# Laboratorium 5

# Emocje w głosie i języku

Informatyka afektywna

dr inż. Mateusz Pomianek

23 września 2026

# Część I
# Część organizacyjna

## 1 Cel i rezultat

Laboratorium prowadzi od poprawnie zdefiniowanego problemu do wyniku, który można zweryfikować i zinterpretować. Rezultatem nie jest sam działający kod: student ma przedstawić dane wejściowe, konfigurację, wynik ilościowy, przypadek błędny oraz ograniczenia wnioskowania.

## 2 Zasady pracy

| Obszar | Wymaganie |
|---|---|
| Raport | Uzupełniaj `evidence/lab05.md`; szablon jest osadzony w PDF. |
| Reprodukowalność | Zapisz wersję danych, Pythona i bibliotek, seed, parametry oraz sposób podziału danych. |
| Dane | Nie modyfikuj danych źródłowych; artefakty i wyniki zapisuj osobno. |
| Walidacja | Nie dobieraj parametrów na podstawie finalnego zbioru testowego. |
| Interpretacja | Rozdziel obserwację, cechę obliczoną, predykcję modelu i interpretację psychologiczną. |
| Niepewność | System powinien dopuszczać wynik niepewny lub `unknown`; nie wymuszaj klasyfikacji bez podstaw. |

## 3 Wymagania wstępne

Python, librosa lub openSMILE, pandas, scikit-learn oraz dane audio i/lub transkrypcje. Przed rozpoczęciem sprawdź uruchomienie środowiska, dostęp do danych i możliwość zapisania artefaktów eksperymentu.

## 4 Punkty kontrolne

| | |
|---|---|
| 3.0 | działający baseline, poprawne wejście/wyjście i udokumentowane środowisko |
| 3.5 | kompletna implementacja zadania głównego i poprawnie dobrana metryka |
| 4.0 | kontrolowany eksperyment, wynik ilościowy i właściwa wizualizacja |
| 4.5 | failure case oraz analiza odporności lub generalizacji |
| 5.0 | ablation, analiza niepewności i krytyczna interpretacja ograniczeń |

# Część II
# Część teoretyczna

## 5 Co właściwie mierzymy?

Sygnał mowy niesie informacje akustyczne, prozodyczne i językowe. Pitch, energia, tempo czy MFCC nie są emocjami, lecz cechami korelującymi z ekspresją w określonych danych. Tekst i głos mogą dostarczać sprzecznych przesłanek; taka rozbieżność jest informacją, a nie błędem wymagającym automatycznego ukrycia.

**Zasada interpretacji**

Pipeline dokumentuj jako: **dane** → **preprocessing** → **reprezentacja** → **model/algorytm** → **wynik** → **interpretacja**. Każde przejście w tym łańcuchu wprowadza założenia i potencjalne źródła błędu.

## 6 Walidacja i generalizacja

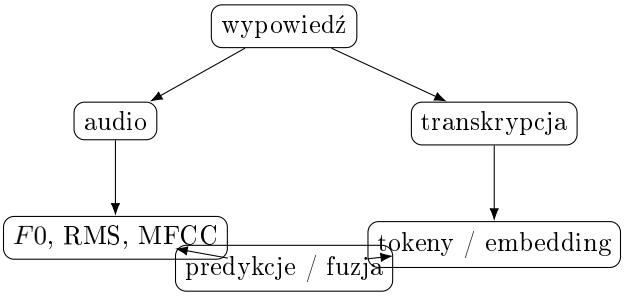
Jeżeli próbki pochodzą od wielu osób, rozróżnij generalizację na nowe próbki tej samej osoby od generalizacji na nową osobę. W drugim przypadku stosuj podział subject-independent, gdy identyfikatory uczestników są dostępne. Raportuj wyniki per klasa lub per warunek, a nie wyłącznie jedną wartość średnią. Zawsze sprawdź niezbalansowanie klas, brakujące dane i możliwość data leakage.

## 7 Niepewność i ground truth

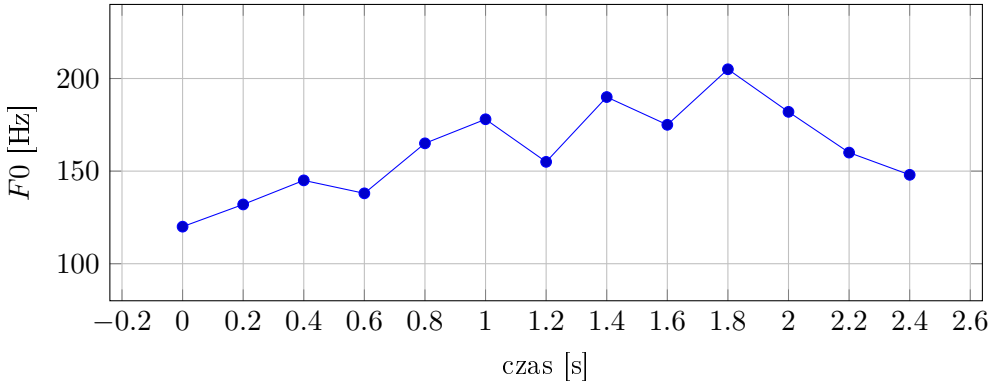
Ground truth w informatyce afektywnej może pochodzić z samoopisu, anotacji obserwatora, etykiety eksperymentalnej albo heurystyki. Te źródła nie są równoważne. W raporcie zapisz, skąd pochodzi etykieta referencyjna, jaka jest jej rozdzielczość i jakie ma ograniczenia. Confidence modelu opisuje zachowanie modelu, a nie pewność co do rzeczywistego stanu psychicznego osoby.

## 8 Diagramy wspomagające interpretację

### 8.1 Dwie ścieżki reprezentacji mowy



### 8.2 Przykład przebiegu $F0$




$F0$ jest cechą sygnału, a nie samodzielnym miernikiem emocji.

## 9 Wzory i miary

Energię krótkookresową sygnału można opisać przez RMS:

$$RMS = \sqrt{\frac{1}{N}\sum_{n=1}^{N} x[n]^2}.$$

Częstotliwość podstawowa $F_0$ opisuje tempo drgań okresowej składowej sygnału, a nie emocję samą w sobie. Dla klasyfikacji stosuj metryki zdefiniowane dla FER, np.

$$F1_{macro} = \frac{1}{C}\sum_{c=1}^{C} F1_c.$$

Przy dwóch modalnościach warto jawnie analizować ich zgodność:

$$A_{modal} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}(\hat{y}_i^{audio} = \hat{y}_i^{text}).$$

# Część III
# Część praktyczna

## 10 Dane wejściowe i oczekiwane artefakty

**Wejście.** Nagrania WAV i/lub transkrypcje z etykietą referencyjną. Jeżeli obie modalności są dostępne, zachowaj wspólny identyfikator próbki.

**Oczekiwany rezultat.** Tabela cech akustycznych/tekstowych, wyniki klasyfikacji, macierz pomyłek i lista próbek, dla których modalności dają różne decyzje.

**Minimalny kontrakt danych**

Przed implementacją sprawdź nazwy pól, jednostki, zakresy wartości, braki oraz identyfikator osoby/próbki. Jeżeli format danych różni się od przykładu, dostosuj kod jawnie i opisz zmianę w raporcie.

## 11 Przykład startowy w Pythonie

Poniższy listing jest punktem startowym, a nie gotowym rozwiązaniem eksperymentu. Student ma go dostosować do rzeczywistego formatu danych i udokumentować zmiany.

```
import librosa
import numpy as np

y, sr = librosa.load("data/sample.wav", sr=None)
rms = float(librosa.feature.rms(y=y).mean())
mfcc = librosa.feature.mfcc(y=y, sr=sr, n_mfcc=13).mean(axis=1)
f0 = librosa.yin(y, fmin=70, fmax=400, sr=sr)

features = np.r_[rms, np.nanmedian(f0), mfcc]
print("feature vector:", features.shape)
```

## 12 Jak czytać wynik

Najpierw sprawdź poprawność techniczną: zakres wartości, braki, liczbę próbek i zgodność wymiarów. Następnie oceń wynik ilościowo. Dopiero na końcu formułuj interpretację afektywną. Nie wyciągaj wniosku psychologicznego wyłącznie z pojedynczej predykcji lub atrakcyjnej wizualizacji.

## 13 1. Przygotowanie eksperymentu

Utwórz katalog wyników i zanotuj wersje środowiska. Ustal pytanie badawcze w formie, na którą można odpowiedzieć pomiarem. Zapisz hipotezę, zmienną niezależną, zmienną zależną oraz kryterium poprawności. Przed analizą sprawdź strukturę danych, braki i etykiety.

## 14 2. Baseline

Uruchom najprostszy wariant rozwiązania, który stanowi punkt odniesienia. Zapisz wejście, parametry i wynik do `evidence/lab05.md` → **Baseline**. Baseline powinien być prosty i możliwy do ponownego uruchomienia bez ręcznych kroków.

**Punkt kontrolny 3.0**

Pokaż działający baseline oraz wskaż dokładnie dane wejściowe, wynik i konfigurację używaną do jego uzyskania.

## 15 3. Zadanie główne

Przygotuj reprezentację akustyczną i tekstową. Dla audio wyznacz wybrane cechy, np. F0, RMS, tempo i MFCC; dla tekstu zbuduj prostą reprezentację leksykalną lub embedding. Uruchom klasyfikator bazowy dla co najmniej jednej reprezentacji i zachowaj predykcje.

**Punkt kontrolny 3.5**

Pokaż kompletny przepływ danych. Wyjaśnij znaczenie parametrów oraz wskaż, które elementy są pomiarem, które cechą, a które inferencją.

## 16 4. Eksperyment ilościowy

Wyznacz macro-F1 oraz macierz pomyłek; dla dwóch reprezentacji także zgodność i przypadki rozbieżne. Zmieniaj jeden kontrolowany czynnik naraz. Dla procedur losowych zachowaj seed i wykonaj powtórzenia pozwalające ocenić zmienność. Wynik przedstaw w tabeli oraz na wykresie z podpisanymi osiami, jednostkami i opisem warunków.

**Punkt kontrolny 4.0**

Wynik ma odpowiadać na pytanie badawcze. Sama wizualizacja bez metryki i warunków eksperymentu nie wystarcza.

## 17 5. Failure case

Znajdź przypadek, w którym pipeline daje wynik błędny, niestabilny lub niepewny. Nie ograniczaj się do pokazania błędu. Ustal, na którym etapie powstaje: akwizycja, preprocessing, reprezentacja, model, próg decyzyjny czy interpretacja. Zaproponuj test, który pozwala odróżnić co najmniej dwie możliwe przyczyny.

**Punkt kontrolny 4.5**

Raport zawiera failure case, diagnozę źródła błędu oraz kontrolowany test odporności albo generalizacji.

## 18 6. Ablation i analiza niepewności

porównaj model akustyczny, tekstowy i – jeśli dane pozwalają – ich prostą fuzję; zinterpretuj rozbieżne predykcje. Użyj tej samej metryki i tego samego zbioru ewaluacyjnego dla obu wariantów. Następnie wskaż warunek, w którym system powinien zwrócić `unknown` lub powstrzymać się od działania.

**Punkt kontrolny 5.0**

Wniosek z ablation jest ilościowy i uwzględnia alternatywne wyjaśnienia, confoundery oraz granice generalizacji.

## 19 7. Raport i artefakty

Raport musi zawierać: pytanie badawcze i hipotezę; pochodzenie danych i ground truth; środowisko i parametry; baseline; opis pipeline'u; metryki; tabelę lub wykres; przypadek poprawny; failure case; ablation; niepewność; ograniczenia i krótki wniosek. Zachowaj kod i artefakty potrzebne do odtworzenia wyniku.

## 20 Pytania kontrolne

1. Co w tym laboratorium jest obserwowane bezpośrednio, a co wywnioskowane?
2. Jak powstał ground truth i jaki błąd może zawierać?
3. Jaki confounder może zmienić wynik bez zmiany badanego konstruktu?
4. Czy zastosowana walidacja pozwala wnioskować o nowej osobie?
5. Jak rozpoznasz data leakage?

6. Kiedy system powinien zwrócić `unknown`?
7. Które parametry są niezbędne do reprodukcji wyniku?

## 21 Warunek zaliczenia

Pipeline musi działać, wynik musi być mierzalny i odtwarzalny, a interpretacja nie może wykraczać poza informacje dostępne w danych. Student powinien umieć wyjaśnić zarówno poprawny wynik, jak i co najmniej jeden przypadek niepowodzenia.