System identyfikacji mówcy metodą niezależnej detekcji jednostek fonetycznych

Tomasz PAŁYS

Zakład Automatyki, Instytut Teleinformatyki i Automatyki,
Wojskowa Akademia Techniczna, ul. Kaliskiego 2, 00-908 Warszawa

STRESZCZENIE: Przedstawiono system identyfikacji mówców metodą niezależnej detekcji jednostek fonetycznych. Etap uczenia polega na wykorzystywaniu technik grupowania w celu wyznaczenia jednostek fonetycznych, charakteryzujących mówcę w przestrzeni cech. Wyznaczone jednostki służą do oceny zgodności z mówcą. Szczególną uwagę zwrócono na metody wyodrębnienia jednostek fonetycznych najlepiej charakteryzujących mówcę.

1. Wprowadzenie

Systemy identyfikacji mówców znajdują zastosowanie w przypadku ochrony dostępu do miejsc (budynki, strefy chronione, systemy dowodzenia itp.) lub usług zastrzeżonych (bankowych, administracyjnych itp.). Na podstawie głosu system dokonuje rozpoznania tożsamości mówcy.

Przedstawiony w artykule system, opracowany przez autora, dokonuje identyfikacji mówcy metodą detekcji jednostek fonetycznych, niezależnie od kontekstu wypowiedzi. Metoda polega na ocenie zgodności badanej próbki z wzorcem, określonym jako zbiór jednostek fonetycznych. Istota metody tkwi w sposobie porównywania wyników oceny zgodności z poszczególnymi jednostkami fonetycznymi. Oceny te dokonywane są niezależnie dla poszczególnych jednostek – według różnych metryk. Stosowane metryki uwzględniają rozproszenie i korelację elementów poszczególnych jednostek fonetycznych. Ujednolicenie dokonywanych ocen jest możliwe poprzez wykorzystanie przekształcenia Karhunena–Loève'a [2].

Metoda identyfikacji polega na realizacji dwóch etapów:

- etap pierwszy odnosi się do analizy dostępnych danych w celu określenia jednostek fonetycznych i reguł klasyfikacji – jest to proces uczenia,
- etap drugi polega na podejmowaniu decyzji zgodnie z nauczoną regułą – jest to proces identyfikacji mówcy.

2. Opis metody

2.1. Ekstrakcja cech

Podstawą ekstrakcji cech jest przyjęcie założenia, że właściwości sygnału mowy nie zmieniają się w krótkim okresie czasu. W związku z tym sygnał mowy dzieli się na mniejsze części tzw. ramki czasowe. W praktyce stosuje się ramki o szerokości od 10 ms do 30 ms.

W wyniku wyznaczania współczynników LPC kolejnych ramek czasowych powstaje zbiór obserwacji:

$$\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_r, ..., \mathbf{x}_N\}, \text{ gdzie: } \mathbf{x}_r = \begin{bmatrix} x_r(1) \\ x_r(2) \\ ... \\ x_r(p) \end{bmatrix}, \tag{1}$$

gdzie:

r – numer ramki,
N – liczba ramek czasowych,
$\mathbf{x}_r$ – wektor współczynników LPC,
p – rząd predykcji.

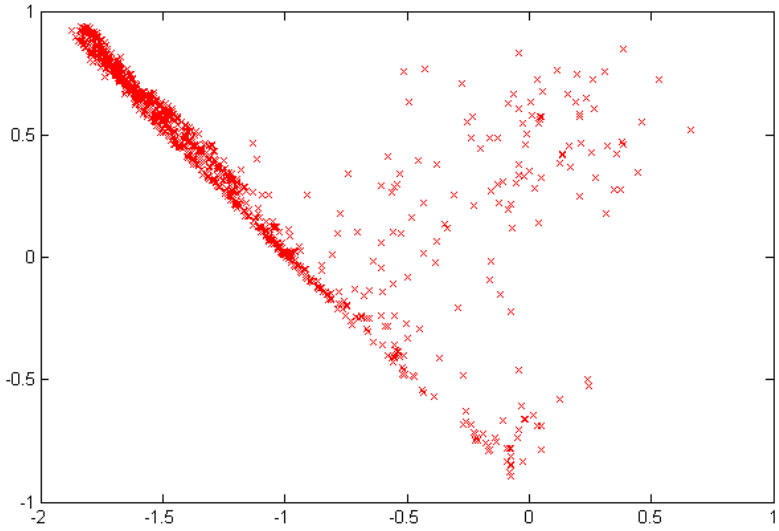
Proces ekstrakcji cech polega na odwzorowaniu sygnału mowy w skończony ciąg (sekwencję) punktów przestrzeni cech. Jako przestrzeń cech przyjęto przestrzeń wartości współczynników LPC (jest to przestrzeń R^p, p – rząd predykcji).

2.2. Wyznaczenie jednostek fonetycznych

Każdy mówca charakteryzowany jest przez zbiór właściwych mu jednostek fonetycznych, rozumianych jako obszary w przestrzeni cech – przestrzeni

wartości współczynników LPC. Obszary te wyznacza się na podstawie analizy skupień.

Przykładowy zbiór uczący przedstawiono na rys. 1. W przedstawianym systemie identyfikacji wydzielenie jednostek fonetycznych realizowane jest metodą grupowania hierarchicznego. Proces grupowania odbywa się przez kolejne łączenie położonych najbliżej siebie grup (w pierwszym kroku punktów), rozumianych jako jednostki fonetyczne. Taki sposób grupowania umożliwia tworzenie drzewa grupowania.



Rys. 1. Punkty zbioru uczącego (dwa współczynniki LPC)

Uzyskanie satysfakcjonujących wyników grupowania zależy od właściwego doboru metryki (odległości). Omawiany system umożliwia zastosowanie wszystkich najczęściej stosowanych metryk w przestrzeni R^p. Są to następujące metryki:

- euklidesowa:

$$d^2(\mathbf{x}_r, \mathbf{x}_s) = (\mathbf{x}_r - \mathbf{x}_s)'(\mathbf{x}_r - \mathbf{x}_s), \tag{2}$$

- standaryzowana euklidesowa:

$$d^2(\mathbf{x}_r, \mathbf{x}_s) = (\mathbf{x}_r - \mathbf{x}_s)'\mathbf{S}^{-1}(\mathbf{x}_r - \mathbf{x}_s), \tag{3}$$

gdzie: **S** – diagonalna macierz wariancji,

- Mahalanobisa:

$$d^2(\mathbf{x}_r, \mathbf{x}_s) = (\mathbf{x}_r - \mathbf{x}_s)' \mathbf{R}^{-1} (\mathbf{x}_r - \mathbf{x}_s), \tag{4}$$

gdzie: $\mathbf{R}$ – macierz kowariancji,

- city block:

$$d(\mathbf{x}_r, \mathbf{x}_s) = \sum_{i=1}^{p} |\mathbf{x}_r(i) - \mathbf{x}_s(i)|, \tag{5}$$

- Minkowskiego

$$d(\mathbf{x}_r, \mathbf{x}_s) = \left\{ \sum_{i=1}^{p} |\mathbf{x}_r(i) - \mathbf{x}_s(i)|^m \right\}^{\frac{1}{m}}, \quad m \geq 1. \tag{6}$$

Podstawę hierarchicznej metody grupowania stanowi określenie odległości między poszczególnymi grupami. W celu przedstawienia tych wielkości wprowadzimy następujące oznaczenia. Niech $\mathbf{G}_k$ oznacza grupę złożoną z N_k punktów ze zbioru $\mathbf{X}$. Przyjmuje się, że razem jest L grup, grupy są rozłączne i wyczerpują zadany zbiór $\mathbf{X}$. Elementy grupy $\mathbf{G}_k$ oznacza się jako punkty $\mathbf{x}_{rk}$, gdzie $r = 1, 2, ..., N_k$, $\sum_{k=1}^{L} N_k = N$, k – numer grupy, r – numer elementu w grupie.

Najczęściej stosuje się jeden z następujących sposobów określania odległości między grupami:

- metoda minimalnego sąsiedztwa:

$$dist(\mathbf{G}_r, \mathbf{G}_s) = \min \left\{ d(\mathbf{x}_{ri}, \mathbf{x}_{sj}),\ i \in \{1, ..., N_r\},\ j \in \{1, ..., N_s\} \right\}, \tag{7}$$

- metoda maksymalnego sąsiedztwa:

$$dist(\mathbf{G}_r, \mathbf{G}_s) = \max \left\{ d(\mathbf{x}_{ri}, \mathbf{x}_{sj}), i \in \{1, ..., N_r\}, j \in \{1, ..., N_s\} \right\}, \tag{8}$$

- metoda średniej odległości:

$$dist(\mathbf{G}_r, \mathbf{G}_s) = \frac{1}{N_r + N_s} \sum_{i=1}^{N_r} \sum_{j=1}^{N_s} d(\mathbf{x}_{ri}, \mathbf{x}_{sj}), \tag{9}$$

- metoda centroidalna:

$$dist(\mathbf{G}_r, \mathbf{G}_s) = d(\overline{\mathbf{x}}_r, \overline{\mathbf{x}}_s), \tag{10}$$

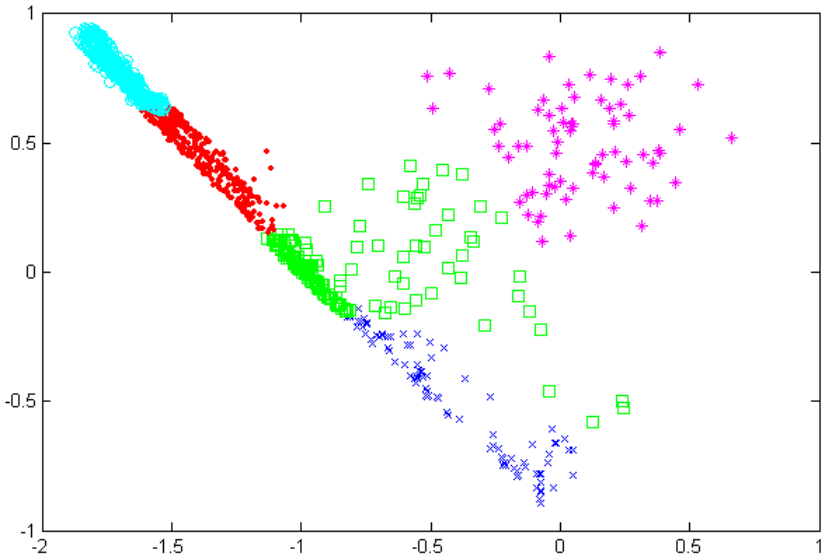
- metoda Warda:

$$dist(\mathbf{G}_r, \mathbf{G}_s) = \frac{N_r N_s}{N_r + N_s} d^2(\overline{\mathbf{x}}_r, \overline{\mathbf{x}}_s), \tag{11}$$

gdzie:

$$\overline{\mathbf{x}}_r = \frac{1}{N_r}\sum_{i=1}^{N_r}\mathbf{x}_{ri}\ ,\ \overline{\mathbf{x}}_s = \frac{1}{N_s}\sum_{j=1}^{N_s}\mathbf{x}_{sj}\ . \tag{12}$$

Każdy krok grupowania metodą hierarchiczną polega na wyszukaniu dwóch grup najbliżej siebie położonych i połączeniu ich w jedną grupę. Na początku przyjmuje się, że każda grupa składa się z jednego punktu, czyli że na początku jest N grup. Grupowanie kończy się po uzyskaniu jednej grupy, złożonej ze wszystkich punktów. Wynik końcowy stanowi drzewo grupowania, na podstawie którego można uzyskać żądaną liczbę grup, albo grupy o zadanych właściwościach [5].



Rys. 2. Wynik podziału zbioru uczącego na pięć grup (metoda Warda z metryką Mahalanobisa)

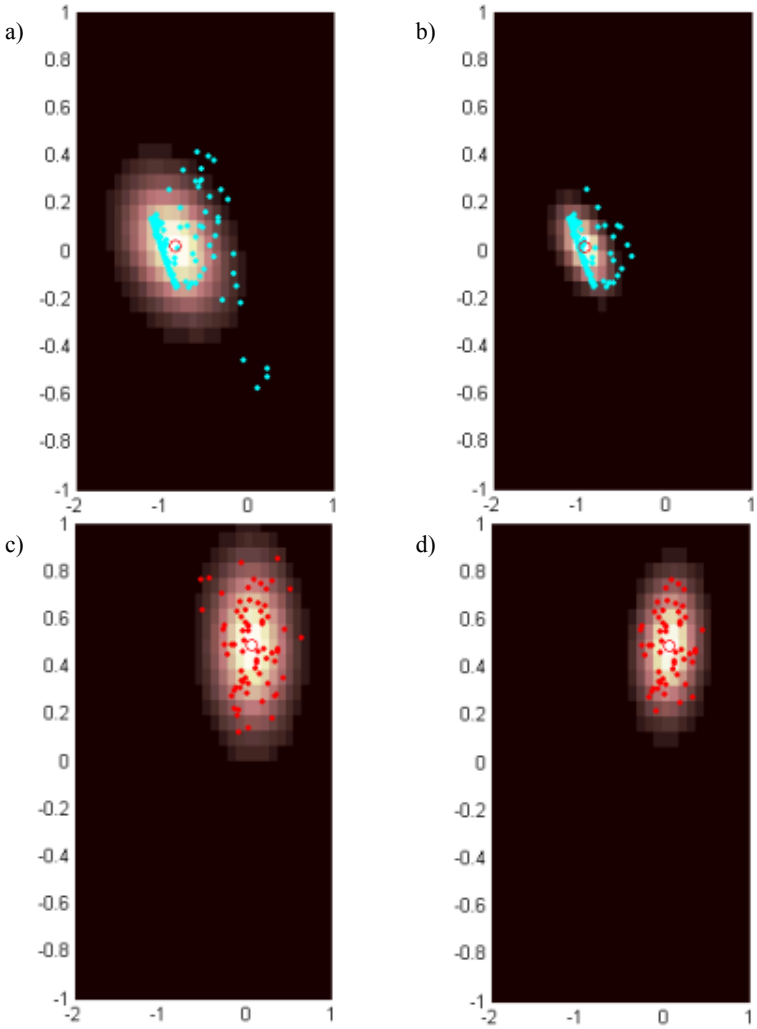
Ocenę jakości grupowania można przeprowadzić na podstawie współczynnika niezgodności grupowania Y_{rs}, który wyznaczany jest według następującego wzoru:

$$Y_{rs} = \frac{dist(\mathbf{G}_r, \mathbf{G}_s) - E(\mathbf{Z})}{\sqrt{V(\mathbf{Z})}}, \tag{13}$$

gdzie:

- $\mathbf{Z} = \mathbf{G}_r \cup \mathbf{G}_s$,
- $E(\mathbf{Z})$ – średnia odległość łączenia grup, od pierwszego do aktualnie rozpatrywanego poziomu grupowania, w wyniku czego otrzymano grupę $\mathbf{Z}$,

- $\sqrt{V(\mathbf{Z})}$ – odchylenie standardowe łączenia grup, od pierwszego do aktualnie rozpatrywanego poziomu grupowania, w wyniku czego otrzymano grupę **Z**.



Rys. 3. a) Grupa nr 1 na tle elipsy rozkładu normalnego;
b) Grupa nr 1 po odrzuceniu punktów niedopasowanych dla $c^2 = 3$;
c) Grupa nr 2 na tle elipsy rozkładu normalnego;
d) Grupa nr 2 po odrzuceniu punktów niedopasowanych dla $c^2 = 3$

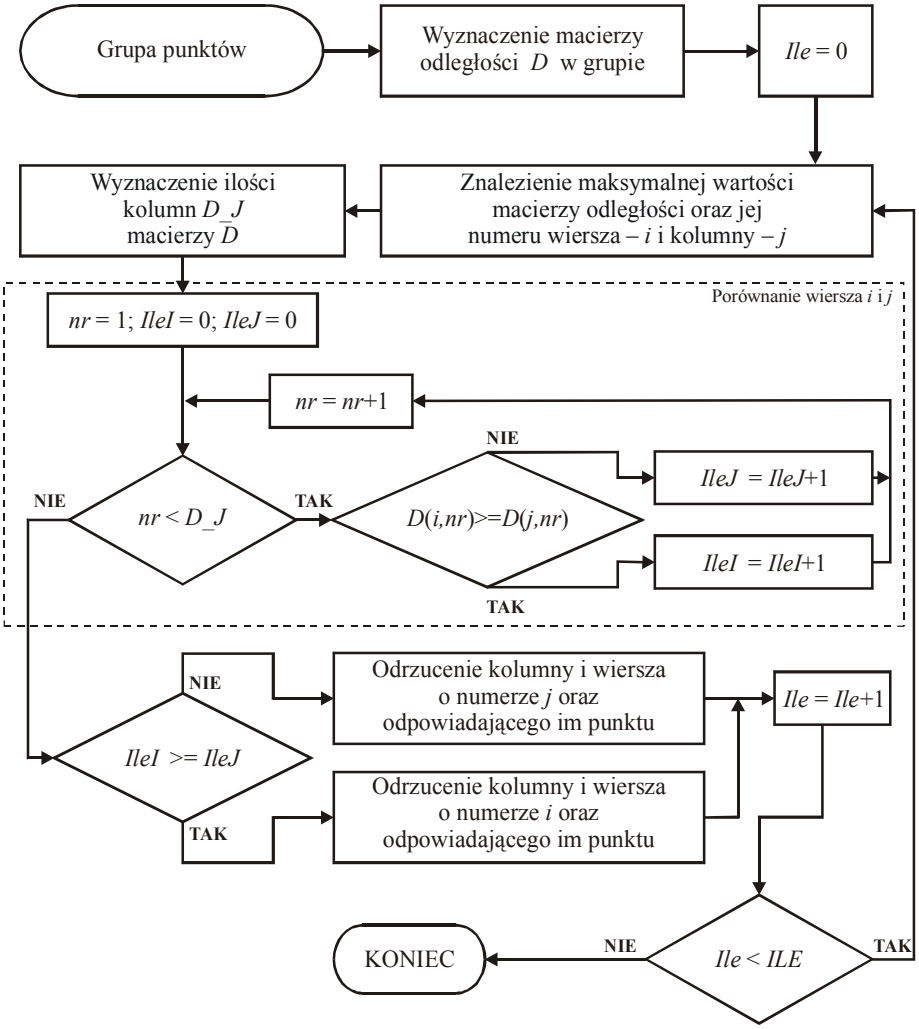
Dokonanie oceny jakości grupowania polega na obliczeniu współczynnika Y_{rs} od pierwszego poziomu drzewa grupowania do poziomu, który odpowiada podziałowi na żądaną liczbę grup. W przypadku określenia jego maksymalnej wartości, podział na grupy wyznacza ten poziom drzewa, który odpowiada maksymalnej wartości współczynnika niezgodności grupowania.

Przykładowy podział zbioru uczącego na pięć grup przedstawiono na rys. 2. Grupę oznaczoną kwadratami trudno jest uznać za jednolite skupienie (co można ocenić przy pomocy współczynnika niezgodności grupowania). W tym przypadku korzystne jest potraktowanie części punktów skupienia jako niedopasowanych i odrzucenie ich. Poniżej przedstawiono dwie metody usuwania punktów niedopasowanych, proponowane przez autora.

1) Metodę tę można zastosować w przypadku, gdy rozkład punktów jednostki fonetycznej jest normalny. W tym przypadku należy odrzucić tę część, która nie spełnia równania:
$$(\mathbf{x}_{rk} - \overline{\mathbf{x}}_k)'\mathbf{V}^{-1}(\mathbf{x}_{rk} - \overline{\mathbf{x}}_k) < c^2, \tag{14}$$
gdzie:
$\mathbf{x}_{rk}$ – element r grupy k,
$\overline{\mathbf{x}}_k$ – wektor wartości średnich w grupie k,
$\mathbf{V}_k$ – macierz kowariancji w grupie k,
c – współczynnik.

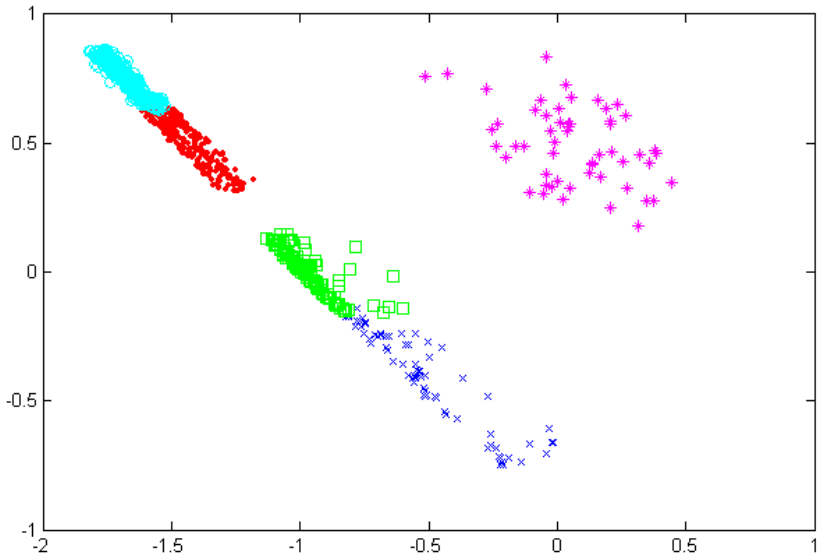
   Przykładowe dwie grupy po odrzuceniu punktów niedopasowanych przedstawiono na rys. 3.
2) Drugi sposób polega na określeniu punktów niedopasowanych jako położonych najdalej od pozostałych punktów grupy. Algorytm przedstawiono na rys. 4.
   Pierwszy krok algorytmu polega na wyznaczeniu macierzy odległości **D** między punktami grupy. Macierz **D** jest kwadratowa i symetryczna względem głównej przekątnej (główna przekątna zawiera tylko elementy zerowe), a element w kolumnie *j* i wierszu *i* opisuje odległość pomiędzy punktem *j* i *m*.
   Drugi krok polega na wyszukaniu maksymalnej wartości elementu macierzy (największą odległość pomiędzy dwoma punktami grupy) i wyznaczeniu jego numeru kolumny *j* i wiersza *i*. Następnie należy porównać elementy kolumny, o tym samym numerze, z wiersza *j* i wiersza *i*. Wyznaczyć wartość *IleJ*, wartość ta określa liczbę kolumn, dla których wartości w wierszu *j*, są większe niż w wierszu *i*. Analogicznie wyznacza się wartość *IleI*. Jeżeli *IleJ* >= *IleI*, to z macierzy **D** należy usunąć kolumnę oraz wiersz o indeksie *j*. W przypadku przeciwnym należy usunąć kolumnę oraz wiersz o indeksie *i*.

Usunięcie wiersza i kolumny o tym samym indeksie jest równoznaczne z odrzuceniem punktu niedopasowanego.



Rys. 4. Algorytm odrzucenia punktów niedopasowanych

Drugi krok algorytmu powtarza się do momentu aż zostanie odrzucona żądana liczba elementów z grupy. Efekt odrzucenia 30% punktów grupy, jako punktów niedopasowanych, przedstawiony został na rys. 5.



Rys. 5. Wynik podziału przestrzeni cech na pięć jednostek fonetycznych (metoda Warda z metryką Mahalanobisa) po odrzuceniu punktów niedopasowanych

2.3. Model mówcy

Model mówcy składa się z L jednostek fonetycznych. Opisywany system umożliwia wyznaczenie następujących parametrów modelu mówcy:

- wartości oczekiwanej jednostki fonetycznej **G**:

$$\overline{\mathbf{x}} = \frac{1}{N}\sum_{\mathbf{x}\in\mathbf{G}} \mathbf{x}\,, \tag{15}$$

  gdzie N – liczba punktów jednostki fonetycznej **G**,
- macierzy kowariancji jednostki fonetycznej **G**:

$$\mathbf{R} = \sum_{\mathbf{x}\in\mathbf{G}} (\mathbf{x}-\overline{\mathbf{x}})(\mathbf{x}-\overline{\mathbf{x}})'\,. \tag{16}$$

- wartości własnej λ_n oraz odpowiadającym im wektorów własnych $\mathbf{t}_n$ macierzy kowariancji **R**, tzn. wektorów spełniających następujące warunki:

$$\mathbf{R}\,\mathbf{t}_n = \lambda_n \mathbf{t}_n, \quad n = 1, 2, ..., p\,. \tag{17}$$

  Wektory własne $\mathbf{t}_n$ są porządkowane według malejących wartości własnych, to znaczy:

$$\lambda_1 > \lambda_2 > \ldots > \lambda_p . \tag{18}$$

- macierzy przekształcenia Karhunena–Loève'a:

$$\mathbf{T} = \begin{bmatrix} \mathbf{t}_1' \\ \mathbf{t}_2' \\ \ldots \\ \mathbf{t}_p' \end{bmatrix}. \tag{19}$$

Parametryczny model jednostki fonetycznej definiuje się następująco:

$$(\overline{\mathbf{x}}, \mathbf{T}, \boldsymbol{\lambda}), \tag{20}$$

gdzie:

- $\overline{\mathbf{x}}$ – wektor wartości oczekiwanej jednostki fonetycznej **G**,
- **T** – macierz przekształcenia Karhunena–Loève'a jednostki fonetycznej **G**,
- $\boldsymbol{\lambda}$ – wektor wartości własnych macierzy kowariancji jednostki fonetycznej **G**.

2.4. Identyfikacja mówcy

W przypadku, gdy w systemie zarejestrowanych jest M mówców, parametryczne modele mówców zapisuje się następująco:

$$(\overline{\mathbf{x}}_m^k, \mathbf{T}_m^k, \boldsymbol{\lambda}_m^k), \quad \text{gdzie } k = 1, 2, ..., L_m; \quad m = 1, 2, ..., M \ . \tag{21}$$

Ocena zgodności rozpoznawanej wypowiedzi z mówcą m dokonywana jest na podstawie przyporządkowania punktów próby do poszczególnych jednostek fonetycznych mówcy. Ocena zgodności $d_m^k(\mathbf{x}_r)$, punktu $\mathbf{x}_r$ próby z jednostką fonetyczną k mówcy m, polega na obliczeniu następującej transformaty Karhunena–Loève'a:

$$\mathbf{y}_m^k = \mathbf{T}_m^k(\mathbf{x}_r - \overline{\mathbf{x}}_m^k), \tag{22}$$

i wyznaczeniu wartości funkcjonału:

$$d_m^k(\mathbf{x}_r) = \sum_{i=1}^{p} \frac{\mathbf{y}_m^k(i)^2}{\boldsymbol{\lambda}_m^k(i)} . \tag{23}$$

Funkcjonał (23) określa kwadrat odległości Mahalanobisa punktu $\mathbf{x}_r$ i $\overline{\mathbf{x}}_m^k$. Zastosowanie wzoru (23) wymaga spełnienia założenia, że $\boldsymbol{\lambda}_m^k(i) \neq 0$.

Rozważmy sytuację, gdy tylko z ($z \le p$) pierwszych współrzędnych wektora wartości własnych $\boldsymbol{\lambda}_k$ jest różnych od zera. W tym przypadku będziemy stosować przekształcenie określone macierzą:

$$\tilde{\mathbf{T}}_k = \begin{bmatrix} \dfrac{\mathbf{t}'_1}{\sqrt{\boldsymbol{\lambda}_k(1)}} \\ \dfrac{\mathbf{t}'_2}{\sqrt{\boldsymbol{\lambda}_k(2)}} \\ \cdots \\ \dfrac{\mathbf{t}'_r}{\sqrt{\boldsymbol{\lambda}_k(z)}} \end{bmatrix}, \tag{24}$$

Dzięki włączeniu informacji o wektorze wartości własnych do macierzy $\tilde{\mathbf{T}}_k$, modele mówców możemy uprościć:

$$(\bar{\mathbf{x}}_m^k, \tilde{\mathbf{T}}_m^k), \text{ gdzie } k = 1, 2, ..., L_m; m = 1, 2, ..., M \ . \tag{25}$$

Ocena zgodności punktu $\mathbf{x}_r$ z jednostką fonetyczną k mówcy m jest obliczana zgodnie z wzorem:

$$d_m^k(\mathbf{x}_r) = \sum_{i=1}^{z} \left(\mathbf{y}_m^k(i)\right)^2 , \tag{26}$$

gdzie:

$$\mathbf{y}_m^k = \tilde{\mathbf{T}}_m^k(\mathbf{x}_r - \bar{\mathbf{x}}_m^k) . \tag{27}$$

W tym przypadku funkcjonał (27) określa kwadrat uogólnionej odległości Mahalanobisa punktu $\mathbf{x}_r$ i $\bar{\mathbf{x}}_m^k$.

Ocena zgodności punktu $\mathbf{x}_r$ z modelem mówcy m wyznaczana jest następująco:

$$d_m(\mathbf{x}_r) = \min_{k=1, ..., L_m} \left\{ d_m^k(\mathbf{x}_r) \right\}, \tag{28}$$

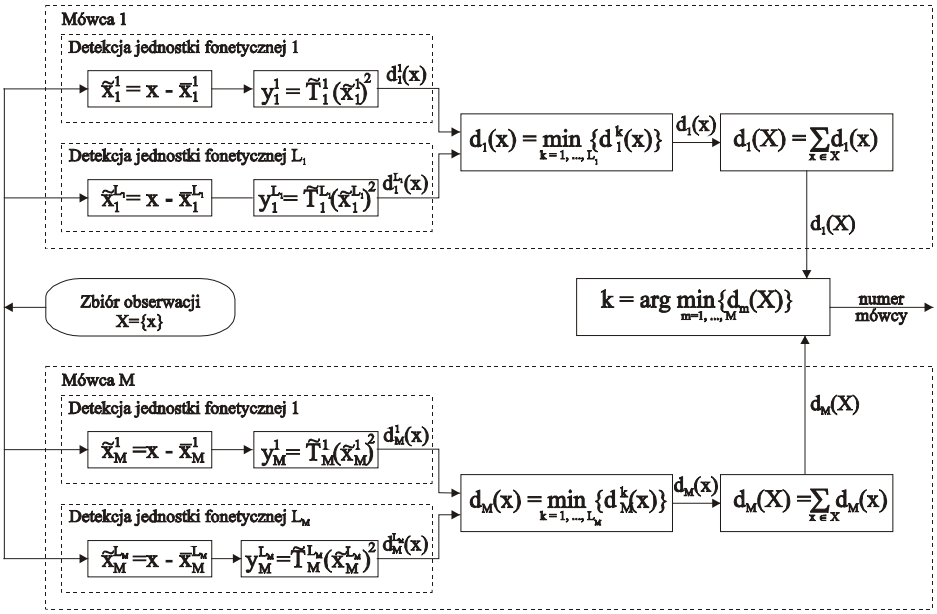
a oceny zgodności całego zbioru obserwacji $\mathbf{X}$ z tym modelem dokonuje się według wzoru:

$$d_m(\mathbf{X}) = \sum_{\mathbf{x}_r \in \mathbf{X}} d_m(\mathbf{x}_r) \tag{29}$$

Stosując opisany powyżej sposób należy dokonać oceny zgodności zbioru obserwacji $\mathbf{X}$ ze wszystkimi modelami mówców (25). Identyfikacja mówcy (wyznaczenie numeru k mówcy) polega na wybraniu tego modelu, który posiada najlepszą ocenę zgodności (29) ze zbiorem obserwacji $\mathbf{X}$, to znaczy:

$$k = \arg \min_{m=1,\ldots,M} \{d_m(\mathbf{X})\}. \tag{30}$$

Opisany algorytm identyfikacji mówcy metodą niezależnej detekcji jednostek fonetycznych przedstawiono na rys. 6.



Rys. 6. Algorytm identyfikacji mówcy metodą niezależnej detekcji jednostek fonetycznych

3. Opis eksperymentu

Opracowany system identyfikacji jest przeznaczony do eksperymentalnego wyznaczenia wartości parametrów algorytmu identyfikacji, a szczególnie liczby *L* jednostek fonetycznych. Opiszemy, przeprowadzony w tym celu, przykładowy eksperyment.

Z zasobu mowy STUDENT wybrano czterech mówców. Modele mówców zostały wyznaczone na podstawie 30 wypowiedzi, po 6 różnych wypowiedzi słów: zero, jeden, dwa, trzy, cztery. Zbiory uczące powstały po podziale wypowiedzi uczących na ramki czasowe o szerokości 20 ms (przy skoku ramki o 0.8 szerokości ramki) i wyznaczeniu z nich 10 współczynników LPC.

Na wszystkich etapach eksperymentu wykorzystywano metrykę Mahalanobisa, a podziału na grupy dokonano metodą Warda. Do testowania użyto 200 wypowiedzi, które nie zostały użyte w procesie uczenia. Każdy mówca był reprezentowany przez 50 wypowiedzi, po 10 różnych wypowiedzi słów: zero, jeden, dwa, trzy, cztery.

Do oceny wyników eksperymentu wykorzystano dwie wielkości:

- stopy niepoprawnej identyfikacji mówców γ_m [6]
- skuteczność systemu.

3.1. Stopy niepoprawnej identyfikacji

Identyfikacja w zamkniętym zbiorze mówców może być opisana za pomocą następującej funkcji:

$$I : \mathbf{X} \rightarrow \{1, 2, ..., M\} \tag{31}$$

gdzie:

$\mathbf{X}$ – zbiór wypowiedzi testowych,
M – liczba mówców.

Błąd niepoprawnej identyfikacji zachodzi, gdy dla wypowiedzi testowej $\mathbf{x}_m^i$ wygenerowanej przez mówcę m o numerze i, zachodzi nierówność $I\left(\mathbf{x}_m^i\right) \neq m$.

Zakładając, że liczba wypowiedzi testowych jest większa od zera, stopę niepoprawnej identyfikacji wyznacza się następująco:

$$\gamma_m = 1 - \frac{1}{N_m} \sum_{i=1}^{N_m} \delta\left[I\left(\mathbf{x}_m^i\right), m \right] \tag{32}$$

gdzie:

m – numer mówcy,
γ_m – stopa niepoprawnej identyfikacji,
N_m – liczba wypowiedzi testowych mówcy m,
$\mathbf{x}_m^i$ – wypowiedź testowa o numerze i mówcy m,
δ – funkcja Kroneckera, gdzie: $\delta\left(k,n\right) = \begin{cases} 1, & k = n \\ 0, & k \neq n \end{cases}$.

Wskaźnik γ_m jest estymatorem prawdopodobieństwa wystąpienia zdarzenia polegającego na błędnej identyfikacji mówcy m.

3.2. Skuteczność identyfikacji mówcy

Skuteczność systemu rozumiana jest jako stosunek liczby poprawnych identyfikacji systemu do liczby wszystkich wypowiedzi testowych wyrażony w procentach, co można zapisać następująco:

$$S = \frac{\sum_{m=1}^{M}\sum_{i=1}^{N_m}\delta\left[I\left(\mathbf{x}_m^i\right), m\right]}{\sum_{m=1}^{M} N_m} \cdot 100\% \tag{33}$$

3.3. Wyniki eksperymentu

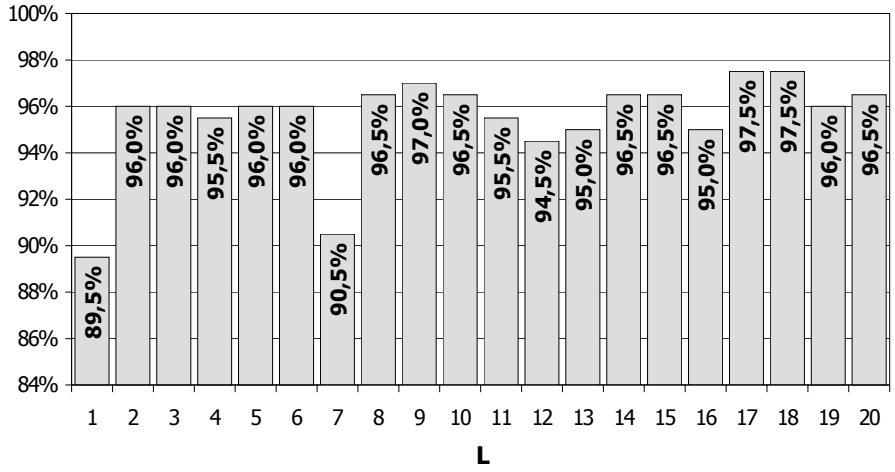
Pierwszy etap eksperymentu polegał na zbadaniu skuteczności identyfikacji mówcy w zależności od podziału zbioru uczącego na zadaną liczbę jednostek fonetycznych. Zależność tę przedstawiono na rys. 7, a w tab. 1 przedstawiono wartości stopy niepoprawnej identyfikacji mówców γ_m. W tym przypadku największą skuteczność (97,5%) uzyskano dla podziału na 17 albo 18 jednostek fonetycznych.

Tab. 1. Wartości stopy niepoprawnej identyfikacji γ_m mówców w zależności od liczby jednostek fonetycznych L

L	Wartości stopy niepoprawnej identyfikacji				L	Wartości stopy niepoprawnej identyfikacji			
	γ_1	γ_2	γ_3	γ_4		γ_1	γ_2	γ_3	γ_4
1	0,02	0,3	0	0,1	11	0	0,1	0	0,08
2	0,02	0,12	0	0,02	12	0	0,14	0	0,08
3	0,02	0,08	0	0,06	13	0	0,12	0	0,08
4	0,02	0,14	0	0,02	14	0	0,12	0	0,02
5	0,02	0,14	0	0	15	0	0,12	0	0,02
6	0,02	0,14	0	0	16	0	0,12	0	0,08
7	0,02	0,26	0	0,1	17	0	0,06	0	0,04
8	0	0,1	0	0,04	18	0,02	0,06	0	0,02
9	0	0,08	0	0,04	19	0,02	0,08	0	0,06
10	0	0,1	0	0,04	20	0,02	0,06	0	0,06

Drugi etap eksperymentu polegał na podziale zbioru uczącego na jednostki fonetyczne według współczynnika niezgodności grupowania (13). W tym przypadku mówcy są reprezentowani przez różną liczbę jednostek fonetycznych,

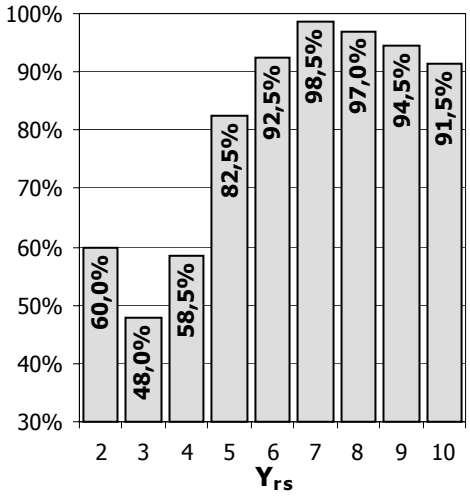
co jest związane z nieodpowiednim pokryciem przestrzeni akustycznej mówcy przez zbiór uczący. Pomimo tego skuteczność identyfikacji zwiększyła się i była równa 98,5% dla (rys. 8, tab. 2).



Rys. 7. Skuteczność identyfikacji mówców w zależności od liczby jednostek fonetycznych *L*

Tab. 2. Wartości stopy niepoprawnej identyfikacji γ_m mówców w zależności od współczynnika Y_{rs}

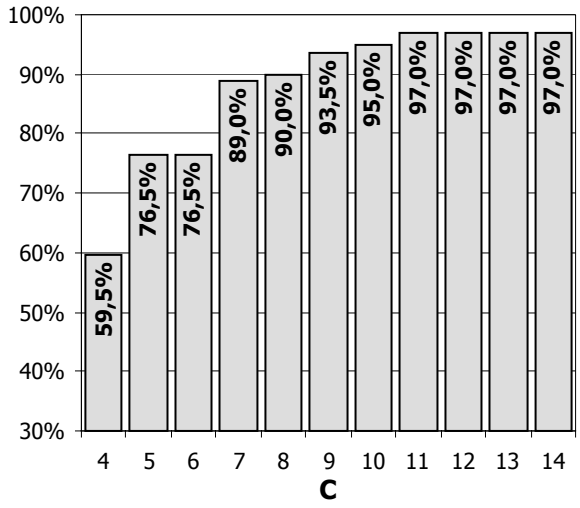
Y_{rs}	Wartości stopy niepoprawnej identyfikacji				Liczba jednostek L_m			
	γ_1	γ_2	γ_3	γ_4	L_1	L_2	L_3	L_4
2	0	0,68	0	0,04	20	12	24	28
3	0,7	0,02	0,7	0,66	53	51	51	53
4	0,44	0,26	0,8	0,18	31	32	36	32
5	0	0,7	0	0	18	20	18	19
6	0	0,3	0	0	13	13	11	14
7	0	0,06	0	0	11	9	10	6
8	0,06	0,06	0	0	8	9	6	6
9	0	0,22	0	0	4	5	4	5
10	0	0,16	0,18	0	3	3	1	4



Rys. 8. Skuteczność identyfikacji mówców w zależności od współczynnika Y_{rs}

Tab. 3. Wartości stopy niepoprawnej identyfikacji γ_m mówców dla Y_{rs} = 7 i odrzuceniu punktów niedopasowanych według wzoru (14)

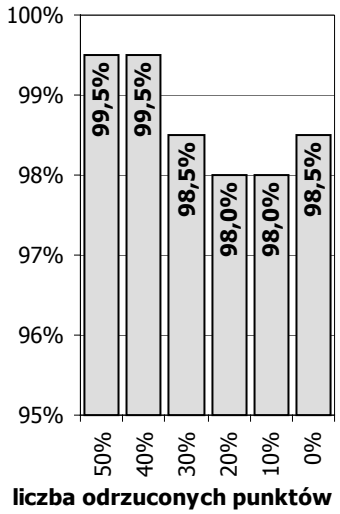
c	Wartości stopy niepoprawnej identyfikacji			
	γ_1	γ_2	γ_3	γ_4
4	1	0,26	0,36	0
5	0,44	0,28	0,22	0
6	0,54	0,02	0,38	0
7	0,38	0,06	0	0
8	0,34	0,06	0	0
9	0,16	0,1	0	0
10	0,1	0,1	0	0
11	0,02	0,1	0	0
12	0,02	0,1	0	0
13	0	0,12	0	0
14	0	0,12	0	0



Rys. 9. Skuteczność identyfikacji mówców dla Y_{rs} = 7 i odrzuceniu punktów niedopasowanych według wzoru (14)

Tab. 4. Wartości stopy niepoprawnej identyfikacji γ_m mówców dla Y_{rs} = 7 i odrzuceniu punktów niedopasowanych według algorytmu przedstawionego na rys. 4

liczba odrzuconych punktów	Wartości stopy niepoprawnej identyfikacji			
	γ_1	γ_2	γ_3	γ_4
50%	0,02	0	0	0
40%	0,02	0	0	0
30%	0,02	0,04	0	0
20%	0,02	0,06	0	0
10%	0	0,08	0	0
0%	0	0,06	0	0



Rys. 10. Skuteczność identyfikacji mówców dla Y_{rs} = 7 i odrzuceniu punktów niedopasowanych według algorytmu przedstawionego na rys. 4

Zbadano również skuteczność systemu przy podziale zbioru uczącego według współczynnika niezgodności grupowania $Y_{rs} = 7$ i przyjęciu założenia, że punkty każdej jednostki fonetycznej wyznaczają rozkład normalny. Skuteczność identyfikacji była największa w przypadku usunięcia punktów niedopasowanych według wzoru (14) dla $c^2 = 11$ i wynosiła 97%. Wyniki przedstawiono na rys. 9 i w tab. 3. Na tym etapie uzyskano najmniejszą skuteczność identyfikacji mówcy, w porównaniu do poprzednich etapów eksperymentu.

Pożądany efekt przynosi dopiero podział zbioru uczącego przy współczynniku niezgodności grupowania $Y_{rs} = 7$ i odrzuceniu punktów niedopasowanych według algorytmu przedstawionego na rys. 4. Po odrzuceniu 50% punktów każdej jednostki fonetycznej, skuteczność identyfikacji mówcy wzrosła do 99,5 (rys. 10, tab. 4).

4. Podsumowanie

W artykule przedstawiono system identyfikacji mówcy metodą niezależnej detekcji jednostek fonetycznych. System służy celom badawczym oraz do eksperymentalnego wyznaczenia wartości parametrów, których nie można wyznaczyć a priori. Konieczność taka zachodzi w przypadku wyznaczania

liczby jednostek fonetycznych, czy określania sposobu usuwania punktów niedopasowanych.

Otrzymane wyniki potwierdzają hipotezę, że wydzielenie wyraźnych, charakterystycznych jednostek fonetycznych dla mówcy zwiększa skuteczność identyfikacji. W przypadku wydzielenia 17 albo 18 jednostek hierarchiczną metodą grupowania osiągnięto skuteczność identyfikacji 97,5%. Z kolei dokonanie podziału według współczynnika niezgodności grupowania $Y_{rs} = 7$ doprowadziło do uzyskania skuteczności równej 98,5%. W celu zwiększenia skuteczności identyfikacji zastosowano usuwanie punktów niedopasowanych. Pierwsza z zaproponowanych metod odrzucenia punktów niedopasowanych, na skutek niespełnienia przyjętych założeń o normalności rozkładów wzorców, nie przyniosła oczekiwanych rezultatów. Natomiast wyższą skuteczność identyfikacji, wynoszącą 99,5%, osiągnięto za pomocą drugiej opracowanej metody, której podstawę stanowiło heurystyczne podejście, polegające na odrzuceniu punktów położonych najdalej od pozostałych punktów grupy.

Wydzielenie jednostek fonetycznych charakteryzujących mówcę, według maksymalnego współczynnika niezgodności grupowania, spowodowało, że każdy z mówców uzyskał ich różną liczbę. Podczas procesu identyfikacji mówcy istnieje możliwość wystąpienia negatywnego zjawiska, polegającego na zawłaszczeniu punktów zbioru obserwacji przez mówcę, który posiada najmniejszą liczbę jednostek fonetycznych. Można tego uniknąć, zapewniając odpowiednie pokrycie przestrzeni cech. Z kolei wydzielenie dużej liczby jednostek fonetycznych prowadzi do uzyskania błędnych ocen zgodności z zorcami i w konsekwencji powoduje zmniejszenie skuteczności identyfikacji.

Poprawne wydzielenie jednostek fonetycznych, które charakteryzują mówcę w systemie, powinno umożliwić uniezależnienie procesu identyfikacji od kontekstu wypowiedzi.

Literatura

[1] Grad L., *Badanie możliwości rozpoznawania mówcy na podstawie reprezentacji LPC sygnału mowy*. Biuletyn IAiR nr 13, 2000.
[2] Grad L., *Metoda rozpoznawania mówcy na podstawie nieuzgodnionej wypowiedzi*. Rozprawa doktorska WAT, 2000.
[3] Grad L., Zastosowanie *transformaty Karhunena–Loève'a do rozpoznawania mówcy*. Biuletyn IAiR nr 13, 2000.
[4] Kwiatkowski W., *Wstęp do cyfrowego przetwarzania sygnałów*. IAiR WAT, 2003.

[5] Kwiatkowski W., *Metody automatycznego rozpoznawania wzorców*. IAiR WAT, 2001.
[6] Wiśniewski A. M., *Metody oceny systemów rozpoznawania mówców*. Biuletyn IAiR nr 13, 2000.

Recenzent: prof. dr hab. inż. Włodzimierz Kwiatkowski

Praca wpłynęła do redakcji: 01.12.2003r.