

# Librosa i przetwarzanie dźwięku w języku Python

Dominik Thiem

# Przetwarzanie dźwięku a obrazu

## Przetwarzanie obrazu

- dane przeważnie w 2D
- wiele zbiorów danych
- liczne zastosowania komercyjne
- popularne konkursy, w których udział biorą olbrzymie korporacje
- relatywnie niski próg wejścia

## Przetwarzanie dźwięku

- dane w 1D
- duże zbiory danych muzycznych, mowy
- mały wybór zbiorów ESR
- mniej znanych konkursów, najbardziej popularne zastosowanie jako rozpoznawanie mowy
- wyższy próg wejścia

# Librosa

Opublikowana w 2015 na Python in science conf.

Open source

Stworzona celem ujednoczenia bibliotek przetwarzania dźwięku w Pythonie

Modularna budowa

Niski próg wejścia dla użytkowników Matlaba

# Librosa - najważniejsze moduły

## Core IO, DSP:

### Audio Processing :

Ładowanie plików dźwiękowych

Konwersja mono, przepróbkowywanie sygnału

Generacja dźwięku

### Spectral representations:

Obliczanie transformat STFT, ISTFT, CQT, ICQT, FMT

Rozciąganie spektrogramu - Phase Vocoder

### Magnitude scaling:

Zamiana amplitudy na wzmocnienie lub moc

# Librosa - najważniejsze moduły

## Core IO, DSP:

Time and frequency conversion :

Konwersje w dziedzinie dyskretnej, częstotliwościowej

Zamiana próbek, ramek, czasu, częstotliwości, nut, MIDI, itp.

## Display:

Wyświetlanie przebiegu czasowego lub częstotliwościowego

Moduły ułatwiające formatowanie osi (dla nut, czasu, częstotliwości)

# Librosa - najważniejsze moduły

## Feature extracion:

Obliczanie MFCC, melowo skalowanego spektrogramu, Chroma, tonnetz, rmse, spektralnej centroidy, itp.  
Obliczanie pochodnej zadanego sygnału

## Pozostałe funkcje:

Wyznaczanie tempa utworu  
Detekcja zdarzeń (wykrywanie nut)  
Dekompozycja hpss  
Zmiana tonu utworu  
DTW - Dynamic Time Warping  
Banki filtrów

# Inne biblioteki przetwarzania dźwięku

**PyAudioAnalysis**

**ThinkDSP**

**SciPy**

**FastDTW**

**Soundfile**

**Audioop**

**Aubio**

**Friture**

...

# Przykład użycia: ESR - Environmental Sound Recognition



Baza danych UrbanSound8k:

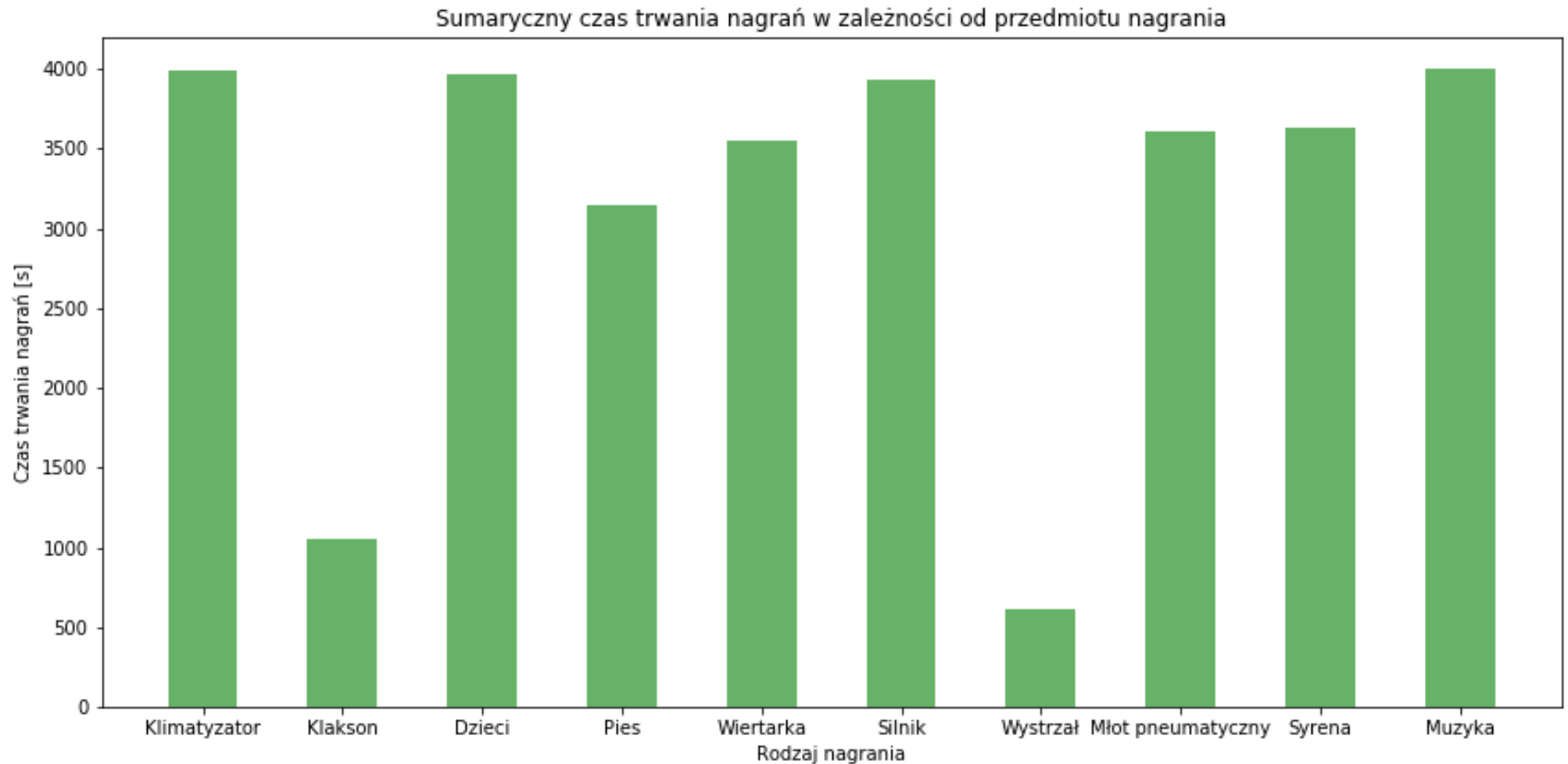
Około 8 tysięcy nagrań, 10 klas,  
nie dłuższe niż 4 sekundy

Różne częstotliwości próbkowania,  
ilość przykładów per klasa.

Niektóre utwory nagrane w stereo, inne mono.

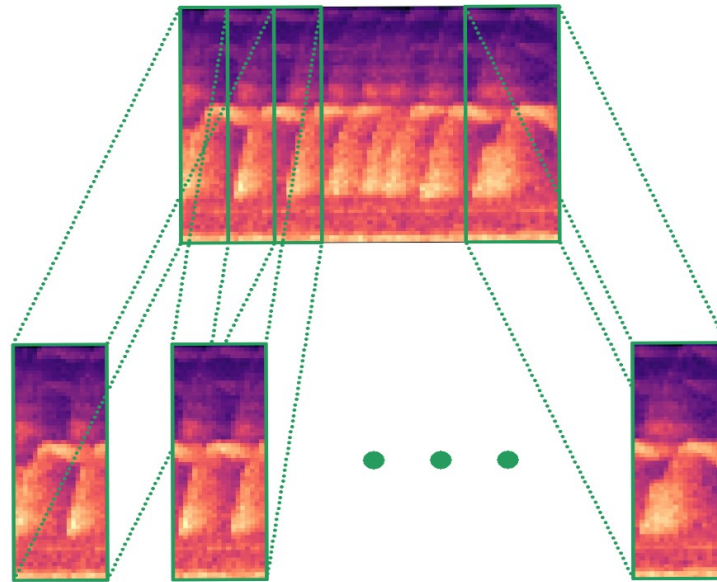


# Przykład użycia: ESR - Environmental Sound Recognition



# Przykład użycia: ESR - Environmental Sound Recognition

Przetwarzanie spektrogramu sliding window  
Klasyfikator - głęboka spoolotowa sieć neuronowa  
Agregacja wvników dla pojedynczych segmentów



# Przykład użycia: ESR - Environmental Sound Recognition

	0	2	4	6	8					
0	84	0	17	3	29	11	0	26	5	21
1	1	52	0	6	0	2	0	0	0	2
2	0	0	164	23	2	1	0	0	3	4
3	5	6	14	164	2	0	0	0	0	4
4	0	2	5	4	134	3	0	33	7	7
5	8	0	2	0	0	154	0	6	0	8
6	0	0	0	16	0	0	45	0	0	0
7	0	0	0	0	34	4	0	136	0	1
8	1	0	2	12	0	5	0	0	141	0
9	0	0	12	0	0	0	0	0	0	186
	Predicted label									

Macierz konfuzji wyników

# Inne zastosowania

Rozpoznawanie mowy

Text to speech

Systemy ostrzegania

Modulacja muzyki

Wyszukiwanie utworów muzycznych

Music prescription

**Dziękuję za uwagę!**