Henryk Kubzdela

# WYZNACZANIE ZAKRESÓW STACJONARNOŚCI WYBRANYCH PARAMETRÓW WIDMOWYCH DLA AUTOMATYCZNEJ LOKALIZACJI FONEMÓW

14/1995



WARSZAWA 1995

ISSN 0208-5658

Praca wpłynęła do Redakcji dnia 19 grudnia 1994 r.



Na prawach rękopisu

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

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

Henryk KUBZDELA
Zakład Fonetyki Akustycznej
IPPT PAN Poznań

# WYZNACZANIE ZAKRESÓW STACJONARNOŚCI WYBRANYCH PARAMETRÓW WIDMOWYCH DLA AUTOMATYCZNEJ LOKALIZACJI FONEMÓW

S t r e s z c z e n i e

Powstały wcześniej podstawowy system spektrografii zorientowany na ekstrakcję oraz wizualizację wybranych parametrów widmowych poddano kolejnej modyfikacji. Zdefiniowano punkt charakterystyczny widma. Odpowiedni moduł systemu zaadaptowano do wyznaczania punktów charakterystycznych w widmie średnim z czterech kolejnych widm "chwilowych". Uśrednianie widm stało się możliwe również dzięki rozbudowaniu programu realizującego wspomniany wyżej system spektrografii. Odpowiednie punkty charakterystyczne (PC) kolejnych widm średnich tworzą ciągi. Postawiono kryterium, według którego wyznacza się stacjonarne segmenty tych ciągów. Ciąg punktów wskazujących na takie segmenty określa stacjonarny fragment ciągu punktów charakterystycznych. Azymuty czasowe początku i końca tych fragmentów naniesione na oś czasu tworzą niekiedy skupienia wskazujące na stadia przejściowe między fonemami. Z analizy sygnału mowy uzyskuje się szereg współbieżnych ciągów PC. Prawie w każdym z nich są fragmenty stacjonarne. Pożyteczną informację dla lokalizacji fonemu przedstawia ciąg czasowy ich liczebności. Na przykładzie wypowiedzi kilku krótkich wyrazów rozważono ułożenie fragmentów stacjonarnych ciągów punktów charakterystycznych względem pozycji fonemów. Podstawę do rozważań stanowiły mapy ciągów PC z zaznaczonymi zakresami fragmentów stacjonarnych. U podstawy większości tych map zamieszczono osie z zaznaczonymi azymutami czasowymi początku i końca fragmentów stacjonarnych. Stwierdzono, że dla segmentacji fonematycznej sygnału mowy korzystniejsza jest informacja o rozkładzie czasowym fragmentów stacjonarnych jedynie wybranych ciągów PC. Są nimi

ciągi PC przebiegające w paśmie częstotliwości dystynktywnym dla występujących w wypowiedzi fonemów. Wyniki badań wskazują, że cecha jaką jest stacjonarność fragmentów niektórych ciągów punktów charakterystycznych widma uśrednionego, powinna być wykorzystana w lokalizacji fonemów w sygnale mowy.

## 1. Wstęp.

Dla rozpoznawania mowy ciągłej pożądany byłby system automatycznego wyznaczania granic czasowych oddzielających segmenty różniące się ze względu na zadane cechy. Wśród wielu parametrów sygnału mowy nie wszystkie w równym stopniu nadają się do wyróżniania odmiennych segmentów. Nie jest też konieczne uwzględnianie w tym celu zbyt wielu parametrów równocześnie. W niniejszej pracy proponuje się w charakterze cechy segmentacyjnej stacjonarność lub niestacjonarność niektórych parametrów częstotliwościowych pochodzących z analizy widmowej akustycznego sygnału mowy. Na tle tej propozycji powstaje pytanie, jak wyznaczone na takiej podstawie segmenty sytuują się względem klasycznych segmentów fonetyczno-akustycznych w mowie i na ile stanowić one mogą obszar identyfikacji tych ostatnich.

## 2. Redefinicja charakterystycznych punktów widmowych sygnału mowy.

W pracach [5] i [6] przedstawiono komputerowy system analizy akustycznej sygnału mowy wyznaczający parametry częstotliwościowe przewidziane do wykorzystania w identyfikacji segmentów fonetycznych. Parametry te zdefiniowano jako miejsca na osi częstotliwości, w których przebieg funkcji modelującej obwiednię widma wykazuje maksymalną wypukłość. Odpowiadające tym miejscom częstotliwości są bliskie częstotliwościom formantów. W niektórych przypadkach miejsca, w których przebieg funkcji modelującej obwiednię widma wykazuje maksymalną wypukłość są równocześnie miejscami wystąpienia ekstremum tej funkcji. Są jednak przypadki nie pokrywania się tych miejsc. Wówczas też miejsce maksymalnej wypukłości lokalnej obwiedni widma gorzej charakteryzuje dane widmo niż miejsce ekstremum lub przegięcia. Z tych dwóch powodów zastąpiono dotychczasowy sposób

charakteryzowania widma za pomocą miejsc maksymalnej wypukłości funkcji modelującej obwiednię widma. Jako charakterystyczne punkty widma przyjęto obecnie te, w których funkcja modelująca obwiednię widma wykazuje minimalne lokalne nachylenie i które jednocześnie występują najbliżej miejsc maksymalnej lokalnej wypukłości tej funkcji. Ten ostatni warunek odnosi się do przypadków, gdy w pobliżu jednej maksymalnej wypukłości lokalnej jest kilka miejsc minimalnego nachylenia funkcji modelującej obwiednię widma. Przypadki takie nierzadko występują dla widma sygnału mowy. Zatem w porównaniu z poprzednią zasadą wyznaczania miejsc charakterystycznych nowa zasada różni się tym, że przyjmuje jako punkt charakterystyczny miejsce minimum nachylenia funkcji modelującej obwiednię widma najbliższe miejscu maksymalnej wypukłości przebiegu tej funkcji. Oddalenie tych dwóch miejsc nie może przekraczać pewnej założonej wartości.

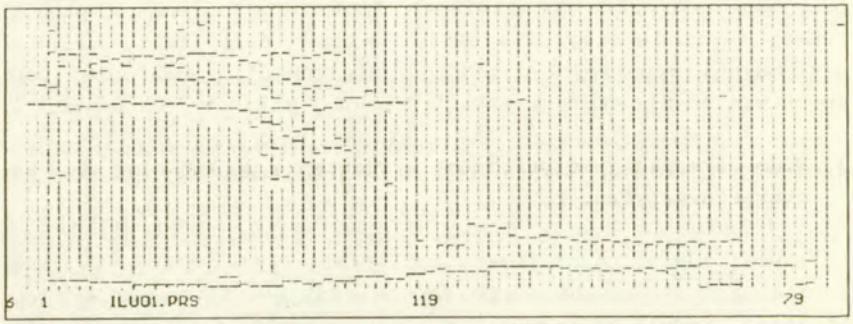
## 3. Widmo średnie jako przedmiot wyznaczania punktów charakterystycznych

Wyżej zdefiniowane punkty charakterystyczne kolejnych widm tworzą w skali czasu ciągi (PC) układające się w mniej lub bardziej regularne przebiegi. Regularność tych przebiegów poprawia się, gdy dotyczą one charakterystycznych punktów nie pojedynczych widm lecz sumy kilku kolejnych widm. Z tego powodu wcześniej powstały system analizy akustycznej sygnału mowy wyposażono obecnie w procedurę wyznaczania średniego widma. Liczba widm składających się na widmo średnie może być teoretycznie dowolna. W praktyce decydują o niej:

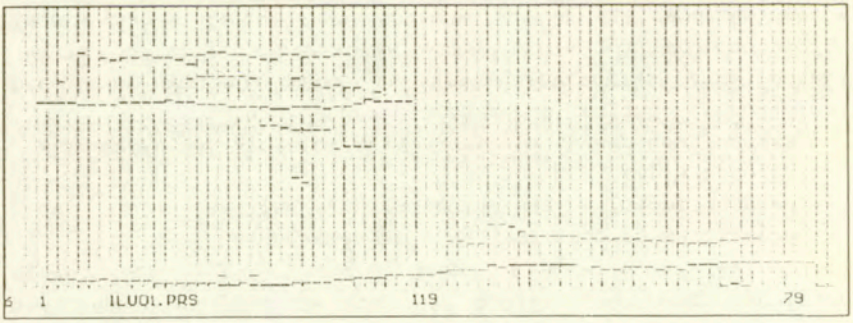
- zakres zachodzenia na siebie fram, dla których wyznaczono kolejne widma,
- częstotliwość, z jaką spróbkowano sygnał mowy oraz
- przeciętna długość segmentu fonetycznego w mowie.

Liczba kolejnych widm składowych nie została definitywnie ustalona. Chwilowo przyjęto ją równą 4, przy 50 procentowym zakresie zachodzenia na siebie fram oraz przy częstotliwości próbkowania sygnału mowy wynoszącej 10 kHz. Gdyby liczba widm składowych była równa ilorazowi długości framy oraz długości zakresu zachodzenia na siebie kolejnych fram, wówczas przy dużych wartościach tego ilorazu widmo średnie nie wykazywałoby uchybów pojawiających się w niektórych widmach składowych wskutek

niekorzystnego usytuowania okna względem okresu podstawowego analizowanego drgania. Przy wartościach tej liczby większych niż wspomniany iloraz, średnie widmo wygładzałoby dodatkowo cechy transsegmentalne sygnału mowy. Na przykładzie wypowiedzi wyrazu *ilu* zilustrowano na rys.1 korzystny wpływ uśredniania widm na wygładzenie przebiegu ciągów punktów charakterystycznych. Rysunek ten w części a) przedstawia ciągi punktów charakterystycznych pochodzących z indywidualnych widm a w części b) analogiczne ciągi wyznaczone z widm uśrednionych.




a)




b)

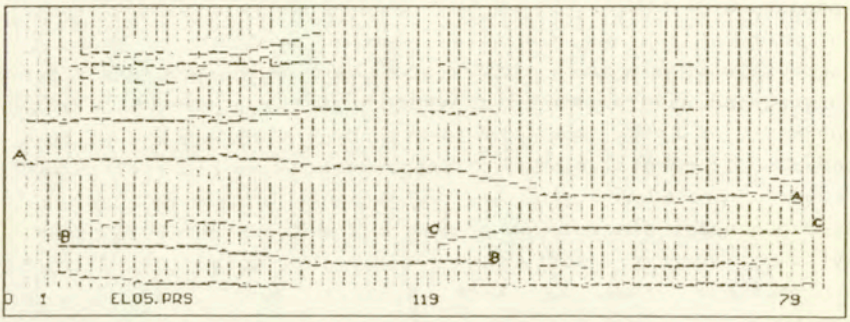
Rys. 1. Ciągi punktów charakterystycznych dla wypowiedzi wyrazu *ilu*; a) z widm indywidualnych, b) z widm uśrednionych.

**4. Wyznaczanie czasowych zakresów stacjonarności ciągów punktów charakterystycznych.**

Dla wspomnianego wyżej systemu analizy akustycznej sygnału mowy opracowano procedurę badania stacjonarności ciągów punktów charakterystycznych. U podstaw algorytmu tej procedury leży zasada, według której o stacjonarności krótkiego segmentu ciągu PC w danym miejscu pola określonego przez częstotliwość i czas decyduje liczba kolejnych punktów ciągu mieszczących się w zadanym oknie. Wymiary tego okna uwzględniać powinny zarówno cechy stacjonarności sygnału mowy jak i parametry jego cyfrowej analizy akustycznej. Pojęcie stacjonarności sygnału mowy jest bardzo szerokie. Stacjonarność sygnału mowy można bowiem rozpatrywać wybiorczo ze względu na jego różne parametry, takie na przykład jak poziom ogólny, poziom w wybranych pasmach, częstotliwość podstawowa lub częstotliwość wybranych formantów. Rozciągłość stacjonarności ze względu na jedne parametry może obejmować aż kilka segmentów fonetycznych a ze względu na inne jedynie fragment segmentu fonetycznego, zwykle centralny. Omawiana procedura umożliwia wyznaczanie czasowych zakresów stacjonarności dla różnych segmentów fonetycznych. Parametrami cyfrowej analizy akustycznej, które decydują o wymiarach wspomnianego okna są krok analizy oraz rozdzielczość częstotliwości. Wyznaczanie czasowych zakresów stacjonarności musi dotyczyć konkretnego parametru oraz opierać się na kryterium oceny jego zmienności. W omawianym przypadku, parametrem jest numer punktu charakterystycznego w widmie. Szereg punktów charakterystycznych pochodzących z kolejnych widm tworzy ciąg. Kryterium zmienności tego ciągu wyrażone jest wymiarami okna. Warunkiem stacjonarności sygnału mowy ze względu na rozpatrywany tutaj parametr jest odpowiednia liczba punktów charakterystycznych objętych oknem. W omawianej procedurze wymiary okna oraz próg stacjonarności są nastawialne. Wstępnie przyjęto okno o wymiarach 8 x 3, czyli o długości 8 kroków czasowych i szerokości 3 punktów częstotliwościowych. Jeden krok czasowy ma długość 12 mS a jednemu punktowi na osi częstotliwości przyporządkowane jest pasmo 33,3 Hz. Próg stacjonarności ustalono na 5. Oznacza to, że aby objęty oknem segment ciągu punktów charakterystycznych mógł być uznany za stacjonarny, wewnątrz okna powinno znaleźć się więcej niż 5 punktów charakterystycznych.

## 5. Mapa ciągów punktów charakterystycznych widma oraz ich fragmentów stacjonarnych.

Wyżej opisane innowacje, o które rozszerzono powstały wcześniej system analizy akustycznej sygnału mowy, umożliwiają prezentację wypowiedzi w formie nałożonych na siebie dwóch prostych obrazów. Ich ilustracją jest rys. 2. Pierwszy z tych obrazów przedstawia ciągi punktów charakterystycznych wyznaczonych według ostatnio zmodyfikowanej definicji punktu charakterystycznego widma sygnału mowy. Uwzględniane są jedynie te punkty, w których widmo wykazuje poziom nie niższy niż 30 dB poniżej poziomu maksymalnego w widmie.




Rys. 2. Mapa ciągów PC z zaznaczonymi fragmentami stacjonarnymi.

Przedmiotem wyznaczenia punktów charakterystycznych jest widmo średnie z czterech kolejnych widm "chwilowych". Na obrazie punkt charakterystyczny zaznaczony jest krótkim odcinkiem, którego długość jest adekwatna do poziomu widma w tym punkcie. Drugi obraz przedstawia pozycje okna prostokątnego o szerokości trzech punktów, w których obejmuje ono punkty charakterystyczne w ponad 62.5 % swojej długości . W celu wyznaczenia tych pozycji wspomniane okno przebiega pole określone przez częstotliwość i czas w skali dyskretnej. Pozycję okna wystarczająco wypełnionego zaznacza na tym obrazie krótki odcinek nieznacznie przesunięty w kierunku pionowym względem poziomu mających ten sam azymut częstotliwościowy odcinków oznaczających punkty charakterystyczne

i w kierunku poziomym względem podstawy podobnych odcinków mających ten sam azymut czasowy. Liczba punktów mogących maksymalnie wypełnić okno równa jest długości okna. Nie uwidacznia się pozycji okna, przy których jego wypełnienie punktami charakterystycznymi w kierunku poziomym obejmuje mniej niż 62.5 % długości okna. Oznaczenie pozycji okna wystarczająco wypełnionego wskazuje na krótki stacjonarny segment ciągu punktów charakterystycznych. Krótkie odcinki oznaczające takie stacjonarne segmenty ciągu punktów charakterystycznych układają się także w ciągi. Ciągi te będą odtąd nazywane skrótem SSCPC utworzonym z pierwszych liter wyrazów tworzących ich pełne określenie: (ciągi) segmentów stacjonarnych ciągu punktów charakterystycznych. Ciągi te są usytuowane bądź równolegle bądź pod małym kątem względem osi czasu i przyjmują różną długość. Brak stacjonarności ciągu punktów charakterystycznych może być spowodowany głównie zmiennym przebiegiem tego ciągu. Prócz tej przyczyny mogą być jeszcze inne powody, na przykład luki w ciągu punktów charakterystycznych albo zbyt mała jego długość. Przerywany lub krótki ciąg punktów charakterystycznych może mieć znamiona stacjonarności wyrażające się w jego ułożeniu równoległym względem osi czasu lecz nie spełniać warunku wystarczającego wypełnienia okna.

## 6. Obserwacja pozycji fragmentów stacjonarnych ciągu punktów charakterystycznych na tle położenia segmentów fonetycznych w wybranych przykładach.

Wyznaczenie granicy między głoskami w akustycznym sygnale mowy jest tak samo problematyczne jak wskazanie początku i końca niektórych głosek. W obrębie danej głoski występuje zwykle kilka faz. W fazach skrajnych dają o sobie znać wpływy głosek sąsiednich. W fazie centralnej głoska posiada zwykle charakter częściowo stacjonarny. O sygnale mowy wiadomo [1], że prawie każdy jego fonem posiada pewne stadium ustalone. Względna długość tego stadium jest różna. Odznaki stacjonarności przejawiać się mogą w przebiegu jedynie niektórych, zwykle tych najbardziej znaczących parametrów. W tym samym czasie inne parametry niekoniecznie muszą pozostawać niezmienne.

Przestrzeń czasową wypowiedzi można posegmentować za pomocą punktów wyznaczających początek i koniec fragmentów stacjonarnych

ciągów punktów charakterystycznych. Powstałe w ten sposób segmenty czasowe dzielą się na takie, które obejmują wyłącznie fragmenty stacjonarne we wszystkich ciągach punktów charakterystycznych, takie które obejmują wyłącznie fragmenty niestacjonarne w tychże samych ciągach oraz takie, które obejmują fragmenty tych ciągów zarówno jednego jak i drugiego rodzaju . W czasie trwania dwóch do trzech głosek występuje przynajmniej raz początek i koniec jednego z ciągów SSCPC. Ze względu na wypełnienie przestrzeni pomiędzy końcem jednego takiego ciągu a początkiem najbliższego z następnych podobnych ciągów możliwe są następujące warianty. Pierwszy wariant cechuje brak wypełnienia tej przestrzeni. W drugim wariancie wspomnianą przestrzeń wypełnia nieprzerwany ciąg punktów charakterystycznych. W wariancie trzecim przestrzeń tę wypełniają punkty charakterystyczne niekoniecznie tworzące nieprzerwany lub regularny ciąg. W przypadku pierwszego wariantu brak jest odznak stadium nieustalonego. W obu pozostałych wariantach są podstawy do wnioskowania, że rozpatrywana przestrzeń leży w segmencie niestacjonarnym. Na podstawie różnych ciągów punktów charakterystycznych widma oraz ciągów SSCPC - wskazujących fragmenty stacjonarne w ciągach punktów charakterystycznych - można zorientować się w rozkładzie stacjonarnych i niestacjonarnych segmentów sygnału mowy. Fragmenty stacjonarne mają wskazywać zakres czasowy głoski, z którego pobrana próba powinna umożliwiać identyfikację głoski. Miejsce czasowe pobrania tej próby powinno przypadać na osi przecinającej wyłącznie fragmenty stacjonarne głównych ciągów punktów charakterystycznych. Miejsc takich może być wiele i tworzyć one mogą różnej długości zakresy. Na główne ciągi składają się te punkty charakterystyczne, w których poszczególne widma wykazują poziom mieszczący się w odpowiednim zakresie. Nie rozpatrywane są punkty charakterystyczne, w których poziom widma jest niski. Segmenty niestacjonarne ciągów punktów charakterystycznych wskazują na fazy pojawiania się i zaniku głoski. Miejsca czasowe, w których oś poprowadzona równolegle do osi częstotliwości przecina same segmenty niestacjonarne głównych ciągów punktów charakterystycznych przypadają w zkresie stadiów nieustalonych głoski.

Na przykładzie kilku wypowiedzi krótkich wyrazów zilustrowane zostaną próby fonematycznej segmentacji sygnału mowy

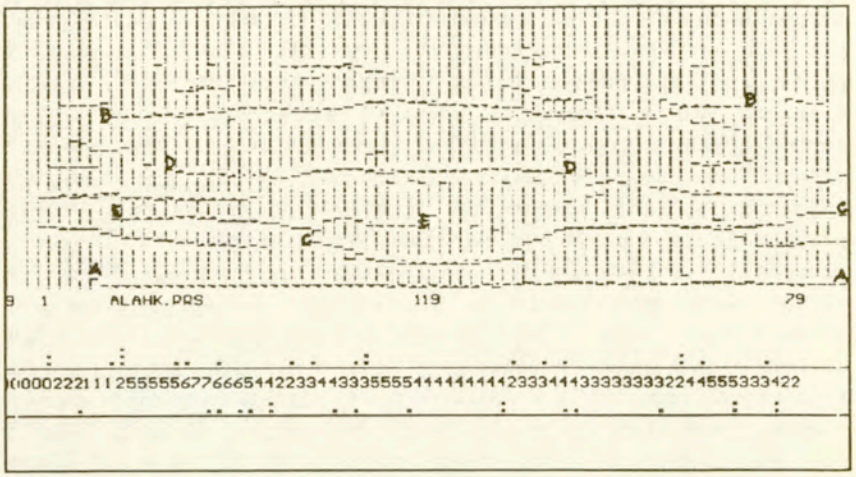
na podstawie stacjonarnych i niestacjonarnych fragmentów ciągów punktów charakterystycznych. Pierwszy przykład pochodzi z wypowiedzi wyrazu *ela*. Rys.2 przedstawia mapę ciągów punktów charakterystycznych wyznaczonych dla kolejnych widm. Podstawa rysunku pokrywa się z osią czasu a jego wysokość z osią częstotliwości. Liczby 1 i 79 wskazują na osi czasu odpowiednio początek i koniec okna czasowego, w którym rozpatrywana jest wypowiedź. Poszczególne widma, dla których wyznaczano punkty charakterystyczne dotyczą miejsc na osi czasu rozstawionych z odstępem 12 ms. Punkty charakterystyczne oznaczono na rysunku krótkimi odcinkami. Odcinki te układają się w różnej długości ciągi. Najdłuższy ciąg oznaczony liter'a A przebiega wzdłuż środkowej części mapy. Jest parę bardzo krótkich ciągów liczących zaledwie po kilka, - 2 do 5 - punktów. Jest także kilka pojedynczych punktów. Zarówno pojedyncze punkty, jak i krótkie ciągi nie będą na razie rozpatrywane. Pod lub nad niektórymi odcinkami oznaczającymi punkty charakterystyczne znajduje się bardzo krótki odcinek wskazujący pozycję środka okna o wymiarach 3 * 8, w której obejmuje ono lokalnie najdłuższy oraz najbliżej osi okna położony segment ciągu punktów charakterystycznych. Znak taki świadczy zatem, że odpowiedni segment ciągu punktów charakterystycznych objętych oknem uważa się za stacjonarny. Ciąg bardzo kótkich odcinków wskazuje na dłuższy fragment stacjonarnego ciągu punktów charakterystycznych. W ciągu punktów charakterystycznych rozpatrywanego przykładu oznaczonym literą A wyróżnione zostały trzy fragmenty wskazujące na charakter stacjonarny ciągu. Fragmenty te sugerować mogą ewntualne segmenty fonetyczne. Potwierdzenia tej sugestii należy szukać oceniając rozkład segmentów stacjonarnych w innych ciągach punktów charakterystycznych. W ciągu oznaczonym literą B wyróżnione zostały trzy fragmenty, w których ciąg ten wykazuje cechy stacjonarności. Środkowy z tych fragmentów jest znacznie krótszy od dwóch pozostałych i jest odgraniczony krótką, zaledwie jednopunktową przerwą od fragmentu pierwszego. Za wyjątkiem swoich kilku skrajnych punktów fragment trzeci jest usytuowany w tym samym przedziale czasu co fragment drugi najdłuższego ciągu w przykładzie. Fragment pierwszy ciągu punktów charakterystycznych oznaczonego literą B położony jest wewnątrz przedziału wyznaczonego przez fragment pierwszy ciągu A. Pozwala to przypuszczać, że pierwszy i drugi fragment stacjonarny ciągu

A oraz pierwszy i trzeci fragment ciągu B wskazują na te same dwa segmenty fonetyczno-akustyczne mowy. Fragment trzeci ciągu punktów charakterystycznych oznaczonego literą A mieści się w przedziale wyznaczonym przez bardzo długi fragment stacjonarny ciągu punktów charakterystycznych oznaczonego literą C. Ta zbieżność lokalizacji stacjonarnych fragmentów obu tych ciągów wskazuje na segment fonetyczny. Fragmenty niestacjonarne rozpatrywanych ciągów sytuują się także w bliskich sobie lub identycznych miejscach na osi czasu i można je z tego powodu traktować jako wskazanie stref przejściowych pomiędzy segmentami fonetycznymi. W ciągu punktów charakterystycznych oznaczonym literą A są 3 fragmenty niespełniające warunku stacjonarności. Jeden z nich jest na końcu tego ciągu będącym jednocześnie końcem wypowiedzi. Dwa pozostałe odgraniczają kolejne fragmenty stacjonarne. Ciąg PC oznaczony literą B ma 2 fragmenty niestacjonarne. Odgraniczają one fragmenty stacjonarne tego ciągu. Drugi z nich jest usytuowany blisko jednego z segmentów niestacjonarnych ciągu A rozpatrywanego przykładu. W ciągu oznaczonym literą C jest tylko jeden fragment niestacjonarny - na początku tego ciągu. Jego miejsce pokrywa się po części z położeniem na osi czasu drugiego z trzech fragmentów niestacjonarnych ciągu A. Fragment niestacjonarny odgradzający dwa sąsiednie segmenty stacjonarne ciągu PC obejmuje fazę zmiany położenia punktów charakterystycznych z pozycji, jaką zajmowały one na końcu ostatniego fragmentu stacjonarnego do pozycji, jaką przyjmują one na początku najbliższego następnego fragmentu stacjonarnego. Jeśli pozycje punktów charakterystycznych w dwóch kolejnych fragmentach stacjonarnych danego ciągu PC odpowiednio na końcu i początku tych fragmentów są wyraźnie różne oraz jeśli między nimi przypada odpowiedniej długości fragment niestacjonarny, dowodzi to, że owe fragmenty stacjonarne wskazują na różne segmenty fonetyczne. Zakres pokrywania się na osi czasu miejsc fragmentów stacjonarnych niektórych ciągów PC można uważać za położenie centralnego segmentu głoski.

W rozpatrywanym przykładzie oprócz wyżej omówionych najdłuższych ciągów PC jest jeszcze 8 innych ciągów. Cztery z nich mają w całości charakter stacjonarny. Ich położenie wzdłuż osi czasu w różnym stopniu pokrywa się z położeniem fragmentów stacjonarnych poprzednio omawianych ciągów w przykładzie. Zawsze jednak ma miejsce przynajmniej częściowa zgodność tych położeń. W trzech

przypadkach na 8 zakres czasowy krótszego fragmentu stacjonarnego przypada w pełni w zakresie czasowym dłuższego fragmentu. W pozostałych przypadkach krótsze fragmenty na znacznej długości mieszczą się w dłuższych. Początki a także końce częściowo zachodzących na siebie w wymiarze czasu fragmentów stacjonarnych różnych ciągów są rozmaicie rozłożone. Są one poza pewnymi wyjątkami blisko siebie położone, lecz prawie nigdy nie przypadają w tym samym punkcie czasu. Wspomniane wyjątki stanowią przypadki krótkich ciągów pojawiających się niespodziewanie. W rozpatrywanym przykładzie są 3 takie ciągi - trzeci od dołu ciąg w lewej części mapy, najwyższy ciąg w części środkowej oraz drugi od dołu ciąg w prawej części mapy. Końce fragmentów stacjonarnych tych ciągów przypadają w wymiarze czasu blisko fragmentów innych ciągów, natomiast początki tych fragmentów sytuują się w oddzielnych punktach czasu nieraz znacznie odległych od początków fragmentów innych ciągów, z których końcami blisko sąsiadują końce fragmentów tych krótkich ciągów.

Spostrzeżenia i wnioski z nich wypływające zebrane na przykładzie wypowiedzi wyrazu *ela* zostały uzupełnione na dwóch następnych przykładach. Przykłady te pochodziły z dwóch wypowiedzi wyrazu *ala* dostarczonych przez dwa odmienne głosy męskie. Mapa ciągów punktów charakterystycznych dla pierwszego z tych dwóch przykładów zamieszczona jest na rysunku 3. Dominują na niej ciągi długie oraz średnio długie. Są także ciągi krótkie. Niektóre z nich ze względu na swą małą długość nie mogły spełnić przyjętego kryterium stacjonarności. Podobnie jak przy rozpatrywaniu poprzedniego przykładu również teraz nie będą brane pod uwagę pojedynczo występujące punkty charakterystyczne oraz występujące niekiedy pary takich punktów. W zamieszczonym przykładzie najdłuższym ciągiem jest ciąg pierwszy czyli najniższy, oznaczony literą A. Ciąg ten w całości został uznany za stacjonarny. Nie zostały wskazane w nim oddzielne fragmenty stacjonarne przedzielone przerwami. Drugi pod względem długości ciąg PC oznaczony na mapie zamieszczonej na rysunku 3 literą B znajduje się w jej górnej części. Ta część mapy dla rozpatrywanego przykładu odnosi się do zakresu częstotliwości trzeciego formantu. Częstotliwość tego formantu w znacznie mniejszym stopniu niż częstotliwości niższych formantów charakteryzuje różne segmenty fonetyczne i z tego powodu mniej nadaje się jako wskaźnik do segmentacji fonematycznej sygnału mowy.




Rys. 3. Mapa ciągów PC z zaznaczonymi fragmentami stacjonarnymi i rozkładem azymutów czasowych początku i końca tych ostatnich oraz wiersz liczebnosci fragmentów stacjonarnych na przestrzeni czasu wypowiedzi. Mapa dotyczy wypowiedzi wyrazu *ala* głosem HK oraz przypadku wyznaczenia fragmentów stacjonarnych ciągów PC dla całego pasma częstotliwości.

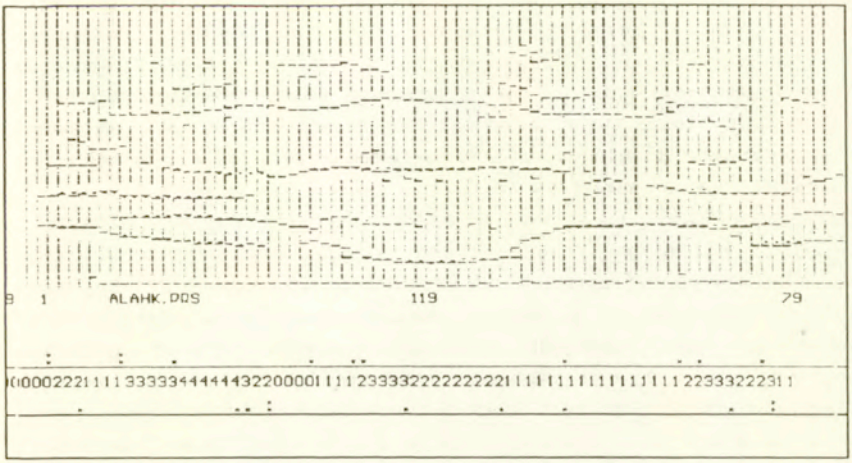
Dlatego fragmenty stacjonarne ciągu B, odzwierciedlającego trzeci formant nie mogą odgrywać głównej roli w próbie przeprowadzenia segmentacji fonematycznej. Jest ich w tym ciągu aż 5. Pierwszy i ostatni z nich przypadają odpowiednio w zakresie pierwszego i drugiego *a* w rozpatrywanej wypowiedzi wyrazu *ala*. Fragmenty drugi i czwarty częściowo także przypadają w obrębie segmentów obu *a* tej wypowiedzi, lecz po części zachodzą także na segment *l*. Różnice w położeniu wszystkich tych fragmentów w wymiarze częstotliwości są znikome i nie sugerują wraźnie, że poszczególne fragmenty ciągu PC wskazywać mogą różne segmenty fonetyczne. Także krótkie odstępy pomiędzy końcem jednego fragmentu a początkiem następnego nie wskazują, iż te dwa fragmenty leżą w obrębie segmentów różnych fonemów.
Pozostałe 3 długie ciągi PC rozpatrywanego przykładu związane są z niższymi formantami wyrażającymi cechy dystynktywne. W ciągach tych są albo dłuższe odstępy pomiędzy kolejnymi fragmentami

niestacjonarnymi, albo położenia tych fragmentów w wymiarze częstotliwości różnią się znacznie, albo zachodzi jedno i drugie jednocześnie. Takie ułożenie fragmentów stacjonarnych ciągu PC wskazuje, iż poszczególne fragmenty stacjonarne przypadają w różnych segmentach fonetycznych. I tak pierwszy ze wspomnianych trzech długich ciągów PC, oznaczony literą C rozciąga się w zakresie czasu głosek *l* oraz *a* i ma dwa fragmenty stacjonarne oddzielone znaczną przerwą. Taki układ fragmentów stacjonarnych wskazuje na dwa segmenty fonetyczne ze strefą przejściową pomiędzy nimi. Drugi z tych trzech ciągów oznaczono literą D. Mieści się on w środku zakresu częstotliwości mapy i zaczyna się w niewielkim oddaleniu od jej lewego brzegu a kończy za połową jej długości. Posiada on także dwa fragmenty stacjonarne oddzielone przerwą na tyle długą, aby uznać, że fragmenty te dotyczą dwóch różnych segmentów fonetycznych. Podobnie zinterpretować można ostatni z trzech wspomnianych długich ciągów. Oznaczono go literą E. Jest on w znacznej części swojej długości trzecim w kolejności - w kierunku wyznaczonym przez oś częstotliwości - ciągiem PC. Ma on dwa fragmenty stacjonarne oddzielone znaczną przerwą, wskazującą na przestrzeń między dwoma segmentami fonetycznymi. Sąsiadujące z sobą fragmenty stacjonarne oraz przerwa między nimi w obu ostatnio omawianych ciągach przypadają w pokrywających się po części zakresach czasu.
Krótkie ciągi PC w rozważanym przykładzie mają przeważnie po jednym fragmencie stacjonarnym. Są też częste przypadki, że ze względu na małą długość ciągu nie został w nim wskazany żaden fragment stacjonarny. Krótkie ciągi PC mogą przyczynić się do wskazania segmentów fonetycznych, jeśli posiadają fragment stacjonarny i jeśli początek lub/i koniec fragmentu stacjonarnego tych ciągów przypadają w pobliżu początku lub/i końca fragmentów stacjonarnych innych ciągów.

Ze względu na złożony obraz mapy ciągów PC rozpatrywanego przykładu zbadano, jak dla tego przykładu przedstawia się rozkład początków i końców fragmentów stacjonarnych we wszystkich ciągach oraz tylko w niektórych, a mianowicie w tych, które mieszczą się w paśmie częstotliwości dystynktywnym dla głosek występujących w przykładzie.
Poniżej mapy ciągów PC zamieszczonej na rys. 3 wykreślono dwie linie równoległe do osi czasu. Nad górną linią naniesiono azymuty czasowe początków fragmentów stacjonarnych a nad dolną azymuty

czasowe końców tychże fragmentów. Punkty zaznaczone na osi początków i końców fragmentów stacjonarnych wszystkich ciągów PC są w rozpatrywanym przykładzie raczej rozproszone. Owszem w niektórych miejscach występują słabe i mało liczne skupienia tych punktów. Jest jednak wiele punktów, które ze względu na znaczne oddalenie od sąsiednich punktów nie mogą być zaliczone do żadnego skupienia. Dwa wzajemnie blisko położone skupienia, z których jedno leży na osi początków a drugie na osi końców fragmentów stacjonarnych, mogą wskazywać na nagłe przejście z jednego segmentu fonetycznego na drugi. Na osiach dotyczących azymutów czasowych fragmentów wszystkich ciągów widoczne są dwie takie pary skupień. Znajdują się one rzeczywiście w zakresach czasu obejmujących stadia przejściowe między głoskami *a* i *l* oraz *l* i *a*. W pobliżu końca mapy na obu osiach znajduje się jeszcze po jednym mało skoncentrowanym skupieniu, przy czym skupienie azymutów czasowych początków fragmentów stacjonarnych nieco wyprzedza rozciągnięte na znacznej długości azymuty czasowe licznych końców tych fragmentów. Świadczy to, że do wcześniej zainicjowanych fragmentów stacjonarnych dochodzą wpierw nowe a następnie wszystkie stopniowo kończą się. Skupienia pierwszej pary następują w zrozumiałej kolejności tzn. wpierw skupienie konców a potem skupienie początków. W przypadku drugiej pary skupienie azymutów początków kończy się wcześniej niż skupienie azymutów końców i zakresy obu tych skupień w znacznym stopniu pokrywają się.

Bardziej jasno przedstawia się sytuacja na osiach z naniesionymi azymutami początków i końców fragmentów stacjonarnych jedynie tych ciągów PC, które odnoszą się do pasma częstotliwości obejmującego dwa pierwsze formanty głosek występujących w badanej wypowiedzi. Skupienia pierwszej pary są oczywiście mniej liczne ale bardziej skoncentrowane niż w przypadku, gdy brane były pod uwagę wszystkie ciągi PC. Fragmenty stacjonarne, których azymuty początku tworzą jedno ze skupień tej pary kończą się stopniowo na względnie długiej przestrzeni czasu. Końce tych fragmentów nie tworzą więc skupienia. Następne fragmenty stacjonarne rozpoczynają się w znacznej odległości czasowej od końca poprzednich i są bardzo krótkie. Na położenie segmentu fonetycznego wskazuje przestrzeń pomiędzy skupieniem azymutów czasowych początku fragmentów stacjonarnych a skupieniem azymutów ich końca.



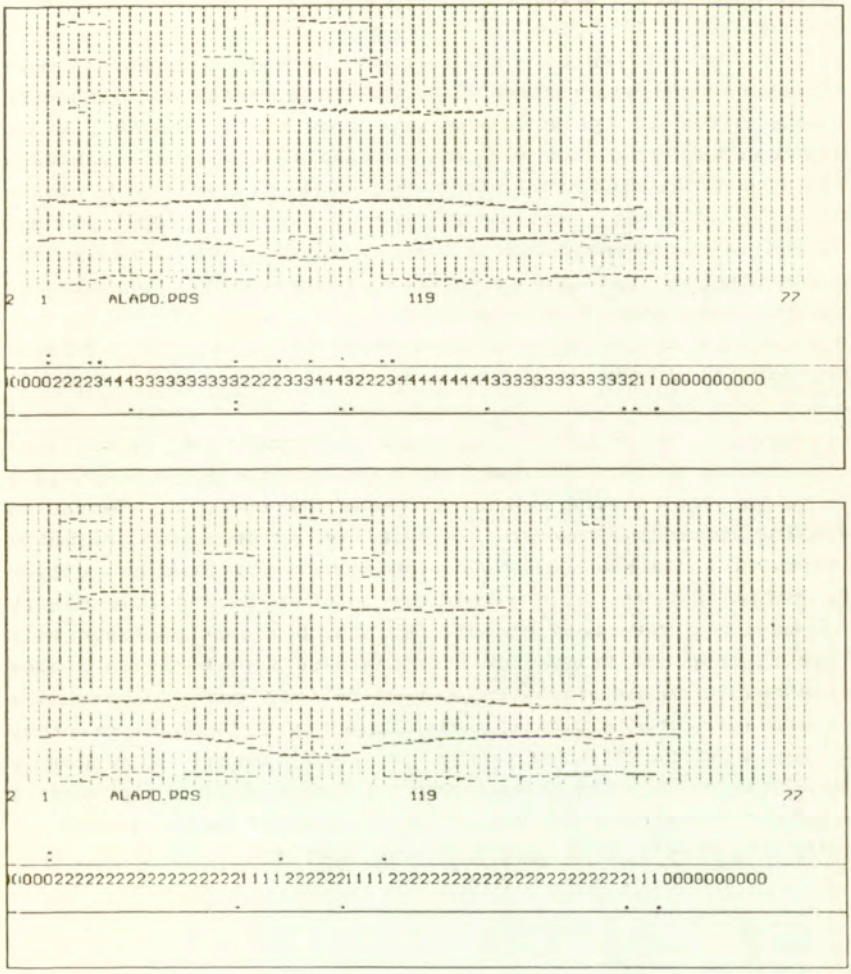

Rys. 4. Mapa ciągów PC z zaznaczonymi fragmentami stacjonarnymi i rozkładem azymutów czasowych początku i końca tych ostatnich oraz wiersz liczebnosci fragmentów stacjonarnych na przestrzeni czasu wypowiedzi. Mapa dotyczy wypowiedzi wrazu *ala* głosem HK oraz przypadku wyznaczenia fragmentów stacjonarnych ciągów PC dla pasma częstotliwości obejmującego dwa pierwsze formanty głosek występujących w badanej wypowiedzi.

W przypadku, jaki ma miejsce w rozpatrywanym przykładzie, gdy nie ma skupienia azymutów czasowych końca fragmentów stacjonarnych, a jest skupienie azymutów początku wnioskować można o istnieniu rozległego przejścia do następnego segmentu fonetycznego. Na początku mapy również istnieje rozproszenie azymutów czasowych początku fragmentów stacjonarnych. Świadczy to o pewnym procesie tworzenia się segmentu fonetycznego. Koniec tego segmentu następuje dość nagle. Wskazuje na to skupienie azymutów końca fragmentów stacjonarnych ciągów PC. Następny segment fonetyczny ma już dość jednoznaczny początek. Świadczy o tym skupienie azymutów czasowych początku fragmentów stacjonarnych ciągów PC. Jednakże, jak już powiedziano wyżej, zakończenie tego segmentu fonetycznego następuje w kilku krokach na dłuższej przestrzeni czasowej. Część ustalona ostatniego

segmentu fonetycznego jest krótka. Za wskazanie jego początku można przyjąć azymuty czasowe początku dwóch fragmentów stacjonarnych ciągów PC. Azymuty czasowe końca tych fragmentów oraz współbieżnego z nimi fragmentu zainicjowanego znacznie wcześniej świadczą, że zakończenie ostatniego segmentu fonetycznego przykładowej wypowiedzi nie jest nagłe, lecz rozciąga się w czasie.

Pomiędzy osiami, na które naniesiono azymuty czasowe początku i końca fragmentów stacjonarnych ciągów PC zapisano ciąg liczebności tych fragmentów we współbieżnych ciągach PC. Zapisana w danym miejscu liczba, będąca zawsze jednocyfrową, wyraża ilość segmentów stacjonarnych różnych ciągów, jakie przecina poprowadzona w tym miejscu oś równoległa do osi częstotliwości. Ciąg liczebności fragmentów stacjonarnych charakteryzuje się tym, że składa się on z krótkich ciągów liczb o jednakowej wartości. Pojedyńczy krótki ciąg odpowiada jakiemuś segmentowi mowy wyznaczonemu przez początek lub/i koniec różnych lub jednego tylko fragmentu stacjonarnego także różnych lub jednego ciągu PC. Dla lokalizacji fonemu w mowie znaczenie szczególne mogą mieć krótkie ciągi liczb małych i dużych w porównaniu z liczbami tworzącymi krótkie ciągi sąsiednie. W rozpatrywanym przykładzie krótki ciąg małych liczb wskazuje na miejsce graniczne między sąsiadującymi z sobą fonemami. Krótki ciąg dużych liczb wskazuje na segment fonemu, z którego próba parametrów widmowych stanowić powinna właściwą reprezentację tego fonemu dla jego klasyfikacji i identyfikacji.

Na rysunku 5 przedstawiono tego samego rodzaju dane, jak na rysunkach 3 i 4. Odnoszą się one również do wypowiedzi wyrazu *ala* wypowiedzianego jednakże przez inny głos męski. Glos ten jest znacznie bardziej spektrogeniczny od głosu, który dostarczył poprzedniego przykładu. W porównaniu z poprzednim przykładem obecny charakteryzuje się znacznie mniejszą liczbą ciągów PC. Fragmenty stacjonarne ciągów PC przebiegających w dystynktywnym paśmie częstotliwości (patrz dolna część rys.5) jednoznacznie wskazują na poszczególne fonemy. Fragmenty stacjonarne niektórych z ciągów przebiegających poza tym pasmem wskazują na więcej fonemów niż jest ich w wypowiedzi. Dodatkowe wskazania wyróżniają różne stadia fonemu trudno rozpoznawalne w obrębie ściśle dystynktywnego pasma częstotliwości.




Rys. 5. Mapa ciągów PC z zaznaczonymi fragmentami stacjonarnymi i rozkładem azymutów czasowych początku i końca tych ostatnich oraz wiersz liczebności fragmentów stacjonarnych na przestrzeni czasu wypowiedzi. Mapa dotyczy wypowiedzi wrazu *ala* głosem PD. Części górna i dolna rysunku odnoszą się odpowiednio do fragmentów stacjonarnych ciągów PC w całym i dystynktywnym pasmie częstotliwości.

## 7. Podsumowanie i wnioski.

W poszukiwaniu cech segmentacyjnych sygnału mowy zainteresowano się rozkładem czasowym fragmentów stacjonarnych ciągów PC. Do przeprowadzenia odpowiednich badań, które pozwoliłyby ustalić, w jakim stopniu znajomość zakresów stacjonarności niektórych ciągów PC pozwala na lokalizację głosek w mowie ciągłej, konieczne jest posiadanie odpowiedniego narzędzia badawczego. Powstało ono w ramach zadania badawczego zakończonego tą pracą. Na przedstawionych przykładach zilustrowano wynik jego działania.

O pozycji głoski w mowie ciągłej można wnioskować na podstawie rozkładu azymutów czasowych początku i końca fragmentów stacjonarnych ciągów PC. Skupienia tych azymutów, jeśli występują, są miejscami charakterystycznymi w przestrzeni czasowej wypowiedzi. Stanowić one mogą informację tym pewniejszą im są bardziej liczne i skoncentrowane. Wskazują one wówczas na granicę pomiędzy fonemami. Niewielkie rozproszenie azymutów czasowych początków lub/i końców fragmentów stacjonarnych ciągów PC może wskazywać na rozleglejsze stadium przejściowe między fonemami. Do wskazania części centralnych głosek posłużyć może ciąg liczebności fragmentów stacjonarnych współbieżnych ciągów punktów charakterystycznych widma. Zakresy dużej liczebności wskazują na stadia głosek kwalifikujące się do pobrania próby głosek w celu ich klasyfikacji lub identyfikacji. Przy pomocy przedstawionego w tej pracy narzędzia będzie można przeprowadzić w szerokim zakresie badania, których przykład na nielicznym materiale fonetycznym zaprezentowano powyżej.

*B i b l i o g r a f i a*

[1]. AINSWORTH, W.A., *Speech Recognition by Machine*, Peter Peregrimus Ltd. London 1988.

[2]. GLASS, J.R., *Finding Acoustic Regularities in Speech: Application to Phonetic Recognition*, Ph. D. Thesis, MIT press, May 1988.

[3]. HOLMES, J.N., *Speech Synthesis and Recognition*, Van Nostrand Reinhold (UK) Co.Ltd.,1988.

[4]. JASSEM, W., KUBZDELA, H., DOMAGAŁA, P., *Automatic acoustic-phonetic Segmentation of the Speech Signal*, Acta Universitatis Umensis, From Sound to Word, ed. by R. Hedquist, Umea 1983.

[5]. KUBZDELA, H.K., *System do badania cech widmowych oraz selekcji i odsłuchu segmentów sygnału mowy*, Prace IPPT 28/1993, W-wa 1993.

[6]. KUBZDELA, H.K., *Automatyczna ekstrakcja wybranych cech widmowych mowy*, Prace IPPT 2/1994, W-wa 1994.

[7]. VIDAL, E.,MARZAL, A., *A Review and New Approach for Automatic Segmentation of Speech Signals*, Signal Processing V : Theories and Applications, L. Torres, E. Masgran and M.A. Lagunas, Elsevier Science Publisher B. V., pp 43-53, 1990.

[8]. ZUE, V.W., GLASS, J., PHILIPS, M. and SENEFF, S., *Acoustic Segmentation and Phonetic Classification in the Summit System*, Proceeding of IEEE ICASSP-89, pp 389-392.