

Henryk Kubziela

**WYZNACZANIE CHARAKTERYSTYCZNEGO
FRAGMENTU SAMOGŁOSKOWEGO
I POMIAR CZĘSTOTLIWOŚCI FORMANTÓW
DLA AUTOMATYCZNEJ KLASYFIKACJI
I IDENTYFIKACJI SAMOGŁOSEK**

41 / 1979

P. 269



WARSZAWA 1979

Praca wpłynęła do Redakcji dnia 15 marca 1979 r.

Zarejestrowana pod nr 44/1979



57212



Na prawach rękopisu

Instytut Podstawowych Problemów Techniki PAN

Nakład 130 egz. Ark.wyd. 1,5. Ark.druk. 0,9.

Oddano do druku w wrześniu 1979 r.

Nr zamówienia 553/0/79

Warszawska Drukarnia Naukowa, Warszawa,
ul. Śniadeckich 8

Henryk Kubzdela
Pracownia Fonetyki Akustycznej IPPT PAN

WYZNACZANIE CHARAKTERYSTYCZNEGO FRAGMENTU SAMOGŁOSKOWEGO
I POMIAR CZĘSTOTLIWOŚCI FORMANTÓW DLA AUTOMATYCZNEJ
KLASYFIKACJI I IDENTYFIKACJI SAMOGŁOSEK.

Streszczenie

Praca przedstawia elementy modelu automatycznej klasyfikacji i identyfikacji polskich samogłosek. Podano opracowaną przez autora metodę identyfikacji środka samogłoski. Omówiono nowe elementy wcześniej opublikowanej metody ekstrakcji pasm formantowych. Jednym z nich jest automatyczna weryfikacja wyników ekstrakcji pasm formantowych bazująca na konfrontacji zmierzonych i możliwych zmian częstotliwości formantów. Zaprezentowano przykłady widm cyfrowych i wydruki wyników ekstrakcji pasm formantowych oraz przebiegu uśrednionego poziomu sygnału w pasmach formantowych.

1. Uwagi o założonym modelu rozpoznawania wyrazów.

Model automatycznego rozpoznawania mowy, którego dotyczy niniejsza praca bazuje na dwóch następujących założeniach :

1/ Jego funkcją ma być rozpoznawanie ograniczonego słownika izolowanych wyrazów,

2/ Rozpoznanie wyrazu następuje poprzez rozpoznanie poszczególnych fonemów tworzących ten wyraz.

Doniosłą rolę w strukturze wyrazów odgrywają samogłoski. Mają też one bardziej niż inne fonemy powtarzalną konfigurację widma i dają się fizycznie opisać przy pomocy niewielkiej liczby parametrów akustycznych. Dwa spośród tych parametrów, a mianowicie częstotliwości pierwszego i drugiego formantu zawierają istotne cechy dystynktywne poszczególnych samogłosek. Te parametry, jak wykazano między innymi w badaniach [1 , 2]

pozwalają sklasyfikować polskie samogłoski na tyle dobrze, że możliwa jest przy ich pomocy poprawna identyfikacja niewiđomych segmentów samogłoskowych występujących w rozpoznawanym wyrazie. Spośród głównych elementów zdefiniowanego wyżej modelu automatycznego rozpoznawania mowy podstawowe funkcje pełnią :

1/ Segmentator dzielący wyraz na segmenty przynależne do odpowiednich grup fonemowych charakteryzujących się swoistymi parametrami klasyfikacyjnymi.

2/ Ekstraktor parametrów klasyfikacyjnych.

W odniesieniu do samogłosek segmentator wskazuje ich położenie w obrębie wyrazu a ekstraktor mierzy częstotliwości pierwszego i drugiego formantu. Na obecnym etapie realizacji takiego modelu automatycznego rozpoznawania wyrazów opracowano metodę wyznaczania położenia środka samogłosek występujących w przetwarzanym wyrazie oraz ekstraktor częstotliwości pierwszych dwóch formantów samogłoskowych. Segmentator i ekstraktor zaprogramowano w minikomputerze Mera 303. Analizę widmową wykonuje układ analogowy. Poniżej omówiona zostanie idea segmentatora oraz innowacje wprowadzone do metody ekstrakcji częstotliwości pierwszych dwóch formantów samogłoskowych przedstawionej w pracy [3]. Danymi wyjściowymi dla ekstrakcji częstotliwości formantów są 43 parametry widmowe odczytywane co 20 ms na skomutowanym wyjściu wielokanałowego analizatora widma. Linie widmowe sygnału wyznaczone przez analizator reprezentują poziomy sygnału w 43 pasmach częstotliwości o jednakowej szerokości 80 Hz pokrywających łącznie zakres częstotliwości od 120 do 3560 Hz. Analizę widmową poprzedza odpowiednia korekcja widmowa sygnału zmniejszająca różnice poziomów formantów. Automatycznie i w czasie rzeczywistym dane z wyjścia wielokanałowego analizatora widma są wprowadzane do minikomputera Mera 303 przy pomocy specjalnie zaadoptowanego do tego celu układu nazwanego kanałem funkcji analogowych [5]. Na etapie wprowadzania ma miejsce programowe wygładzanie widma według następującego wzoru :

$$b_j = \frac{1}{m} \sum_{i=j}^{j+(m-1)} a_i \quad (1)$$

dla $j = 1, 2, \dots, 43-m$,

gdzie b_j i a_i są odpowiednio liniami widmowymi po i przed wygładzeniem. Biorąc pod uwagę szerokość pasm analizatora oraz minimalne odległości między sąsiadującymi ze sobą formantami przyjęto $m = 4$.

Zacytowana wyżej metoda ekstrakcji pierwszych dwóch formantów samogłoskowych polega na znajdowaniu dwóch pasm widmowych, dla których sumaryczny sygnał jest największy. W oparciu o dane o wartościach częstotliwości pierwszych dwóch formantów samogłoskowych ograniczono zbiór możliwych teoretycznie par pasm widmowych do rozmiarów uzasadnionych spodziewanymi kombinacjami pierwszego i drugiego formantu. Algorytm tej metody został oprogramowany w języku wewnętrznym minikomputera Mera 303 dzięki czemu wykonanie programu ekstrakcji częstotliwości formantów przebiega w czasie rzeczywistym. Dla zweryfikowania wyników ekstrakcji i oceny samej metody oraz w celu znalezienia cech widmowych dających się wykorzystać do określenia położenia samogłosek w obrębie izolowanych wyrazów, opracowano dodatkowo programy wydruku następujących informacji :

- 1/ Spektrogramów cyfrowych.
- 2/ Numerów wyekstrahowanych pasm formantowych pierwszego i drugiego formantu i grafu ich przebiegu w czasie.
- 3/ Liczb reprezentujących **sygnał w pasmach formantowych** pierwszego i drugiego formantu.

Pasmem formantowym nazwano tutaj umownie zakres czterech kolejnych pasm analizatora, w którym podaną wyżej metodą stwierdzona została obecność formantu. Poniżej zamieszczono kilka wydruków, na podstawie których omówiono zasady znajdowania położenia samogłosek w wyrazach oraz procedury automatycznej weryfikacji wyników ekstrakcji pasm formantowych z myślą o zastosowaniu w klasyfikacji i identyfikacji samogłosek polskich.

2. Wyznaczanie położenia samogłosek w obrębie wyrazu.

Na rys. 1 zamieszczono przykładowy wydruk informacji widmowych uzyskanych z analizy akustycznej wybranego wyrazu. Dane w kolumnach od 1 do 29 tworzą spektrogram cyfrowy. Każdy wiersz danych zawiera parametry widmowe sygnału dla próbki sygnału przypadającej w przedziale 20 ms. Parametry te reprezentują poziomy sygnał w pasmach częstotliwości o numerach kolejnych od 2 do 30. Z powodu zastosowania wygładzania widma, szerokości tych pasm są czterokrotnie większe niż szerokości pasm stosowanego analizatora widma. Numerację, zakresy i środki poszczególnych pasm przypadających w strefie pojawiania się pierwszych dwóch formantów samogłoskowych zamieszczono w tabelicy nr 1:

Tablica Nr 1. Numeracja, zakresy i środki pasm analizy.

Nr pasma	Zakres pasma [Hz]	Srodek pasma [Hz]	Nr pasma	Zakres pasma [Hz]	Srodek pasma [Hz]
2	200 - 520	360	17	1400 - 1720	1560
3	280 - 600	440	18	1480 - 1800	1640
4	360 - 680	520	19	1560 - 1880	1720
5	440 - 760	600	20	1640 - 1960	1800
6	520 - 840	680	21	1720 - 2040	1880
7	600 - 920	760	22	1800 - 2120	1960
8	680 - 1000	840	23	1880 - 2200	2040
9	760 - 1080	920	24	1960 - 2280	2120
10	840 - 1160	1000	25	2040 - 2360	2200
11	920 - 1240	1080	26	2120 - 2440	2280
12	1000 - 1320	1160	27	2200 - 2520	2360
13	1080 - 1400	1240	28	2280 - 2600	2440
14	1160 - 1480	1320	29	2360 - 2680	2520
15	1240 - 1560	1400	30	2440 - 2760	2600
16	1320 - 1640	1480	31	2520 - 2840	2760

Pierwsza z trzech ostatnich kolumn zawiera dane liczbowe o sumarycznej wartości sygnału w pasmach pierwszego i drugiego formantu wyekstrahowanych w procesie analizy. Numery tych pasm podane są w dwóch ostatnich kolumnach. Sumaryczny sygnał w umownych pasmach formantowych nazwano parametrem SPF.

NUMERY PASM →

2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	SPF	F ₁	F ₂						
13	13	9	5															1	1	1	1				2	2	3	3	16	3	20						
15	15	9	6														4	4	9	9	4	4				3	6	10	10	25	3	29					
18	16	10	6														3	7	13	13	9	5				2	2	7	7	5	31	2	19				
18	16	10	5	1	1	1	1									2	4	11	14	12	10	3					1	1	32	2	19						
15	14	8	6													1	1	6	11	16	18	13	8	3	1	1	1	4	4	3	33	2	20				
18	24	26	24	16	12	6	6	8	8	7	11	11	11	11	11	12	20	25	25	15	9	7	9	14	17	20	15	51	4	21							
11	17	23	22	16	12	4	4	5	8	10	15	16	14	9	9	10	16	23	27	29	24	17	11	8	11	13	17	16	52	4	22						
10	15	11	22	18	14	6	5	4	8	11	15	19	17	12	12	12	18	25	29	26	19	12	7	8	9	14	14	53	5	22							
11	15	21	23	16	12	5	3	5	6	10	11	15	14	9	8	8	13	21	26	29	25	19	15	11	11	12	14	15	52	5	22						
13	17	22	20	16	12	7	7	4	5	6	7	12	6	5	4	7	16	21	25	23	16	11	5	5	7	8	11	47	4	22							
15	17	20	18	13	11	5	5	2	2	2	2	7	7	6	6	5	5	12	15	1	21	13	10	2	2	6	8	11	41	1	22						
16	16	18	18	6	6	1	1	1	1								4	5	12	12	10	9	2	2	3	3	3	30	4	22							
14	14	13	11	3	3												2	2	7	7	7	7	2	2	3	3	6	21	5	24							
9	5	5																														9	3				
5	5	2	2																														5	3			
5	5	2	2																															5	3		
4	1																																	4	3		
3	3																																		3	3	
2	2																																			2	3
1	1																																			1	3
1	1																																			1	3

l

e

n

Rys. 1. Przykład wydruku widma cyfrowego, wartości parametru SPF i wyników ekstrakcji częstotliwości pierwznych dwóch formantów.
 /Wyniki te wyrażone są, numerami kolejajzj pasa analizy, w których pojawiał się formant oraz zilustrowane odpowiednią grafem/.

W wyniku oceny podobnych wydruków dla szeregu polskich wyrazów zawierających wszystkie polskie samogłoski w różnych kontekstach spółgłoskowych ustalono jako regułę, iż w segmentach samogłoskowych izolowanych wyrazów parametr SPF przyjmuje wartości większe niż w spółgłoskowym otoczeniu tych segmentów.

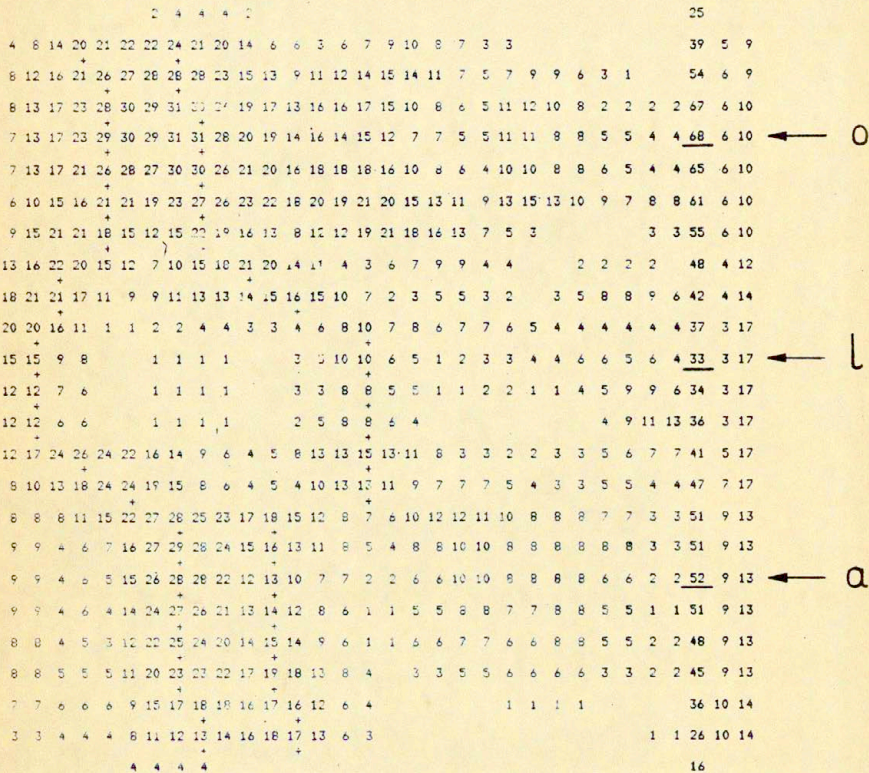
Przebieg P_j średniej z pięciu kolejnych wartości p_i parametru SPF wyznaczony na podstawie wzoru :

$$P_j = \frac{1}{m} \sum_{i=j-\frac{m-1}{2}}^{j+\frac{m-1}{2}} p_i \quad (2)$$

charakteryzuje się istnieniem wyraźnych ekstremów, których liczba dla danego wyrazu równa jest łącznej ilości występujących w tym wyrazie fonemów samogłoskowych **otoczonych** obustronnie spółgłoskami i nieskrajnych fonemów spółgłoskowych. Maksima tego przebiegu przypadają z dużym przybliżeniem w środkach segmentów samogłoskowych a minima w środkach nieskrajnych segmentów spółgłoskowych. Na rys. 2 i 3 zamieszczono dla przykładu wydruki dla innych dwóch wyrazów zawierające w trzeciej kolumnie z prawej strony przebieg uśrednionego parametru SPF. Przykłady te ilustrują korzystną rolę uśrednianego parametru SPF w identyfikacji środków fonemów. Procedura wyznaczania uśrednionego parametru SPF jest w znacznym stopniu integralną częścią ekstrakcji pasm formantowych. Jedynie obliczanie średniej i szukanie ekstremum wymaga dodatkowych bardzo zresztą prostych i krótkich działań. W procesie klasyfikacji lub identyfikacji samogłosek łącznie z ekstrakcją pasm formantowych szukane jest maksimum uśrednionego parametru SPF. Po jego znalezieniu następują operacje bądź to tworzenia lub korekty obszaru klasy w wymiarach częstotliwości formantowych bądź też identyfikacji rozpoznawanego segmentu samogłoskowego. Bieżącymi danymi w tych działaniach są trzy pary pasm formantowych wyekstrahowane w obrębie środkowego wycinka samogłoski w trzech kolejnych chwilowych widmach. Przyjmuje się jako środek tego wycinka położenie maksimum uśrednionego parametru SPF. Za maksimum uśrednionego parametru SPF przyjmuje się tę rzędną P_j , która spełnia następujący warunek:

$$P_j \geq P_{j \pm 1} \quad , \text{ gdzie } i=1,2 \quad (3)$$

j jest numeracją bieżących danych o uśrednionym parametrze SPF.



Rys. 2. Przykład wdruku widma cyfrowego, wyników ekstrakcji częstotliwości pierwszych dwóch formantów oraz uśrednionego parametru SPP.

2	3	...	→	SPF	
1				1	
2				2	
3				3	
4				4	
5				5	
6				6	
7				7	
8				8	
9				9	
10				10	
11				11	
12				12	
13				13	
14				14	
15				15	
16				16	
17				17	
18				18	
19				19	
20				20	
21				21	
22				22	
23				23	
24				24	
25				25	
26				26	
27				27	
28				28	
29				29	
30				30	
31				31	
32				32	
33				33	
34				34	
35				35	
36				36	
37				37	
38				38	
39				39	
40				40	
41				41	
42				42	
43				43	
44				44	
45				45	
46				46	
47				47	
48				48	
49				49	
50				50	
51				51	
52				52	
53				53	
54				54	
55				55	
56				56	
57				57	
58				58	
59				59	
60				60	
61				61	
62				62	
63				63	
64				64	
65				65	
66				66	
67				67	
68				68	
69				69	
70				70	
71				71	
72				72	
73				73	
74				74	
75				75	
76				76	
77				77	
78				78	
79				79	
80				80	
81				81	
82				82	
83				83	
84				84	
85				85	
86				86	
87				87	
88				88	
89				89	
90				90	
91				91	
92				92	
93				93	
94				94	
95				95	
96				96	
97				97	
98				98	
99				99	
100				100	

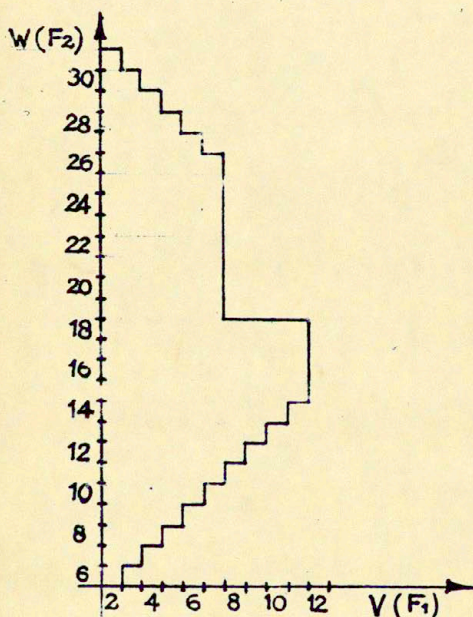
Uwaga ! Krzyżyk w kolumnie danych oznacza także SPF oznacza obecność foranentu drugiego w pamięci nr 01. Na tle tryb graficzny z przyczyn technicznych nie zamieszczono informacji o amplitudzie symbolu w tym paśmie. Uwaga ta dotyczy także innych wydruków w tej pracy.

Str. 7. Przykład wyznaczenia cyfrowego wyniku ekstrakcji określonej wielkości danych i kodowania 32 uśrednionego w składowym SPF.

3. Udoskonalenia pierwotnej metody ekstrakcji częstotliwości pierwszych dwóch formantów samogłoskowych.

Na podstawie pierwszych wyników ekstrakcji częstotliwości formantów metodą podaną w pracy [3] ustalono następujące warunki ekstrakcji :

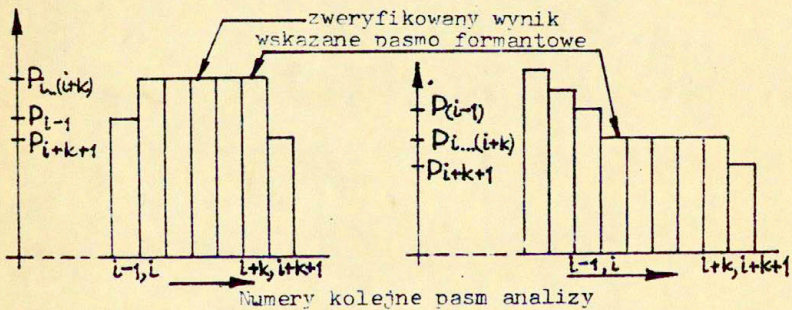
- a/ Wyeliminowano z analizy pasmo pierwszego kanału analizatora widma.
- b/ Obszar oczekiwanych par częstotliwości pierwszych dwóch formantów samogłoskowych ograniczono do rozmiarów nakreślonych na rys. 4.
- c/ Najmniejsza odległość środków pasm, w których w danym widmie obecne mogą być pierwsze dwa formanty samogłoskowe ustalono jako równą odległości między środkami czterech kolejnych pasm analizatora widma.



Rys. 4. Obszar prawdopodobnych par pasm formantowych ustalony dla pierwszych dwóch formantów samogłosek polskich.

Na rys. 6.7.9 zamieszczono spektrogramy cyfrowe wraz z wynikami ekstrakcji pasm formantowych otrzymanymi za pomocą wyżej metody z uwzględnieniem wyszczególnionych wyżej warunków. Na wydrukach wstępują przypadki, ze względu na które należało rozbudować pierwotną metodę ekstrakcji o szereg nowych elementów. Zostaną one omówione kolejno poniżej.

1. W widmach stanowiących punkt wyjścia do ekstrakcji pasm formantowych niektóre słabsze formanty charakteryzują się jednakowym poziomem w kilku kolejnych pasmach analizy (Przykład A na rys. 6). Wynikiem ekstrakcji formantu metodą pierwotną jest w takich przypadkach pasmo skrajne zamiast środkowego lub jednego z dwóch środkowych. Aby tego uniknąć opracowano odpowiednią procedurę. Warunkiem koniecznym ale niewystarczającym zadziałania tej procedury jest, aby wstępnie wyekstrahowane pasmo formantowe było jednym z grupy co najmniej 3-ech, w których poziom sygnału jest identyczny. Dodatkowo spełniony musi być drugi warunek, aby niższy był poziom sygnału w najbliższym paśmie poprzednim i następnym graniczącym z tymi kilkoma o jednakowym poziomie. Oba te warunki zilustrowano na rys. 5.

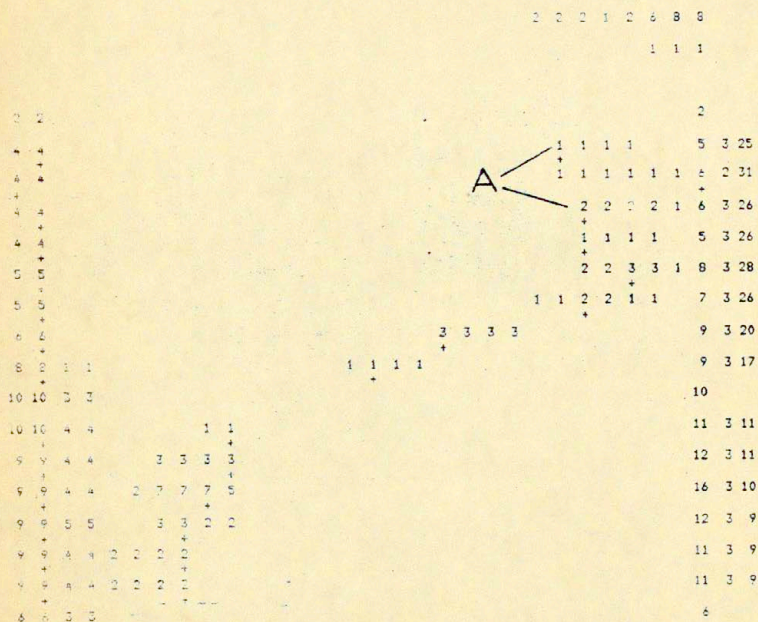


$$P_{i...i+k} > P_{i-1} \text{ i } P_{i+k+1}$$

$$P_{i-1} > P_{i...i+k} > P_{i+k+1}$$

Rys. 5. Ilustracja warunków ekstrakcji pasm formantowych w przypadku rozmytych wierzchołków formantowych.

2. Wśród uzyskanych wyników ekstrakcji pasm formantowych napotkano przypadki błędnego wyekstrahowania częstotliwości formantu 3-go zamiast 2-go w samogłosce \ddot{a} . Przykład tego zawiera wydruk na rys. 7.



P

i

W

Fig. 6. Wydruk wyniku ekstrakcji pasm formantowych z przewle-
dani spłaszczonego formantów.

24	23	17	10		2	2	4	4	1	1	2	2	4	3	7	12	17	26	37	13	11	7	9	9	9	9	4	44	2	21		
25	26	21	14	2	5	4	7	7	3	2				3	3	3	3	4	6	14	15	10	10	2	7	11	14	17	13	43	3	29
24	24	18	11	2	7	2	2						1	1	1	1	4	4	8	8	7	7	5	6	8	12	14	15	79	3	30	
21	21	15	10														3	4	10	13	10	8	5	2	6	7	10	10	34	3	22	
18	19	13	8														2	4	8	12	13	12	8	5	7	6	11	15	34	3	30	
14	14	8	7														3	3	7	7	7	7	5	8	10	12	14	18	3	30		
12	12	8	6														1	1	5	5	7	7	4	5	7	7	9	11	3	30		
10	10	4	4														1	1	3	3	4	4	3	3	4	4	3	14	3	24		

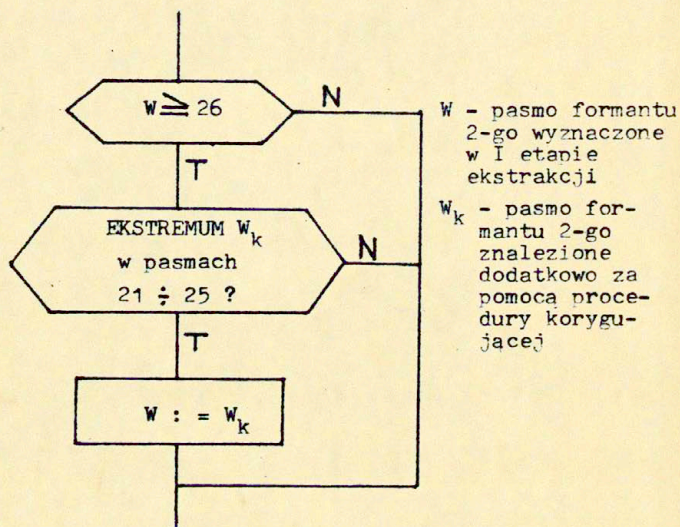
b

t

k

Rys. 7. Wydruk wyniku ekstrakcji pasm formantowej; przykładem bieżącego wyekstranowania częstotliwości drugiego formantu samogłoski /e/.

Dla uniknięcia tego rodzaju błędów zastosowano procedurę korygującą, której algorytm zamieszczony jest na rys. 8. Uruchamia się ona wówczas, gdy wynikiem ekstrakcji częstotliwości 2-go formantu jest jedno z sześciu najwyższych pasm analizy o numerach kolejnych od 26 do 31. Korekta polega na sprawdzeniu, czy nie istnieje dodatkowe ekstremum w pasmach od 20-go do 25-go. Jeżeli tak, wówczas anuluje się pierwotny wynik ekstrakcji częstotliwości 2-go formantu i w zamian akcentuje się wynik uzyskany przy pomocy procedury korygującej.



Rys. 8. Algorytm procedury korygującej błędny wynik ekstrakcji częstotliwości 2-go formantu samogłoski /ɛ/.

3. Inspiracją do wprowadzenia kolejnej procedury korygującej było stwierdzenie przypadków błędnej ekstrakcji częstotliwości obu formantów w samogłosce /o/. Przykład tego znajduje się na wydruku na rys. 9. Sporadycznie następuje wykstrahowanie częstotliwości formantu 2-go zamiast 1-go oraz 3-go w miejsce 2-go.

8 6 1
+
7 7
+
6 6
+
6 6
+
4 4
+

1 4 6 9 11 14 22 3 30
+
3 3 5 5 4 3 12 3 27
+
3 3 3 3 1 9 3 28
+
1 1 2 2 2 2 2 3 28
+
4 3

P

i

1 1 1 1

1 17

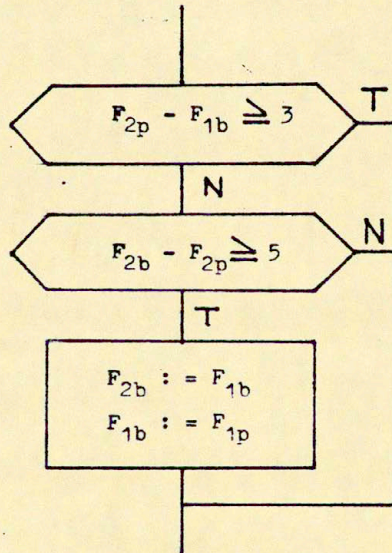
k

16 25 25 24 15 9 9 10 15 18 15 13 8 6 5 3 3 3 2 2 43 4 11
+
14 14 17 17 8 8 8 11 22 24 19 14 8 6 11 13 14 14 8 5 3 3 6 6 3 3 2 2 6 41 5 11
+
10 10 13 13 11 12 12 16 25 25 22 18 12 11 13 13 13 13 5 5 5 5 11 11 5 5 5 5 9 38 11 17 ←
+
10 10 13 13 12 13 14 16 23 23 19 16 12 11 13 14 13 13 5 5 5 5 10 10 5 5 4 4 7 37 11 17 ←
+
10 10 14 14 13 15 15 16 24 23 19 15 11 10 12 12 11 11 5 5 5 5 10 10 5 5 3 3 4 36 7 11
+
10 10 14 14 14 15 15 17 22 22 17 14 10 9 12 13 11 11 5 4 4 4 9 9 4 4 1 1 1 37 7 11
+
10 11 14 14 14 15 16 19 23 21 15 12 8 8 11 11 10 10 4 4 3 3 6 6 3 3 37 6 10
+
9 9 13 13 13 14 13 17 22 21 16 12 5 5 7 7 8 8 3 3 1 1 4 4 2 2 35 7 11
+
7 7 11 11 11 11 10 13 18 18 14 11 1 1 2 2 4 4 2 2 1 1 1 1 29 5 11
+
2 3 7 7 5 6 6 6 6 4 13 7 10

o

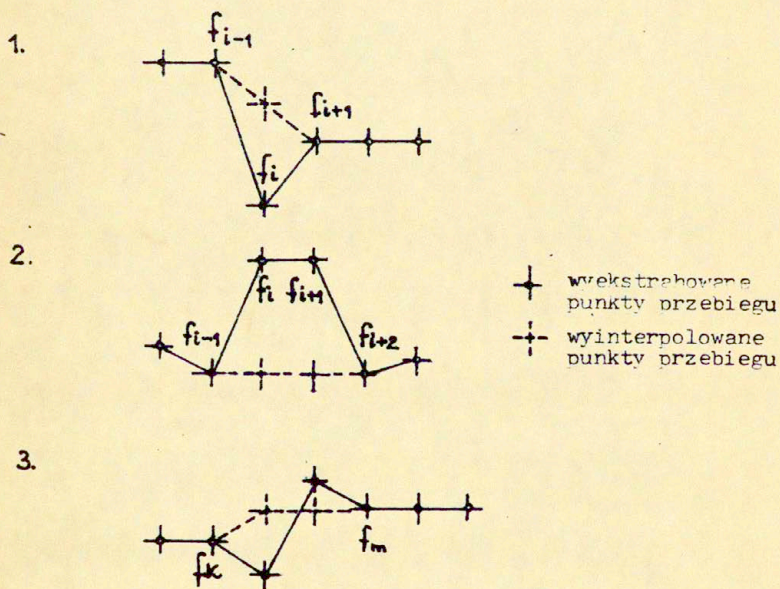
Rys. 9. Wdruk wyniku ekstrakcji pasm formantowych z przykładami błędnego wyekstrahowania częstotliwości pierwszych dwóch formantów samogłoski /o/.

Algorytm korygujący tego typu błąd przedstawiono na rys. 10. Bieżące wyniki ekstrakcji częstotliwości formantu 1-go F_{1b} i 2-go F_{2b} są porównywane z poprzednim wynikiem ekstrakcji częstotliwości formantu 2-go F_{2p} . Gdy różnica między bieżącym i poprzednim wynikiem ekstrakcji częstotliwości formantu 2-go wyrażonym numerami kolejnymi pasm, w których obecność tego formantu została stwierdzona, przekracza założoną wartość 4, a bezwzględna różnica między bieżącym wynikiem ekstrakcji częstotliwości formantu 1-go F_{1b} i poprzednim formantu 2-go F_{2p} podobnie wyrażonymi nie przekracza 2 wówczas przyjmuje się, że jako formant pierwszy mylnie wyekstrahowany został formant drugi a jako drugi - trzeci. Korekta powstałego błędu polega na anulowaniu bieżącego wyniku ekstrakcji częstotliwości formantu drugiego F_{2b} i zastąpieniu go bieżącym wynikiem ekstrakcji częstotliwości formantu pierwszego F_{1b} . W miejsce brakującego bieżącego wyniku dla formantu pierwszego zostaje wpisany wynik poprzedni F_{1p} .



Rys. 10. Algorytm procedury korygującej błędny wynik ekstrakcji pasm formantowych powstały z błędnej identyfikacji kolejnych formantów.

4. Na rys. 11 zilustrowano zaobserwowane w wynikach ekstrakcji przypadki odchyień naruszających regularność przebiegu częstotliwości formantu. W przykładzie pierwszym nieuzasadniona jest duża różnica wartości pomiędzy danymi f_i , f_{i-1} - reprezentującymi numery kolejne wwekstrahowanych pasm formantowych - tym bardziej, że następną daną f_{i+1} bliższa jest wartości danej f_{i-1} . Zastosowana w programie ekstrakcji częstotliwości formantów dodatkowa procedura anuluje w takim przypadku daną f_i i zastępuje ją nową wyliczoną jako średnia z f_{i-1} i f_{i+1} . W przykładzie drugim korekty wymaga przebieg wyniku ekstrakcji w obrębie danych f_{i-1} , f_i , f_{i+1} , f_{i+2} . Występująca bowiem w tym przebiegu nieciągłość nie znajduje uzasadnienia w rzeczywistym przebiegu częstotliwości formantu. Korekta polega na anulowaniu danych f_i i f_{i+1} i interpolacji w ich miejsce nowych danych zachowujących ciągłość przebiegu.



Rys. 11. Ilustracja przypadków nieregularnego przebiegu wyniku ekstrakcji i sposobu jego korekty.

Przykład trzeci ilustruje zygzakowaty kształt wykstrahowanego przebiegu częstotliwości formantu otrzymywany sporadycznie w niektórych segmentach samogłoskowych. Z uwagi na przeznaczenie wykstrahowanych danych do klasyfikacji i identyfikacji samogłosek zdecydowano zmodyfikować tego rodzaju przebieg tak, aby nadać mu charakter monotoniczny. Osiąga się to interpolując dane w zakresie nieregularnego przebiegu w oparciu o następującą formułę :

$$f_i = \frac{f_{i-1} + f_{i+1}}{2}$$

dla $i = k, k + 1, \dots, m - 1, m$.

gdzie f_k i f_m są odpowiednio danymi z początku i końca zakresu w którym przebieg wykstrahowanej częstotliwości formantu ma naruszoną regularność.

Na rys. 12 zamieszczono wydruk będący przykładem wyniku ekstrakcji pasm formantowych przy zastosowaniu kompletu **wszystkich** opisanych wyżej nowych elementów uzupełniających pierwotną metodę ekstrakcji. Nowa metoda której ocenę przeprowadzono na podstawie materiału fonetycznego złożonego z 21 wyrazów zawierających wszystkie polskie samogłoski przeważnie w kontekście spółgłosek zwartych, nosowych i bocznych, daje wyniki dobre. Zniknęły błędy, jakie miały miejsce przy stosowaniu pierwotnej metody. Programy ekstrakcji częstotliwości formantów nową metodą są wykonywane przez minikomputer Mera 303 w czasie rzeczywistym. Wynik ekstrakcji pobierany jest do dalszego przetwarzania z chwilą podania informacji, że od pojawienia się środka samogłoski minęło 60 ms, co jest równoznaczne z wykonaniem trzech cykli wpisywania danych z analizatora a następnie przetwarzania ich według programów ekstrakcji częstotliwości formantów.

Pełna ocena nowej metody możliwa będzie jednak dopiero po masowych próbach automatycznej klasyfikacji polskich samogłosek w rozmaitych kontekstach i wymawianych przez różne głosy. Omówienie jej wyników będzie przedmiotem osobnej publikacji.

									8
9	5	3						2	7
+								8	7
10	5	4	1					4	10
+								13	14
19	22	21	19	14	8	8	7	7	6
+								7	6
21	23	25	21	12	9	5	5	8	6
+								9	12
20	21	21	16	8	4	3	3	4	3
+								3	3
17	17	16	11	6	2	2	2		
+								2	2
8	6	5	1	1					
+									
3	2	2							
3	2	2							
+									
3	3	2	1						
6	6	4	2						
+									
8	8	5	3	1					
+									
7	7	4	4	1	1	1			
+									
7	7	3	3		1	1			
+									
8	8	3	3	1	1	1	1		
+									
7	6	4	3						
+									

l
e
n
i

v

									15
16	14	9	6	1	1	1			
+									
23	21	16	10	2	4	4	2	2	
+									
23	22	18	13	4	4	4	1	1	
+									
22	21	16	13	5	5	3	1	1	1
+									
18	16	13	10	4	4	4	1	1	1
+									
14	14	10	8	3	2	3	3	1	1
+									
10	10	6	6	1					
+									
4	3	2	2						
+									

t

Rys. 12. Wydruk wyniku ekstrakcji pasm formantowych poszerzona metoda.

BIBLIOGRAFIA

- [1] JASSEM, W., KRZYŚKO, M., DYCZKOWSKI, A., Klasyfikacja i identyfikacja samogłosek polskich na podstawie częstotliwości formantów, Prace IPPT 64/1972, Warszawa, 1972.
- [2] JASSEM, W., SZYBISTA, D., DYCZKOWSKI, A., Rozpoznawanie samogłosek polskich w typowych zdaniach, Prace IPPT 43/1975, Warszawa, 1975.
- [3] KUBZDELA, H., Techniczna realizacja formantowej metody rozpoznawania samogłosek polskich, Prace IPPT 90/1975, Warszawa, 1975.
- [4] HESS, W., A Pitch-Synchronous Digital Feature Extraction System for Phonemic Recognition of Speech, IEEE Trans. Acoust. Speech and Sign. Proc., vol. ASSP-24, No. 1, February 1976.
- [5] MYTKOWSKI, K., Kanał funkcji analogowych typ KF-01, Prace IPPT 39/1976, Warszawa, 1976.