JOURNAL of COMPUTER SCIENCES INSTITUTE

JCSI 13 (2019) 323-331
Wysłane: 2019-09-07
Przyjęte: 2019-09-17

Ekstrakcja parametrów z próbek danych biometrycznych

Paweł Danek*, Krzysztof Ćwirta, Piotr Kopniak
Politechnika Lubelska, Instytut Informatyki, Nadbystrzycka 36B, 20-618 Lublin, Polska

Streszczenie. W artykule opisano możliwe sposoby ekstrakcji parametrów z próbek danych biometrycznych, takich jak odcisk palca czy nagranie głosu. Zweryfikowano wpływ konkretnych sposobów obróbki na skuteczność algorytmów obróbki próbek biometrycznych oraz ich porównania. Wykonano badania polegające na przetworzeniu dużej liczby próbek z użyciem wybranych algorytmów. W przypadku odcisku palca wykorzystano normalizację obrazu, filtr Gabora i porównanie z użyciem deskryptorów. Dla autoryzacji głosowej analizowano algorytmy LPC i MFCC. W przypadku obu rodzajów autoryzacji uzyskano zadowalającą skuteczność rzędu 60-80%.

Słowa kluczowe: biomteria; odcisk; głos; autoryzacja; normalizacja; gabor; deskryptor; lpc; mfcc

*Autor do korespondencji.

Adresy e-mail: pawel.danek@pollub.edu.pl, krzysztof.cwirta@pollub.edu.pl, p.kopniak@pollub.pl

Extraction of parameters from biometric data samples

Paweł Danek*, Krzysztof Ćwirta*, Piotr Kopniak
Institute of Computer Science, Lublin University of Technology, Nadbystrzycka 36B, 20-618 Lublin, Poland

Abstract. This article describes possible ways to extract parameters from biometric data samples, such as fingerprint or voice recording. Influence of particular approaches to biometric sample preparation and comparision algorithms accuracy was verified. Experiment involving processing big ammount of samples with usage of particular algorithms was performed. In fingerprint detection case the image normalization, Gabor filtering and comparision method based on descriptors were used. For voice authorization LPC and MFCC alghoritms were used. In both cases satisfying accuracy (60-80%) was the result of the surveys.

Keywords: biometrics; fingerprint; voice; authorization; normalization; gabor; descriptor; lpc; mfcc

*Corresponding author.

E-mail addresses: pawel.danek@pollub.edu.pl, krzysztof.cwirta@pollub.edu.pl p.kopniak@pollub.pl

1. Wstęp

W ramach pracy przeanalizowano działanie algorytmów autoryzacji biometrycznej i ekstrakcji danych z próbek pod kątem odcisku palca oraz głosu. Nie ulega wątpliwości fakt, że w dzisiejszych czasach spora część codziennego życia przenosi się do Internetu. Codziennością jest kontakt ze znajomymi czy współpracownikami z wykorzystaniem *mediów społecznościowych*, robienie zakupów *on-line*, praca czy rozrywka z wykorzystaniem sieci. W tym, ogólnodostępnym medium znajduje się coraz więcej danych dotyczących każdego człowieka oraz dostęp do wielu, wrażliwych funkcjonalności - takich jak np. bankowość *on-line*. W dobie miniaturyzacji oraz przenoszenia tych funkcjonalności na ekrany *smartfonów* - coraz większy nacisk kładzie się na zabezpieczenia.

Jednym z najdokładniejszych i najtrudniejszych do złamania mechanizmów zabezpieczeń jest obecnie biometryka. Dzieje się tak dlatego, że ciało człowieka, jego budowa i cechy charakterystyczne dają świetne modele autoryzacyjne, które dla każdego człowieka są unikalne i niepowtarzalne. Nie da się ich *pożyczyć* lub *ukraść*, a sklonowanie ich jest dużo trudniejsze niż odgadnięcie hasła do serwisu. Każda próbka danych biometrycznych charakteryzuje się swoją budową i listą cech, które można z nich ekstrahować. Proces wyodrębniania poszczególnych parametrów jest jednym z najważniejszych, które budują skuteczne systemy autoryzacji biometrycznej.

W celu analizy sposobów obróbki próbek stosowanych w takich mechanizmach rozpoznawania użytkownika wybrano dwa, najbardziej popularne ze sposobów – autoryzacja z wykorzystaniem odcisku palca i głosu. W ramach badań porównano po dwa różne sposoby ekstrakcji próbek – z wykorzystaniem detektora Harrisa (wykrywającego duże zmiany częstotliwości) oraz algorytmu wykrywającego minucje na podstawie liczby przecięć linii papilarnych.. W przypadku weryfikacji odcisku palca – wykorzystano mechanizm porównania oparty o *deskryptory obrazu* – natywny mechanizm bibliotek do przetwarzania obrazu, z użyciem któregoś można zweryfikować podobieństwo dwóch grafik. Wejściem dla deskryptora jest lista punktów kluczowych, na podstawie której zbudowany będzie deskryptor. Od określenia źródła tych punktów zależy skuteczność i szybkość działania algorytmu – w jednym algorytmie wykorzystano punkty wykryte przez detektor Harrisa, natomiast w drugim wyznaczono *minucje* – punkty charakterystyczne odcisku. W ramach badań wykorzystano bazę odcisków międzynarodowego konkursu FVC (publicznie dostępną w Internecie). Skrypt, napisany w języku Python, kolejno porównywał odciski na zasadzie *każdy z każdym* z wykorzystaniem obu algorytmów. Każdy z przypadków miał swój oczekiwany rezultat, na bazie którego generowano

końcowy wynik. Z wykorzystaniem tych danych obliczono procentowe skuteczności algorytmów (ogólne, przypadku pozytywnej autoryzacji i pozytywnej odmowy dostępu).

W przypadku autoryzacji z wykorzystaniem próbki głosowej porównano skuteczność dwóch algorytmów ekstrakcji próbek – MFCC oraz LPC. Algorytm trenował się z wykorzystaniem *n* próbek dla każdego mówcy – w tym procesie wyznaczał oba rodzaje cech dla każdej z próbek treningowych mówcy. Na tej podstawie powstała książka kodowa. Następnie – dla *m* innych próbek każdego z mówców wyszukiwano największy stopień podobieństwa (najmniejszy dystans) do próbek zapisanych w książkach kodowych. Jeśli nazwy mówców się zgadzały – porównanie zaliczano jako udaną próbę. Na podstawie tych wyników określono procentową skuteczność autoryzacji z wykorzystaniem obu sposobów ekstrakcji cech z próbek.

2. Autoryzacja z wykorzystaniem odcisku palca

Jednym z najpopularniejszych rozwiązań w temacie biometryki jest wykorzystanie linii papilarnych. Obróbka skanu odcisku palca, choć jest procesem bardzo skomplikowanym, to na wysokim poziomie abstrakcji składa się z kilku prostych kroków. Wśród nich można wyróżnić [1]:

- przygotowanie obrazu
- uwydatnienie linii
- odchudzenie linii
- ekstrakcję cech
- porównanie modeli

Każdy z tych punktów realizowany jest z użyciem różnych, charakterystycznych dla swojego przebiegu algorytmów. Największą różnorodność wśród zastosowanych rozwiązań można spotkać w pierwszym z punktów. Spotyka się on bowiem z najtrudniejszym zadaniem. Skany odcisków mogą pochodzić z różnych skanerów (niekoniecznie optycznych) lub nawet być wygenerowane przez odpowiednie oprogramowanie. Skutkuje to dużą rozbieżnością w jakości analizowanych próbek - różne mogą być nie tylko takie parametry jak rozdzielczość obrazów, ale także szczegółowe informacje - na przykład jasność pikseli w miejscu nacisku lub pustym polu.

Z tego powodu algorytmy wykorzystane przy przygotowaniu odcisku są często kluczem do stworzenia dobrej jakości oprogramowania. Celem tego punktu jest obróbka obrazu w takich sposób, aby otrzymać wyraźny obraz, bez szumów, z odpowiednim kontrastem pomiędzy liniami papilarnymi a pustymi polami. Jednym z ciekawych i skutecznych rozwiązań do osiągnięcia tego celu jest wykorzystanie mechanizmu *Contrast Limited Adaptive Histogram Equalization* - w skrócie *CLAHE* [2-3]. Jest to zoptymalizowany proces normalizacji histogramu. Jego założeniem jest analiza natężenia jasności punktów obrazu oraz wyrównanie i *wygładzenie* dużych skoków w histogramie. Efektem tego procesu jest eliminacja prześwietleń obrazu (zmniejszenie liczby zbyt jasnych punktów) oraz uwydatnienie kontrastu (poprzez wzmocnienie ciemnych obszarów). W przypadku obrazów, w których kontrast nie jest *globalny* - czyli nie powinien być zbliżony dla każdego regionu obrazu, zaczęto wykorzystywać mechanizmy adaptacyjnej normalizacji – *AHE* [4]. Zakładają one podział obrazu na bloki o określony rozmiarze oraz uwydatnieniu kontrastu w danym regionie. Jednak nie zawsze skutkuje to oczekiwanym rezultatem - małe bloki składające się z samych jasnych punktów mogą być po takim procesie prześwietlone, a bloki zawierające szum - wzmocnione. Stąd mechanizm *CLAHE*, którego przedrostek - *Contrast Limited* - zapewnia redukcję punktów, które przekraczają zadany limit kontrastu. Są one przesuwane do innych miejsc histogramu. Dopiero po tym procesie - w każdym bloku przeprowadzana jest normalizacja histogramu [2].

Innym, prostszym podejściem jest wykorzystanie progu jasności [5]. Zakładając, że przetwarzany jest obraz w skali szarości analizie poddawana jest jasność z zakresu od 0 do 255. Punktem zainteresowania są najciemniejsze piksele - czyli miejsca, w których przyciśnięta została linia papilarna. Warto wyznaczyć zakres analizy - czyli odrzucić wszystkie punkty o jasności wyższej niż - przykładowo - 30. Następnie należy poznać najczęściej występujące wartości jasności w tym zakresie - idealnie sprawdzi się tutaj mediana. Po jej poznaniu oraz wykorzystaniu odpowiedniej tolerancji - 5% lub 10% wartości - wszystkie punkty niższe od określonej wartości uznaje się za fragment linii papilarnej. Wszystkim punktom odcisku przypisuje się minimalną wartość jasności - 0, natomiast wszystkie inne powinny być tłem - czyli otrzymać maksymalną wartość 255. To podejście sprawdzi się dobrze w przypadku dobrej jakości skanów. Podobny algorytm można zastosować w oparciu o odchylenie standardowe parametru jasności.

W przypadku uwydatnienia linii odcisku najczęściej wykorzystywany jest filtr Gabora [5,6]. Proces zaproponowany przez Dennis'a Gabor'a [7] znalazł szerokie zastosowanie w medycznym przetwarzaniu obrazu [8]. Wykorzystywany jest przy obróbce obrazów uzyskanych z aparatury medycznej. Jest on również stosowany w większości projektów zajmujących się autoryzacją z wykorzystaniem odcisku palca. Głównym założeniem wykorzystania tego filtru jest ekstrakcja fragmentów obrazu o konkretnej częstotliwości, skierowanych w konkretnym kierunku [7]. Takie fragmenty obrazu oznaczone będą wysoką wartością jasności. Do zastosowania tego filtra trzeba znać kierunki linii papilarnych na obrazie oraz częstotliwości tych fragmentów. Do określenia kierunków linii można skorzystać z gradientów funkcji Gaussa (których wartości narastają odpowiednio z położeniem w danym kierunku) lub transformat Sobela w oparciu o matematyczne metody wyznaczania kierunku wektorów oparte o *arcus tangens* [9]. Częstotliwość bloków obrazu należy rozumieć jako ilość wysokich zmian poziomów jasności na konkretnym obszarze. Można liczyć ją poprzez analizę każdego z pikseli lub ponownie wykorzystać wyliczenie odchylenia standardowego. W większości jednak będą to wartości na tyle zbliżone, że można również przyjąć ich średnią za wystarczającą. Znając częstotliwość obrazu oraz orientację każdego z małych bloków obrazu można przystąpić do wyliczenia wartości filtra Gabora. Te z kolei wylicza się z użyciem wzoru matematycznego (1), według którego oryginalnie zdefiniowany został ten filtr [7].

$$g(x,y;\lambda,\theta,\phi,\gamma)=\exp\left(-\left(\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}\right)\right)\cos\left(2\pi\frac{x'}{\lambda}+\phi\right) \quad (1)$$

gdzie [7]: λ - długość fali, θ - kąt fali, ϕ - przesunięcie falowe, γ - współczynnik proporcjonalności, σ- odchylenie standardowe obwiedni fali.

Na każdy z bloków obrazu należy nakładać jego odpowiednią wersję [5,6] - zgodną z kierunkiem linii w danym bloku. Efektem wykonania tego procesu będzie skan odcisku z *uwydatnionymi* liniami papilarnymi. Efekt tego procesu widoczny jest na rysunku [1]. Dzięki temu procesowi są one proste i jednolite - znikają poszarpania oraz puste pola, które pozostawiają po sobie pory w skórze. Wykonanie tego procesu jest jednym z najważniejszych kroków prowadzących do prawidłowego wyznaczenia cech charakterystycznych odcisku.

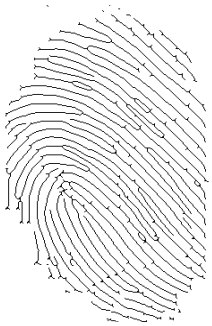
Rys. 1. Skan odcisku po zastosowaniu filtra Gabora [10]

Kolejnym z ważnych procesów przygotowujących obraz do wykrywania punktów charakterystycznych odcisku jest *odchudzanie* odcisku palca [6]. Proces ten polega na cieniowaniu pogrubionych linii [11] - do momentu kiedy ich grubość będzie możliwie najniższa. Jednym z rozwiązań tego problemu jest wykorzystanie jąder przekształceń, których nałożenie na obraz cieniuje jego krawędzie [11]. Każdy blok obrazu (o rozmiarach równych macierzy) jest przekształcony zgodnie z wagą każdego z pikseli jądra. W celu wykonania procesu cieniowania odpowiednio są to jądra zwiększające wagi zewnętrznych krawędzi linii. Przykładowo, taka macierz [6] zmniejsza jasność dolnych krawędzi bloku.

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0.1 & 0.1 & 0.1 \end{bmatrix} \quad (2)$$

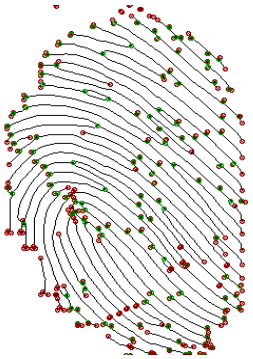
Po zdefiniowaniu odpowiedniej liczby operatorów należy wykorzystywać je jako jądro przekształcenia do momentu, aż wykonanie przekształceń przestanie dawać jakiekolwiek efekty [6]. Wynikiem przekształcenia obrazu poprzez operatory *cieniowania* jest skan odcisku, w którym każda linia ma grubość rzędu jednego piksela, jak widać na rysunku [2].

Dzieje się tak ze względu na rozmiar analizowanego obszaru – w przypadku ograniczenia się do najprostszych typów (przy założeniu że oczekiwana grubość linii to 1 piksel) wystarczy analiza drobnych bloków o rozmiarze 3x3. Takie podejście jest znane jako *Crossing number concept* i powszechnie stosowane w algorytmach autoryzacji biometrycznej z użyciem odcisku palca [6,14].



Rys. 2. Skan odcisku po procesie cieniowania [10]

Jeśli w takim obszarze znajdują się co najmniej dwa ciemne punkty obok siebie, których najbliższy sąsiad jest jasnym punktem - jest to zakończenie lub rozpoczęcie linii. Jeśli w tym samym obszarze znajduje się punkt, który ma dwóch, bezpośrednich sąsiadów, to jest to rozwidlenie linii [6,14]. Już przy tak prostym algorytmie, analizując kolejne bloki obrazów, można wykryć bardzo dużą liczbę punktów kluczowych w pojedynczym skanie odcisku. Ich położenie (w obrazie oraz względem siebie) oraz rodzaj są kluczowe dla algorytmów porównawczych. Efekt działania tego algorytmu widać na rysunku [3].



Rys. 3. Zakończenia linii oraz rozwidlenia wykryte w skanie odcisku [10]

Ostatnim i najważniejszym etapem procesu przetwarzania odcisków jest porównanie dwóch zestawów minucji względem siebie. Jednym z ciekawych sposobów porównania dwóch odcisków jest wykorzystanie deskryptorów obrazów [5] oraz ich komparatorów. Mechanizmy (takie jak *ORB - Oriented Rotated Brief* czy *bruteforce matcher*) są wbudowane w popularne biblioteki (np. *OpenCV*) służące do przetwarzania obrazu. Posiadają one implementacje w wielu językach programowania. Deskryptor budowany jest na podstawie punktów kluczowych obrazu (którymi mogą być minucje oraz ich okolice), a ich porównanie odbywa się na zasadzie odnalezienia *trafień*, czyli korespondujących punktów w drugim obrazie. Pomiędzy korespondującymi punktami kluczowymi wyznaczona jest odległość - im mniejsza, tym trafienie jest dokładniejsze [3]. Dzięki wykorzystaniu takiego mechanizmu - procesy zaimplementowane w popularnych bibliotekach automatycznie odnajdą korespondujące układy współrzędnych dla dwóch odcisków. Wśród innych, ciekawych i skuteczniejszych rozwiązań prym wiodą te, które analizują kierunek oraz typ minucji [13]. Jednym z ciekawszych podejść jest wstępne obrócenie obu

skanów (tak aby znajdowały się w jednym układzie odniesienia), odnalezienie minucji, po czym odnalezienie korespondujących punktów poprzez założenie że dzieli je najmniejsza odległość [15]. Po wydobyciu takich informacji należy porównać typy minucji oraz ich kierunki. Ostatnim z ciekawych i skutecznych rozwiązań jest ekstrakcja z obrazu wszystkich fragmentów poza okolicami minucji [13]. Po wykonywaniu obróceń i przesunięć obu obrazów, nakładając je na siebie, w pewnym punkcie powinno otrzymać się dużą liczbę pokryć. Świadczy to o tym, że dane odciski są zbieżne. Wszystkie z opisanych powyżej rozwiązań spotykane są w amatorskich projektach zajmujących się autoryzacją biometryczną z użyciem odcisku palca.

3. Badania skuteczności

W celu zbadania skuteczności algorytmów przygotowano skrypt wykorzystujący projekty oparte o wykrywanie punktów kluczowych na bazie detektora Harrisa oraz ręczne wykrywanie minucji. Zbiór punktów wykorzystywano do zbudowania deskryptora obrazu. Następnie, z wykorzystaniem rodzimych mechanizmów biblioteki *OpenCV* porównano podobieństwo obu obrazów. Badanie skuteczności wykonano w oparciu o publicznie dostępną bazę skanów ogólnoświatowego konkursu FVC z 2004 roku [10]. Zbiór plików – w strukturze dziesięć różnych odcisków, po osiem próbek każdy – porównano na zasadzie *każdy z każdym*. W przypadku porównania różnych próbek tego samego odcisku – oczekiwano poprawnej autoryzacji. Jeśli wykorzystane zostały skany innych odcisków – algorytm powinien odmówić dostępu. Jako parametry wyjściowe określono procentową skuteczność autoryzacji (jako procentowy stosunek liczby udanych autoryzacji do liczby prób, w których porównano skany tego samego odcisku) oraz procentową skuteczność odmowy dostępu (analogicznie do poprzedniego przypadku – procent przypadków w których uzyskano poprawny wynik negatywny). Obliczenia wykonano na maszynie VPS. Sumarycznie przypadków porównania było nieco ponad 3100, z czego zdecydowana większość to przypadki odmowy dostępu. W tabeli wynikowej projekt działający na bazie detektorów Harrisa oznaczony został jako *A*, natomiast implementacja z wykrywaniem minucji jako *B*.

Tabela 1. Wynik badania skuteczności algorytmu porównującego odciski palców

Parametr	Projekt A	Projekt B
Ogólna skuteczność procentowa	86%	87%
Skuteczność przypadków autoryzacji	44%	60%
Skuteczność przypadków odmowy dostępu	90%	90%
Średni czas porównania	13s	57s

Jako próg referencji, przy dostosowywaniu minimalnego progu autoryzacji, wykorzystano utrzymanie skuteczności odmowy dostępu na poziomie 90%. Po dostosowaniu algorytmów do tego parametru – oba mają wysoką ogólną skuteczność procentową (procent wszystkich przypadków, w których oczekiwany wynik zgadzał się z faktycznym). Wysoka wartość tego wyniku spowodowana jest faktem, że w całym badaniu przypadków odmowy dostępu było dużo więcej niż przypadków autoryzacji. Dużo bardziej interesującym parametrem jest procentowa skuteczność w przypadku porównania dwóch, tych samych odcisków. Projekt definiujący punkty kluczowe jako wynik działania detektora Harrisa autoryzował poprawnie tylko 44% przypadków. Wynik jest bardzo niski i rozwiązanie zdecydowanie nie nadaje się na produkcyjne rozwiązanie. Duży wzrost – do 60% - odnotowano w przypadku zdefiniowania punktów kluczowych deskryptora jako minucji odcisku. Pomimo sporego skoku – procent autoryzacji jest wciąż niski. Winnym jest sposób porównania – deskryptory obrazu, chociaż jest to bardzo kreatywne rozwiązania, nie dają dobrej skuteczności w przypadku porównania tak szczegółowych *obrazów*. W celu poprawienia tego wyniku warto zmienić sposób porównania danych otrzymanych na podstawie obróbki skanu. Wśród najpopularniejszych rozwiązań często definiuje się sposób oparty o odnalezienie najbliższej, sąsiadującej minucji (wraz z zapamiętaniem ich typów) czy założenie, że w odpowiednio obróconych skanach odcisków odpowiadające sobie minucje, to te które dzieli najmniejsza odległość. Sposoby te wymieniono w poprzednim rozdziale.

W badaniach określono także średni czas porównania dwóch skanów odcisków. W projekcie B, wykrywającym minucje, widać duży (blisko pięciokrotny) wzrost czasu wykonania. Jest to spowodowane procesem odchudzania odcisku, który wykonuje się kilka razy dla każdego ze skanów (aż przestanie dawać jakikolwiek efekt). Warto brać pod uwagę jakość kodu przy implementacji takich projektów, ponieważ nieoptymalne zaprojektowanie nawet jednego, prostego procesu może przynieść duże straty w czasie wykonania algorytmu. Odchudzanie obrazu, chociaż konieczne do wykrycia minucji, znacznie spowolniło pojedyncze wykonanie. Konsekwencją zysku skuteczności był duży wzrost czasu wykonania.

Ogólna skuteczność procentowa rzędu 80% to zadowalający wynik, należy pamiętać jednak, że zdecydowana większość przypadków testowych to przypadki odmowy dostępu (jest dużo więcej danych do porównania). Przy utrzymaniu tego współczynnika na poziomie 90% udało się uzyskać maksymalnie 60% skuteczności w przypadku prawidłowej autoryzacji. Wynik – jak na prosty projekt opublikowany w Internecie – jest zadowalający, jednak dla komercyjnego rozwiązania z pewnością potrzebna byłaby dużo większa skuteczność. Kolejnym etapem badań w tym kierunku oraz sposobem na ulepszenie algorytmu powinna być modyfikacja sposobu porównania. Wykorzystanie prostego mechanizmu na bazie deskryptorów obrazów jest bardzo czytelne i optymalne, jednak nie daje stuprocentowej skuteczności. Przegląd literatury dowiódł, że w tej dziedzinie częściej stosowane są inne rozwiązania, które opisano we wcześniejszych fragmentach artykułu.

4. Rozpoznanie mówcy

Innym, coraz popularniejszym sposobem autoryzacji z wykorzystaniem cech budowy organizmu

człowieka jest rozpoznanie osoby mówiącej - biometria głosowa. Metoda automatycznego rozpoznawania mówcy bazuje przede wszystkim na wykorzystaniu cech behawioralnych mówcy, które z biegiem lat zostały nabyte przez mówcę (zostały świadomie wyuczone lub nabyte nieświadomie) [16].

Główną zasadą obowiązująca w biometryce jest obowiązek, aby charakterystyka behawioralna była unikatowa i uniwersalna, trwała i prosta do ilościowej oceny- tylko w takim przypadku wspomniana charakterystyka będzie zapewniała wymagany stopień złożoności konieczny do uwidocznienia wymaganych zróżnicowań cech indywidualnych mówcy [16].

Głos każdej osoby jest odmienny, ponieważ głos jest zależny od fundamentalnych cech fizycznych, które mają ogromny wpływ na rodzaj powstałych sygnałów dźwiękowych, generowanych przez człowieka podczas mówienia [17].

Każdy mówca ma wpływ na wiele czynników, które z kolei stanowią podstawę do wygenerowania głosu. Są to m.in atrybuty [17]:

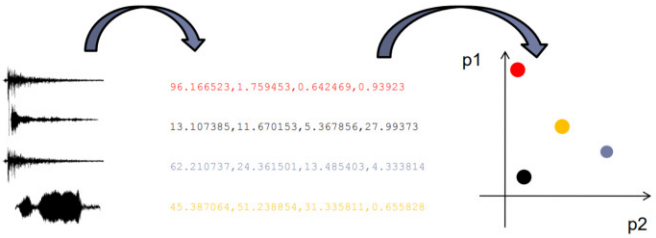
- częstotliwość podstawowa (zależy m.in od długości kanału głosowego) – zwana wysokością tonu. Dla mężczyzn zakres częstotliwości podstawowej wynosi 100-120Hz, dla kobiet 200-250Hz [18],
- dźwięk nosowy – dźwięk mowy, w którym strumień powietrza przechodzi przez nos wskutek obniżenia podniebienia miękkiego z tyłu jamy ustnej. Podczas wymawiania spółgłosek nosowych, usta są w pewnym momencie zamknięte (przez wargi lub język), a strumień powietrza jest całkowicie wydalany przez nos [19],
- kadencja (ton opadający) – następstwo dźwięków nadające frazie charakter zakończenia [20],
- przegięcie – sposób mówienia, w którym głośność lub wysokość tonu jest modyfikowana [21].

Poziom skuteczności systemu, który zajmuje się rozpoznawaniem mówcy jest podporządkowany temu, w jakim stopniu badane parametry fizyczne sygnału mowy odpowiadały będą za przenoszenie cech osobniczych mówcy [22].

W celu rozpoznania osoby, w biometryce głosowej wykorzystywane są następujące parametry i wielkości [22]:

- parametry wyznaczone bezpośrednio z przebiegu czasowego: względne długości czasu wypowiedzi poszczególnych elementów fonetycznych; obwiednia czasowa amplitudy dźwięku; parametry analizy przejść przez zero sygnału mowy; rozkład interwałów czasowych;
- parametry wyznaczone z widma sygnału mowy: uśrednione widmo amplitudowe; widmo mocy; częstotliwość podstawowa tonu krtaniowego; częstotliwości, stosunki amplitudowe oraz szerokości pasm formantów; widmo krótkoterminowe; momenty widmowe
- parametry liniowego kodowania predykcyjnego;
- inne, jak np. charakterystyki prozodyczne, parametry cepstralne.

Proces parametryzacji jest jednym z podstawowych etapów rozpoznawania mowy [23]. Dzięki zastosowaniu procesu parametryzacji, a więc wykorzystaniu parametrów i ich przeanalizowaniu system pozwala na zauważenie różnic, o istnieniu, których człowiek mógł nie zdawać sobie sprawy [24].

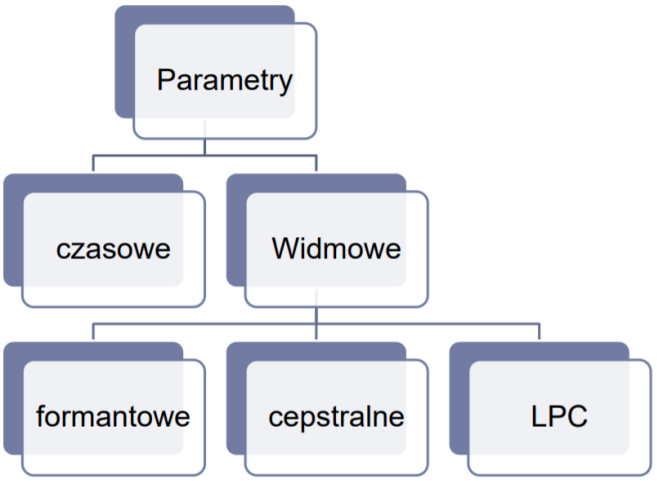


Rys. 4. Ogólny schemat parametryzacji [25]

Celem parametryzacji jest [24]:

- umożliwienie rozróżniania obiektów różnych klas,
- rozpoznanie obiektu nieznanej klasy,
- zweryfikowanie czy dany obiekt należy do danej klasy.

Dokonano następującej klasyfikacji parametrów [24]:



Rys. 5. Schemat klasyfikacji parametrów [26]

Metody czasowe z powodu zbyt wysokiego rozproszenia informacji użytecznych nie są wykorzystywane jako efektywny sposób opisu mowy. Z kolei najczęściej wykorzystywanymi metodami są metody widmowe [27].

Najczęściej stosowanymi metodami parametryzacji sygnału mowy są [16,23,24,27]:

- współczynniki predykcji liniowej (LPC)
- współczynniki analizy cepstralnej w skali mel (MFCC)

Zaletami LPC jest wysoka precyzja estymacji parametrów mowy i szybkie działanie [28].

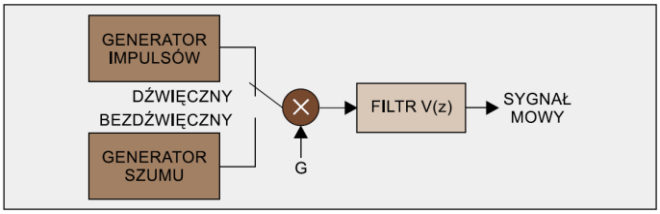
Pierwsze wykorzystanie LPC notowane jest na 1966 rok. Metody tej użyli Saito i Itakura [29].

MFCC pierwszy raz opisano w artykule autorstwa Davisa i Mermelsteina. Artykuł powstał w 1980 roku, a więc 14 lat później niż pierwsze wykorzystanie LPC [30].

MFCC opiera się na metodzie parametryzacji mowy, która wykorzystuje analizę podpasmową sygnału poprzez filtry pasmowo-przepustowe, które są rozłożone równomiernie na melowej skali częstotliwości [30].

Współczynniki predykcji liniowej (z ang. Linear Prediction Coefficients - LPC). W metodzie, która

wykorzystuje współczynniki predykcji liniowej dokonywane jest założenie, iż mowa jest to sygnał, który powstał w wyniku splotu danego pobudzenia i wolnozmiennego filtru, który został skojarzony z transmitancją toru głosowego człowieka [22].



Rys. 6. Uogólniony schemat powstawania sygnału mowy [31]

Na przedstawionym schemacie symbolem G oznaczono pobudzenie generowane przez krtań oraz płuca. Filtr V(z) odpowiedzialny jest za modelowanie traktu głosowego człowieka (dotyczy jamy ustnej, głosowej i gardłowej).

Podczas mówienia człowiek generuje głoski. Wytwarzane przez człowieka głoski podzielono na dźwięczne i bezdźwięczne. W przypadku głosek bezdźwięcznych filtr pobudzany będzie sygnałem szumowym, natomiast dla głosek dźwięcznych filtr będzie pobudzany ciągiem impulsów.

Ciąg impulsów generuje pewną częstotliwość, którą nazwano *częstotliwością tonu krtaniowego*. Częstotliwość tonu krtaniowego definiuje częstotliwość drgania strun głosowych człowieka. Z powodu odmienności wspominanego parametru dla każdego mówcy, parametr ten może być wykorzystywany do rozpoznawania głosu [22].

Jeśli zostanie ustalone założenie upraszczające (dla głosek nosowych) to filtr, który jest odpowiedzialny za modelowanie traktu głosowego może zostać przedstawiony w postaci modelu 'same bieguny'. Przy takim założeniu wspominany filtr będzie filtrem o nieskończonej odpowiedzi impulsowej.

Finalnie, aby obliczyć transmitancję H(z) omawianego modelu należy skorzystać ze wzoru [22]:

$$H(z)=G(z)V(z)\frac{G}{A(z)}=\frac{G}{1-\sum_{k=1}^{p}a_k z^{-k}} \tag{3}$$

gdzie: H(z) – transmitancja modelu, G(z) – pobudzenie, V(z) – filtr, a_k – współczynniki predykcji, p – rząd predykcji, G – intensywność pobudzenia traktu głosowego.

Przedstawiony powyżej wzór opisuje otrzymany sygnał mowy. Aby opisany sygnał mowy stanowił wiarygodny model mówcy, należy nieustannie aktualizować utworzony model. Głos człowieka jest zmienny, natomiast sygnał mowy człowieka w pewnych, krótkich przedziałach 10-20 [ms] jest stacjonarny. Stacjonarność sygnału mowy człowieka powoduje, iż właśnie w krótkich przedziałach współczynniki omawianego modelu H(z) nie podlegają zmianie.

W celu aktualizacji modelu mówcy korzysta się z metody liniowej predykcji sygnału mowy. Metoda ta pozwala na obliczenie parametrów konkretnego modelu dla kolejnych ramek sygnału (bloków dźwięku o określonej długości).

We wspominanej metodzie zakłada się, iż próbkę sygnału mowy można uzyskać wskutek kombinacji liniowej p poprzednich próbek. Można rzec, iż metoda LPC pozwala na 'przewidzenie' wartości sygnału mowy na podstawie p poprzednich wartości sygnału mowy [22].

Wartość rzeczywista, która opisuje próbkę sygnału mowy jest inna niż wartość próbki, ponieważ w metodzie LPC nie jest brany pod uwagę wpływ sygnału pobudzenia. Wskutek pominięcia wpływu sygnału pobudzenia pojawia się 'błąd', który stanowi różnicę pomiędzy dwiema wartościami: rzeczywistą oraz przewidzianą.

Ideą LPC jest znalezienie zbioru współczynników predykcji. Współczynniki predykcji można otrzymać podczas minimalizowania średniokwadratowej wartości błędu E dla całej ramki. Do obliczenia błędu można użyć metody autokorelacji [22].

Po dokonaniu obliczeń pochodnych błędu względem kolejnych współczynników predykcji, a następnie po przyrównaniu ich do zera otrzymujemy ciąg równań danych przedstawiony w postaci iloczynu macierzy [22]. Następnym krokiem jest rozwiązanie układu równań liniowych, gdzie wynikiem jest wyznaczona wartość a_k

Gdy wszystkie parametry LPC dla wszystkich ramek zostaną obliczone to efektem jest ciąg wielowymiarowych wektorów (każdy wektor to p współczynników a_k) [22].

Otrzymane w ten sposób wektory finalnie wykorzystywane są do obliczenia podobieństwa pomiędzy analizowaną wypowiedzią mówcy, a zarejestrowanym modelem mówcy. Analizowana wypowiedź mówcy zostaje opisana za pomocą sekwencji wektorów testowych [22].

Współczynniki mel-cepstralne (z ang. Mel-frequency Cepstral Coefficents - MFCC) są parametrami, które wykorzystuje się w akustyce mowy i kompresji sygnałów fonicznych. Parametry mel-cepstralne powstają z cepstrum sygnału prezentowanego w skali melowej [24].

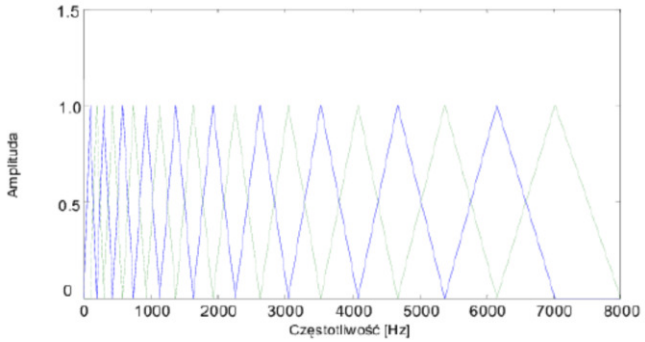
Podstawowa różnica pomiędzy MFCC, a rzeczywistym cepstrum sygnału jest taka, iż w MFCC wykorzystywane jest nieliniowane skalowanie częstotliwości. Do wspomnianego skalowania częstotliwości wykorzystywana jest skala mel [23].

Dla porównania poniżej przedstawiono zależność występującą pomiędzy częstotliwością w skali mel, a częstotliwością opisaną przy pomocy Hz [23]

$$f_{mel}(f_{Hz})=2595\log_{10}\left(1+\frac{f_{Hz}}{700}\right) \tag{4}$$

gdzie: f_{mel} – częstotliwość w skali mel, f_{Hz} – częstotliwość w skali Hz.

Skala melowa uzyskiwana jest wskutek filtracji sygnału bankiem filtrów o charakterystyce trójkątnej. K-ty współczynnik mel-cepstralny odpowiada zawartości k-tego pasma. Liczba pasm mieści się w przedziale od 12 do 20 [24].



Rys. 7. Filtracja sygnału bankiem filtrów o charakterystyce trójkątnej [32]

Obliczanie współczynników MFCC zostało podzielone na kilka kroków [23].

Krok 1: Poddanie otrzymanego sygnału mowy procesowi preemfazy. We wspomnianym procesie dokonywana jest filtracja formująca. Podczas filtracji formującej następuje osłabienie składowych (dotyczy tylko składowych o małych częstotliwościach), składowe o wysokich częstotliwościach zostają wzmocnione [23].

Wzór preemfazy określono jako:

$$x'_n = x_n - a x_{n-1} \tag{5}$$

gdzie: x_n - sygnał przed procesem preemfazy, x'_n - sygnał po procesie preemfazy, a - wartość zazwyczaj ustalana jako 0,97.

Krok 2: Następnie dokonywany jest proces ramkowania sygnału. Nazwa procesu wywodzi się z faktu, iż w omawianym kroku dokonywane jest podzielenie sygnału na krótkie fragmenty. Fragmenty te nazwano ramkami (z ang. frames). Dopuszcza się, aby kolejne ramki nakładały się na siebie [23].

Krok 3: Gdy ramkowanie sygnału zostało wykonane, rozpoczyna się proces okienkowania (z ang. windowning). Okienkowanie wykonywane jest przy użyciu tzw. okna Hamminga [23].

$$Ham(N) = 0{,}54 - 0{,}46\cos\left(2\pi\frac{n-1}{N-1}\right) \tag{6}$$

gdzie: Ham(N) – okno Hamminga, N – długość ramki, n = 1, 2,..., N.

Krok 4: Na wszystkich ramkach przeprowadzany zostaje proces szybkiej transformacji Fouriera (FFT). W zależności od zaprogramowanego algorytmu, uzyskuje się widmo mocy $|FFT|^2$ albo widmo amplitudowe |FTT| [23].

Krok 5: Do zestawu filtrów H_m wprowadzone zostają widma opisane w kroku 4. Liczba filtrów jest zróżnicowana, a definiuje ją sposób zaprogramowania algorytmu. Najpopularniejszymi filtrami są filtry trójkątne.

Efektem zastosowania filtrów jest wygenerowanie na wyjściu każdego z filtrów energii pasma S_m.

Krok 6: Za pomocą logarytmu energii dokonywane jest osłabienie czułości filtrów na bardzo głośne oraz bardzo ciche dźwięki. Przeprowadzony również zostaje proces modelowania nieliniowej amplitudowej wrażliwości ucha ludzkiego.

Dzięki zastosowaniu logarytmu energii polepsza się jakość rozpoznawania [23].

Krok 7: W ostatnim kroku przeprowadzana zostaje dyskretna transformata kosinusowa (z ang. Discrete Cosine Transform - DCT).

Końcowe wartości współczynników MFCC określone zostały za pomocą wzoru [23]:

$$c_i = \sqrt{\frac{2}{M}}\sum_{m=1}^{M}\log(S_m)\cos\left(\frac{\pi i}{M}\right)(m-0{,}5) \tag{7}$$

gdzie: c_i – wartość współczynnika MFCC, i – numer współczynnika (zazwyczaj wartość i>1), M – liczba użytych filtrów, S_m – energia pasma, m – numer filtra.

5. Badania skuteczności

W ramach badania skuteczności wybrano dwa zestawy próbek głosu – treningowy i testowy. Do testu wybrano nagrania dziesięciu mówców – po 30 w zestawie treningowym i po 10 w testowym. Sumarycznie – 400 próbek głosu – średnio każda o długości ok. 2s. W procesie trenowania algorytmu wszystkie z 300 próbek zostało przetworzonych pod kątem cech MFCC i LPC oraz stworzono na ich podstawie książki kodowe. W ramach testu dla każdej ze 100 próbek testowych wyliczono dystans euklidesowy w stosunku do każdego z wpisów w książce kodowej. Wpis z najmniejszą odległością od testowanej próbki traktowany był jako pasujący. Każda z próbek oznaczona była kodem mówcy, dzięki któremu można było jednoznacznie identyfikować autora wypowiedzi. Jeśli w trakcie testu dopasowano wytrenowaną próbkę tego samego autora – zaliczano test jako udany. Na podstawie ilości prawidłowych dopasowań (osobno dla ekstrakcji cech MFCC i LPC) oraz ilości wszystkich prób wyliczono procentowe współczynniki skuteczności.

Poniżej przedstawiono wynik badania skuteczności algorytmu porównującego próbki głosowe:

Tabela 2. Wynik badania skuteczności algorytmu porównującego próbki głosowe

Parametr	MFCC	LPC
Skuteczność procentowa	34%	61%

Zdecydowanie większą skuteczność można zauważyć w przypadku porównania cech ekstrahowanych z użyciem algorytmu LPC. Skuteczność weryfikacji mówców przy użyciu algorytmu LPC wyniosła 61%, tym samym osiągając wynik lepszy o 27% aniżeli skuteczność algorytmu MFCC. Biorąc pod uwagę fakt, iż metoda LPC osiągnęła wynik 61%, a MFCC 34% oznacza to, iż LPC jest prawie dwa razy skuteczniejszą metodą weryfikacji mówców. W przypadku próby wdrożeń wybranego algorytmu do danych rozwiązań technologicznych z pewnością lepszym rozwiązaniem będzie wdrożenie algorytmu LPC. Pod uwagę należy wziąć fakt, iż skuteczność weryfikacji mówcy przy użyciu metod biometrycznych na poziomie 61% może nie być wystarczająca do zastosowań na szeroką skalę. Kolejnym kierunkiem badań w celu usprawnienia działania rozpoznawania mówcy powinna być analiza algorytmu LPC pod kątem rodzaju analizowanych próbek – jakości dźwięku, zaszumienia otoczenia, płci mówcy

czy wieku. Warto również sprawdzić jak ten algorytm zachowa się w stosunku do różnej liczby próbek treningowych oraz testowych.

6. Podsumowanie

Przeprowadzono badania, w których określono skuteczność obu z tych rodzajów autoryzacji. Fakt dużego wpływu procesu ekstrakcji parametrów z próbek danych dowodzą badania, w których ten sam sposób porównania próbki zestawiono w stosunku do dwóch różnych repozytoriów (badania autoryzacji odcisku palca). Wyniki znacznie się różnią – zależnie od sposobu analizy próbki i ekstrakcji danych. Dużo większą skutecznością może pochwalić się algorytm, który wykrywał minucje w liniach papilarnych. Znacznie niższe wyniki zaobserwowano w przypadku podania punktów kluczowych z detektora Harrisa do deskryptora obrazu. Warto również zauważyć fakt, że sam mechanizm porównania odcisków bazujący na deskryptorach daje wysokie, lecz nie stuprocentowe wyniki. Kolejnym kierunkiem badań w tym kontekście powinna być modyfikacja algorytmu porównania modeli próbek, uzyskanych w wyniku ekstrahowania cech. Dzięki badaniom wiadomo, że dużo większą skuteczność da proces ekstrakcji bazujący na wykrywaniu minucji.

W przypadku autoryzacji głosowej na podstawie badań udowodniono, że dużo lepszą skuteczność autoryzacji daje algorytm LPC – jego wynik był dwukrotnie lepszy niż w przypadku MFCC. Procentowa skuteczność autoryzacji jest na zadowalającym poziomie, lecz w celu wykorzystania tego rozwiązania w produkcyjnym środowisku należałoby ją znacznie zwiększyć. Kolejnym kierunkiem badań w tym aspekcie powinna być analiza algorytmu pod kątem trenowanych i badanych próbek – ich jakości oraz cech dotyczących mówcy (takich jak wiek, płeć itd.).

Literatura

[1] Thai R. - Fingerprint ImageEnhancement and Minutiae Extraction - https://www.peterkovesi.com/studentprojects/raymondthai/RaymondThai.pdf - 8.06.2019

[2] Dokumentacja OpenCV - adaptacyjna normalizacja histogramu https://docs.opencv.org/3.1.0/d5/daf/tutorial_py_histogram_equalization.html - 28.05.2019

[3] Repozytorium python-fingerprint-recognition https://github.com/kjanko/python-fingerprint-recognition - 28.05.2019

[4] TheAILearner - Adaptive Histogram Equalization (AHE) - https://theailearner.com/2019/04/14/adaptive-histogram-equalization-ahe/ - 08.06.2019

[5] WaveMetrics - progowanie obrazu https://www.wavemetrics.com/products/igorpro/imageprocessing/thresholding - 28.05.2019

[6] Repozytorium biometrics - GitHub https://github.com/rtshadow/biometrics - 28.05.2019

[7] Shah A. - Through The Eyes of Gabor Filter https://medium.com/@anuj_shah/through-the-eyes-of-gabor-filter-17d1fdb3ac97 - 28.05.2019

[8] Błaszczyk Ł. - Filtry Gabora i ich zastosowanie w obrazowaniu medycznym http://pages.mini.pw.edu.pl/~blaszczykl/nauka/prace/inz_biomed.pdf - 28.05.2019

[9] Vector magnitude and direction review - https://www.khanacademy.org/math/precalculus/vectors-precalc/component-form-of-vectors/a/vector-magnitude-and-direction-review - 08.06.2019

[10] Bazy odcisków FVC2004 http://bias.csr.unibo.it/fvc2004/databases.asp - 28.05.2019

[11] Khanyile N.P., Tapamo J.R., Dube E. - A Comparative Study of Fingerprint Thinning Algorithms https://pdfs.semanticscholar.org/f0e3/947dae712e8c27ee297e2ef569a369dcccbe.pdf - 28.05.2019

[12] Gońda S., Juszczak D. - Identyfikacja osób poprzez ich odciski palców http://sequoia.ict.pwr.wroc.pl/~witold/aiarr/2009_projekty/odciski/ - 28.05.2019

[13] Więcław Ł.- A minutiae-based matching algorithms in fingerprint recognition systems https://www.researchgate.net/publication/228644313_A_minutiae-based_matching_algorithms_in_fingerprint_recognition_systems - 28.05.2019

[14] Chaudhari A. S., Dr. Girish K. Patnaik, Patil S. S. - Implementation of Minutiae Based Fingerprint Identification Systemusing Crossing Number Concept https://pdfs.semanticscholar.org/49ad/2473ecc1dcbfdfc9a981f5191a1224107ef1.pdf - 28.05.2019

[15] Paulino A. A., Feng J., Jain - Latent A. K. - Fingerprint Matching using Descriptor-Based Hough Transform http://biometrics.cse.msu.edu/Publications/Fingerprint/PaulinoFengJain_LatentFPMatching_DescriptorBasedHoughTransform_IJCB11.pdf - 28.05.2019

[16] Ślot K., Wybrane zagadnienia biometrii, Wydawnictwa Komunikacji i Łączności, Warszawa, 2008

[17] Bolle R.M., Connell J. H., Pankanti S., Ratha N. K., Senior A. W., - Biometria, Wydawnictwa Naukowo-Techniczne, Warszawa, 2008

[18] Zrozumiałość mowy - https://livesound.pl/tutoriale/4629-zrozumialosc-mowy - 28.05.2019

[19] Nasal speech sound - https://www.britannica.com/topic/nasal-speech-sound - 28.05.2019

[20] Termin ‘kadencja’ - https://sjp.pl/kadencja - dostęp 28.05.2019

[21] Termin ‘inflection’ - https://www.vocabulary.com/dictionary/inflection , [28.05.2019]

[22] Dustor A., Izydorczyk J., Rozpoznawanie mówców - http://www.przegladtelekomunikacyjny.pl/archive/WWW/artrec/dustor2-3'2003.pdf - 28.05.2019

[23] Kacprzak S., Inteligentne metody rozpoznawania dźwięku, Politechnika Łódzka, Łódź, 2010 http://www.dsp.agh.edu.pl/_media/pl:homepage:msc_s_kacprzak.pdf - dostęp 28.05.2019

[24] Parametryzacja sygnału mowy. Perceptualne skale częstotliwości - https://sound.eti.pg.gda.pl/student/amowy/am_04_parametryzacja.pdf - 28.05.2019

[25] Ogólny schemat parametryzacji [rys.] https://sound.eti.pg.gda.pl/student/amowy/AM_04_parametryzacja.pdf - 28.05.2019

[26] Schemat klasyfikacji parametrów [rys.] https://sound.eti.pg.gda.pl/student/amowy/AM_04_parametryzacja.pdf - 28.05.2019

[27] Gałka J., Optymalizacja parametryzacji sygnału w aspekcie rozpoznawania mowy polskiej, Akademia Górniczo–Hutnicza im. Stanisława Staszica w Krakowie, Kraków, 2008 https://docplayer.pl/24752866-Optymalizacja-parametryzacji-

sygnalu-w-aspekcie-rozpoznawania-mowy-polskiej.html - 28.05.2019

[28] Rabiner L. R., Schafer R. W. - Digital procesing of speech signal. Prentice Hall, Englewood Cliffs, NJ, 1978.

[29] Basztura C. - Źródła, sygnały i obrazy akustyczne. Wydawnictwo Komunikacji i Łacznosci, Warszawa, 1988.

[30] Davis S., Mermelstein P. - Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. Acoustics, Speech and Signal Processing, IEEE Transactions on, Sier. 1980.

[31] Uogólniony schemat powstawania sygnału mowy [rys.] http://www.przegladtelekomunikacyjny.pl/archive/WWW/artrec/dustor2-3'2003.pdf - 28.05.2019

[32] Filtracja sygnału bankiem filtrów o charakterystyce trójkątnej [rys.] https://sound.eti.pg.gda.pl/student/amowy/AM_04_parametryzacja.pdf - 28.05.2019

[33] Wanat I., Iwaniec M., Tworzenie modelu akustycznego na potrzeby weryfikacji mówcy przy użyciu Ukrytych Modeli Markowa - http://www.kms.polsl.pl/mi/pelne_9/31.pdf - 28.05.2019