze sobą wyniki analiz czterech czytań oraz tekstu mówionego u większości osób widoczne są najbardziej skrajne wartości parametrów dla tekstu mówionego. Biorąc pod uwagę odchylenia średnich dla wszystkich osób możemy powiedzieć, że parametry statystyczne wahają się między jedną rejestracją a inną w warunkach określonych uprzednio średnio o około 6%. W Tablicy 6 podano parametry rozkładów wartości  $F_0$  przedstawionych na ryc. 8. Parametry w pierwszym wierszu Tablicy odnoszą się do nieemocjonalnego sposobu czytania, a drugim do czytania ze zróżnicowaną intonacją. Analizując podane wyniki

oraz ryc. 8, jako najbardziej charakterystyczną różnicę zauważa się 2,5-krotny wzrost wariancji w przypadku czytania emocjonalnego w stosunku do czytania obojętnego i konsekwentnie zmianę współczynnika ekscesu z wartości  $d_4 = -1$  do wartości  $d_4 = 1$ .

Tablica 6.  
Parametry statystyczne rozkładów  $F_0$ .

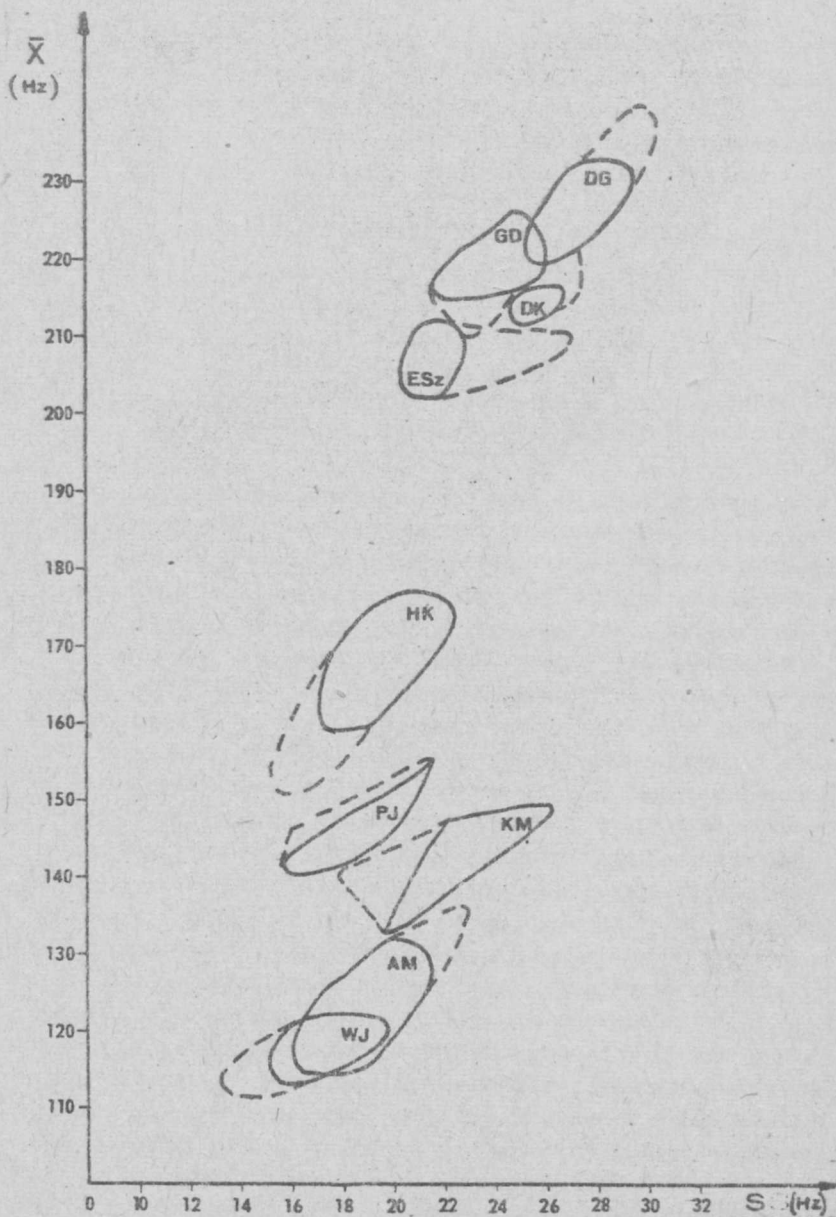
	N	$\bar{x}$	$s^2$	s	$d_3$	$d_4$
1	2992	197.5	514.7	22.6	0.9	1.16
2	3430	210	1319	36.3	0.27	-1.02

Na ryc. 13 przedstawiono obliczone wartości średniej arytmetycznej oraz odchylenia standardowego dla wszystkich prób każdej z badanych osób. Linia przerywaną zaznaczono obszar, w którym znajdują się parametry wyznaczone z tekstu mówionego.

#### 7. Analiza statystyczna $F_0$ w pracach innych autorów.

W języku polskim zagadnieniem statystycznych własności częstotliwości podstawowej zajmował się W. Jassem (por. [7], [8], [12]). Prace te z braku odpowiedniej aparatury wykonane zostały znacznym nakładem pracy. Sugeruje się w nich, poprzez stwierdzenie różnic międzyosobniczych oraz wewnątrzosobniczych na podstawie  $F_0$ , znalezienie inwariantów nadających się do wykorzystania przy identyfikacji głosów i automatycznym rozpoznawaniu mowy. Porównując otrzymane w przedstawionej pracy wyniki pod względem obliczonych parametrów statystycznych z wynikami podanymi w powyższych pracach nie zauważono rozbieżności.

W języku francuskim podobnymi zagadnieniami zajmował się Boë (por. [2]). Analizie poddano różne rodzaje mowy: proza, poezja, zdania oderwane, z górnym czasem trwania każdego zapisu równym 50 sekund. Rozkład częstotliwości podstawowej został oszacowany jako normalny. Rozpiętość wartości średnich i odchyłeń standardowych jest tutaj znacznie większa ze względu na różnorodność badanego materiału. Dla przykładu Boë podaje rozpiętość współczynnika zmienności - dla mężczyzn 7-26, dla kobiet od 9 do 29. W niniejszej pracy rozpiętość ta jest mniejsza, dla mężczyzn



Ryc. 13. Rozrzut wartości średniej  $F_0$  oraz odchylenia standardowego dla badanych głosów.



wynosi 10-17, dla kobiet 9-13.

Bardzo dokładnie statystyczną analizę dla języka rosyjskiego przeprowadził Ramiszwili (por. [11]). Został zanalizowany obszerny materiał z 26 godzin tekstu dla 102 osób. Porównanie tych rozkładów wskazało, że przybliżenie logarytmiczno normalne jest lepsze niż normalne. Charakter funkcji rozkładu zależy tak od mówcy, jak i od tekstu. Emocjonalna wymowa wykazuje w większości przypadków bimodalne rozkłady, natomiast narracyjny sposób czytania daje zwykle jednomodalne z prawostronną asymetrią. Jako podstawowy parametr posiadający największą wartość informacyjną przyjęto wartość średnią rozkładów prawdopodobieństwa. Porównując przeprowadzone prace w języku polskim, francuskim, rosyjskim widoczna jest zbieżność poglądów, wyników badań i wniosków. W związku z tym należy przypuszczać, że niezależnie od języka, podstawowe prawa związane z samą "produkcją" tonu krtaniowego pozostają stałe. Niepokojący jest jednak fakt traktowania w cytowanych pracach rozkładów  $F_0$  jako normalnych lub logarytmiczno normalnych. Biorąc pod uwagę próby dostosowania tych rozkładów do typu gamma (por. [11]) oraz fakt, iż przeprowadzone statystyczne teksty wskazują istotne odchylenia od normalności konieczne jest dokładniejsze zbadanie hipotezy normalności rozkładów wartości  $F_0$ .

#### 8. Wnioski.

Analizę matematyczną częstotliwości podstawowej należy przeprowadzać traktując ten sygnał jako losowy o stopniu zdeterminowania zależnym głównie od przyjętych warunków badania. Zakładając nieemocjonalny sposób mówienia, rozkłady wartości chwilowych z lingwistycznie ekwiwalentnych wypowiedzi, wykazują większą zmienność międzysobniczą niż wewnątrzsobniczą, pod warunkiem analizy zbioru złożonego z kilku parametrów statystycznych. Możliwości dyskryminacyjne za pomocą średniej  $F_0$  są ograniczone, niemniej jednak można dokonać podziału głosów na kilka nie pokrywających się grup. Dodatkowo uwzględniając inne parametry - odchylenie standardowe, miary asymetrii i ekscesu (często bardzo charakterystyczne dla niektórych głosów) można przeprowadzić bardziej dokładnie rozdział głosów. Taki podział może się okazać przydatny w systemach identyfikacji osób. Biorąc pod uwagę pracę [8] postulującą normalizację

czasowo-częstotliwościową przebiegów  $F_0$  na podstawie rozkładów wartości chwilowych tego parametru, należy stwierdzić, że proponowany sposób analizy statystycznych własności częstotliwości podstawowej jest przydatny w znalezieniu inwariantnych cech badanego sygnału. Określenie natomiast efektywności przyjętych metod wymaga dokładniejszego zbadania sposobu analizy  $F_0$  poprzez poszerzenie materiału doświadczalnego. Statystyczne własności  $F_0$  mogą być również pomocne w zagadnieniach rehabilitacji osób z uszkodzonym układem krtańowym (por. [6]). Biorąc pod uwagę stosunkowo niewielki nakład pracy związany z przetwarzaniem tego parametru i jego duży stopień użyteczności należy spodziewać się, że odegra on istotną rolę w rozwiązywaniu szeregu podstawowych problemów automatycznego rozpoznawania mowy.

BIBLIOGRAFIA

- [1] BENDAT, J., PIERSOL, A. : Metody analizy i pomiaru sygnałów losowych, PWN, Warszawa, 1976.
- [2] BOE, L., CONTINI, J. : A statistical Study of Laryngeal Frequency, *Phonetica*, 32/1975.
- [3] DIXON, W., MASSEY, F. : Intruduction to statistical analysis, Tokyo, 1969.
- [4] FREUND, J. : Podstawy nowoczesnej statystyki, PWE, Warszawa, 1968.
- [5] GORALSKI, A. : Metody opisu i wnioskowania statystyczne w psychologii, PWN, Warszawa, 1976.
- [6] GUBRYNOWICZ, R., MIKIEL, W., ŻARNECKI, P. : Acoustical analysis for the evaluation of laryngeal dysfunction in the case of vocal cord paralysis, *Speech Analysis and Synthesis*, vol. V, w druku.
- [7] JASSEM, W. : Parametry statystyczne rozkładów średnich wartości  $F_0$  jako cech osobniczych głosu, *Prace IPPT*, 58/1970.
- [8] JASSEM, W., KUDELA-DOBROGOWSKA, K. : Inwarianty w przebiegach parametru  $F_0$ , *Prace IPPT*, 25/1973.
- [9] KUBZDELA, H. : Automatyczna ekstrakcja częstotliwości tonu podstawowego oraz pierwszych trzech formantów sygnału mowy, *Prace IPPT*, 51/1973.
- [10] MYTKOWSKI, K. : Kanał funkcji analogowych typ KF-01 do wprowadzania i wyprowadzania informacji w systemie "On-line" do/z pamięci minikomputera Momik 8B/100, *Prace IPPT*, 39/1976.
- [11] RAMISZWILI, G. : Rieczewoj signal i individualnost' gołosa, Tbilisi, 1976.
- [12] STEFFEN-BATOG, M., JASSEM, W., GRUSZKA-KOSCIELAK, H. : Statistical distribution of short-term  $F_0$  values as a personal voice characteristic, *Speech Analysis and Synthesis*, vol. II, Warszawa, 1970.

