СӘТБАЕВ УНИВЕРСИТЕТІ

SATBAYEV UNIVERSITY

Основы искусственного интеллекта. Лекция 13. Обработка звука. Основы обработки звуковых сигналов, библиотеки для обработки звука (Librosa)

Преподаватель: **Қасенхан Арай Мейрамбайқызы, доктор PhD, ассоциированный профессор кафедры «Программная инженерия»**

a.kassenkhan@satbayev.university

Содержание

1. Как компьютер представляет звук
2. Частота и амплитуда
3. Дискретизация
4. Sample Rate и Bit Depth
5. Анализ звука
6. Waveform
7. Спектр
8. Спектрограмма
9. Признаки аудио
10. Librosa
11. Загрузка аудио
12. Визуализация сигнала

По завершению урока Вы будете знать:

1. **Объяснить, как звук представляется в цифровом виде**
2. **Различать частоту и амплитуду**
3. **Понимать Sample Rate и Bit Depth**
4. **Объяснить назначение waveform и спектрограммы**
5. **Понимать, зачем извлекать признаки из аудио**
6. **Загружать и анализировать аудио с помощью Librosa**

Что такое звук?



ИСТОЧНИК
(например, удар по барабану)



ВИБРАЦИЯ
(колебания объекта)



ЗВУКОВАЯ ВОЛНА
(распространяется в воздухе)



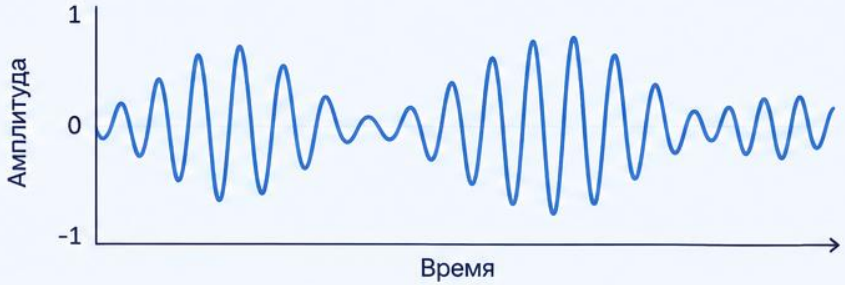
МИКРОФОН
(преобразует волну в электрический сигнал)



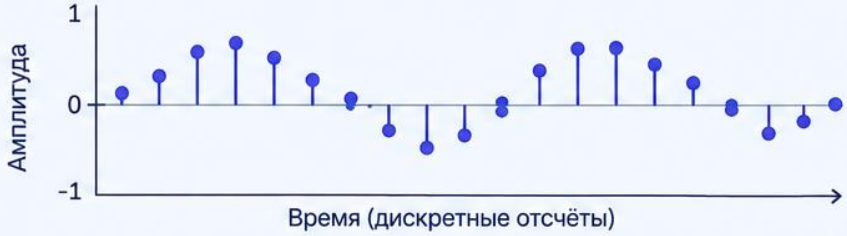
ЦИФРОВОЙ СИГНАЛ
(последовательность чисел)

[0.12, 0.18, 0.31, 0.25, ...]

Звуковая волна (в воздухе)



Цифровой сигнал (после оцифровки)



[0.12, 0.18, 0.31, 0.25, 0.10, -0.05, -0.28, -0.32, -0.18, 0.06, ...]



Изображение → 
массив пикселей



Звук → 
массив значений сигнала,

Для компьютера звук — последовательность числовых значений амплитуды во времени.

Частота и амплитуда

Frequency — частота

Сколько колебаний происходит за секунду.

Единица: **Hz — Герц**

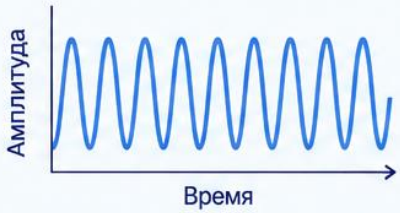
Пример:

Низкая частота



БАС

Высокая частота



ВЫСОКИЙ ТОН

Пример: Нота Ля ≈ **440 Hz**

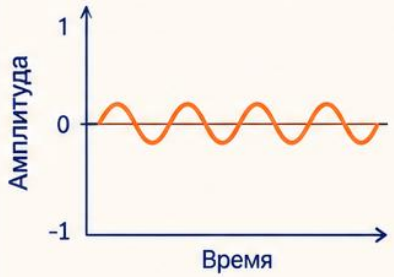
Amplitude — амплитуда

Насколько сильны колебания.

Единица: (относительные единицы, например 0 ... 1)

Пример:

Тихий звук



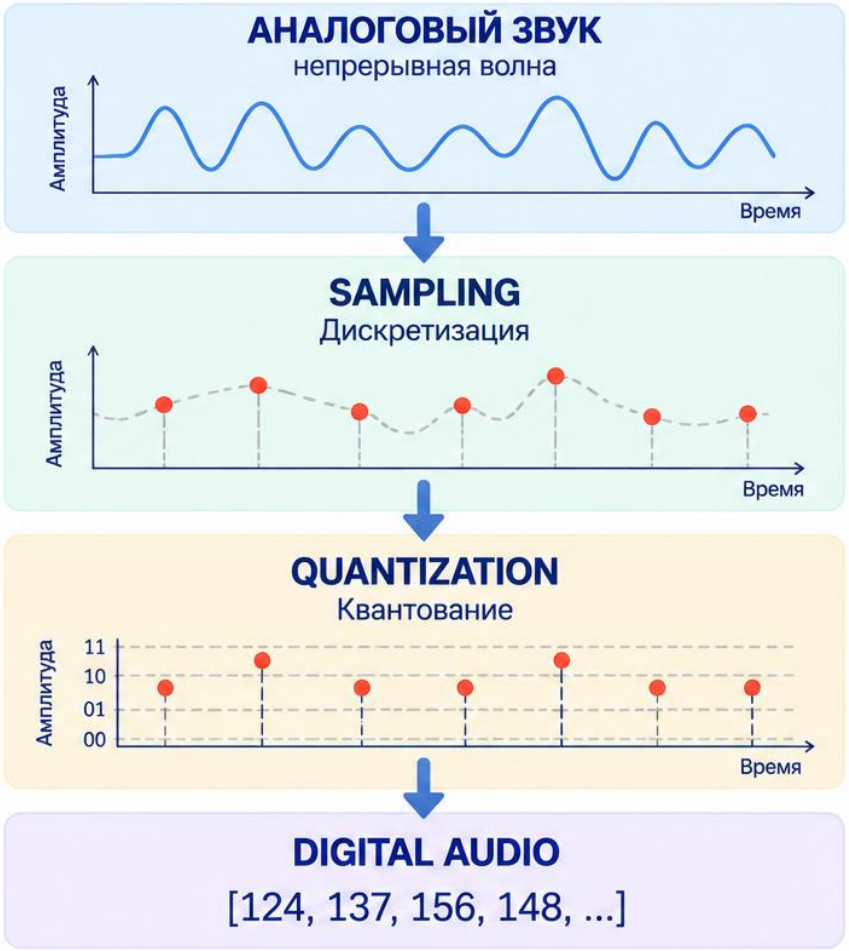
Громкий звук



Frequency → высота звука

Amplitude → интенсивность сигнала / воспринимаемая громкость

Как аналоговый звук становится цифровым?



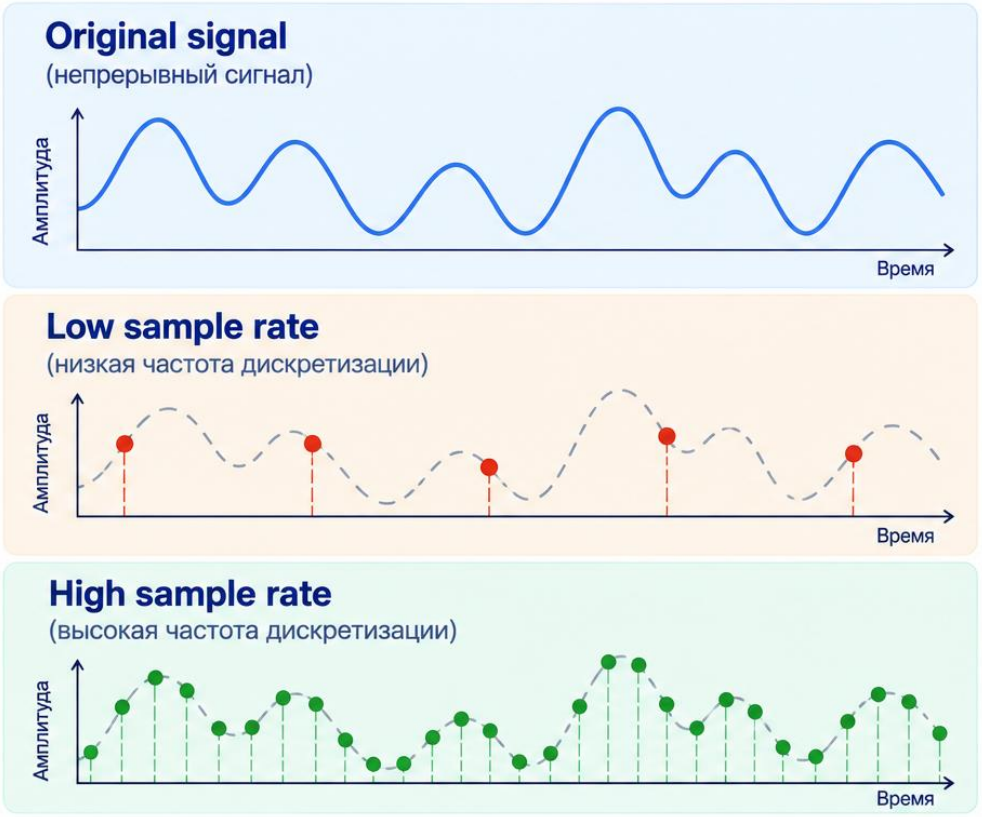
Дискретизация
→ когда измеряем сигнал

Квантование
→ насколько точно записываем значение

Аналоговый звук → измерения → числа → цифровой звук

Sample Rate

Sample Rate - количество измерений звукового сигнала за одну секунду.



8 kHz → речь / некоторые телекоммуникационные задачи

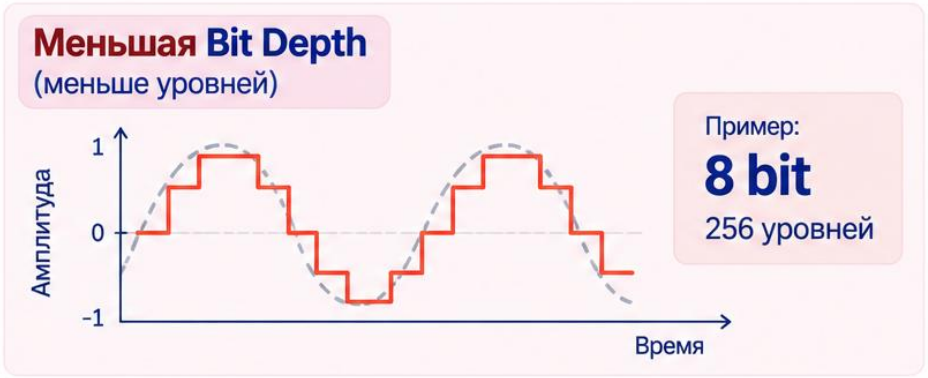
44.1 kHz → CD audio

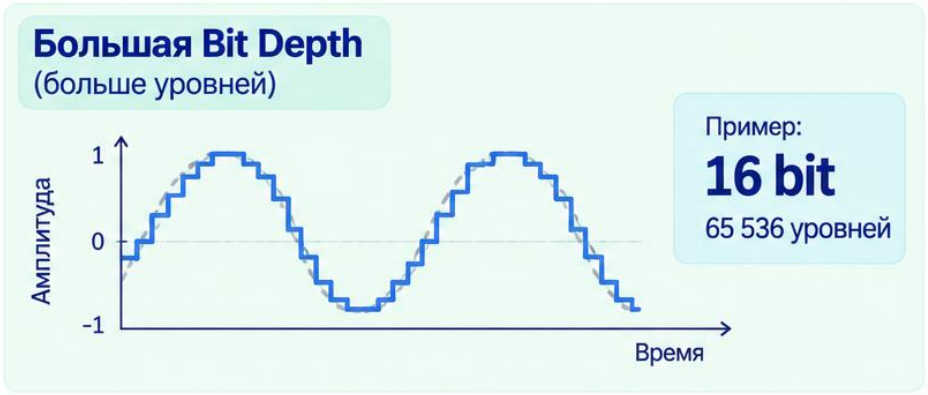
48 kHz → распространено для видео/профессионального аудио

Чем выше Sample Rate, тем чаще измеряется сигнал.

Bit Depth

Bit Depth - определяет количество возможных уровней амплитуды.





Примеры:
8 bit → 256 уровней
16 bit → 65 536 уровней
24 bit → 16 777 216 уровней

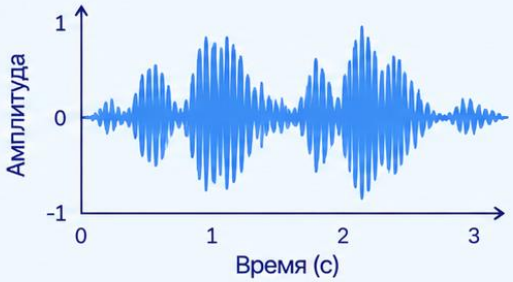
Sample Rate → точность по времени

Bit Depth → точность представления амплитуды

Как можно посмотреть на звук?

1 Waveform

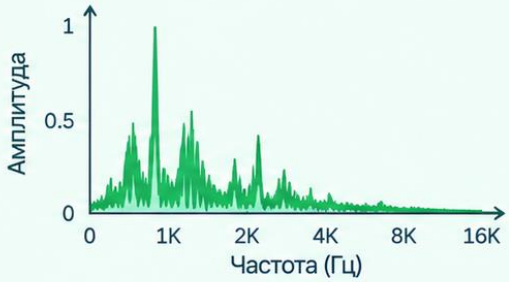
Amplitude × Time



Показывает:
Как изменяется сигнал во времени?

2 Spectrum

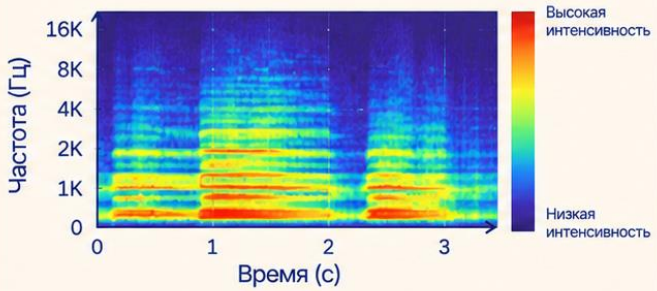
Amplitude × Frequency



Показывает:
Какие частоты присутствуют?

3 Spectrogram

Time × Frequency × Intensity

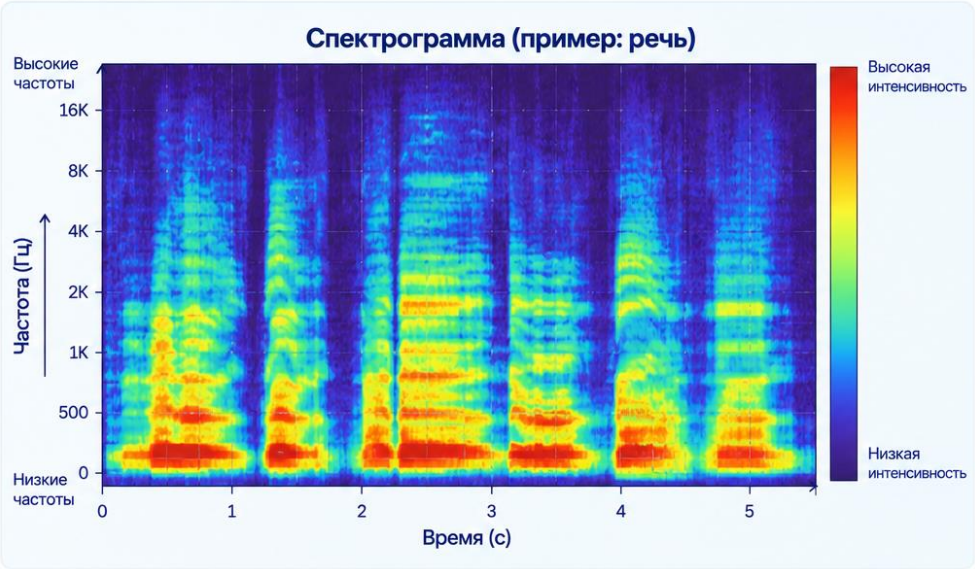


Показывает:
Какие частоты присутствуют и когда?

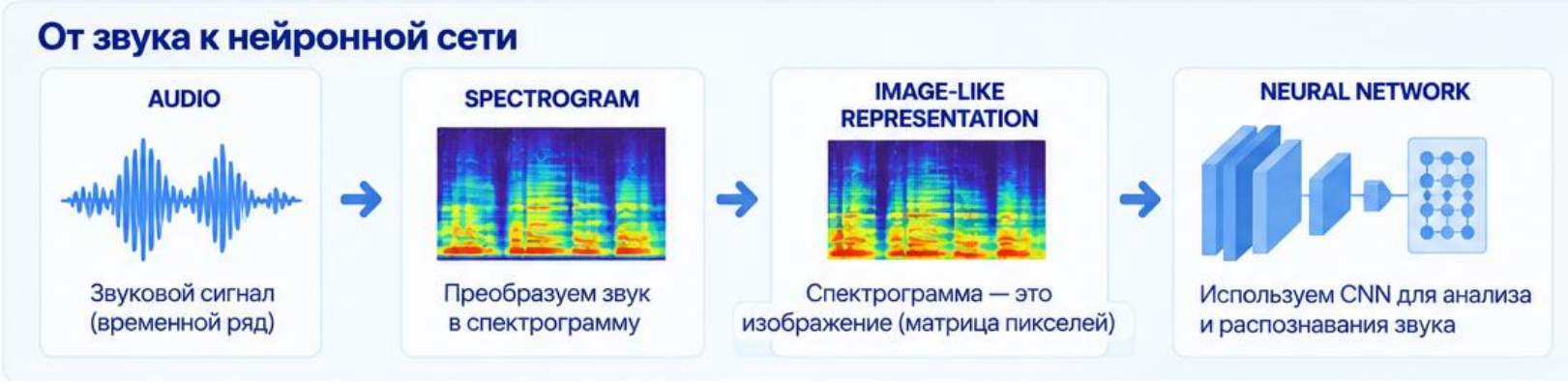
Один и тот же звук можно увидеть по-разному: во времени, в частотной области и в виде спектрограммы.

Что такое спектрограмма?

Спектрограмма превращает звук в визуальное представление.

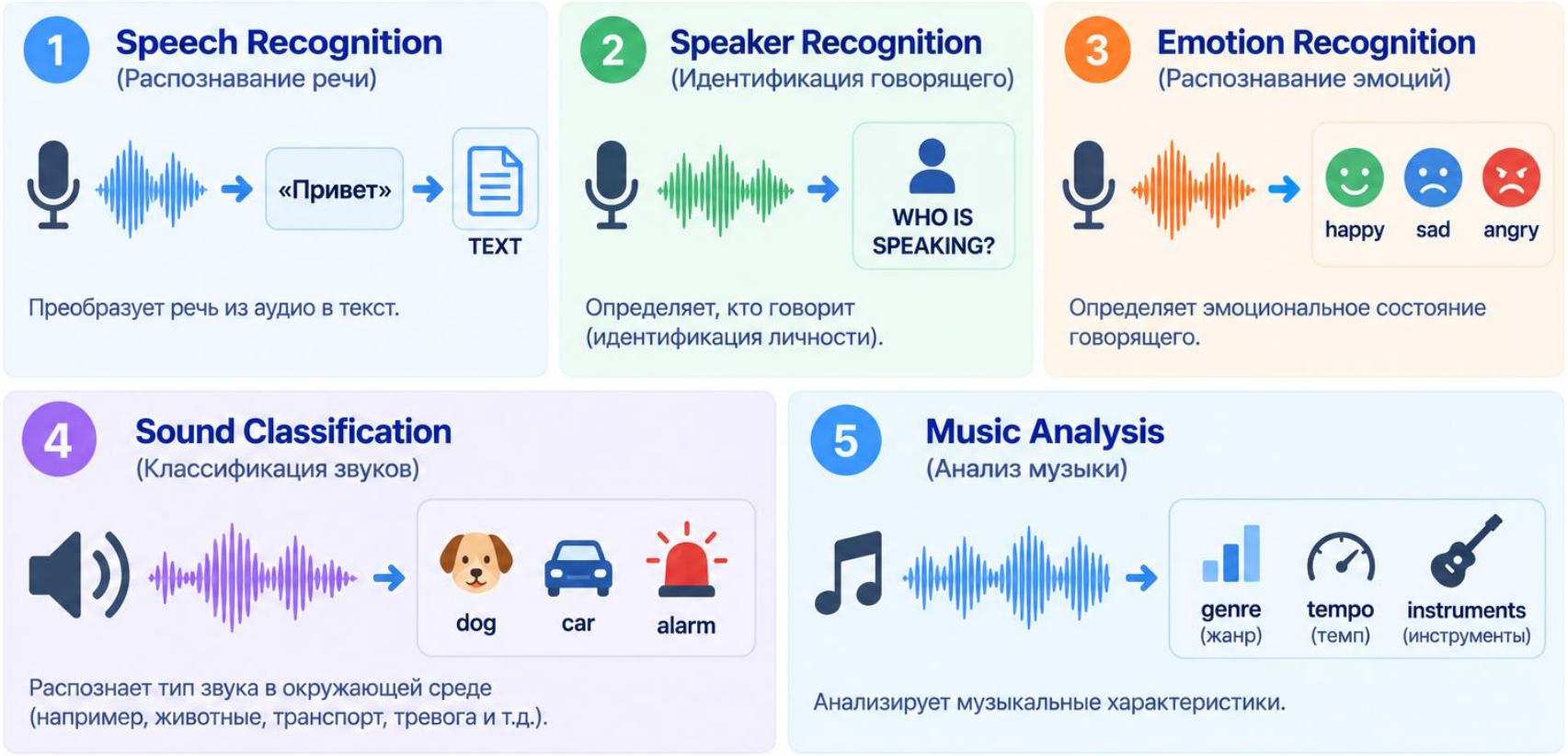


X → время
Y → частота
Color → интенсивность



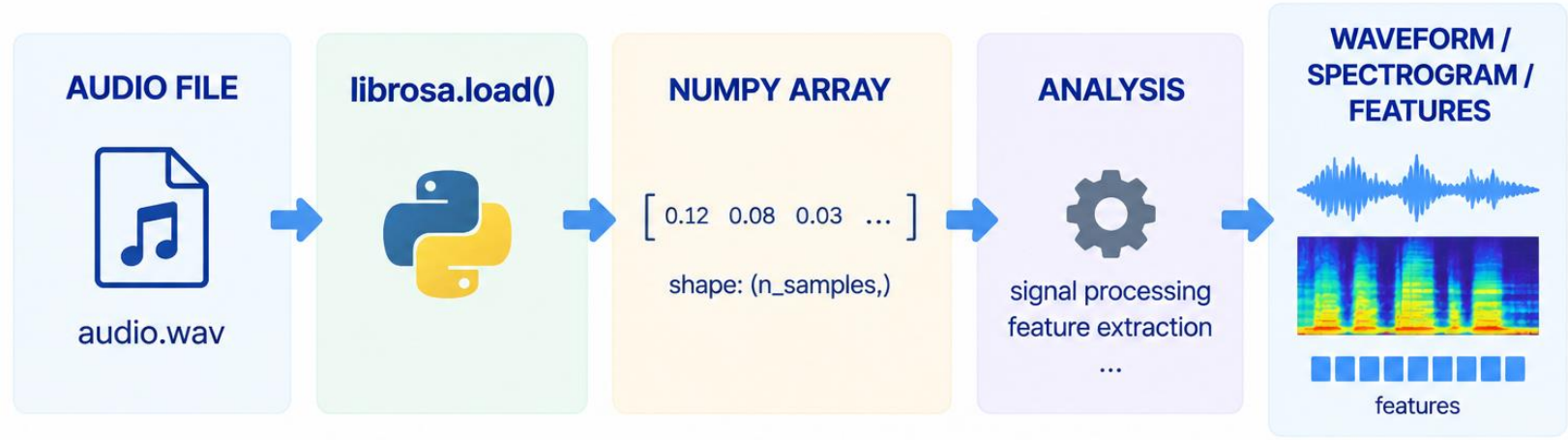
Где AI работает со звуком?

AI учится находить закономерности в аудиосигналах так же, как в изображениях или тексте.



Librosa

Librosa - Python-библиотека для анализа аудио.



Установка:

pip install librosa

Импорт:

import librosa

import librosa.display

Librosa загружает аудио как NumPy-массив.

Загружаем и визуализируем аудио

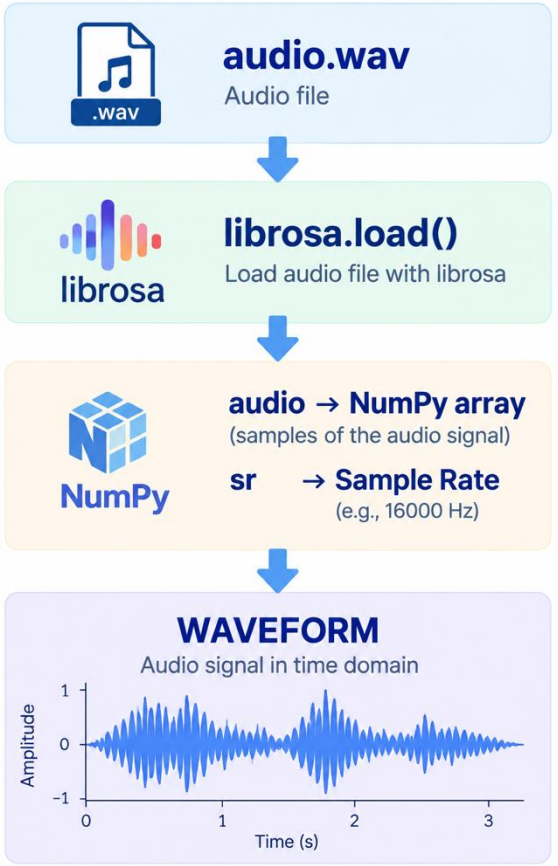
```
import librosa
import librosa.display
import matplotlib.pyplot as plt

audio, sr = librosa.load(
    "audio.wav",
    sr=22050
)

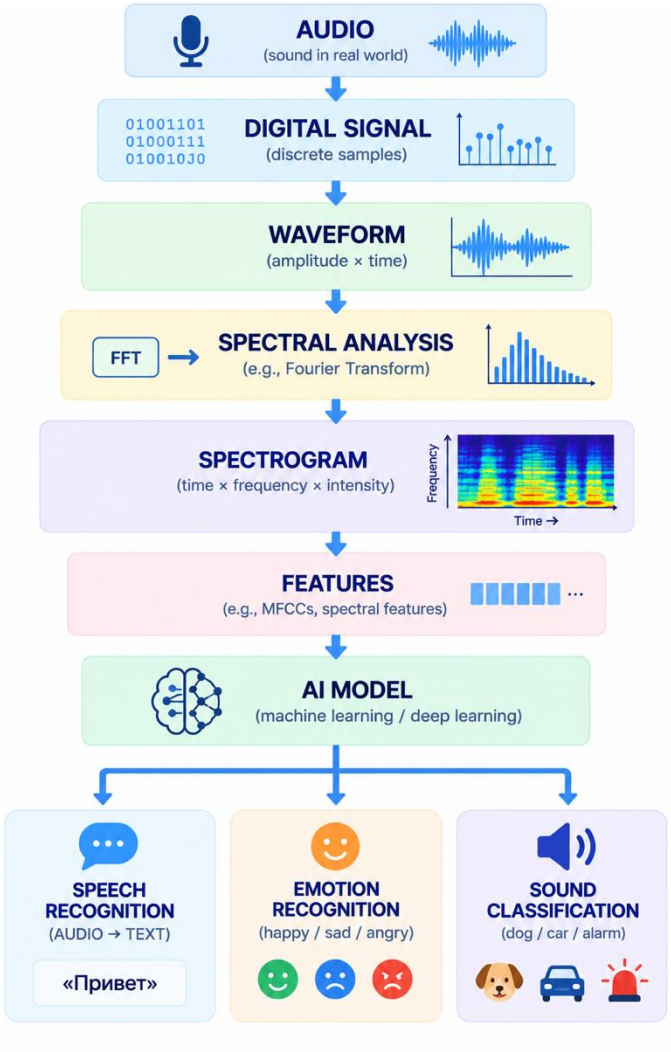
print(audio.shape)
print("Sample rate:", sr)

librosa.display.waveshow(
    audio,
    sr=sr
)

plt.xlabel("Time")
plt.ylabel("Amplitude")
plt.show()
```



Заключение. От звука к признакам для AI



Frequency
→ высота тона
Amplitude
→ интенсивность сигнала
Sample Rate
→ сколько измерений в секунду
Bit Depth
→ сколько уровней амплитуды
Spectrogram
→ частоты во времени
Librosa
→ анализ аудио в Python

Для человека звук — это речь, музыка или шум. Для искусственного интеллекта звук сначала становится числами, из которых модель учится извлекать полезные признаки.