Sathvik Reddy V

BL.EN.U4AIE23127

SPEECH PROCESSING LAB 2

```
!pip install soundfile
!pip install simpleaudio
!pip install librosa
!pip install scipy
```

```
Requirement already satisfied: soundfile in
/usr/local/lib/python3.12/dist-packages (0.13.1)
Requirement already satisfied: cffi>=1.0 in
/usr/local/lib/python3.12/dist-packages (from soundfile) (2.0.0)
Requirement already satisfied: numpy in
/usr/local/lib/python3.12/dist-packages (from soundfile) (2.0.2)
Requirement already satisfied: pycparser in
/usr/local/lib/python3.12/dist-packages (from cffi>=1.0->soundfile)
(2.23)
Collecting simpleaudio
  Downloading simpleaudio-1.0.4.tar.gz (2.0 MB)
━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 2.0/2.0 MB 23.2 MB/s eta
0:00:00
etadata (setup.py) ... pleaudio
  error: subprocess-exited-with-error

  × python setup.py bdist_wheel did not run successfully.
  │ exit code: 1
  ╰─> See above for output.

  note: This error originates from a subprocess, and is likely not a
problem with pip.
  Building wheel for simpleaudio (setup.py) ...   ERROR: Failed
building wheel for simpleaudio
pleaudio
Failed to build simpleaudio
ERROR: ERROR: Failed to build installable wheels for some
pyproject.toml based projects (simpleaudio)
Requirement already satisfied: librosa in
/usr/local/lib/python3.12/dist-packages (0.11.0)
Requirement already satisfied: audioread>=2.1.9 in
/usr/local/lib/python3.12/dist-packages (from librosa) (3.1.0)
Requirement already satisfied: numba>=0.51.0 in
/usr/local/lib/python3.12/dist-packages (from librosa) (0.60.0)
Requirement already satisfied: numpy>=1.22.3 in
/usr/local/lib/python3.12/dist-packages (from librosa) (2.0.2)
Requirement already satisfied: scipy>=1.6.0 in
/usr/local/lib/python3.12/dist-packages (from librosa) (1.16.3)
Requirement already satisfied: scikit-learn>=1.1.0 in
```

```
/usr/local/lib/python3.12/dist-packages (from librosa) (1.6.1)
Requirement already satisfied: joblib>=1.0 in
/usr/local/lib/python3.12/dist-packages (from librosa) (1.5.3)
Requirement already satisfied: decorator>=4.3.0 in
/usr/local/lib/python3.12/dist-packages (from librosa) (4.4.2)
Requirement already satisfied: soundfile>=0.12.1 in
/usr/local/lib/python3.12/dist-packages (from librosa) (0.13.1)
Requirement already satisfied: pooch>=1.1 in
/usr/local/lib/python3.12/dist-packages (from librosa) (1.8.2)
Requirement already satisfied: soxr>=0.3.2 in
/usr/local/lib/python3.12/dist-packages (from librosa) (1.0.0)
Requirement already satisfied: typing_extensions>=4.1.1 in
/usr/local/lib/python3.12/dist-packages (from librosa) (4.15.0)
Requirement already satisfied: lazy_loader>=0.1 in
/usr/local/lib/python3.12/dist-packages (from librosa) (0.4)
Requirement already satisfied: msgpack>=1.0 in
/usr/local/lib/python3.12/dist-packages (from librosa) (1.1.2)
Requirement already satisfied: packaging in
/usr/local/lib/python3.12/dist-packages (from lazy_loader>=0.1-
>librosa) (25.0)
Requirement already satisfied: llvmlite<0.44,>=0.43.0dev0 in
/usr/local/lib/python3.12/dist-packages (from numba>=0.51.0->librosa)
(0.43.0)
Requirement already satisfied: platformdirs>=2.5.0 in
/usr/local/lib/python3.12/dist-packages (from pooch>=1.1->librosa)
(4.5.1)
Requirement already satisfied: requests>=2.19.0 in
/usr/local/lib/python3.12/dist-packages (from pooch>=1.1->librosa)
(2.32.4)
Requirement already satisfied: threadpoolctl>=3.1.0 in
/usr/local/lib/python3.12/dist-packages (from scikit-learn>=1.1.0-
>librosa) (3.6.0)
Requirement already satisfied: cffi>=1.0 in
/usr/local/lib/python3.12/dist-packages (from soundfile>=0.12.1-
>librosa) (2.0.0)
Requirement already satisfied: pycparser in
/usr/local/lib/python3.12/dist-packages (from cffi>=1.0-
>soundfile>=0.12.1->librosa) (2.23)
Requirement already satisfied: charset_normalizer<4,>=2 in
/usr/local/lib/python3.12/dist-packages (from requests>=2.19.0-
>pooch>=1.1->librosa) (3.4.4)
Requirement already satisfied: idna<4,>=2.5 in
/usr/local/lib/python3.12/dist-packages (from requests>=2.19.0-
>pooch>=1.1->librosa) (3.11)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.12/dist-packages (from requests>=2.19.0-
>pooch>=1.1->librosa) (2.5.0)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.12/dist-packages (from requests>=2.19.0-
```

```
>pooch>=1.1->librosa) (2026.1.4)
Requirement already satisfied: scipy in
/usr/local/lib/python3.12/dist-packages (1.16.3)
Requirement already satisfied: numpy<2.6,>=1.25.2 in
/usr/local/lib/python3.12/dist-packages (from scipy) (2.0.2)
```

1

1

```
import numpy as np
import matplotlib.pyplot as plt
import soundfile as sf
from IPython.display import Audio
import librosa
import librosa.display

y, sr = sf.read("/content/LJ037-0171.wav")
Audio(y, rate=sr)
```

```
<IPython.lib.display.Audio object>
```

2

```
num_samples = len(y)
duration = num_samples / sr

print("Sample rate (Hz):", sr)
print("Number of samples:", num_samples)
print("Total duration (sec):", round(duration, 3))
```

```
Sample rate (Hz): 22050
Number of samples: 167226
Total duration (sec): 7.584
```

```
!pip install transformers datasets torchaudio
```

```
Requirement already satisfied: transformers in
/usr/local/lib/python3.12/dist-packages (4.57.6)
Requirement already satisfied: datasets in
/usr/local/lib/python3.12/dist-packages (4.0.0)
Requirement already satisfied: torchaudio in
/usr/local/lib/python3.12/dist-packages (2.9.0+cpu)
Requirement already satisfied: filelock in
/usr/local/lib/python3.12/dist-packages (from transformers) (3.20.3)
Requirement already satisfied: huggingface-hub<1.0,>=0.34.0 in
/usr/local/lib/python3.12/dist-packages (from transformers) (0.36.0)
Requirement already satisfied: numpy>=1.17 in
/usr/local/lib/python3.12/dist-packages (from transformers) (2.0.2)
Requirement already satisfied: packaging>=20.0 in
```

```
/usr/local/lib/python3.12/dist-packages (from transformers) (25.0)
Requirement already satisfied: pyyaml>=5.1 in
/usr/local/lib/python3.12/dist-packages (from transformers) (6.0.3)
Requirement already satisfied: regex!=2019.12.17 in
/usr/local/lib/python3.12/dist-packages (from transformers)
(2025.11.3)
Requirement already satisfied: requests in
/usr/local/lib/python3.12/dist-packages (from transformers) (2.32.4)
Requirement already satisfied: tokenizers<=0.23.0,>=0.22.0 in
/usr/local/lib/python3.12/dist-packages (from transformers) (0.22.2)
Requirement already satisfied: safetensors>=0.4.3 in
/usr/local/lib/python3.12/dist-packages (from transformers) (0.7.0)
Requirement already satisfied: tqdm>=4.27 in
/usr/local/lib/python3.12/dist-packages (from transformers) (4.67.1)
Requirement already satisfied: pyarrow>=15.0.0 in
/usr/local/lib/python3.12/dist-packages (from datasets) (18.1.0)
Requirement already satisfied: dill<0.3.9,>=0.3.0 in
/usr/local/lib/python3.12/dist-packages (from datasets) (0.3.8)
Requirement already satisfied: pandas in
/usr/local/lib/python3.12/dist-packages (from datasets) (2.2.2)
Requirement already satisfied: xxhash in
/usr/local/lib/python3.12/dist-packages (from datasets) (3.6.0)
Requirement already satisfied: multiprocess<0.70.17 in
/usr/local/lib/python3.12/dist-packages (from datasets) (0.70.16)
Requirement already satisfied: fsspec<=2025.3.0,>=2023.1.0 in
/usr/local/lib/python3.12/dist-packages (from
fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (2025.3.0)
Requirement already satisfied: torch==2.9.0 in
/usr/local/lib/python3.12/dist-packages (from torchaudio) (2.9.0+cpu)
Requirement already satisfied: typing-extensions>=4.10.0 in
/usr/local/lib/python3.12/dist-packages (from torch==2.9.0-
>torchaudio) (4.15.0)
Requirement already satisfied: setuptools in
/usr/local/lib/python3.12/dist-packages (from torch==2.9.0-
>torchaudio) (75.2.0)
Requirement already satisfied: sympy>=1.13.3 in
/usr/local/lib/python3.12/dist-packages (from torch==2.9.0-
>torchaudio) (1.14.0)
Requirement already satisfied: networkx>=2.5.1 in
/usr/local/lib/python3.12/dist-packages (from torch==2.9.0-
>torchaudio) (3.6.1)
Requirement already satisfied: jinja2 in
/usr/local/lib/python3.12/dist-packages (from torch==2.9.0-
>torchaudio) (3.1.6)
Requirement already satisfied: aiohttp!=4.0.0a0,!=4.0.0a1 in
/usr/local/lib/python3.12/dist-packages (from
fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (3.13.3)
Requirement already satisfied: hf-xet<2.0.0,>=1.1.3 in
/usr/local/lib/python3.12/dist-packages (from huggingface-
```

```
hub<1.0,>=0.34.0->transformers) (1.2.0)
Requirement already satisfied: charset_normalizer<4,>=2 in
/usr/local/lib/python3.12/dist-packages (from requests->transformers)
(3.4.4)
Requirement already satisfied: idna<4,>=2.5 in
/usr/local/lib/python3.12/dist-packages (from requests->transformers)
(3.11)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/usr/local/lib/python3.12/dist-packages (from requests->transformers)
(2.5.0)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.12/dist-packages (from requests->transformers)
(2026.1.4)
Requirement already satisfied: python-dateutil>=2.8.2 in
/usr/local/lib/python3.12/dist-packages (from pandas->datasets)
(2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in
/usr/local/lib/python3.12/dist-packages (from pandas->datasets)
(2025.2)
Requirement already satisfied: tzdata>=2022.7 in
/usr/local/lib/python3.12/dist-packages (from pandas->datasets)
(2025.3)
Requirement already satisfied: aiohappyeyeballs>=2.5.0 in
/usr/local/lib/python3.12/dist-packages (from aiohttp!=4.0.0a0,!
=4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (2.6.1)
Requirement already satisfied: aiosignal>=1.4.0 in
/usr/local/lib/python3.12/dist-packages (from aiohttp!=4.0.0a0,!
=4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (1.4.0)
Requirement already satisfied: attrs>=17.3.0 in
/usr/local/lib/python3.12/dist-packages (from aiohttp!=4.0.0a0,!
=4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (25.4.0)
Requirement already satisfied: frozenlist>=1.1.1 in
/usr/local/lib/python3.12/dist-packages (from aiohttp!=4.0.0a0,!
=4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (1.8.0)
Requirement already satisfied: multidict<7.0,>=4.5 in
/usr/local/lib/python3.12/dist-packages (from aiohttp!=4.0.0a0,!
=4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (6.7.0)
Requirement already satisfied: propcache>=0.2.0 in
/usr/local/lib/python3.12/dist-packages (from aiohttp!=4.0.0a0,!
=4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (0.4.1)
Requirement already satisfied: yarl<2.0,>=1.17.0 in
/usr/local/lib/python3.12/dist-packages (from aiohttp!=4.0.0a0,!
=4.0.0a1->fsspec[http]<=2025.3.0,>=2023.1.0->datasets) (1.22.0)
Requirement already satisfied: six>=1.5 in
/usr/local/lib/python3.12/dist-packages (from python-dateutil>=2.8.2-
>pandas->datasets) (1.17.0)
Requirement already satisfied: mpmath<1.4,>=1.1.0 in
/usr/local/lib/python3.12/dist-packages (from sympy>=1.13.3-
>torch==2.9.0->torchaudio) (1.3.0)
```

```
Requirement already satisfied: MarkupSafe>=2.0 in
/usr/local/lib/python3.12/dist-packages (from jinja2->torch==2.9.0-
>torchaudio) (3.0.3)
```

3

```
import torch
import librosa
import soundfile as sf
from transformers import Wav2Vec2Processor, Wav2Vec2ForCTC

processor = Wav2Vec2Processor.from_pretrained("Bluecast/wav2vec2-
Phoneme")
model = Wav2Vec2ForCTC.from_pretrained("Bluecast/wav2vec2-Phoneme")

if sr != 16000:
    y = librosa.resample(y.astype(float), orig_sr=sr, target_sr=16000)
    sr = 16000

input_values = processor(y, sampling_rate=sr,
return_tensors="pt").input_values

with torch.no_grad():
    logits = model(input_values).logits

predicted_ids = torch.argmax(logits, dim=-1)
phoneme_sequence = processor.batch_decode(predicted_ids)

print("Recognized phoneme sequence:")
print(phoneme_sequence[0])
```

```
WARNING:torchao.kernel.intmm:Warning: Detected no triton, on systems
without Triton certain kernels will not work
/usr/local/lib/python3.12/dist-packages/huggingface_hub/utils/_auth.py
:94: UserWarning:
The secret `HF_TOKEN` does not exist in your Colab secrets.
To authenticate with the Hugging Face Hub, create a token in your
settings tab (https://huggingface.co/settings/tokens), set it as
secret in your Google Colab and restart your session.
You will be able to reuse this secret in all of your notebooks.
Please note that authentication is recommended but still optional to
access public models or datasets.
  warnings.warn(
```


```
{"model_id":"634c1dd98fa64378813928a61f9000b3","version_major":2,"vers
ion_minor":0}
```

```
{"model_id":"4985537c9f26421cadeac97384cccb68","version_major":2,"vers
ion_minor":0}
```

{"model_id":"a0a727078a354919a2b7749ef620ba50","version_major":2,"version_minor":0}

{"model_id":"eeb72f3f481643459aa6b836395f6712","version_major":2,"version_minor":0}

{"model_id":"efc3002b9ec743d6a95525eecf201625","version_major":2,"version_minor":0}

{"model_id":"36a0c10b830b4133aa524640ec64a212","version_major":2,"version_minor":0}

{"model_id":"8aedc3d449b04535b96d344e2c3d5df2","version_major":2,"version_minor":0}

```
Recognized phoneme sequence:
ah ih g z ae m ah n ey sh ah n ah n d t eh s t ah m ow n iy ah v dhah
eh k s p er t s eh n ey b ah l d dh ah k ah m ih sh ah n t uw k ah n k
l uw d dh ae t f ay v sh aa t s m ey hh ae v b ih n f ay er d
```

4

```
num_frames = logits.shape[1]

total_duration = len(y) / sr
time_per_frame = total_duration / num_frames

frame_ids = torch.argmax(logits, dim=-1)[0].numpy()

token_intervals = []
prev_id = frame_ids[0]
start_frame = 0

for i in range(1, len(frame_ids)):
    if frame_ids[i] != prev_id:
        token_intervals.append((
            prev_id,
            start_frame * time_per_frame,
            i * time_per_frame
        ))
        prev_id = frame_ids[i]
        start_frame = i

token_intervals.append((
    prev_id,
    start_frame * time_per_frame,
    len(frame_ids) * time_per_frame
))

phoneme_intervals = []
```

```
current_phoneme = ""
phoneme_start = None
prev_end = None

for pid, start_t, end_t in token_intervals:
    token = processor.tokenizer.convert_ids_to_tokens([int(pid)])[0]

    if token in ["|", "<pad>", "[PAD]"]:
        if current_phoneme != "":
            phoneme_intervals.append(
                (current_phoneme, phoneme_start, prev_end)
            )
            current_phoneme = ""
        continue

    if current_phoneme == "":
        current_phoneme = token
        phoneme_start = start_t
    else:
        current_phoneme += token

    prev_end = end_t

if current_phoneme != "":
    phoneme_intervals.append(
        (current_phoneme, phoneme_start, prev_end)
    )

print("Estimated phoneme time intervals:\n")

for phoneme, start_t, end_t in phoneme_intervals:
    print(
        f"Phoneme: {phoneme:>4} | "
        f"Start: {start_t:.3f}s | End: {end_t:.3f}s"
    )
```

```
Estimated phoneme time intervals:

Phoneme:   ah | Start: 0.020s | End: 0.080s
Phoneme:   ih | Start: 0.120s | End: 0.161s
Phoneme:    g | Start: 0.201s | End: 0.221s
Phoneme:    z | Start: 0.241s | End: 0.281s
Phoneme:   ae | Start: 0.341s | End: 0.401s
Phoneme:    m | Start: 0.461s | End: 0.482s
Phoneme:   ah | Start: 0.522s | End: 0.562s
Phoneme:    n | Start: 0.602s | End: 0.622s
Phoneme:   ey | Start: 0.642s | End: 0.702s
Phoneme:   sh | Start: 0.742s | End: 0.782s
Phoneme:   ah | Start: 0.803s | End: 0.863s
Phoneme:    n | Start: 0.883s | End: 0.903s
```

```
Phoneme:   ah | Start: 0.923s | End: 0.963s
Phoneme:    n | Start: 0.983s | End: 1.003s
Phoneme:    d | Start: 1.023s | End: 1.043s
Phoneme:    t | Start: 1.083s | End: 1.103s
Phoneme:   eh | Start: 1.144s | End: 1.184s
Phoneme:    s | Start: 1.224s | End: 1.244s
Phoneme:    t | Start: 1.284s | End: 1.304s
Phoneme:   ah | Start: 1.344s | End: 1.384s
Phoneme:    m | Start: 1.404s | End: 1.425s
Phoneme:   ow | Start: 1.465s | End: 1.525s
Phoneme:    n | Start: 1.545s | End: 1.585s
Phoneme:   iy | Start: 1.625s | End: 1.665s
Phoneme:   ah | Start: 1.705s | End: 1.746s
Phoneme:    v | Start: 1.766s | End: 1.786s
Phoneme: dhah | Start: 1.806s | End: 1.926s
Phoneme:   eh | Start: 1.986s | End: 2.026s
Phoneme:    k | Start: 2.067s | End: 2.107s
Phoneme:    s | Start: 2.127s | End: 2.147s
Phoneme:    p | Start: 2.207s | End: 2.227s
Phoneme:   er | Start: 2.267s | End: 2.327s
Phoneme:    t | Start: 2.408s | End: 2.448s
Phoneme:    s | Start: 2.528s | End: 2.548s
Phoneme:   eh | Start: 2.969s | End: 3.010s
Phoneme:    n | Start: 3.050s | End: 3.070s
Phoneme:   ey | Start: 3.110s | End: 3.170s
Phoneme:    b | Start: 3.230s | End: 3.250s
Phoneme:   ah | Start: 3.270s | End: 3.310s
Phoneme:    l | Start: 3.351s | End: 3.371s
Phoneme:    d | Start: 3.411s | End: 3.431s
Phoneme:   dh | Start: 3.451s | End: 3.511s
Phoneme:   ah | Start: 3.531s | End: 3.571s
Phoneme:    k | Start: 3.591s | End: 3.611s
Phoneme:   ah | Start: 3.631s | End: 3.692s
Phoneme:    m | Start: 3.732s | End: 3.752s
Phoneme:   ih | Start: 3.792s | End: 3.832s
Phoneme:   sh | Start: 3.872s | End: 3.912s
Phoneme:   ah | Start: 3.953s | End: 3.993s
Phoneme:    n | Start: 4.013s | End: 4.033s
Phoneme:    t | Start: 4.093s | End: 4.113s
Phoneme:   uw | Start: 4.153s | End: 4.213s
Phoneme:    k | Start: 4.253s | End: 4.274s
Phoneme:   ah | Start: 4.294s | End: 4.354s
Phoneme:    n | Start: 4.374s | End: 4.414s
Phoneme:    k | Start: 4.454s | End: 4.474s
Phoneme:    l | Start: 4.534s | End: 4.554s
Phoneme:   uw | Start: 4.595s | End: 4.675s
Phoneme:    d | Start: 4.795s | End: 4.815s
Phoneme:   dh | Start: 5.217s | End: 5.277s
Phoneme:   ae | Start: 5.317s | End: 5.357s
```

```
Phoneme:    t | Start: 5.397s | End: 5.417s
Phoneme:    f | Start: 5.517s | End: 5.538s
Phoneme:   ay | Start: 5.598s | End: 5.678s
Phoneme:    v | Start: 5.758s | End: 5.778s
Phoneme:   sh | Start: 5.859s | End: 5.919s
Phoneme:    a | Start: 5.959s | End: 5.979s
Phoneme:    a | Start: 6.019s | End: 6.039s
Phoneme:    t | Start: 6.079s | End: 6.099s
Phoneme:    s | Start: 6.159s | End: 6.180s
Phoneme:    m | Start: 6.240s | End: 6.260s
Phoneme:   ey | Start: 6.300s | End: 6.360s
Phoneme:    h | Start: 6.400s | End: 6.440s
Phoneme:    h | Start: 6.460s | End: 6.481s
Phoneme:   ae | Start: 6.501s | End: 6.561s
Phoneme:    v | Start: 6.581s | End: 6.621s
Phoneme:    b | Start: 6.641s | End: 6.661s
Phoneme:   ih | Start: 6.681s | End: 6.721s
Phoneme:    n | Start: 6.761s | End: 6.781s
Phoneme:    f | Start: 6.842s | End: 6.882s
Phoneme:   ay | Start: 6.942s | End: 7.002s
Phoneme:   er | Start: 7.143s | End: 7.223s
Phoneme:    d | Start: 7.283s | End: 7.303s
```

5

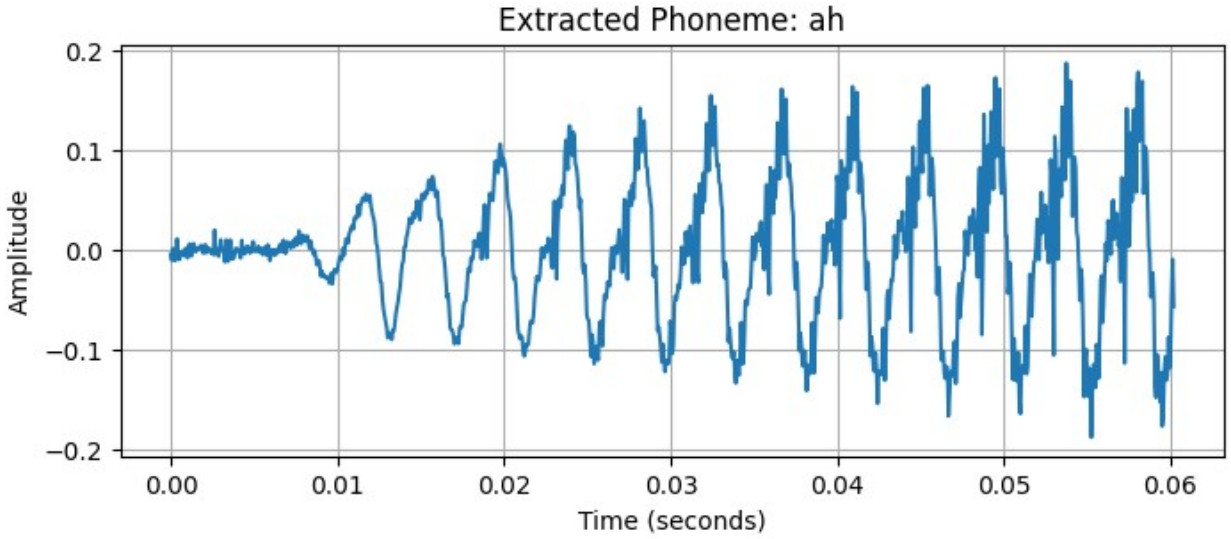
```
labels = list(set([p[0] for p in phoneme_intervals]))

print("Recognized phonemes:")
print(labels)
print()

for phoneme, start_t, end_t in phoneme_intervals:
    if phoneme.strip() != "":
        selected_phoneme = phoneme
        start_time = start_t
        end_time = end_t
        break

start_sample = int(start_time * sr)
end_sample = int(end_time * sr)

phoneme_signal = y[start_sample:end_sample]

print("Selected phoneme:", selected_phoneme)
print(f"Start time: {start_time:.3f} seconds")
print(f"End time: {end_time:.3f} seconds")
print(f"Extracted samples: {len(phoneme_signal)}")
```

```
Recognized phonemes:
['p', 'f', 'uw', 'h', 'k', 'ae', 'g', 'eh', 'er', 'ay', 't', 'b', 'n',
```

```
's', 'iy', 'ow', 'dhah', 'd', 'm', 'ah', 'dh', 'a', 'ey', 'sh', 'ih',
'l', 'z', 'v']

Selected phoneme: ah
Start time: 0.020 seconds
End time: 0.080 seconds
Extracted samples: 963
```

6

```
sf.write("extracted_phoneme.wav", phoneme_signal, sr)

t = np.linspace(0, len(phoneme_signal) / sr, len(phoneme_signal))

plt.figure(figsize=(8, 3))
plt.plot(t, phoneme_signal)
plt.xlabel("Time (seconds)")
plt.ylabel("Amplitude")
plt.title(f"Extracted Phoneme: {selected_phoneme}")
plt.grid(True)
plt.show()
```



7

```
vowels = ['aa', 'ae', 'ah', 'ao', 'eh', 'er', 'ih', 'iy', 'uh', 'uw']

print("Phoneme-wise source analysis:\n")

for phoneme, start_t, end_t in phoneme_intervals:

    if phoneme.strip() == "":
        continue
```

```
    if phoneme in vowels:
        source_type = "Voiced sound (vocal cord vibration)"
    else:
        source_type = "Unvoiced or consonantal sound (airflow
obstruction)"

    print(
        f"Phoneme: {phoneme:>4} | "
        f"Start: {start_t:.3f}s | End: {end_t:.3f}s | "
        f"Source: {source_type}"
    )
```

```
Phoneme-wise source analysis:

Phoneme:   ah | Start: 0.020s | End: 0.080s | Source: Voiced sound
(vocal cord vibration)
Phoneme:   ih | Start: 0.120s | End: 0.161s | Source: Voiced sound
(vocal cord vibration)
Phoneme:    g | Start: 0.201s | End: 0.221s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:    z | Start: 0.241s | End: 0.281s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:   ae | Start: 0.341s | End: 0.401s | Source: Voiced sound
(vocal cord vibration)
Phoneme:    m | Start: 0.461s | End: 0.482s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:   ah | Start: 0.522s | End: 0.562s | Source: Voiced sound
(vocal cord vibration)
Phoneme:    n | Start: 0.602s | End: 0.622s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:   ey | Start: 0.642s | End: 0.702s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:   sh | Start: 0.742s | End: 0.782s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:   ah | Start: 0.803s | End: 0.863s | Source: Voiced sound
(vocal cord vibration)
Phoneme:    n | Start: 0.883s | End: 0.903s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:   ah | Start: 0.923s | End: 0.963s | Source: Voiced sound
(vocal cord vibration)
Phoneme:    n | Start: 0.983s | End: 1.003s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:    d | Start: 1.023s | End: 1.043s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:    t | Start: 1.083s | End: 1.103s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:   eh | Start: 1.144s | End: 1.184s | Source: Voiced sound
(vocal cord vibration)
Phoneme:    s | Start: 1.224s | End: 1.244s | Source: Unvoiced or
```

```
consonantal sound (airflow obstruction)
Phoneme:    t | Start: 1.284s | End: 1.304s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:   ah | Start: 1.344s | End: 1.384s | Source: Voiced sound
(vocal cord vibration)
Phoneme:    m | Start: 1.404s | End: 1.425s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:   ow | Start: 1.465s | End: 1.525s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:    n | Start: 1.545s | End: 1.585s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:   iy | Start: 1.625s | End: 1.665s | Source: Voiced sound
(vocal cord vibration)
Phoneme:   ah | Start: 1.705s | End: 1.746s | Source: Voiced sound
(vocal cord vibration)
Phoneme:    v | Start: 1.766s | End: 1.786s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme: dhah | Start: 1.806s | End: 1.926s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:   eh | Start: 1.986s | End: 2.026s | Source: Voiced sound
(vocal cord vibration)
Phoneme:    k | Start: 2.067s | End: 2.107s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:    s | Start: 2.127s | End: 2.147s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:    p | Start: 2.207s | End: 2.227s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:   er | Start: 2.267s | End: 2.327s | Source: Voiced sound
(vocal cord vibration)
Phoneme:    t | Start: 2.408s | End: 2.448s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:    s | Start: 2.528s | End: 2.548s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:   eh | Start: 2.969s | End: 3.010s | Source: Voiced sound
(vocal cord vibration)
Phoneme:    n | Start: 3.050s | End: 3.070s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:   ey | Start: 3.110s | End: 3.170s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:    b | Start: 3.230s | End: 3.250s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:   ah | Start: 3.270s | End: 3.310s | Source: Voiced sound
(vocal cord vibration)
Phoneme:    l | Start: 3.351s | End: 3.371s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:    d | Start: 3.411s | End: 3.431s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:   dh | Start: 3.451s | End: 3.511s | Source: Unvoiced or
consonantal sound (airflow obstruction)
```

```
Phoneme:   ah | Start: 3.531s | End: 3.571s | Source: Voiced sound
(vocal cord vibration)
Phoneme:    k | Start: 3.591s | End: 3.611s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:   ah | Start: 3.631s | End: 3.692s | Source: Voiced sound
(vocal cord vibration)
Phoneme:    m | Start: 3.732s | End: 3.752s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:   ih | Start: 3.792s | End: 3.832s | Source: Voiced sound
(vocal cord vibration)
Phoneme:   sh | Start: 3.872s | End: 3.912s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:   ah | Start: 3.953s | End: 3.993s | Source: Voiced sound
(vocal cord vibration)
Phoneme:    n | Start: 4.013s | End: 4.033s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:    t | Start: 4.093s | End: 4.113s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:   uw | Start: 4.153s | End: 4.213s | Source: Voiced sound
(vocal cord vibration)
Phoneme:    k | Start: 4.253s | End: 4.274s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:   ah | Start: 4.294s | End: 4.354s | Source: Voiced sound
(vocal cord vibration)
Phoneme:    n | Start: 4.374s | End: 4.414s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:    k | Start: 4.454s | End: 4.474s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:    l | Start: 4.534s | End: 4.554s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:   uw | Start: 4.595s | End: 4.675s | Source: Voiced sound
(vocal cord vibration)
Phoneme:    d | Start: 4.795s | End: 4.815s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:   dh | Start: 5.217s | End: 5.277s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:   ae | Start: 5.317s | End: 5.357s | Source: Voiced sound
(vocal cord vibration)
Phoneme:    t | Start: 5.397s | End: 5.417s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:    f | Start: 5.517s | End: 5.538s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:   ay | Start: 5.598s | End: 5.678s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:    v | Start: 5.758s | End: 5.778s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:   sh | Start: 5.859s | End: 5.919s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:    a | Start: 5.959s | End: 5.979s | Source: Unvoiced or
```

```
consonantal sound (airflow obstruction)
Phoneme:    a | Start: 6.019s | End: 6.039s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:    t | Start: 6.079s | End: 6.099s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:    s | Start: 6.159s | End: 6.180s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:    m | Start: 6.240s | End: 6.260s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:   ey | Start: 6.300s | End: 6.360s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:    h | Start: 6.400s | End: 6.440s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:    h | Start: 6.460s | End: 6.481s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:   ae | Start: 6.501s | End: 6.561s | Source: Voiced sound
(vocal cord vibration)
Phoneme:    v | Start: 6.581s | End: 6.621s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:    b | Start: 6.641s | End: 6.661s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:   ih | Start: 6.681s | End: 6.721s | Source: Voiced sound
(vocal cord vibration)
Phoneme:    n | Start: 6.761s | End: 6.781s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:    f | Start: 6.842s | End: 6.882s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:   ay | Start: 6.942s | End: 7.002s | Source: Unvoiced or
consonantal sound (airflow obstruction)
Phoneme:   er | Start: 7.143s | End: 7.223s | Source: Voiced sound
(vocal cord vibration)
Phoneme:    d | Start: 7.283s | End: 7.303s | Source: Unvoiced or
consonantal sound (airflow obstruction)
```