

UNIwersYTET RZESZOWSKI
WYDZIAŁ NAUK ŚCISŁYCH I TECHNICZNYCH
INSTYTUT INFORMATYKI



Julia Chmura, Dominik Dziadosz
jc131414, dd131428

Informatyka

Projekt zaliczeniowy - "Weryfikacja głosowa na stronie internetowej"

Projekt
Grupa: 02

Prowadzący
dr Zbigniew Gomółka

Rzeszów 2025

Spis treści

1. Informacje ogólne	6
1.1. Opis ogólny projektu	6
1.2. Wstęp teoretyczny	6
1.2.1. Rejestracja.....	6
1.2.2. Weryfikacja	7
1.3. Opis wykorzystywanego oprogramowania i narzędzi	8
1.4. Działanie programu	8
1.5. Instrukcja nagrywania próbki głosu	10
1.6. Wnioski.....	10
1.7. Bibliografia	11
Bibliografia	12
Spis rysunków	13
Spis listingów	14
Oświadczenie studenta o samodzielności pracy	15

1. Informacje ogólne

Biometryczne Systemy Zabezpieczeń	
Projekt	
Temat	Weryfikacja głosowa na stronie internetowej
Technologie	Python, JavaScript, framework - Flask, narzędzie FFmpeg
Biometryka	Voice
Autorzy	Julia Chmura, Dominik Dziadosz
Grupa laboratoryjna	02

1.1. Opis ogólny projektu

Celem projektu jest przedstawienie komponentu biometrycznego - voice. Głos człowieka może być zastosowany jako unikalne zabezpieczenie (np. podczas logowania, lub innego zabezpieczenia zasobów). Komponent ten został zaimplementowany na stronie internetowej jako zabezpieczenie konta użytkownika. Użytkownik ma możliwość nagrać próbkę swojego głosu z wypowiadając daną frazę, a następnie przechodząc do logowania nagrywa ponownie próbkę głosu z tą samą frazą, po czym następuje weryfikacja. Obsługa komponentu biometrycznego została napisana w pythonie z użyciem odpowiednich bibliotek, a prosta witryna internetowa do jego wykorzystania została napisana we frameworku pythona - Flask.

1.2. Wstęp teoretyczny

Głos jest jednym z wyjątkowych i bardzo ciekawym narzędziem, które może mieć bardzo szeroki zakres zastosowań w zabezpieczeniach biometrycznych. Łączy on w sobie dwa aspekty, które wpływają na jego unikalność:

- Cechy fizyczne głosu - w tym punkcie zawarte są na przykład kształt i rozmiar traktu głosowego, strun głosowych i płuc. Podobnie jak inne części ciała wykorzystywane w biometrii (np. tęczówka, linie papilarne), te cechy głosu są również dla każdego człowieka unikalne.
- Cechy behawioralne - na te cechy wpływa akcent, tonacja, rytm czy tempo. Są to cechy które wykształciły się na przestrzeni lat u każdej osoby.

W projekcie wykorzystywana jest kombinacja obu tych aspektów, aby stworzyć koncepcję systemu, który mógłby zostać częściej wykorzystywany jako główne zabezpieczenie zasobów.

Projekt ten działa w dwóch fazach:

1.2.1. Rejestracja

W tej fazie system najpierw potrzebuje wzorca biometrycznego, który będzie później wykorzystywany w głównej fazie projektu.

Pobranie danych

Użytkownik wypowiada swoje hasło (frazę potrzebną później do weryfikacji). Mikrofon w komputerze przetwarza fale ciśnienia akustycznego na sygnał elektryczny. Następnie przeglądarka przy pomocy skryptu (JavaScript) digitalizuje go i koduje na format **.webm**

Przetwarzanie wstępne

W tym kroku surowy plik **.webm** jest przesyłany do analizy. Wykorzystywana jest biblioteka **pydub**, aby przekształcić ten plik do formatu **.wav** z częstotliwością 16000Hz i kanałem mono. Jest to działanie istotne by zachować spójność.

Wydobycie cech biometrycznych

Jest to najważniejszy krok w projekcie. System nie porównuje bezpośrednio plików dźwiękowych, ponieważ byłoby to nieefektywne. Z tego powodu z sygnału audio wydobywane są jego cechy charakterystyczne. W przypadku tego projektu są to **Współczynniki Cepstralne w Skali Mel (MFCC)**.

Współczynniki Cepstralne w Skali Mel (MFCC) - jest to zbiór kilkunastu liczb które w połączeniu tworzą numeryczną reprezentację barwy dźwięku w krótkim fragmencie czasu.

W programie najpierw następuje wczytanie pliku **.wav** jako tablicy liczb (numpy). Następnie z sygnału usuwana jest cisza, by system skupił się tylko na głównej części czyli wypowiedzianej frazie. Na końcu biblioteka Librosa analizuje sygnał klatka po klatce, w bardzo krótkich fragmentach i dla każdej z nich oblicza 13 współczynników MFCC. Liczby te opisują kształt widma, czyli barwę dźwięku w sposób, który jest odporny na zmiany głośności i tonu.

Efektom tych działań jest wydobyty wzorzec biometryczny - macierz cech. Jest to tabela liczb, w której wiersze to kolejne ramy czasowe a kolumny to współczynniki MFCC.

Zapisanie wzorca

W tym projekcie dla uproszczenia zapisujemy cały plik **.wav**. Za każdym razem gdy system potrzebuje wzorca, cechy są z niego wydobywane na nowo

1.2.2. Weryfikacja

W tej fazie następuje weryfikacja, czy nowa próbka głosu jest podobna do zapisanego wzorca.

Dla próbki do weryfikacji punkty: pobranie danych, przetwarzanie wstępne i wydobycie cech biometrycznych są powtarzane.

Teraz system mając dwie macierze cech z obu próbek musi je porównać. Na tym etapie jednak pojawia się pewien problem, gdyż użytkownik, mimo że będzie to ta sama osoba, nigdy nie powtórzy idealnie i w tym samym tempie wypowiedzianej frazy. By rozwiązać ten problem w projekcie został wykorzystany algorytm **Dynamicznego Dopasowania Czasu - DTW**.

Algorytm Dynamicznego Dopasowania Czasu - DTW - jest to metoda służąca porównaniu dwóch sekwencji (w tym przypadku czasowych), która potrafi poradzić sobie z różnicami w tempie. Głównym zadaniem tego algorytmu jest znalezienie optymalnego dopasowania między elementami jednej sekwencji a drugiej, by łączna różnica między dopasowanymi elementami była jak najmniejsza.

W pierwszej kolejności tworzona jest mapa kosztów lokalnych. W tym celu wykorzystywana jest funkcja `cdist` z biblioteki `scipy`. Tworzy ona macierz, gdzie każda komórka zawiera informację jak bardzo różni się i -ta ramka z pierwszego nagrania od j -tej ramki z drugiego nagrania. Dzieje się to poprzez obliczanie dystansu euklidesowego między współczynnikami MFCC.

Następnie algorytm DTW przechodzi przez macierz kosztów i znajduje ścieżkę o najmniejszym łącznym koszcie. Im ta wartość jest mniejsza tym głosy są bardziej podobne. Efektem jego działania jest skumulowany dystans. Jest to finalny wynik porównania biometrycznego.

Ostatni punkt to określenie na podstawie prostego warunku logicznego czy obliczona odległość jest wyższa lub niższa od skalibrowanego wcześniej progu decyzyjnego na podstawie którego użytkownik jest zweryfikowany lub nie.

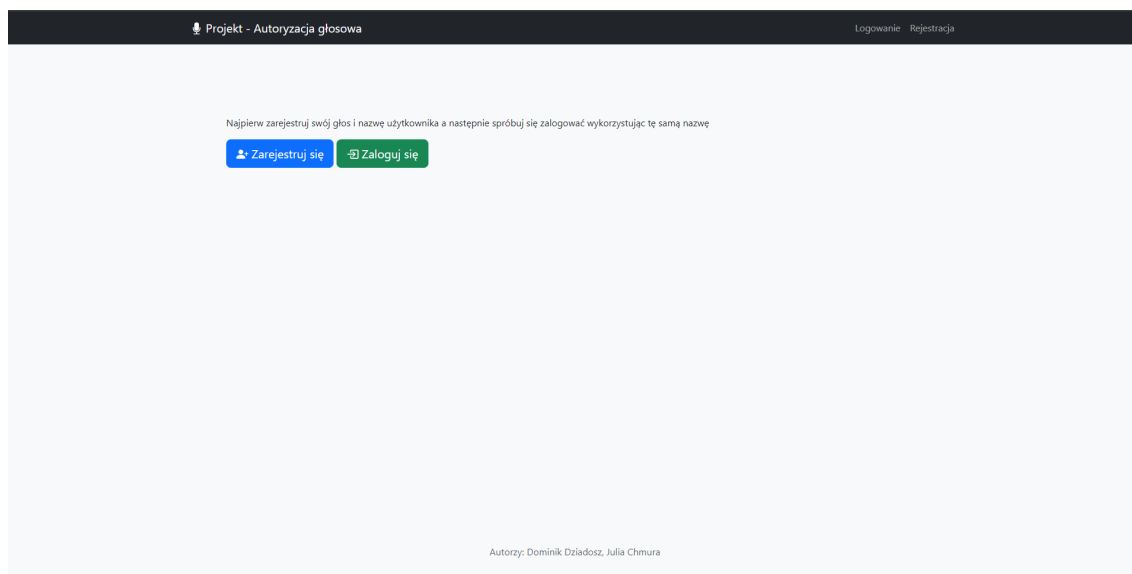
1.3. Opis wykorzystywanego oprogramowania i narzędzi

- **Python** - główny język programowania na którym opiera się cały projekt
- **JavaScript** - język skryptowy, który posłużył do stworzenia systemu nagrywania głosu (dynamicznie działające przyciski, zapisywanie głosu w danym formacie)
- **Flask** - framework Pythona który posłużył do stworzenia szaty graficznej projektu (strony internetowej na której odbywa się weryfikacja głosowa)
- Biblioteki Pythona
 - **numpy** - jest jedną z głównych elementów projektu. Wykorzystana została do stworzenia tablicy jednowymiarowej w której przechowywane są liczby zmiennoprzecinkowe stworzone przez inną bibliotekę `librosa`, reprezentujące głos. Potrzebna jest również do uproszczenia operacji matematycznych oraz pozwala na zarządzanie macierzami zawartymi w kodzie
 - **librosa** - biblioteka ta w projekcie ma szczególnie zadanie, odpowiada ona za cały aspekt analizy dźwięku. Dzięki niej wczytywany jest głos oraz dekodowany, wstępnie oczyszcza dźwięk (np. usuwa sekundy ciszy z nagrania) oraz najważniejsze odpowiada za wydobycie cech biometrycznych z głosu.
 - **SciPy moduł `scipy.spatial.distance`** - biblioteka ta to olbrzymi zbiór gotowych maszyn i narzędzi matematycznych, inżynierskich i naukowych
 - **pydub** - ta biblioteka posłużyła jako narzędzie do manipulacji i przetwarzania plików audio. W kontekście projektu jej główna rola to przyjmowanie plików głosowych, a następnie konwertuje je do formatu WAV.
 - **os** - biblioteka ta posłużyła do obsługi plików i folderów

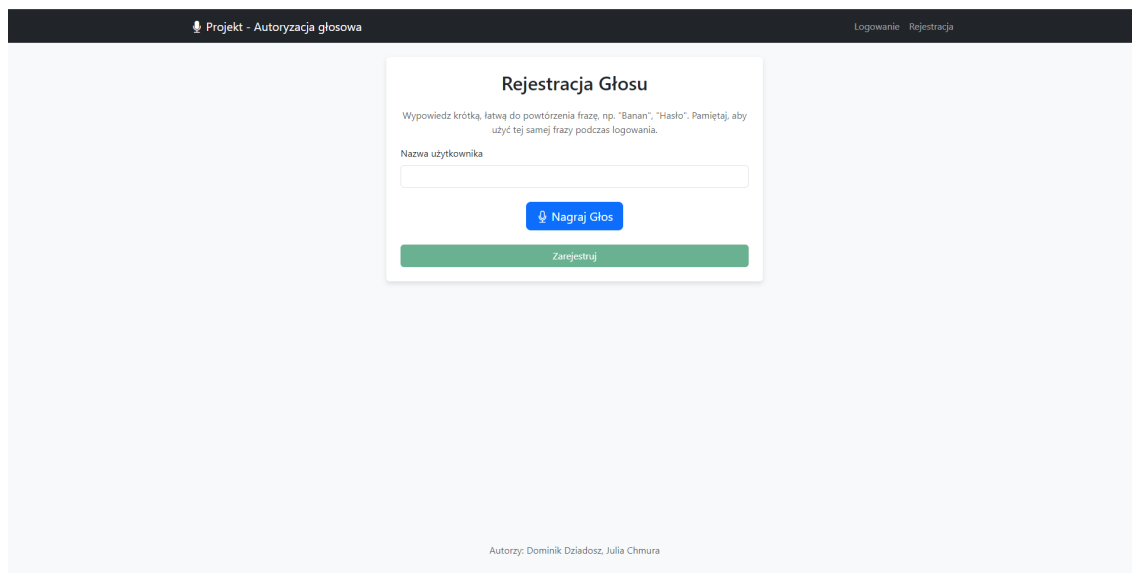
1.4. Działanie programu

Najważniejszą rzeczą przed uruchomieniem projektu jest pobranie i dodanie do zmiennych środowiskowych narzędzia FFMPEG. Jest ono potrzebne do poprawnego działania biblioteki `pydub`.

Działanie projektu zaczyna od strony głównej na której jest rejestracja i logowanie. W pierwszym kroku użytkownik rejestruje się nagrywając próbkę głosu.

**Rys. 1.1.** Obraz strony głównej

Po kliknięciu "Zarejestruj się" przechodzimy do strony rejestracji. Tutaj podawana jest nazwa użytkownika i nagrywana próbka głosu (około 2 do 3 sekundy).

**Rys. 1.2.** Obraz strony rejestracji

Po zarejestrowaniu się należy przejść do logowania. Tutaj ponownie podajemy nazwę użytkownika i nagrywana jest kolejna próbka, która jest porównywana do tej podanej podczas rejestracji. Jeżeli próbki będą zgodne, użytkownik przechodzi poprawnie przez weryfikację lub nie przechodzi, w każdym przypadku zostaje o tym poinformowany stosownym komunikatem.

Projekt - Autoryzacja głosowa Logowanie Rejestracja

Logowanie Głosowe

Nazwa użytkownika

Nagraj Głos

Zaloguj się

Autorzy: Dominik Dziadosz, Julia Chmura

Rys. 1.3. Obraz strony logowania

1.5. Instrukcja nagrywania próbki głosu

Aby poprawnie nagrać próbkę głosu należy spełnić dane warunki:

- Przygotować odpowiedni sprzęt: poprawnie działający mikrofon,
- Mówić jednolitym tonem,
- Wypowiedzieć proste hasło, najlepiej składające się z jednego lub dwóch słów,
- Przy rejestracji i logowaniu posłużyć się tym samym zwrotem, mówiąc w takim samym tonie i tempie,
- Upewnić się że nagrania nie zakłcają odgłosy z otoczenia,
- Nagrać próbkę trwającą około dwie lub trzy sekundy

1.6. Wnioski

Kluczowe wnioski z realizacji projektu to:

- Głos, który charakteryzuje się unikalnymi właściwościami fizycznymi i behawioralnymi, jest na tyle skutecznym i użytecznym elementem biometrycznego zabezpieczenia tożsamości użytkownika. Stanowi on unikalny identyfikator, który jest bardzo trudno podrobić i może okazać się alternatywą dla tradycyjnych metod logowania lub uzupełnieniem.
- Weryfikacja dokonywana na podstawie obliczonego dystansu pomiędzy macierzami cech daje wiarygodny wynik, ponieważ przesłanki wstępne dla tej metody są spełniane. Próbkę głosową którą zostały nagrane w odpowiednich warunkach lub zgodnie z instrukcjami, system użytkownika rozpoznaje jako poprawnie.
- Za pomocą algorytmu DTW doskonale kompensujemy naturalne różnice w tempie i intonacji tej samej frazy wypowiedzianej przez tę samą osobę, co umożliwia precyzyjne porównanie sekwencji głosowych.

- Współczynniki Cepstralne w Skali Mel pozwalają na numeryczne reprezentowanie barwy dźwięku, a my możemy ten ton zareprezentować i porównywać próbki głosowe. To zabezpiecza nas przede wszystkim przed zmiennością głośności i barwy dźwięków.
- Projekt podkreśla znaczenie warunków nagrywania (jednolity ton, brak zakłóceń, stałe tempo i fraza), co jest kluczowe dla poprawnego działania i dokładności weryfikacji biometrycznej.

1.7. Bibliografia

- MATLAB Image Processing Toolbox Documentation [1],
- Rafael Gonzalez, Richard Woods, Digital Image Processing Global Edition [3].
- Flask Documentation. Pallets. [6]
- Pydub Documentation. Read the Docs. [7]
- Librosa Documentation [4]
- SciPy Documentation [9]
- NumPy Documentation [5]
- scikit-learn Documentation [8]
- FFmpeg Documentation [2]

Bibliografia

- [1] Matlab image processing toolbox documentation, 2024. Accessed: 2025-02-15.
- [2] FFmpeg. FFmpeg Documentation. <https://ffmpeg.org/documentation.html>. Accessed: 2025-06-14.
- [3] Rafael Gonzalez and Richard Woods. *Digital Image Processing Global Edition*. Pearson Deutschland, 2017.
- [4] Librosa. Librosa Documentation. <https://librosa.org/doc/latest/index.html>. Accessed: 2025-06-14.
- [5] NumPy. NumPy Documentation. <https://numpy.org/doc/stable/>. Accessed: 2025-06-14.
- [6] Pallets. Flask Documentation. <https://flask.palletsprojects.com/en/latest/>. Accessed: 2025-06-14.
- [7] Read the Docs. Pydub Documentation. <https://pydub.readthedocs.io/en/stable/>. Accessed: 2025-06-14.
- [8] scikit-learn. scikit-learn Documentation. <https://scikit-learn.org/stable/documentation.html>. Accessed: 2025-06-14.
- [9] SciPy. SciPy Documentation. <https://docs.scipy.org/doc/scipy/reference/>. Accessed: 2025-06-14.

Spis rysunków

1.1	Obraz strony głównej	9
1.2	Obraz strony rejestracji	9
1.3	Obraz strony logowania	10

Spis listingów

Załącznik nr 2 do Zarządzenia nr 228/2021 Rektora Uniwersytetu Rzeszowskiego z dnia 1 grudnia 2021 roku w sprawie ustalenia procedury antyplagiatowej w Uniwersytecie Rzeszowskim

OŚWIADCZENIE STUDENTA O SAMODZIELNOŚCI PRACY

.....Julia Chmura, Dominik Dziadosz.....

Imię (imiona) i nazwisko studenta

Wydział Nauk Ścisłych i Technicznych

.....Informatyka.....

Nazwa kierunku

.....jc131414, dd131428.....

Numer albumu

1. Oświadczam, że moja praca projektowa pt.: Projekt zaliczeniowy - "Weryfikacja głosowa na stronie internetowej"
 - 1) została przygotowana przeze mnie samodzielnie*,
 - 2) nie narusza praw autorskich w rozumieniu ustawy z dnia 4 lutego 1994 roku o prawie autorskim i prawach pokrewnych (t.j. Dz.U. z 2021 r., poz. 1062) oraz dóbr osobistych chronionych prawem cywilnym,
 - 3) nie zawiera danych i informacji, które uzyskałem/am w sposób niedozwolony,
 - 4) nie była podstawą otrzymania oceny z innego przedmiotu na uczelni wyższej ani mnie, ani innej osobie.
2. Jednocześnie wyrażam zgodę/nie wyrażam zgody** na udostępnienie mojej pracy projektowej do celów naukowo-badawczych z poszanowaniem przepisów ustawy o prawie autorskim i prawach pokrewnych.

(miejscowość, data)

(czytelny podpis/y studenta/ów)

* Uwzględniając merytoryczny wkład prowadzącego przedmiot

** – niepotrzebne skreślić