

Bogusław Twaróg, Ewa Żesławska

Mechanizm predykcji liniowej w zadaniu biometrycznej identyfikacji mówcy

Edukacja - Technika - Informatyka nr 3(13), 271-277

2015

Artykuł został opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej bazhum.muzhp.pl, gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach dozwolonego użytku.

Bogusław TWARÓG

Uniwersytet Rzeszowski, Polska

Ewa ŻEŚLAWSKA

Wyższa Szkoła Informatyki i Zarządzania w Rzeszowie, Polska

Mechanizm predykcji liniowej w zadaniu biometrycznej identyfikacji mówcy

Wstęp

Wykorzystując szybki rozwój branży IT oraz dostęp do nowych technologii, w szybki i łatwy sposób można przechowywać i przesyłać poufne dane, komunikować się z innymi, robić zakupy przez internet czy realizować transakcje bankowe. Obecnie stosowane zabezpieczenia, takie jak: hasła, kody PIN, kody jednorazowe oraz karty magnetyczne lub mikroprocesorowe, nie gwarantują całkowitego bezpieczeństwa użytkownika. Weryfikacja biometryczna nie wymaga od użytkownika zapamiętywania wielu haseł lub posiadania kart. Uwierzytelnienie przeprowadzane jest na podstawie indywidualnych cech badanej osoby, takich jak: tęcza oka, głos czy odciski palców. Systemy biometryczne zapewniają najwyższy poziom bezpieczeństwa oraz są wygodniejsze i tańsze w eksploatacji.

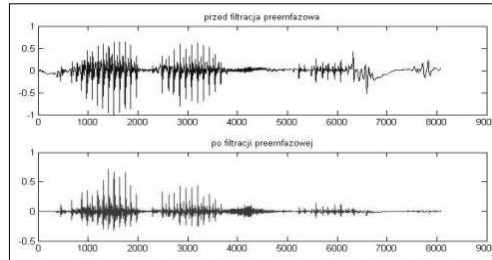
Klasyfikacja metod biometrycznych

Biometryka to nauka o mierzalnych cechach biologicznych, które dla każdego człowieka są unikalne i umożliwiają automatyczne rozpoznanie danej osoby lub też weryfikację, czy dany użytkownik jest tym, za kogo się podaje. Identyfikacja polega na automatycznym rozpoznaniu danego użytkownika przez zbadanie jednej lub kilku jego cech biometrycznych. System porównuje ze wzorcami zapisanymi w bazie danych aktualne informacje pobrane i zapisane przez odpowiednie urządzenia. W biometryce analizowane są cechy fizyczne i behawioralne. Do cech fizycznych zalicza się m.in. metodę Bertillona, odciski palców, skanowanie siatkówki i tęczy oka, kod DNA, rozpoznawanie twarzy, geometrie ust i dłoni. Natomiast do cech behawioralnych zaliczamy: głos, rozpoznawanie odręcznego pisma, charakterystykę ruchu gałki ocznej i ust, rytmikę pisania na klawiaturze, chód itp.

System identyfikacji mówcy wykorzystujący mechanizm predykcji

Danymi wejściowymi do systemu oraz procesu identyfikacji mówcy jest pakiet pobranych próbek głosu zapisanych w postaci cyfrowej. Wstępne przetwa-

rzanie sygnału cechuje się taką obróbką sygnału wejściowego, by wydobyć maksymalną ilość informacji z danej wypowiedzi. Także pozwala zminimalizować błędy systemu oraz szумы i wady sprzętowe. W procesie tym można wyróżnić etapy: eliminacji stałej składowej, normalizacji sygnału, wycinania ciszy.



Rys. 1. Wynik działania filtracji preemfazowej dla głoski dźwięcznej

Selekcja i ekstrakcja cech to procesy, w których sygnał jest poddawany operacjom matematycznym mającym za zadanie wyznaczenie takich parametrów, które posiadają możliwie jak najwięcej informacji określających danego mówcę. W procesie tym wyróżniamy trzy składowe: preemfazę, autokorelację oraz współczynniki LPC (*linear predictive coding*). Przykładowy wynik filtracji preemfazowej przedstawiono na rys. 1.

Korelacja wzajemna sygnałów dyskretnych sprecyzowana jest jako iloczyn skalarny sygnału i jego wiernej kopii przesuniętej w czasie rozpatrywany jako funkcja przesunięcia; charakteryzuje wzajemne rozmieszczenie energii w czasie między sygnałem i sygnałem przesuniętym. Funkcja ta odnajduje zastosowanie m.in. w detekcji zaszumionych sygnałów, określaniu podobieństwa do danych wzorców oraz w pomiarach opóźnień. Miarą korelacji wzajemnej jest współczynnik korelacji liniowej Pearsona opisany zależnością:

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}, \quad r_{xy} \in [-1, 1], \quad (1)$$

gdzie: x, y – badane wektory; x_i, y_i – wartości poszczególnych elementów wektora $i = 1, 2, \dots, n$, \bar{x}, \bar{y} – wartości średnie wektorów.

Specyficznym przypadkiem funkcji korelacji wzajemnej jest funkcja autokorelacji, która badany sygnał porównuje sam ze sobą. Dzięki temu można przeprowadzić analizę właściwości sygnału w dziedzinie czasu. Dla sygnałów dyskretnych korzysta się z definicji autokorelacji w postaci:

$$r(k) = \sum_{n=1}^{N-1} s(n)s(n+k), \quad (2)$$

gdzie: $k = 1, 2, \dots, n$ $s(n)$ – fragment sygnału składający się z N próbek, utworzony z mowy oryginalnej poddanej wcześniej filtracji premfazowej (rys. 1).

Funkcja ta jest zawsze rzeczywista i parzysta oraz ma maksimum w punkcie odpowiadającym zerowemu przesunięciu. Przybierać też może wartości dodatnie i ujemne. Funkcja autokorelacji ma specyficzne znaczenie w analizie sygnałów losowych. Głównym jej przeznaczeniem jest badanie na jej podstawie, w jakim stopniu wartości sygnału w danej chwili mają wpływ na wartości sygnału w przyszłości.

Współczynniki LPC są powszechnie stosowane w kompresji sygnału mowy. Główną zaletą tej metody jest to, że za pomocą niewielkiej liczby współczynników można z dużą dokładnością opisać dany sygnał mowy. Wynika to z faktu, że głos ludzki charakteryzuje się dużą redundancją, co w tej sytuacji oznacza, że każdą kolejną próbkę można zaprezentować jako kombinację liniową poprzednich próbek. Struny głosowe mogą generować sygnał dźwięczny i bezdźwięczny. Sygnał ten następnie jest kształtowany w trakcie głosowym, który porównać można do filtra IIR o pewnej transmitancji $H(z)$. Transmitancja ta jest cechą osobniczą każdego mówcy. Wynika to z tego, że struktura traktu głosowego jest odmienna, indywidualna i niepowtarzalna u poszczególnych ludzi. Współczynniki LPC w tym przypadku są współczynnikami takiego filtra IIR. W następstwie tego na ich podstawie można z bardzo dużą wiarygodnością identyfikować i rozpoznawać konkretnego mówcę.

Predykcja liniowa bazuje na tym, że wartość sygnału $a(p)$ jest wyznaczana na podstawie poprzednich wartości. Jednym ze sposobów wyliczania współczynników LPC jest rekursywny algorytm Durбина-Levinsona, który działa następująco:

- **wejście:** $r_0, r_1, r_2, \dots, r_p$ – estymata (współczynniki) funkcji autokorelacji sygnału ($r_i = r(i)$),
- **wyjście:** $a_1, a_2, a_3, \dots, a_p$ – współczynniki LPC, $\gamma_1, \gamma_2, \gamma_3, \dots, \gamma_p$ – współczynniki odbicia (*parcor*), E_p – błąd średniokwadratowy,

$$E_0 = r_0 \quad (3)$$

dla: $i = 1, 2, 3, \dots, p$:

$$\gamma_i = \frac{r_i + a_{(i-1)}r_{i-1} + a_{(i-1)2}r_{i-2} + \dots + a_{(i-1)(i-1)}r_1}{E_{i-1}}, \quad E_i = (1 - \gamma_i^2)E_{i-1} \quad (4)$$

$$a_{ii} = -\gamma_i \quad (5)$$

dla $j = 1, 2, 3, \dots, i-1$:

$$a_{ij} = a_{(i-1)j} - \gamma_i a_{(i-1)(i-j)} \quad (6)$$

Klasyfikacja identyfikowanej osoby jest wykonywana przy użyciu metody minimalnoodległościowej. Niech a^x reprezentuje wektor współczynników LPC:

$$a^x = [a(1) \ a(2) \ a(3) \ \dots \ a(p)]^T. \quad (7)$$

Równość (8) będzie definiowała macierz utworzoną z powyższych wektorów obliczonych z $n = 8$ powtórzeń hasła przez osobę X :

$$A^x = a_1^x \ a_2^x \ a_3^x \ \dots \ a_p^x \quad (8)$$

czyli:

$$A^x = \begin{bmatrix} A^x(1,1) & A^x(1,2) & \dots & A^x(1,n) \\ A^x(2,1) & A^x(2,2) & \dots & A^x(2,n) \\ \dots & \dots & \dots & \dots \\ A^x(p,1) & A^x(p,2) & \dots & A^x(p,n) \end{bmatrix}, \quad (9)$$

gdzie: p – liczba współczynników LPC, n – liczba powtórzeń hasła.

Wektor wartości średnich współczynników jako uśrednienie wierszy macierzy można zapisać:

$$a^{\bar{x}}(i) = \frac{1}{n} \sum_{j=1}^n a^x(i, j), \quad (10)$$

w którym: $i, j = 1, 2, \dots, p$.

Wyznaczenie odchylenia dla powtórzeń hasła wypowiedzianych przez inne osoby od powyższego uśrednionego wektora osoby X wynosi:

$$\Delta_x^y = |a_i^y - a_i^{\bar{x}}|, \quad (11)$$

gdzie: $i = 1, 2, \dots, p$; Y – inna osoba.

Decyzja podejmowana jest na podstawie wyboru wektora użytkownika zarejestrowanego, którego odległość od wektora testowanego jest najmniejsza. Odległość tę wyznaczamy, sumując odległości (wynik wyrażony w %):

$$\Delta_z^y = \sum_{j=1}^p \Delta_x^y(i). \quad (12)$$

Projekt biometrycznej aplikacji identyfikacji mowy

Wykorzystując środowisko MATLAB oraz dostępne biblioteki, stworzono aplikację umożliwiającą identyfikację mowcy. Aplikacja została podzielona na panel bazy danych umożliwiający tworzenie i modyfikowanie bazy danych oraz panel identyfikujący mowcę.

Wykonano szereg badań w celu sprawdzenia skuteczności zaprojektowanego systemu identyfikacji mowcy. Przeprowadzono serię prób, wprowadzając różnego rodzaju zniekształcenia sygnału wejściowego, zmieniając modulację głosu oraz parametry programu. Eksperymenty polegały na 10-krotnym powtórzeniu procesu identyfikacji jednej osoby, wypowiedzi słowa, a także sekwencji słów. W bazie danych znajduje się 6 osób, dla każdej z nich zarejestrowano 4 próbki głosu.

Wyniki badań skuteczności działania programu w warunkach domowych przedstawiono w tabeli 1.

Tabela 1

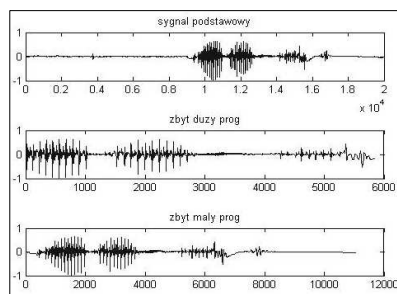
Wyniki badań programu „Identyfikacja mówcy” w warunkach domowych

Nr powtórzeń	1	2	3	4	5	6	7	8	9	10
Trafność (1/0)	1	1	1	1	0	1	1	1	1	1
Zgodność z bazą danych (%)	74	100	81,61	79,96	75,54	100	69,99	94,10	100	91,80
Średnia zgodność z bazą danych (%)	90,05									

Tabela 2

Wyniki badań programu „Identyfikacja mówcy” dla niezgodnionej wypowiedzi

Nr powtórzeń	1	2	3	4	5	6	7	8	9	10
Trafność (1/0)	1	0	1	1	1	0	1	1	1	1
Zgodność z bazą danych (%)	68,72	43,03	74,80	66,89	49,94	34,90	30,29	82,1	81,23	69,31
Średnia zgodność z bazą danych (%)	60,12									



Rys. 2. Przypadki błędnego określenia progu ciszy.

Przeanalizowano skuteczność działania aplikacji dla niezgodnionej wypowiedzi (zob. tabela 2). Założenia programu określają największą skuteczność działania dla uzgodnionej wypowiedzi. Przeprowadzone badania dowodzą, że program równie dobrze identyfikuje mówcę na podstawie próbek głosu losowo wybranych pojedynczych słów. Potwierdza to trafność zastosowania metody LPC w systemach identyfikacji mówcy. Kolejne eksperymenty dotyczyły analizy sekwencji słów – mowy ciągłej, której czas wypowiedzi pokrywał się z cza-

sem nagrywania głosu w programie. 50% skuteczności poprawnej identyfikacji świadczy o tym, że przyjęte rozwiązania nie są wystarczające w rozpoznawaniu mowy ciągłej dla niezgodnionej wypowiedzi. Przeprowadzono serię testów na wypowiedzi obarczonej szumem, którego energia przekraczała próg filtra wycinającego ciszę. W wyniku otrzymano średnią zgodność z bazą danych równą 69,35%, co potwierdziło dużą skuteczność działania aplikacji. Wynika to z faktu zastosowania filtra preemfazowego, który w znacznym stopniu redukuje poziom szumów i trzasków oraz uwydatnia wysokie częstotliwości w zarejestrowanym sygnale mowy. Próg ciszy to jeden z najbardziej istotnych elementów zaprojektowanego programu. Od jego poprawnego ustalenia w dużej mierze zależy skuteczność procesu identyfikacji. Istnieje kilka sprawdzonych algorytmów określania tego progu, jednak zdecydowano się na statyczne ustalanie progu metodą doświadczalną. Rysunek 2 obrazuje przypadki błędnego określenia progu. Zbyt duży próg powoduje, że w próbie głosu oprócz ciszy zostaje wycięta część sygnału mowy. W krańcowych przypadkach algorytm może wyciąć cały sygnał, pozostawiając tylko maksymalną wartość amplitudy, co uniemożliwi poprawny proces identyfikacji. Zbyt mały próg może spowodować, że algorytm zatrzyma się, gdy natrafi na krótki skok amplitudy. Spowoduje to brak usunięcia ciszy oraz brak skrócenia wektora, co ma wpływ na liczbę wykonywanych obliczeń i jakość identyfikacji.

Podsumowanie

Przedstawiona w artykule metoda oparta na predykcji liniowej (LPC) oraz klasyfikacji minimalnoodległościowej to tylko jedno z licznych rozwiązań. Aplikacja została zaprojektowana w taki sposób, aby możliwe było dołączenie do niej nowych algorytmów zwiększających skuteczność i zakres jej działania. Aplikacje podobne do tej mogą znaleźć wkrótce szerokie zastosowanie głównie w systemach weryfikacji tożsamości. Pomimo tego, że nie gwarantują maksymalnej pewności, w połączeniu z innymi metodami pomiaru cech biometrycznych pozwolą stworzyć systemy gwarantujące znacznie większy poziom bezpieczeństwa niż te stosowane obecnie.

Literatura

- Ashbourne J. (2000): *Biometrics: Advanced Identity Verification*, Springer-Verlag UK August.
- Gomółka Z. (2008): *Identyfikacja mówcy z wykorzystaniem współczynników predykcji liniowej*,
- Gomółka Z., Lewandowski T. (2008): *The Biometric Signals Processing*, Advances in soft computing.
- Lyons R.G. (1999): *Wprowadzenie do cyfrowego przetwarzania sygnałów*, Warszawa.
- Szabatin J. (2003): *Przetwarzanie sygnałów*, Warszawa.
- Tadeusiewicz R. (1988): *Sygnal mowy*, Warszawa.
- Zieliński T.P. (2005): *Cyfrowe przetwarzanie sygnałów od teorii do zastosowań*, Warszawa.

Streszczenie

W artykule przedstawiono mechanizm predykcji liniowej w zadaniu biometrycznej identyfikacji mówcy. Przedstawiono zagadnienie metody opartej na predykcji liniowej LPC (*linear predictive coding*). Zaprezentowano otrzymane wyniki badań zaprojektowanej aplikacji.

Słowa kluczowe: identyfikacja, biometria, rozpoznawanie mowy, LPC (*linear predictive coding*).

The Mechanism of Linear Prediction in the Task Speaker Identification

Abstract

The article presents the linear prediction mechanism in the task of biometric identification speaker. The problems method based on linear prediction LPC (Linear Predictive Coding). Also the preliminary results has been presented too.

Keywords: identification, biometrics, voice recognition, LPC (Linear Predictive Coding).