

20/1985

Piotr Domagała

TWORZENIE WZORCÓW
JEDNOSTEK SEGMENTALNYCH SYGNAŁU MOWY
DO CELÓW
JEJ AUTOMATYCZNEGO ROZPOZNAWANIA

20/1985

P. 269



WARSZAWA 1985

ISSN 0208-5658

Praca wpłynęła do Redakcji dnia 28 listopada 1984 r.



56928



Na prawach rękopisu

Instytut Podstawowych Problemów Techniki PAN

Nakład 140 egz. Ark.wyd. 0,75 Ark.druk. 1,25

Oddano do drukarni w kwietniu 1985 r.

Nr zamówienia 277/85

Warszawska Drukarnia Naukowa, Warszawa,
ul. Śniadeckich 8

Piotr Domagała

Pracownia Fonetyki Akustycznej IPPT PAN

TWORZENIE WZORCÓW JEDNOSTEK SEGMENTALNYCH SYGNAŁU MOWY DO CELÓW JEJ AUTOMATYCZNEGO ROZPOZNAWANIA

Streszczenie.

Zbiór 34 logatów, zbudowanych z 5 fonemów i wypowiedzianych przez 1 głos, poddano trzem procedurom automatycznego rozpoznawania segmentów sygnału mowy. Celem procedur jest rozpoznanie segmentu sygnału mowy poprzez detekcję maksymalnej wartości miary podobieństwa z elementami zbioru wzorców uzyskanych w sposób przypadkowy.

1. Wstęp.

Wyniki wieloletnich prac H. Kubzdela nad techniką otrzymywania spektrogramów binarnych i zastosowaniem ich do globalnego rozpoznawania wyrazów ([8], [9], [10], [11]) oraz dobre, wstępne wyniki automatycznej segmentacji sygnału mowy uzyskane w Pracowni Fonetyki Akustycznej IPPT PAN ([2], [6]) stały się zachętą do podjęcia prac nad automatycznym rozpoznawaniem fonemów w oparciu o ten typ spektrogramów. Niniejsza praca jest sprawozdaniem z wstępnych badań nad automatycznym rozpoznawaniem segmentów sygnału mowy, wyodrębnionych ze spektrogramów binarnych i o rozciągłości zbliżonej do głoski.

Problemy związane z otrzymywaniem spektrogramów binarnych i ich automatyczną segmentacją zostały celowo pominięte. Autor korzystał z programów otrzymywania spektrogramów binarnych, opracowanych i życzliwie udostępnionych przez Pana dr H. Kubzdela,

¹ Praca wykonana w ramach problemu międzyresortowego MR.I-24.

któremu pragnie w tym miejscu podziękować. Segmentacji sygnału mowy dokonywano na drodze wizualnej, wg zasad ogólnie stosowanych ([4], [5]).

Pierwsze badania nad zastosowaniem spektrogramów binarnych do wizualizacji sygnału mowy wskazują na segmentalny sposób analizy widm binarnych przez respondentów [12]. Celem przedstawianej pracy jest próba odpowiedzi na pytanie, czy można w sposób automatyczny rozpoznawać segmenty uzyskane ze spektrogramów binarnych.

2. Opis doświadczeń.

Spektrogram binarny sygnału mowy jest binarną funkcją dwóch zmiennych dyskretnych będącą wynikiem przekształceń dokonanych na obrazie widmowym tego sygnału. W niniejszej pracy przeprowadzono doświadczenia na spektrogramach złożonych z 16 bitowych widm binarnych otrzymywanych w odstępach 23 ms, które opisano w pracy [11]. Na rys. 1 pokazano przykładowy spektrogram binarny wyrazu "statut". Po dokonaniu wstępnej analizy można jednoznacznie wskazać fragmenty spektrogramu charakterystyczne dla samogłosek, głosek zwartych i głoski trącej. Na szczególne podkreślenie zasługuje wysoki stopień redukcji informacji : do zapamiętania takiego spektrogramu wystarczą 64 bajty.

Analizując spektrogramy wielu wyrazów, logatomów i fonemów można zauważyć, że istnieją fonemy, których spektrogramy są tak odmienne, że ich zróżnicowanie nie przedstawia większej trudności, np. pary fonemów /s,u/, /i,a/, /o,t/. Nietrudno wskazać również spektrogramy fonemów o takim stopniu podobieństwa, że zróżnicowanie ich jest bardzo utrudnione lub niemożliwe, np. pary /i,j/, /u,w/, /ʃ,ʧ/. W pracy [7] podano miary podobieństwa samogłosek obliczone metodą analizy dyskryminacyjnej. Dla celów niniejszej pracy utworzono zbiór 3 samogłosek /i,a,u/ o możliwie największym stopniu zróżnicowania [7]. Zbiór wybranych samogłosek uzupełniono a priori głoską trącą /s/ i zwartą /t/. Wizualna analiza spektrogramów binarnych wybranych fonemów potwierdziła dobrą, wzajemną rozróżnialność.

Przeprowadzono 3 eksperymenty, w których należało rozpoznać w sposób automatyczny 34 logatomy zbudowane z głosek należących

do wyżej wymienionego zbioru (Tablica 1) i wypowiedziane przez jeden głos. Opracowana lista logatomów zapewnia w przybliżeniu równomierne rozkłady udziałów poszczególnych fonemów i diad (Tablica 2 i 3). Utworzona tak lista nie jest reprezentatywna dla zasad fonotaktyki polskiej. Poszczególne logatomy odczytano w sposób naturalny, głosem męskim, niewyćwiczonym, o stosunkowo słabym zróżnicowaniu wartości średnich formantów samogłosek - głos PD w pracy [7].

Rys. 2. przedstawia przykłady widm binarnych wybranych logatomów.

Pierwszy eksperyment polegał na rozpoznawaniu segmentów, o stałej rozciągłości czasowej, wyodrębnionych tak, że ich ośrodki pokrywały się z granicami między głoskami. Na podstawie pracy [3] oraz licznych obserwacji przyjęto stałą szerokość segmentu równą 7 widmom binarnym (ok. 140 ms). Istotnym problemem, z przyczyn technicznych rozwiązaniem doraźnie, było wyznaczenie wzorców stanów przejściowych między głoskami. Kształty przebiegów formantów w tym obszarze są silnie determinowane kinetyką narządów artykulacyjnych. Ostatecznie zdecydowano się utworzyć listę diad, wielokrotnie wypowiadając każdą z nich przy jednoczesnej obserwacji uzyskiwanych widm i wyodrębnieniu, z realizacji uznanej za najbardziej typową, segmentu difonemicznego. W ten sposób otrzymano zbiór 20 wzorców difonemicznych (rys. 3).

Następny problem wymagający rozwiązania to znalezienie możliwie prostej i skutecznej miary podobieństwa dwóch segmentów. Istnieje wiele sposobów obliczenia miary podobieństwa dwóch obrazów (1). Będąc w dyspozycji środki techniczne (system MERA 303 wspomagany w trybie "off line" mikrokomputerem ZX81) oraz dążenie do minimalizacji czasów procesów przetwarzania są podstawowymi determinantami przedstawionego niżej sposobu obliczania miary podobieństwa.

Niech S i W będą segmentami binarnymi, natomiast \underline{S} i \underline{W} macierzami $n \times k$ (tu 16×7) o kolumnach tożsamościowo równych odpowiednim widmom binarnym.

Niech s_{ij} i w_{ij} będą elementami odpowiednich macierzy, stąd $s_{ij} = 0$ lub 1 gdzie $i = 1, 2, \dots, n$
 $w_{ij} = 0$ lub 1 $j = 1, 2, \dots, k$

Założono, że przynajmniej jedna z macierzy jest niezerowa. Niech $M(\underline{S}, \underline{W})$ oznacza miarę podobieństwa segmentów S i W .

$$M(\underline{S}, \underline{W}) = \frac{\sum_{i=1}^n \sum_{j=1}^k s_{ij} w_{ij}}{\sum_{i=1}^n \sum_{j=1}^k s_{ij} + w_{ij}} \quad (1)$$

Tak zdefiniowana miara ma następujące własności :

- 1) $\bigwedge_i \bigwedge_j s_{ij} = w_{ij} \Leftrightarrow M(\underline{S}, \underline{W}) = 0,5$
- 2) $\bigwedge_i \bigwedge_j s_{ij} w_{ij} = 0 \Leftrightarrow M(\underline{S}, \underline{W}) = 0$
- 3) $M(\underline{S}, \underline{W}) = M(\underline{W}, \underline{S})$ (symetria)

Prawo przechodniości nie jest zachowane. Iloczyn $s_{ij} w_{ij}$ można realizować jako iloczyn logiczny, co jest istotne dla oprogramowań w językach wewnętrznych.

Każdy segment będący przedmiotem rozpoznania, wyodrębniony ze spektrogramów binarnych logatomów przedstawiono w postaci macierzowej, oznaczonej \underline{S}^0 . Na podstawie \underline{S}^0 utworzono 4 macierze $\underline{S}^1, \underline{S}^2, \underline{S}^3, \underline{S}^4$ według poniższych zasad :

$$1 \quad s_{ij}^1 = s_{ij-1}^0 \quad \text{gdzie } j=2\dots7, i=1\dots16$$

$$\text{oraz} \quad s_{11}^1 = s_{11}^0$$

$$2 \quad s_{ij}^2 = s_{ij+1}^0 \quad \text{gdzie } j=1\dots6, i=1\dots16$$

$$\text{oraz} \quad s_{17}^2 = s_{17}^0$$

$$3 \quad s_{ij}^3 = s_{i+1j}^0 \quad \text{gdzie } i=1\dots15, j=1\dots7$$

$$\text{oraz} \quad s_{16j}^3 = s_{16j}^0$$

$$4 \quad s_{ij}^4 = s_{i-1j}^0 \quad \text{gdzie } i=2\dots16, j=1\dots7$$

$$\text{oraz} \quad s_{1j}^4 = s_{1j}^0$$

Macierze $\underline{S}^1, \underline{S}^2, \underline{S}^3, \underline{S}^4$ można utożsamiać z segmentami przesunię-

tymi o jedną pozycję, wzdłuż osi czasu i częstotliwości, w stosunku do pierwotnego.

Niech (W_m) będzie zbiorem segmentów wzorcowych, gdzie $m = 1 \dots 20$.

Niech $(M(W_m, S^1))$ będzie zbiorem wartości miar podobieństw segmentów W_m do S^1 ($m = 1 \dots 20, l = 0 \dots 4$).

Segment S^0 zostanie rozpoznany (utożsamiony) z wzorcem W_r wtedy, gdy

$$M(W_r, S^1) = \max(M(W_m, S^1)) \quad (2)$$

gdzie

$$r \in \langle 1, 20 \rangle$$

W drugim eksperymencie przeprowadzono próbę rozpoznawania segmentów o takiej samej rozciągłości czasowej, ale wyodrębnionych tak, ażeby maksymalnie obejmowały stan ustalony głoski.

Wzorcowe segmenty dla /i, a, u, s/ uzyskano analizując spektrogramy binarne tych głosek wypowiedzianych w sposób przeciągły (1+1,5s), bez zmiany cech prozodycznych. Opracowane wzorce pokazano na rys. 5. Dla głoski zwartej /t/ nie przewidziano wzorcowego segmentu, a detekcji podlega jedynie zwarcie o czasie trwania min. 70 ms (4 kolejne zerowe widma binarne). Zastosowano taką samą technikę rozpoznawania jak w poprzednim eksperymencie.

W trzecim doświadczeniu próbowano rozpoznać głoski bezpośrednio dekodując widma binarne. Kody detekcyjne opracowano na podstawie rozkładów jedynek w dwóch skrajnych widmach binarnych wzorców segmentów difonemicznych i uzupełniono eksperymentalnie (rys. 6). Szkic algorytmu rozpoznawania przedstawia rys. 7.

3. Opracowanie wyników.

3.1. Doświadczenie I.

Rys. 2a przedstawia spektrogram binarny logatomu /asta/ z zaznaczonym segmentem /as/. Przykładowy wydruk miar podobieństwa między tym difonemem a wzorcami pokazano na rys. 4. Tablice 1 i 2 zawierają wyniki rozpoznań logatomów i segmentów według kryterium maksymalnej wartości miary podobieństwa. Bez błędnie rozpoznano 15 logatomów (44%), z jednym błędem 11 (32%), z dwoma 7 (22%). W logatomie /tsui/ nie rozpoznano ani jednego segmentu. Analiza wyników rozpoznawania diał (74 po-

prawne rozpoznania, co stanowi 73 %) pozwala również ocenić wzorce segmentów difonemicznych. Niektóre z nich (np. /iu/, /us/) powinny zostać zmodyfikowane. Program obliczania miar podobieństwa, początkowo realizowany w systemie MERA 303, został przeniesiony w języku BASIC na mikrokomputer ZX81. Do rozpoznania każdego difonemu należało obliczyć $5 \times 20 = 100$ miar podobieństw, co trwało ok. 1 godz. Dodatkowe utrudnienie stanowiło wprowadzenie danych (z klawiatury) do mikrokomputera. Przy zastosowaniu języka wewnętrznego, rezygnacji z heksadecymalnego kodowania wprowadzanych widm binarnych i wykorzystania bezpośredniego dostępu do pamięci, czas przetwarzania ulegnie redukcji do ok. 2s.

Stworzone zostaną wówczas warunki do weryfikacji wzorców segmentów.

Wyniki przedstawione powyżej należy uznać za niezadawalające. W czasie rozpoznawania rejestrowano, oprócz wzorca maksymalizującego miarę podobieństwa, dwa wzorce, dla których miara była najbardziej zbliżona do maksymalnej (Tablica 4). Oczywisty wymóg zgodności końcowej części segmentu poprzedzającego z początkową częścią następnego oraz przynależności rozpoznanego logatomu do zbioru rozpoznawanych haseł stwarza możliwości dokonania weryfikacji syntaktyczno-leksykalnej (Tablica 4). Z grup wzorców najbardziej podobnych (kol. 2) utworzono logatomy, podkreślając te, które należą do zbioru haseł. Jeśli wynikiem weryfikacji jest alternatywne rozpoznanie, należy porównać miary podobieństwa diad - np. /iuit/, /iuis/, /itut/. W wyniku weryfikacji, z 19 wstępnie nierozpoznanych logatomów 9 rozpoznano poprawnie, 2 niejednoznacznie, 3 błędnie a 5 wcale nie rozpoznano. Poprawność rozpoznania logatomów wzrosła do 70 %, a diad do 82 % (o 9 więcej). Dalsza weryfikacja, np. uwzględniająca kontekst lub semantykę, wykracza poza założone ramy tej pracy.

3.2. Doświadczenie II.

Wyniki rozpoznawania segmentów wyodrębnionych z głosek podane są w tablicach 1,2,3, Rozpoznano bezbłędnie 26 haseł (76 %), 90 segmentów difonemicznych (88 %), 126 głosek (93 %). Wyniki takie należy uznać za dobre, tym niemniej możliwa jest

weryfikacja leksykalna (błędne rozpoznania nie należą do zbioru haseł), polegająca na wprowadzeniu miary podobieństw haseł. Czas przetwarzania danych jest 5-krotnie krótszy niż poprzednio.

3.3. Doświadczenie III.

Przedstawione w tablicach wyniki uzyskane w ostatnim eksperymencie są szczególnie interesujące, ponieważ bezpośredniego dekodowania widm binarnych dokonano w czasie rzeczywistym. Poprawność rozpoznawania logatomów wynosi tylko 44 %, pomimo to identyfikacja błędnych wyników nie przedstawia żadnych trudności w wielu przypadkach, np. /iasu/-/asu/, /uau/-/uauu/, /asat/-/asats/ (Tablica 1). Potwierdza to wysoki procent rozpoznanych głosek (89 %). Można przypuszczać, że rozpoznawanie odpowiednio zbudowanego zbioru haseł nie będzie wymagało weryfikacji na poziomie leksykalnym. Doświadczenie przeprowadzono na minikomputerze MERA 303.

4. Wnioski.

4.1. Uzyskane w doświadczeniach wyniki odpowiadają (w ograniczonym zakresie) pozytywnie na postawione we wstępie pytanie.

4.2. Metody rozpoznawania segmentów należy uzupełnić o automatyczną segmentację sygnału mowy.

4.3. Celowe jest określenie miar podobieństw między spektrogramami binarnymi (nie tylko 16-bitowymi) wszystkich fonemów.

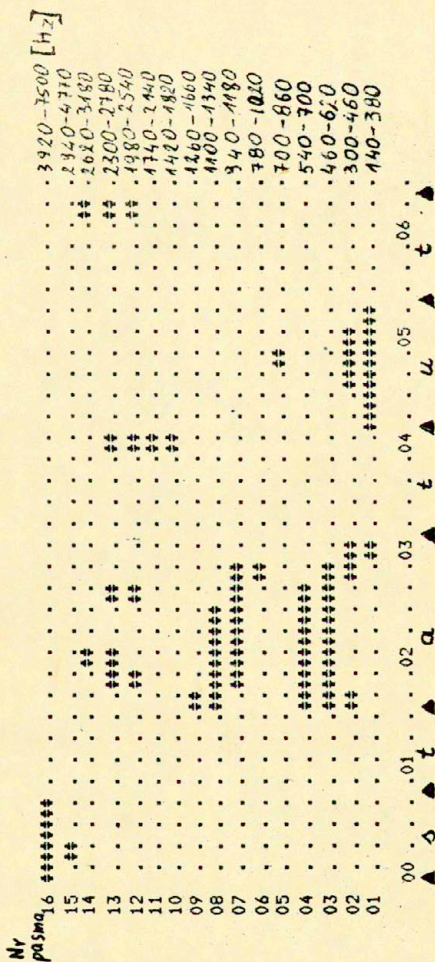
4.4. Zbiory haseł do rozpoznawania powinny być tworzone w oparciu o dobrze rozróżnialne fonemy. Dotyczy to szczególnie tych zastosowań, gdzie istotny jest czas przetwarzania. Hasła nie powinny zaczynać się od głosek zwartych (por. [14]).

4.5. Celowy jest dalszy rozwój prac nad binarną prezentacją sygnału mowy, zmierzający do optymalizacji metod kodowania fonetyczno-akustycznego (por. [13]).

4.6. Dla zbiorów haseł zbudowanych na kilku dobrze rozróżnialnych fonemach korzystne jest stosowanie wzorców fonematycznych. Zbadania wymaga połączenie metod przedstawionych w doświadczeniach I i II ze szczególnym uwzględnieniem możliwości wzajemnej weryfikacji.

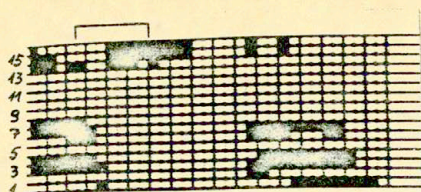
4.7. Zbiór wzorców zwłaszcza difonemicznych należy zoptymalizować.

4.8. Skuteczne rozpoznawanie segmentów jest istotnym elementem procesu rozpoznawania mowy na podstawie widm binarnych.



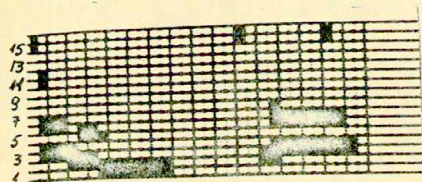
Rys. 1. Spektrogram binarny wyrazu /statut/.

a/



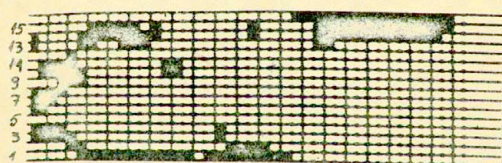
a s t a

b/



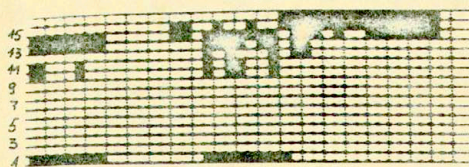
a u t a

c/



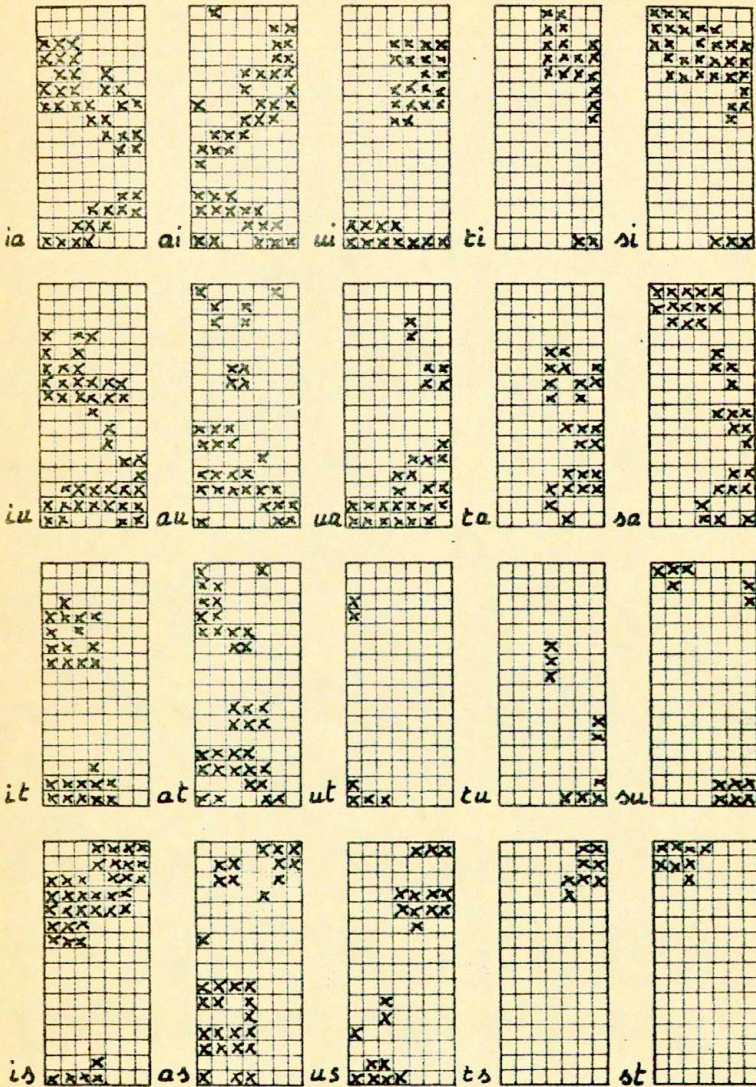
a i u s

d/



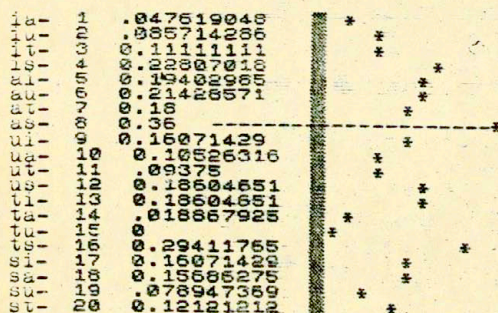
i t i s

Rys. 2. Przykłady spektrogramów binarnych logatomów.

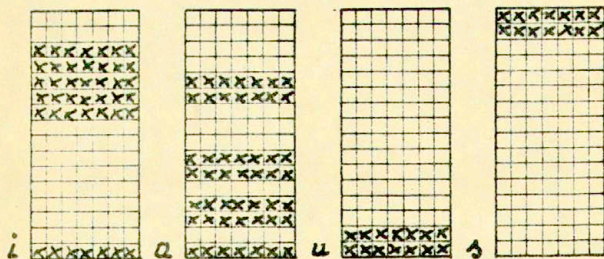


x - oznacza 1

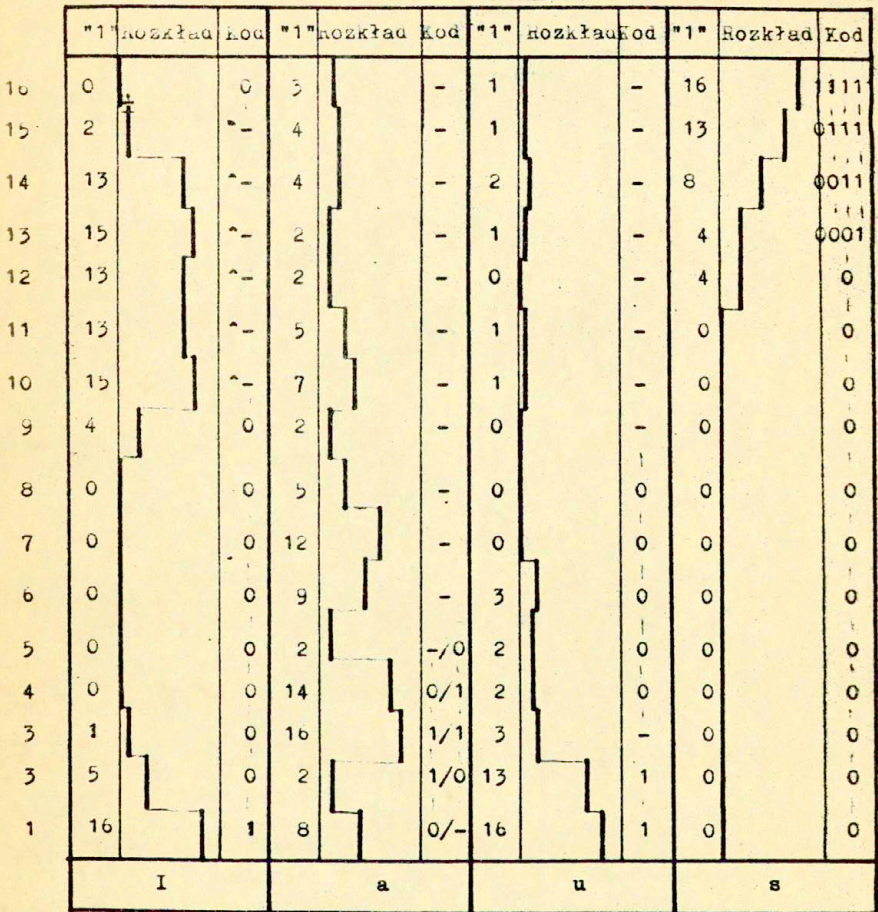
Rys. 3. Wzorce segmentów difonemicznych.



Rys. 4. Wyniki rozpoznania segmentu /as/ logotomu /a s t a/.

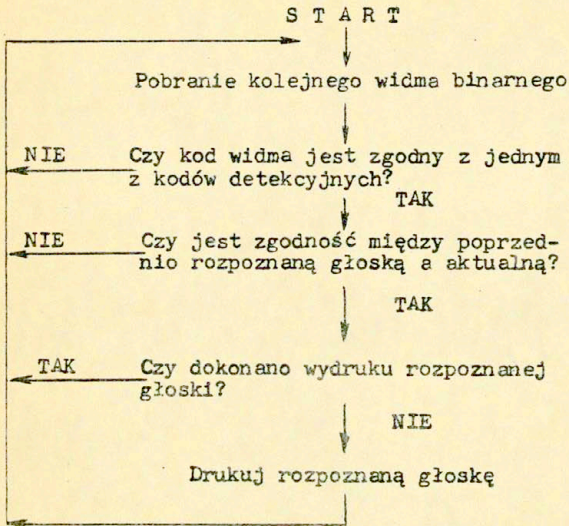


Rys. 5. Wzorce segmentów fonematycznych.



^ W pasmach 10-15 występuje co najmniej jedna 1

Rys. 6. Częstości występowania 1 w dwóch skrajnych widmach binarnych wzorców segmentów difonematycznych oraz kody detekcyjne.



Rys. 7. Algorytm rozpoznawania fonemów metodą bezpośredniego dekodowania widm binarnych.

Nr	Logatom	Wyniki rozp. w dośw.		
		I	II	III
1	asta	+	+	+
2	auta	+	+	+
3	iasu	+	+	asu
4	susa	su-us-su	+	+
5	stia	+	+	stiaa
6	sias	+	+	siuas
7	tiaa	si-ia-at	siat	suat
8	uiat	+	uuat	uat
9	iuit	it-ui-ut	iuut	ut
10	tius	ti-it-ut	iius	+
11	aius	ai-ut-ut	+	+
12	uauu	ua-aüt	+	uauü
13	iuis	it-ui-is	+	iuist
14	itü	ut-si-is	+	+
15	itua	+	+	+
16	itsi	+	+	+
17	itut	is-is-ut	+	+
18	isau	+	+	+
19	ists	+	+	+
20	auai	+	+	auai
21	stai	st-ta-au	+	staut
22	atau	at-ts-au	+	+
23	asai	+	+	sai
24	auti	+	+	+
25	esat	at-sa-at	+	asats
26	atua	+	+	stua
27	tsui	si-si-is	siü	ui
28	stua	st-su-ua	+	stu
29	susi	su-ts-si	suss	+
30	usut	iu-tu-ut	u(i,s) ut	ut
31	tuts	su-us-ts	-uts	uts
32	suis	si-ui-is	+	+
33	sisa	+	+	susa
34	siut	si-it-ut	+	siuts

+ oznacza bezbłądne rozpoznanie

Tablica 1. Logatomy i ich rozpoznanie.

Diada	Wyniki rozp. w dośw.			Liczba wystąpień
	I	II	III	
ia	5	5	0	5
iu	0	5	5	6
it	2	4	4	5
is	6	6	6	6
ai	4	5	1	5
au	5	5	5	5
at	5	5	4	5
as	4	5	4	5
ui	4	1	3	5
ua	5	5	4	5
ut	5	6	6	6
us	1	5	4	5
ti	3	3	3	5
ta	3	4	4	4
tu	2	4	4	5
ts	3	3	3	4
si	5	4	4	5
sa	4	5	5	5
su	3	5	4	6
st	5	5	5	5
razem	74	90	78	102

Tablica 2. Diady i ich rozpoznanie.

Głoska	Liczba wystąpień	Wyniki rozp. w dośw.	
		I	II
i	29	26	22
a	28	28	26
u	25	24	24
s	29	27	27
t	25	21	22
razem	136	126	121

Tablica 3. Głoski i ich rozpoznanie.

Tablica 4. Weryfikacja syntaktyczno-leksykalna wyników doświadczenia I.

Logatom	Wzorce najbardziej podobne	Weryfikacja	Rozpoznanie
1	2	3	4
susa	su us su tu ut sa st is st	susu, <u>susa</u> , tusu, tusa	susa
tiat	si ia at su iu au ti it as	siat, siau, <u>sias</u> , <u>tiat</u> , tiau, tias	sias tiat
iuit	it ui ut iu ua it ia ut is	<u>iuit</u> , <u>iuis</u> ,	iuit
tius	ti it ut si iu us ui ia ts	tits, sits, uits, tiut, <u>siut</u> , uiut <u>tius</u> , sius, uius	tius siut
aius	ai ut ut au it us ui ia is	-	-
uaiu	ua ui ut ui iu su ta ui st	uiut	-
iuis	it ui is iu ua it is su ia	<u>iuis</u> , <u>iuit</u> , iuia	iuis
itis	ut si is it ui ts is ti si	isis, utis, <u>itis</u>	itis
itut	is is ut it tu it ia ua ua-us	<u>itut</u> , <u>itua</u> , itus	itut
stai	st ta au su sa ui sa ua ua	stau	-
atau	at ts au as tu ut au su ui	atut, asui, atui, asut	-
asat	at sa at as ta au	atat, <u>atau</u> , atas, <u>asat</u> , asau, asas	asat

1	2	3	4
tsui	si si is us ti it is sa at	usis,usit,isis,isit, isat	-
susi	su ts si st is su ts us ti	<u>susi</u> ,susu,stsi, stsu	susi
usut	iu tu ut is sa us it si as	<u>itut</u> ,itus,isas	itut-błąqd
tuts	su us ts ua as ti ui ut sa	<u>susa</u> ,suts,usage	susa-błąqd
suis	si ui is su us ut is su as	isut, <u>suis</u> ,	suis
siut	si it ut ti ie us sa is as	<u>sias</u> ,tias	sias-błąqd
stua	st su ua su ua ta sa tu ui	<u>stua</u> ,stui	stua

Do segmentu /su/ logatomu /susa/ najbardziej podobnymi wzorcami są kolejno : /su/, /tu/, /st/.

BIBLIOGRAFIA

- (1) BLIJNEN, E.J., Cluster analysis, Tilburg University Press, The Netherlands, 1973.
- (2) DOMAGAŁA, P., Automatyzacja procesu segmentacji sygnału mowy w układzie analogowo-cyfrowym, Prace IPPT 5/84, Warszawa, 1984.
- (3) FRACKOWIAK-RICHTER, L., The duration of Polish vowels, Speech Analysis and Synthesis 3, W.Jassem, ed., 87-115, Warsaw, 1973.
- (4) JASSEM, W., Podstawy fonetyki akustycznej, PWN, W-wa, 1973.
- (5) JASSEM, W., Fonetyczno-akustyczne założenia rozpoznawania fonemów, Prace IPPT 14/70, Warszawa, 1970.
- (6) JASSEM, W., KUBZDELA, H., DOMAGAŁA, P., Segmentacja sygnału mowy na podstawie zmian rozkładu energii w widmie, Prace IPPT 13/83, Warszawa, 1983.
- (7) JASSEM, W., KRZYSKO, M., STOLARSKI, P., Częstotliwości formantowe samogłosek jako cechy fonematyczne i osobnicze w świetle statystycznej analizy dyskryminacyjnej, Prace IPPT 27/84, Warszawa, 1984.
- (8) KUBZDELA, H., Metoda automatycznego rozpoznawania wyrazów w oparciu o spektrogramy binarne, Prace IPPT 14/80, Warszawa, 1980.
- (9) KUBZDELA, H., Weryfikacja i optymalizacja metody rozpoznawania wyrazów w skończonych zbiorach hasłowych w oparciu o spektrogramy binarne, Prace IPPT 10/82, W-wa, 1982.
- (10) KUBZDELA, H., Badania nad udoskonaleniem spektrogramów binarnych, Prace IPPT 24/83, Warszawa, 1983.
- (11) KUBZDELA, H., Próby automatycznego rozpoznawania wyrazów wymawianych przez różne głosy w oparciu o grupowe zbiory wzorcowych spektrogramów binarnych, Prace IPPT 47/83, Warszawa, 1983.
- (12) RICHTER, L., Wizualne rozpoznawanie samogłosek polskich w prostych kontekstach spółgłoskowych na podstawie spektrogramów komputerowych, Prace IPPT (w druku).
- (13) DREYFUS-GRAF, J.A., Codes phonotigues (Phonocodes) of telecommunication, Proc. FASE 75, Paris, 1975.
- (14) KÖSTER, J.P., Phonocodes und die Perzeption Konstruierter Sprachen, Hamburger Phonetische Beiträge, 17/1976, 35-82 (phonocode STOINAKEMUS).