

Фрагмент 3 – Фрагмент кода визуализации спектрограммы

```
# Визуализируем спектрограмму массива аудио на логарифмической шкале
plt.figure(figsize=(15, 7))
librosa.display.specshow(mfccs, sr=sr, x_axis='time', y_axis='log')
```

Все полученные признаки могут быть интегрированы в единый массив для подачи на вход нейронной сети для последующей обработки. В качестве такой обработки может быть выбрана задача мультиклассовой классификации звуков с целью распознавания жанра музыки, пения птиц и пр. Примером набора данных для идентификации пения птиц является «Cornell Birdcall Identification» на портале Kaggle.

Опыт автора показывает, что подготовка четырехсекундных фреймов указанного датасета из 88 классов на основе шести спектральных параметров в среде Colaboratory требует около семнадцати минут. Тогда как обучение в нейронной сети из семи слоев на 150 эпохах требует около пяти минут. Точность такой нейронной сети на проверочной выборке достигает около 67 процентов.

Рассмотренные методы пакета librosa для представления звука позволяют значительно повысить эффективность нейронных сетей на обучающих, тестовых и проверочных наборах данных в задачах обработки звука.

ЛИТЕРАТУРА

1. Librosa. Audio and music processing in Python [Электронный ресурс]. – Режим доступа: <https://librosa.org>. – Дата доступа: 12.01.2023.
2. SpeechRecognition 3.9.0 [Электронный ресурс]. – Режим доступа: <https://pypi.org/project/SpeechRecognition/>. – Дата доступа: 12.01.2023.
3. Base64 – Base16, Base32, Base64, Base85 Data Encodings [Электронный ресурс]. – Режим доступа: <https://docs.python.org/3/library/base64.html#module-base64>. – Дата доступа: 12.01.2023.
4. SciPy [Электронный ресурс]. – Режим доступа: <https://scipy.org/> – Дата доступа: 12.01.2023.
5. Добро пожаловать в Colaboratory! [Электронный ресурс]. – Режим доступа: <https://colab.research.google.com/>. – Дата доступа: 02.01.2023.

А.А. КОЗИНСКИЙ, А.Л. МИХНЯЕВ
УО БрГТУ, (г. Брест, Беларусь)

ОСНОВЫ РАСПОЗНАВАНИЯ ЗВУКА СРЕДСТВАМИ БИБЛИОТЕКИ SPEECH RECOGNITION

Для языка Python существует множество пакетов, которые могут быть использованы для распознавания небольших по длительности звуковых фрагментов. Одним из них является Speech Recognition [1]. Названный пакет отличает простота использования, поддержка в режимах онлайн и офлайн. Областью применения пакета может быть разработка проектов для «Умного дома» или «Интернет вещей». Пакет эффективен в распознавании диалогов из коротких фраз или команд управления устройствами.

Speech Recognition основан на выделении кепстральных коэффициентов, которые показывают зависимость мощности звукового сигнала от его частоты. Для восприятия звука органами слуха человека используется единица «мел» (см., например, [2]). Необходимо отметить, что субъективное восприятие человеком громкости звука, который определяется его амплитудой, нелинейно зависит от частоты аудио. Такая зависимость определена эмпирически [3].

Остановимся на программной реализации на языке Python применения мел-кепстральных коэффициентов, положенных в основу библиотеки Speech Recognition.

Фрагмент 1 демонстрирует использование Speech Recognition для распознавания короткой фразы, записанной в виде звукового файла.

Фрагмент 1 – Распознавание звука, записанного в аудио файле

```
# установка и подключение библиотеки распознавания речи
!pip install SpeechRecognition
import speech_recognition as sR
# установка и подключение модуля метрики качества в распознавании речи
!pip install jiwer
from jiwer import wer
# Подключение дополнительных библиотек
...
# создание объекта класса Recognizer
r = sR.Recognizer()
# считываем аудиофайл
audio = r.record('Путь к файлу', duration=duration)
# собственно распознавание с установкой используемого языка
result = r.recognize_google(audio, language='ru')

nuzna_fraza = 'вот такой текст' #ожидаем, что этот текст записан в звуковом файле

print('Оригинал: ', nuzna_fraza)
print('Результат распознавания:', result)
# вывод метрики качества
print('WER:', wer(nuzna_fraza.lower(), result.lower()))
```

Указанный фрагмент не включает код с подключением необходимых библиотек. Однако приведенный фрагмент кода с использованием пакета Speech Recognition при его лаконичности с достаточной полнотой демонстрирует распознавание короткой фразы из звукового файла. Использование готовых инструментов для передачи речевого взаимодействия в разрабатываемых приложениях от ведущих компаний (в их числе Google, MicroSoft, Sound Hound, IBM, PocketSphinx) избавляет разработчика от проектирования собственной нейронной сети: сбора и обработки датасета, создания нейросети и ее обучения и других ресурсоемких мероприятий.

ЛИТЕРАТУРА

1. SpeechRecognition 3.9.0 [Электронный ресурс]. – Режим доступа: <https://pypi.org/project/SpeechRecognition/>. – Дата доступа: 12.01.2023.
2. Первичный анализ речевых сигналов [Электронный ресурс]. – Режим доступа: <https://alphacerpei.com/ru/lecture1.pdf>. – Дата доступа: 18.01.2023.
3. Мел-кепстральные коэффициенты (MFCC) и распознавание речи [Электронный ресурс]. – Режим доступа: <https://habr.com/ru/post/140828/>. – Дата доступа: 18.01.2023.

И.А. КОЛЕСНИКОВ, А.А. ГОЛУБ
УО МГПУ им. И.П. Шамякина (г. Мозырь, Беларусь)

ОБЪЕКТНО-ОРИЕНТИРОВАННОЕ ПРОГРАММИРОВАНИЕ В JS

В настоящее время разработка web-сайтов все чаще осуществляется не одним человеком, а командой программистов, поэтому использование классических подходов