

Audio Data

Kaggle study
2019. 8. 4

Applications

Music Generation



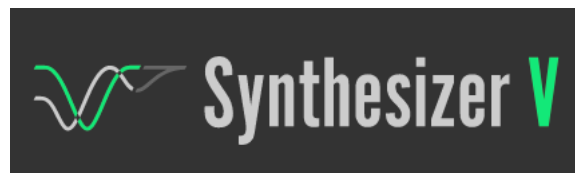
MuseNet

<https://openai.com/blog/musenet/>

Speech Synthesis

 **Tacotron** <https://google.github.io/tacotron/publications/tacotron2/index.html>

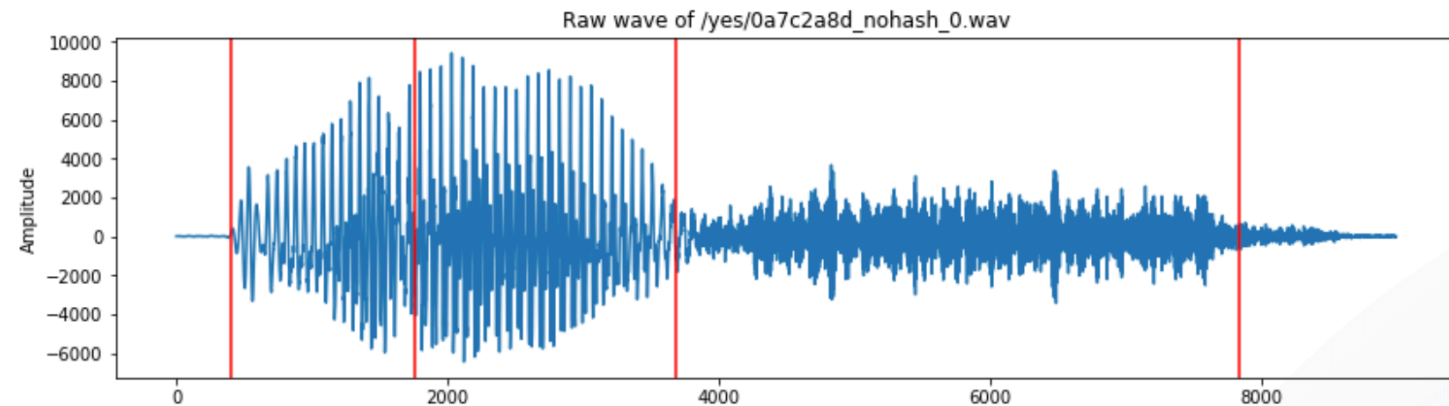
Synthesizer



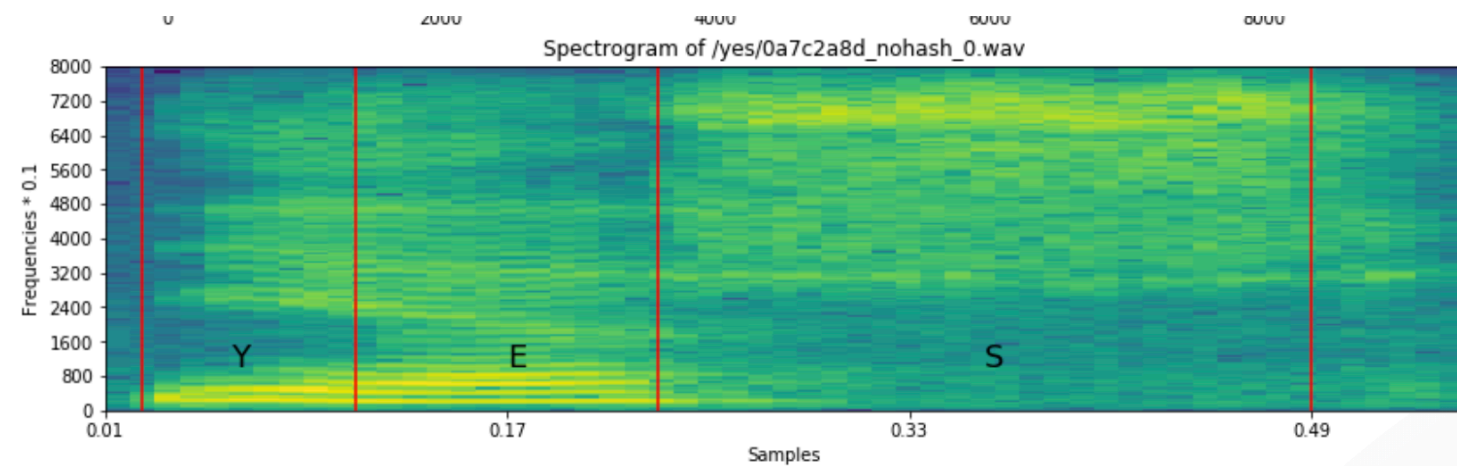
<https://www.youtube.com/watch?v=dddiw3eqwIU>

Data Representation

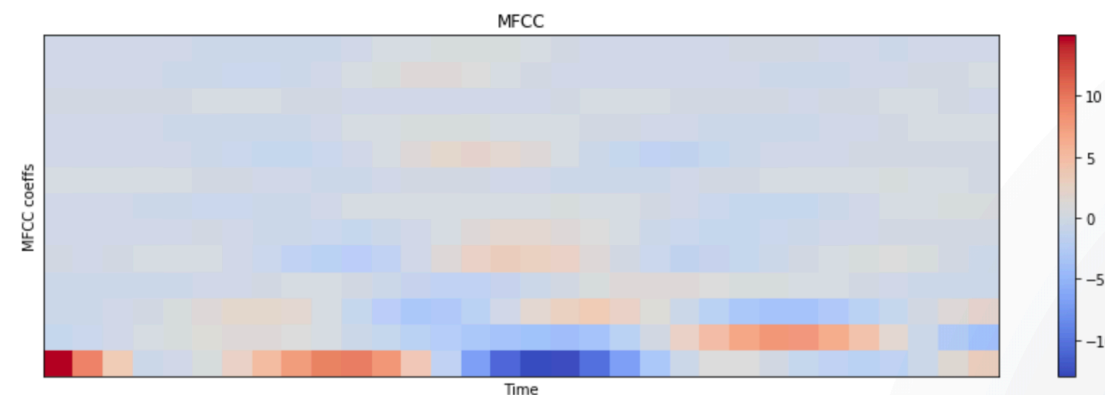
Raw data



Spectrogram



MFCC



Libraries

```
from scipy.io import wavfile
```

```
import librosa
```

```
Import lpython.display
```

```
from scipy.fftpack import fft
```

Wav File format

The Canonical WAVE file format

endian	File offset (bytes)	field name	Field Size (bytes)	
big	0	ChunkID	4	The "RIFF" chunk descriptor
little	4	ChunkSize	4	
big	8	Format	4	
big	12	Subchunk1 ID	4	The "fmt" sub-chunk describes the format of the sound information in the data sub-chunk
little	16	Subchunk1 Size	4	
little	20	AudioFormat	2	
little	22	NumChannels	2	
little	24	SampleRate	4	
little	28	ByteRate	4	
little	32	BlockAlign	2	
little	34	BitsPerSample	2	
big	36	Subchunk2 ID	4	The "data" sub-chunk Indicates the size of the sound information and contains the raw sound data
little	40	Subchunk2 Size	4	
little	44	data	Subchunk2Size	

Visualization and Data Preprocessing

<https://www.kaggle.com/davids1992/speech-representation-and-data-exploration/notebook>