Henryk Kubzdela

METODA AUTOMATYCZNEGO ROZPOZNAWANIA WYRAZÓW W OPARCIU O SPEKTROGRAMY BINARNE

WARSZAWA 1980

ISSN 0208-5658

Praca wpłynęła do Redakcji dnia 19 grudnia 1979 r.

Zarejestrowana pod nr 14/1980

IPPT PAN

Na prawach rękopisu

Instytut Podstawowych Problemów Techniki PAN

Nakład 140 egz. Ark.wyd. 0,9. Ark.druk. 1,5.
Oddano do drukarni w maju 1980 r.
Nr zamówienia 367/80

Warszawska Drukarnia Naukowa, Warszawa,
ul.Śniadeckich 8

Henryk Kubzdela
Pracownia Fonetyki Akustycznej IPPT PAN

METODA AUTOMATYCZNEGO ROZPOZNAWANIA WYRAZÓW W OPARCIU O SPEKTROGRAMY BINARNE[1].

1. Wstęp.

Powstanie zautomatyzowanego systemu rozpoznawania mowy ciągłej poprzedzają liczne próby automatycznego rozpoznawania ograniczonego słownika wyrazów wymawianych oddzielnie. Jednej z takich prób dotyczy niniejsza pracy. Przedstawiony w niej model adaptacji i rozpoznawania obrazów fonetyczno-akustycznych o rozciągłości wyrazu posługuje się binarną reprezentacją widma sygnału mowy tworzoną za pomocą układu analogowo-cyfrowego, w skład którego wchodzą : wielokanałowy, analogowy analizator widma, minikomputer MERA 303 oraz urządzenie wprowadzania sygnału analogowego do tego komputera opisane w pracy [4]. Omówiono sposób otrzymywania widm binarnych, cyfrową metodę tworzenia wzorcowych spektrogramów binarnych dla poszczególnych wyrazów w procesie adaptacji i cyfrową metodę ich automatycznego rozpoznawania. Publikacja ta dotyczy kolejnego etapu długofalowego programu badań nad automatycznym rozpoznawaniem mowy ujętego w planach badawczych Pracowni Fonetyki Akustycznej IPPT PAN.

2. Binarny spektrogram sygnału mowy.

2.1. Definicja, wyznaczanie i parametry.

Idea binarnego spektrogramu sygnału mowy przedstawiona została przez G. Ruske [5] . Spektrogram binarny powstaje z przekształcenia spektrogramu klasycznego w oparciu o badanie

---

[1] Praca wykonana w ramach problemu międzyresortowego I-24

znaku wyrażenia :

$$\Delta s(f,\ t) = s(f,t) - \frac{1}{\Delta f \cdot \Delta t} \int_{t-\frac{\Delta t}{2}}^{t+\frac{\Delta t}{2}} \int_{f-\frac{\Delta f}{2}}^{f+\frac{\Delta f}{2}} s(f,t)\ df\ dt, \quad (1)$$

reprezentującego różnicę pomiędzy składową widmową o częstotliwości f w momencie czasu t sygnału mowy a średnią wartością tegoż sygnału w paśmie Δf, mieszczącym w swym środku częstotliwość f, na przestrzeni czasu Δt, w którego środku przypada moment t. Spektrogram binarny przyjmuje wartość 1 w każdym z tych punktów na płaszczyźnie częstotliwości i czasu, w którym

$$\Delta s(f,t) = 0. \quad (2)$$

Spektrogram pierwotny dany jest zwykle nie w postaci funkcji ciągłej $s(f,t)$ lecz zbioru skończonej liczby danych $s(f_i, t_j)$ wyrażających średnie wartości sygnału w pasmach częstotliwości $(\Delta f)_q$ na które podzielony jest zakres analizy widmowej, w kolejnych widmach sygnału przypadających w odstępach czasowych $(\Delta t)_q$.
Wyrażenie (1) upraszcza się wtedy do postaci :

$$\Delta s(f_i, t_j) = s(f_i, t_j) - \frac{1}{(2k+1)\cdot(2l+1)} \sum_{m=0}^{k} \sum_{n=0}^{l} s(f_{i\pm m}, t_{j\pm n}) \quad (3)$$

W pracy niniejszej zakres uśredniania zawężono do jednego widma. Spowodowało to dalsze uproszczenie wyrażenia, w oparciu o które wyznacza się widmo binarne, do następującej postaci :

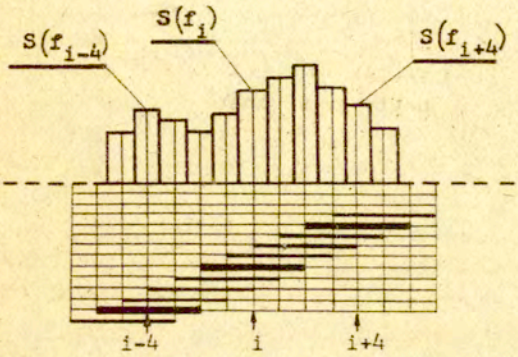
$$\Delta s(f_i, t_j) = s(f_i, t_j) - \frac{1}{(2n+1)} \sum_{n=0}^{1} s(f_{i\pm 4n}, t_j). \quad (4)$$

Dane widmowe $s(f_i)$ każdego z kolejnych widm sygnału wyrażają średnią wartość sygnału w pasmach o szerokości 320 Hz. Sąsiadujące ze sobą pasma częstotliwości pokrywają się na przestrzeni 240 Hz a ich środki oddalone są od siebie o 80 Hz. Do wyznaczenia średniej w wyrażeniu (4) brane są jedynie 3 dane $s(f_i)$, wyrażające wartość sygnału w trzech sąsiednich pasmach, jak to

zilustrowano na rys. 1. Stąd przyjęto w wyrażeniu (4) :

$$l = 1,$$

a zmienna f otrzymała indeks $i \pm 4n$ zamiast $i \pm n$.
Wobec powyższego średnia, którą stanowi odjemnik w wyrażeniu(4) odnosi się do zakresu o szerokości 960 Hz.



Rys. 1. Ilustracja układu pasm częstotliwości z których pochodzą dane widmowe $s(f_i, t_j)$. Dla przykładu uwypuklono grubszymi kreskami odcinki reprezentujące 3 stykające się pasma częstotliwości.

Wartość binarną 1 otrzymuje widmo w miejsce każdej danej $s(f_i)$ wyrażającej wartość sygnału w paśmie Δf_i o szerokości 320 Hz, jeżeli ta dana jest równa lub większa od średniej wartości sygnału w paśmie o szerokości 960 Hz mieszczącym dokładnie w swym środku pasmo Δf_i, co zapisać można następująco :

$$WB(f_i) = 1 \quad \text{dla} \quad s(f_i) \geqslant \frac{s(f_{i-4}) + s(f_i) + s(f_{i+4})}{3} . \qquad (5)$$

W przeciwnym razie

$$WB(f_i) = 0 .$$

Nierówność (5) można przekształcić do postaci :

$$2\, s(f_i) \geqslant s(f_{i-4}) + s(f_{i+4}) \qquad (6)$$

wygodniejszej do obliczeń przy pomocy komputera MERA 303.

Na początku widma, tzn. gdy $i \leq m$ [patrz wyrażenie (3)] można, jak proponuje G.Ruske [5], posługiwać się zwierciadlanym odbiciem widma. W pracy niniejszej przyjęto podobną zasadę przekształcenia początku widma w postać binarną. Zakres uśredniania zawężono o jedno z pasm bocznych, czyli nierówność (5) uproszczono do postaci :

$$s(f_i) \geqslant \frac{s(f_i) + s(f_{i+4})}{2},$$

która jest równoznaczna z

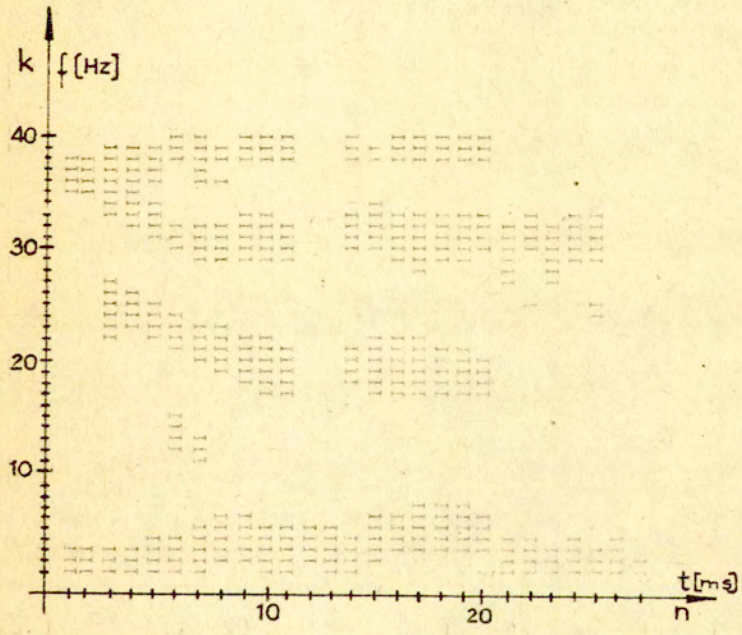
$$s(f_i) \geqslant s(f_{i+4}) . \qquad (7)$$

Nierówność (7) obowiązuje dla $i \leq 4$.

Jak wynika z powyższego, wyznaczenie widma binarnego wiąże się z bardzo prostymi działaniami matematycznymi, dającymi się łatwo wykonać na komputerze MERA 303 w czasie 15-milisekundowej przerwy pomiędzy kolejnymi ładowaniami danych widmowych z analogowego analizatora widma. Jeśli przyjąć za wystarczającą liczbę 50 przekrojów widmowych sygnału na sekundę, wówczas można uznać, że wyznaczanie spektrogramu binarnego następuje w czasie rzeczywistym. Na rys. 2 zamieszczono przykład spektrogramu binarnego wyrazu [jeden] wymówionego przez głos męski. Kolejne widma binarne na tym spektrogramie dotyczą kilkunastomilisekundowych fragmentów sygnału mowy oddalonych od siebie w czasie o 20 ms. Każde widmo na spektrogramie posiada 40 danych binarnych 0 lub 1. Na wydruku zamieszczonym na rys. 2 znak ⊢ reprezentuje wartość 1, a brak tego znaku wartość 0.

2.2. Cechy widma binarnego i sposoby jego zakodowania w pamięci komputera MERA 303.

Widmo binarne zawiera informacje o zakresach częstotliwości, w których w określonym momencie sygnał przyjmuje wartość ekstremalną. Nie ma w nim natomiast informacji o wartości sygnału w tych zakresach. Widmo binarne pozostaje prawie niezależne od korektur widmowych sygnału o charakterystykach monotonicznych z umiarkowanym nachyleniem. Podobnie tego rodzaju korekcje sygnału mowy jak wiadomo nie mają istotnego wpływu

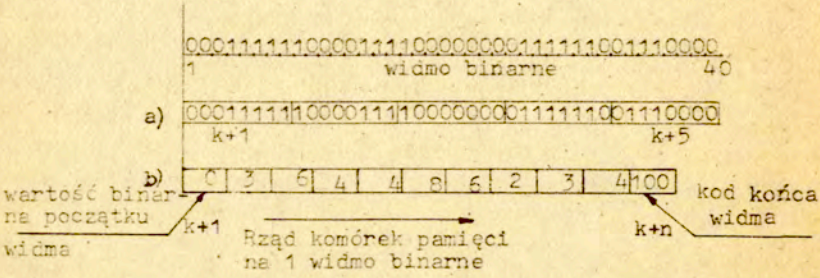


j e d e n

$$f = k \cdot \left(\frac{200}{k} + 80 \right) \ [\mathrm{Hz}] , \quad t = 20 \cdot n \ [\mathrm{ms}] .$$

Rys. 2. Przykład spektrogramu binarnego.

na percepcję mowy przez człowieka. Ta okoliczność sugeruje, że widmo binarne może być dogodną formą zakodowania mowy w systemie jej automatycznego rozpoznawania. Dalszą korzystną cechą widma binarnego jest jego niewielka objętość informacyjna. Możliwe są dwa sposoby zakodowania widma binarnego i zapamiętania go w pamięci minikomputera MERA 303 zilustrowane na rys. 3.



Rys. 3. Sposób kodowania widma binarnego ze stałą (a) i ze zmienną (b) liczbą komórek pamięci na 1 widmo binarne.

Pierwszy z tych sposobów polega na przedstawieniu widma binarnego jako ciągu liczb wyrażających liczebności elementów w przemiennych ciągach zer i jedynek tworzących widmo binarne. Liczby te zapamiętane zostają w rzędzie komórek pamięci, który poprzedza jedna komórka zawierająca informację o tym, czy widmo zaczyna się zerem czy jedynką, a kończy inna dodatkowa komórka z kodem końca widma. Przy takim sposobie zakodowania widma binarnego zajętość pamięci na jedno widmo binarne jest niestała i teoretycznie może wynosić od 3 do N+3 komórek bajtowych. N jest liczebnością danych jednego widma binarnego. Przeciętnie jednak dla 40-to parametrycznego widma binarnego sygnału mowy o zakresie częstotliwości od 120 do 3560 Hz wynosi ona 8 bajtów na 1 widmo binarne i jest większa niż dla drugiego sposobu

zakodowania widma binarnego w pamięci komputera. Opisany wyżej sposób jest jednak prostszy w realizacji od drugiego i korzystniejszy ze względu na rygor wyznaczania i magazynowania tegoż widma w czasie jak najkrótszym. Drugi sposób zakodowania widma binarnego polega na wyrażeniu go ciągiem kilku liczb ośmio-bitowych, utworzonych z ośmioelementowych ciągów na jakie w prostej kolejności pogrupowane zostają dane widma binarnego. Przy takim sposobie zakodowania jedno widmo o objętości informacyjnej 40 bitów zajmuje stały obszar pamięci wielkości pięciu komórek bajtowych. Ten sposób jest więc korzystniejszy z punktu widzenia zajętości pamięci przez 1 widmo binarne. Widmem binarnym tak zakodowanym łatwiej także operować w działaniach związanych z adaptacją i rozpoznawaniem. Z powyższych względów, na etapie tworzenia spektrogramu binarnego widmo kodowane jest według pierwszego z dwóch omówionych sposobów natomiast dla adaptacji i rozpoznawania zostaje ono następnie przekodowane zgodnie z drugim sposobem.

3. Model adaptacji i rozpoznawania wyrazów w oparciu o ich spektrogramy binarne.

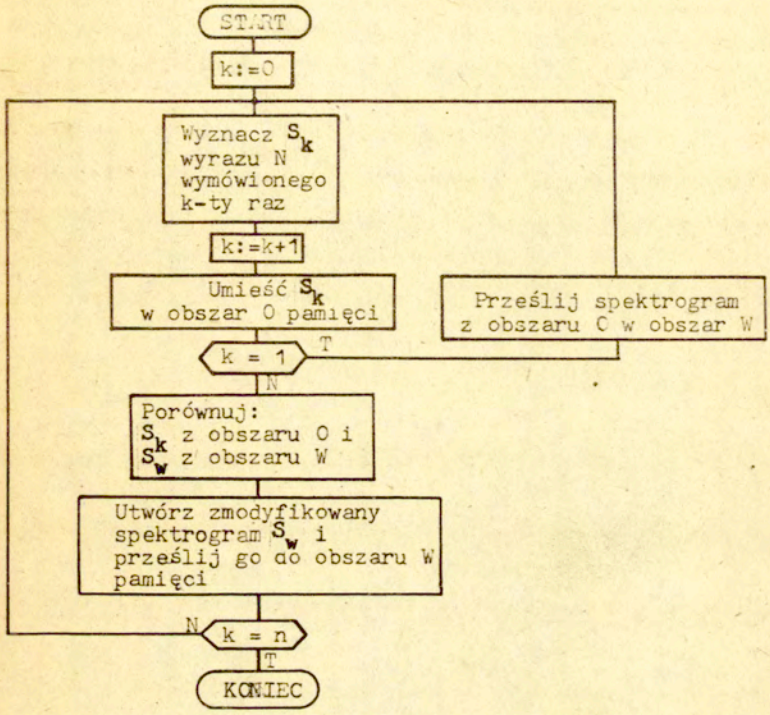
Spektrogram binarny jednego wyrazu posiada średnio objętość informacyjną około 1200 bitów i zajmuje w pamięci komputera zaledwie około 150 komórek bajtowych przy założeniu, że na jedno widmo binarne przypada 40 bitów a częstość wykonywania przekrojów widmowych sygnału wynosi $50[s]^{-1}$. Przekształcenie widma oryginalnego, otrzymanego z wielokanałowego analizatora w widmo binarne jest w realizacji cyfrowej działaniem prostym, dającym się, jak już wspomniano, wykonać w czasie rzeczywistym. Widmo binarne zawiera informacje o najistotniejszych cechach sygnału a mianowicie o zakresach częstotliwości, w których poziom sygnału jest ekstremalny. Spektrogramy binarne danego wyrazu wymówionego wielokrotnie przez ten sam głos są na ogół podobne, natomiast różnią się zauważalnie spektrogramy binarne różnych wyrazów. Umiarkowana zmienność tempa wypowiedzi tego samego wyrazu rzutuje na ogół jedynie na długość czasową tych fragmentów sygnału, w których trwa stan ustalony, nie zmieniając wcale lub w niewielkim stopniu obraz spektrograficz-

ny sygnału mowy w stadiach nieustalonych [3]. Wszystkie te okoliczności skłoniły autora do opracowania modelu automatycznego rozpoznawania wyrazów w oparciu o spektrogramy binarne. Sugeruje się także, iż wykorzystując podaną w pracach [1] i [2] cechę akustyczną samogłosek w mowie ciągłej możliwa będzie automatyczna segmentacja mowy np. na półsylaby. Przedstawioną w niniejszej pracy metodę rozpoznawania wyrazów będzie można wówczas zaadaptować do rozpoznawania tych elementów segmentalnych.

3.1. Algorytm adaptacji.

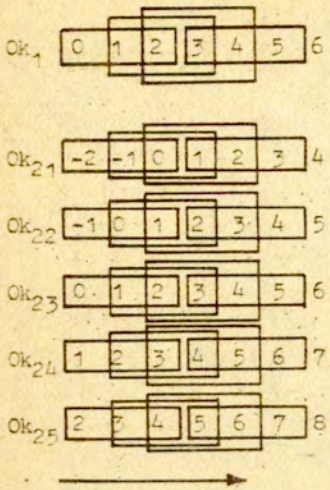
W procesie adaptacji wyraz N wymówiony zostaje n-krotnie. Po każdym kolejnym k-tym wymówieniu spektrogram binarny tego wyrazu umieszczony zostaje w obszarze O pamięci komputera. Po pierwszym wymówieniu wyrazu N jego spektrogram binarny umieszczony zostaje także w obszarze W, przeznaczonym do zapamiętania wzorcowego spektrogramu binarnego dla k wariantów wyrazu N. Spektorgram binarny każdego kolejnego wariantu wyrazu N jest porównywany z tymże wzorcowym spektrogramem binarnym w celu jego zmodyfikowania z uwzględnieniem kolejnego wariantu wyrazu N. Porównanie przebiega według algorytmu, którego uproszczony schemat zamieszczony jest na rys. 4. Kolejność konfrontacji posczególnych wycinków porównywanych ze sobą spektrogramów binarnych zilustrowano na rys. 5. Wzdłuż osi czasu wzorcowego spektrogramu binarnego S_w wyrazu N przemieszcza się okno Ok_1 obejmujące w każdej swej pozycji trzy kolejne widma binarne. Jeden krok przemieszczenia równy jest odstępowi między dwoma kolejnym widmami. Równocześnie pięć identycznych okien Ok_{21} - Ok_{25}, z których każde następne jest przesunięte względem poprzedniego także o odległość pomiędzy dwoma kolejnymi widmami, przemieszcza się wzdłuż osi czasu spektrogramu binarnego S_k wyrazu N wypowiedzianego k-ty raz. Gdy liczby widm binarnych L_w i L_k przypadających odpowiednio na spektrogramy S_w i S_k są równe, wtedy okno Ok_1 i pięć okien Ok_2 przemieszczają się na całej długości obu spektrogramów współbieżnie. Gdy

$$L_w > L_k ,$$



n - założona liczba powtórzeń wyrazu N

Rys. 4. Algorytm tworzenia wzorca w procesie adaptacji.



Rys. 5. Ilustracja kolejności konfrontacji poszczególnych wycinków porównywanych ze sobą spektrogramów binarnych.

Liczby w konturze Ok_1 są numerami kolejnymi objętych tym oknem widm spektrogramu wzorcowego S_w wyrazu N, a liczby w konturach okien Ok_{21} - Ok_{25} numerami kolejnymi objętych tymi oknami widm spektrogramu S_k wyrazu N wymówionego k-ty raz. Na rysunku pokazano 4 kolejne pozycje okien Ok_1 i Ok_{21} - Ok_{25}.

wówczas co liczbę widm binarnych L_{w1} równą części całkowitej ilorazu :

$$\frac{L_w}{L_w - L_k}$$

wypada jedno przemieszczenie grupy pięciu okien Ok_2, natomiast gdy

$$L_w < L_k ,$$

wtedy co liczbę widm binarnych L_{w2} równą części całkowitej ilorazu :

$$\frac{L_w}{L_k - L_w}$$

wypada jedno przemieszczenie okna Ok_1.

Dla każdej pozycji okna Ok_1 i grupy pięciu okien Ok_2 mierzone jest podobieństwo wycinka spektrogramu S_w objętego oknem Ok_1 z wycinkami spektrogramu S_k objętych oknami Ok_{21} - Ok_{25}.

Jako miarę podobieństwa przyjęto stosunek :

$$r = \frac{a}{a + b} , \qquad (8)$$

w którym a i b są odpowiednio liczbami odpowiadających sobie w obu oknach miejsc, na których występują jedynki zgodnie i niezgodnie. Spośród pięciu wycinków spektrogramu S_k objętych oknami Ok_{21} - Ok_{25}, najbardziej podobnym do wycinka spektrogramu S_w objętego oknem Ok_1 jest ten, dla którego r jest największe. Środkowe widmo binarne tego wycinka posłuży do zmodyfikowania środkowego widma binarnego wycinka spektrogramu S_w objętego oknem Ok_1. Modyfikacja polega na sumowaniu logicznym obu wymienionych widm binarnych, a jej wynikiem jest poszerzone o nowe szczegóły jedno z widm binarnych wzorcowego spektrogramu wyrazu N. W podobny sposób zmodyfikowane zostają wszystkie widma binarne wchodzące w skład tegoż spektrogramu. W ten sposób po każdym ponownym wymówieniu wyrazu N następuje uzupełnienie jego obrazu spektrograficznego nowymi szczegółami. Po odpowiednio dużej liczbie wypowiedzi tego samego wyrazu N jego wzorcowy spektrogram binarny ustali się ostatecznie i dalsze wypowiedzi nie będą już zmieniały jego obrazu. Opisaną wyżej metodą zgromadzić można wzorcowe spektrogramy binarne każdego z wyrazów przewidzianych do automatycznego rozpoznawania. Ilość tych wyrazów musi być rozsądnie tak dobrana, aby sumaryczna objętość informacyjna ich wzorcowych spektrogramów binarnych nie przekraczała pojemności pamięci komputera i aby czas rozpoznawania jednego wyrazu nie był zbyt długi. Rząd komórek pamięci pamiętających spektrogram binarny jednego wyrazu nazwano umownie spektrosłowem. Każde spektrosłowo poprzedzane jest informacją o swojej długości. Poszczególnym wzorcowym spektrogramom binarnym przypisane są w pamięci komputera stałe miejsca, następujące po sobie w kolejności odpowiadającej wzrastającej długości spektrosłów. Nasuwa się w tym momencie uwaga, iż mogłoby być pożyteczne redukowanie spektrogramu binarnego o te fragmenty, które reprezentują stan ustalony. Tę operację wykonywać by należało zawsze bezpośrednio po przekształceniu widma oryginalnego w widmo binarne. Polegałaby ona na wyeliminowaniu z powstającego spektrogramu binarnego każdego widma, które byłoby przystające do np. dwóch widm poprzednich. Dzięki temu zmniejszyłaby się objętość informacyjna inwentarza wzorcowych spektrogramów binarnych i tym samym w ograniczonej

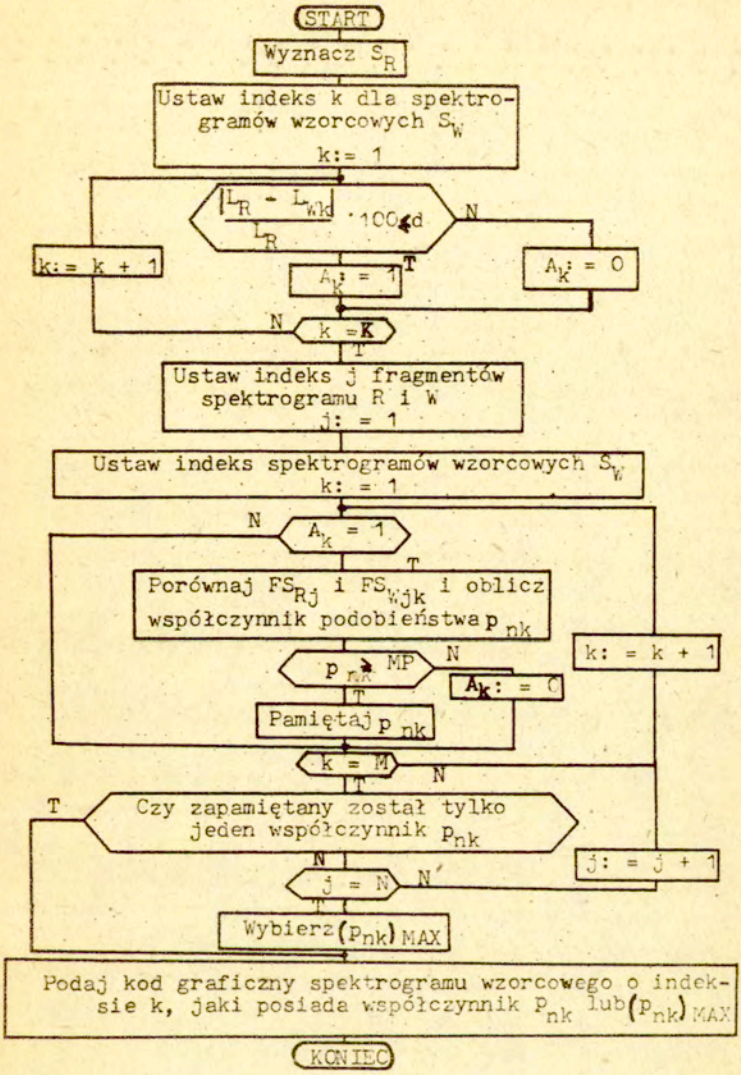
pamięci komputera zmieściłyby się wzorce większej liczby wyrazów. Miałoby to też korzystny wpływ na przebieg a także prawdopodobnie i wyniki rozpoznawania. Koncepcja ta wymaga praktycznego wypróbowania. Przewiduje się, iż pewną trudność stwarzać może dobór trafnego kryterium przystawania dwóch widm binarnych.

3.2. Rozpoznawanie wyrazów.

Po zgromadzeniu w pamięci komputera wzorcowych spektrogramów binarnych pewnej liczby wyrazów niosących np. określone informacje dla zautomatyzowanego procesu, zamknięty zostaje etap adaptacji układu. Poniżej przedstawiono metodę rozpoznawania tych wyrazów. Schemat uproszczonego algorytmu tej metody zamieszczono na rys. 6.

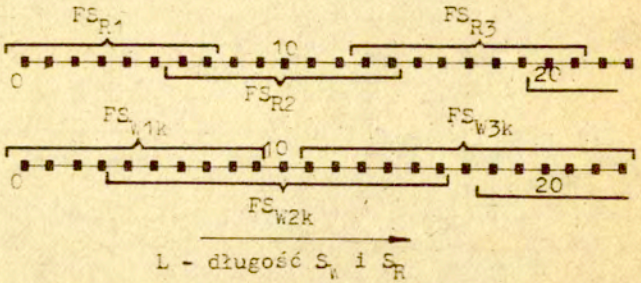
Pierwszym etapem procesu rozpoznawania jest utworzenie spektrogramu binarnego rozpoznawanego wyrazu i umieszczenie go w odpowiednim miejscu pamięci według kodu o stałej ilości liczb na 1 widmo binarne. Ten spektrogram ma być następnie porównany ze spektrogramami wzorcowymi zmagazynowanymi w pamięci komputera. Jak wspomniano wyżej, każdy spektrogram binarny poprzedza informacja o jego długości wyrażonej liczbą widm binarnych.

W kolejnym etapie wyeliminowane zostają z porównania wszystkie spektrogramy wzorcowe o długości różniącej się od długości spektrogramu wyrazu rozpoznawanego o więcej niż d %. Wartość liczby d ustalona będzie empirycznie na podstawie danych o jej wpływie na poprawność rozpoznawania. Spektrogram wyrazu rozpoznawanego S_R dzielony jest następnie na części według schematu na rys. 7. Każda część obejmuje co najmniej siedem widm binarnych i jest w osobnym kroku obliczeniowym porównywana z odpowiadającymi jej częściami wzorcowych spektrogramów binarnych S_W w celu określenia podobieństwa do nich. Porównanie dwóch odpowiadających sobie części, z których jedna jest fragmentem spektrogramu binarnego S_R bieżącego wyrazu a druga fragmentem spektrogramu binarnego jednego z wzorców S_W przebiega w kilku etapach. W etapie pierwszym, tą samą metodą jak w procesie adaptacji, wyznaczane jest podobieństwo wycinka spektrogramu S_R objętego oknem Ok_1 z wycinkami spektrogramu S_{Wk} objętymi oknami Ok_{21} - Ok_{25}. Okno Ok_1 przesuwa się wzdłuż



i - indeks widm przypadających na fragment spektrogramu S_R,
k - indeks spektrogramów wzorcowych,
j - indeks fragmentów na jakie podzielony jest spektrogram,
A_k - wskaźnik zakwalifikowania lub odrzucenia widm S_W.

Rys. 6. Schemat uproszczonego algorytmu rozpoznawania wyrazu.



Rys. 7. Sposób podziału spektrogramu binarnego na części.

fragmentu spektrogramu S_R natomiast grupa pięciu okien Ok_{21} - Ok_{25} wzdłuż fragmentu jednego ze spektrogramów wzorcowych S_W. Dla każdej pozycji okna Ok_1 wybrany zostaje ten z pięciu wycinków objętych oknami Ok_{21} - Ok_{25}, który wykazuje największe podobieństwo z wycinkiem objętym oknem Ok_1. Miara podobieństwa jest taka sama jak w procesie adaptacji. Środkowe spośród trzech widm binarnych wybranego wycinka oraz wycinka objętego oknem Ok_1 są ponownie porównywane. Podobieństwo ich wyrażone zostaje liczbami a_{ij} i b_{ij} zgodności i niezgodności tych odpowiadających sobie danych obu widm, z których przynajmniej jedna jest równa 1. Liczby a_{ij} i b_{ij} wyznaczone dla poszczególnych pozycji okna Ok_1 w obrębie fragmentu FS_{Rj} spektrogramu binarnego rozpoznawanego wyrazu oraz grupy pięciu okien Ok_{21} - Ok_{25} w obrębie fragmentu FS_{Wj} jednego z wzorcowych spektrogramów binarnych są sumowane zgodnie ze wzorami :

$$\delta_{ajk} = \sum_{i=1}^{m} a_{ij} \quad \text{i} \quad \delta_{bjk} = \sum_{i=1}^{m} b_{ij}, \qquad (9)$$

w których m oznacza liczbę widm binarnych przypadających na jeden fragment spektrogramu.
Sumy δ_{ajk} i δ_{bjk} otrzymane w poszczególnych krokach obliczeniowych j są akumulowane dla każdego rozpatrywanego wzorca, co zapisano wzorami :

$$(\delta_a)_{nk} = \sum_{j=1}^{n} \delta_{ajk}, \quad (\delta_b)_{nk} = \sum_{j=1}^{n} \delta_{bjk}, \qquad (10)$$

w których n oznacza liczbę fragmentów spektrogramu binarnego rozpoznawanego wyrazu porównywanych z tyleż samo fragmentami k-tego wzorcowego spektrogramu binarnego. Po skończeniu każdego kolejnego kroku obliczeniowego, w ramach którego wyznaczone zostają sumy δ_{aj} i δ_{bj} oraz $(\delta_a)_n$ i $(\delta_b)_n$ dla wszystkich zakwalifikowanych do porównania wzorców binarnych, obliczone zostają następnie dla tychże wzorców według wzoru :

$$p_n = \frac{(\delta_a)_n}{(\delta_a)_n + (\delta_b)_n} \qquad (11)$$

współczynniki ich podobieństwa ze spektrogramem binarnym rozpoznawanego wyrazu. Po pierwszym kroku podobieństwo to odnosi się do pierwszego fragmentu porównywanych wyrazów. Do porównania w każdym następnym kroku obliczeniowym j dopuszcza się tylko te spektrogramy wzorcowe, dla których współczynnik p_n uzyskany po poprzednim kroku (j - 1) spełnia warunek :

$$p_n \geqslant MP. \qquad (12)$$

W ten sposób dla wyłonienia wyniku rozpoznawania nie jest konieczne porównywanie spektrogramu binarnego rozpoznawanego wyrazu z całymi spektrogramami binarnymi wszystkich wyrazów wybranego słownika. Możliwe są przypadki, że po jednym z kolejnych kroków j, tj. po porównaniu n fragmentów (a nie całego rozpoznawanego wyrazu) z odpowiednimi fragmentami wyrazów słownika

a/ tylko jeden z wzorców spełnia warunek (12)
b/ żaden z wzorców nie spełnia warunku (12).

W przypadku (a) wyraz, do którego ten wzorzec należy, uznany zostaje za tożsamy z wyrazem rozpoznawanym i tym samym proces rozpoznawania zostaje zakończony.
W przypadku (b) należy zakwalifikować do dalszego kroku wszystkie te wzorce, które uczestniczyły w poprzednim kroku. Jeśli dojdzie do skończenia ostatniego kroku tzn. do porównania całego rozpoznawanego wyrazu z wyselekcjonowanymi w poprzednich krokach wyrazami słownika, wówczas wynikiem rozpoznawania jest ten wyraz słownika, dla którego współczynnik podobieństwa p z wyrazem rozpoznawanym jest największy.
Podczas porównywania fragmentu spektrogramu binarnego z odpowiadającymi mu fragmentami wzorcowych spektrogramów binarnych konieczne jest uwzględnienie różnic długości obu porównywanych spektrogramów binarnych. Czyni się to podobnie jak w procesie adaptacji. Gdy długość L_R spektrogramu binarnego rozpoznawanego wyrazu jest większa niż długość L_W danego spektrogramu wzorcowego, wtedy co liczbę widm binarnych L_{R1} będącą częścią całkowitą ilorazu :

$$\frac{L_R}{L_R - L_W}$$

wypada jedno przemieszczenie grupy pięciu okien Ok_2. Oznacza to, że każdy fragment wzorcowego spektrogramu binarnego odpowiadający temu fragmentowi spektrogramu rozpoznawanego wyrazu, w obrębie którego przypada widmo o numerze kolejnym

$$l = n\,L_{R1}\ ,\ \text{gdzie } n = 1, 2, \ldots, (L_R - L_W)$$

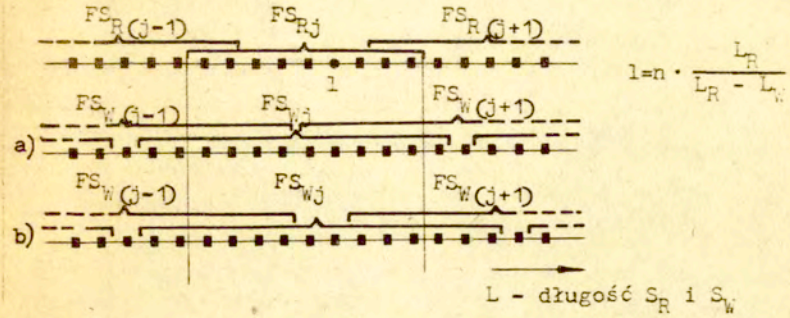
jest zawsze krótszy o jedno widmo w porównaniu z długością pozostałych fragmentów, jak pokazano na rys. 8.
Gdy natomiast

$$L_R < L_W\ ,$$

wtedy co liczbę widm binarnych L_{R2} stanowiącą część całkowitą ilorazu :

$$\frac{L_R}{L_W - L_R}$$

wypada jedno przemieszczenie okna Ok_1.



a) dla $L_R > L_W$,

b) dla $L_W > L_R$.

Rys. 8. Uwzględnienie różnic długości porównywanych spektrogramów binarnych przy podziale ich na fragmenty.

Jeżeli w obrębie konfrontowanego fragmentu spektrogramu binarnego rozpoznawanego wyrazu przypada widmo o numerze kolejnym :

$$l = n\,L_{R1}\ ,\ \text{gdzie}\ n = 1, 2, \ldots, (L_W - L_R)$$

wówczas biorący udział w tej konfrontacji fragment wzorcowego spektrogramu binarnego posiada 1 widmo binarne więcej niż inne fragmenty tego spektrogramu. Zilustrowano tę zasadę na rys. 8.

4. Uwagi końcowe.

Przedstawiony wyżej model automatycznego rozpoznawania wyrazów jest obecnie realizowany w Pracowni Fonetyki Akustycznej IPPT PAN. Czynne są już programy wyznaczania widm binarnych. Opracowane i w stadium uruchamiania są też programy adaptacji. W dalszej kolejności oprogramowane zostaną algorytmy rozpoznawania wyrazów. Ocena przedstawionych tutaj algorytmów zostanie

dokonana na podstawie doświadczeń będących w toku. Spodziewać się należy, że na podstawie pierwszych wyników rozpoznawania wyłonią się postulaty wprowadzania określonych poprawek do przedstawionego tutaj modelu. Eksperymentalnej weryfikacji wymagają np. : szerokość okien Ok_1 i Ok_{21} - Ok_{25} założona obecnie na 3 widma binarne, liczba okien Ok_2 wynosząca według założenia 5, wartości stałych progowych d i MP chwilowo jeszcze nie ustalonych. Zbadania wymagać też będzie zależność wyników rozpoznawania od głosu, indywidualnych cech i struktur fonetyczno akustycznych wyrazu. Poszerzyć również należy w przyszłości zakres częstotliwości widm binarnych ograniczony obecnie stosunkowo wąskim zakresem analizy wielokanałowego analizatora widma sięgającym zaledwie 3520 Hz. Celowe byłoby także ustalić eksperymentalnie, jaki wpływ na wyniki rozpoznawania ma sposób podziału widma na pasma i dokonać wyboru optymalnego podziału. Okazać się także może, że trzeba będzie przyjąć inną miarę podobieństwa porównywanych spektrogramów.

Model tutaj przedstawiony nie wymaga złożonych operacji matematycznych ani obszernej pamięci komputera. Uwzględniono przy jego opracowywaniu posiadane do dyspozycji środki techniczne Pracowni nie rezygnując jednocześnie z dążenia, aby zagwarantować poprawność jego działania a także praktyczną przydatność. Jako przykłady zastosowań takiego modelu można by wymienić : hasłowe wywoływanie informacji z pamięci komputera, dyktowanie głosem informacji uzyskiwanych w warunkach zajętości wzroku np. podczas obserwacji mikroskopowych, sterowanie głosem procesów kontrolowanych wzrokiem.

BIBLIOGRAFIA

[1] KUBZDELA, H., Wyznaczanie charakterystycznego fragmentu samogłoskowego i pomiar częstotliwości formantów dla automatycznej klasyfikacji i identyfikacji samogłosek, Prace IPPT, /oddano do druku/.

[2] KUBZDELA, H., A method of determining the characteristic fragment of vowels and measuring their F_1 and F_2 frequencies in running speech, Speech Analysis and Synthesis, vol. V, PWN, Warszawa, /oddano do druku/.

[3] ŁOBACZ, P., Wpływ tempa mowy na przebieg formantów samogłosek polskich, Prace IPPT 67/74, Warszawa, 1974.

[4] 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 39/76, Warszawa, 1976.

[5] RUSKE, G., An Efficient binary representation of Sonograms, Acustica, vol. 34, No. 4, s. 234-239, Stuttgart, 1976.

[6] RUSKE, G., Real-time information reduction in digital sound spectrograms of Speech, IEEE Int. Conf. an Acoust., Speech and Sign. Process., Philadelphia, 1976.