pyAudioProcessing

Audio Processing, Feature Extraction, and Machine Learning Modeling

Jyotika Singh

Director of Data Science, Placemakr

Introduction

PyAudioProcessing is a Python based library for processing audio data, constructing and extracting numerical features from audio, building and testing machine learning models, and classifying data with existing pre-trained audio classification models or custom user-built models.

PyAudioProcessing contains features built in Python that were originally published in MATLAB. Some implementations were built from scratch, and some were developed on top other open-source functions. This library also creates wrapper functions using pydub and sklearn libraries with the goal to become a one-stop-shop solution for audio processing and classification.

Dependencies (v 1.1.13): numpy, sklearn, matplotlib, scipy, pydub

Audio format conversion

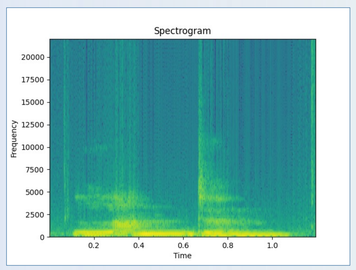
- Convert the audio from one format to another
- Includes mp4, mp3, m4a, aac -> wav

Audio de-noising/cleaning

Remove low-activity regions from your audio clip

Audio visualization

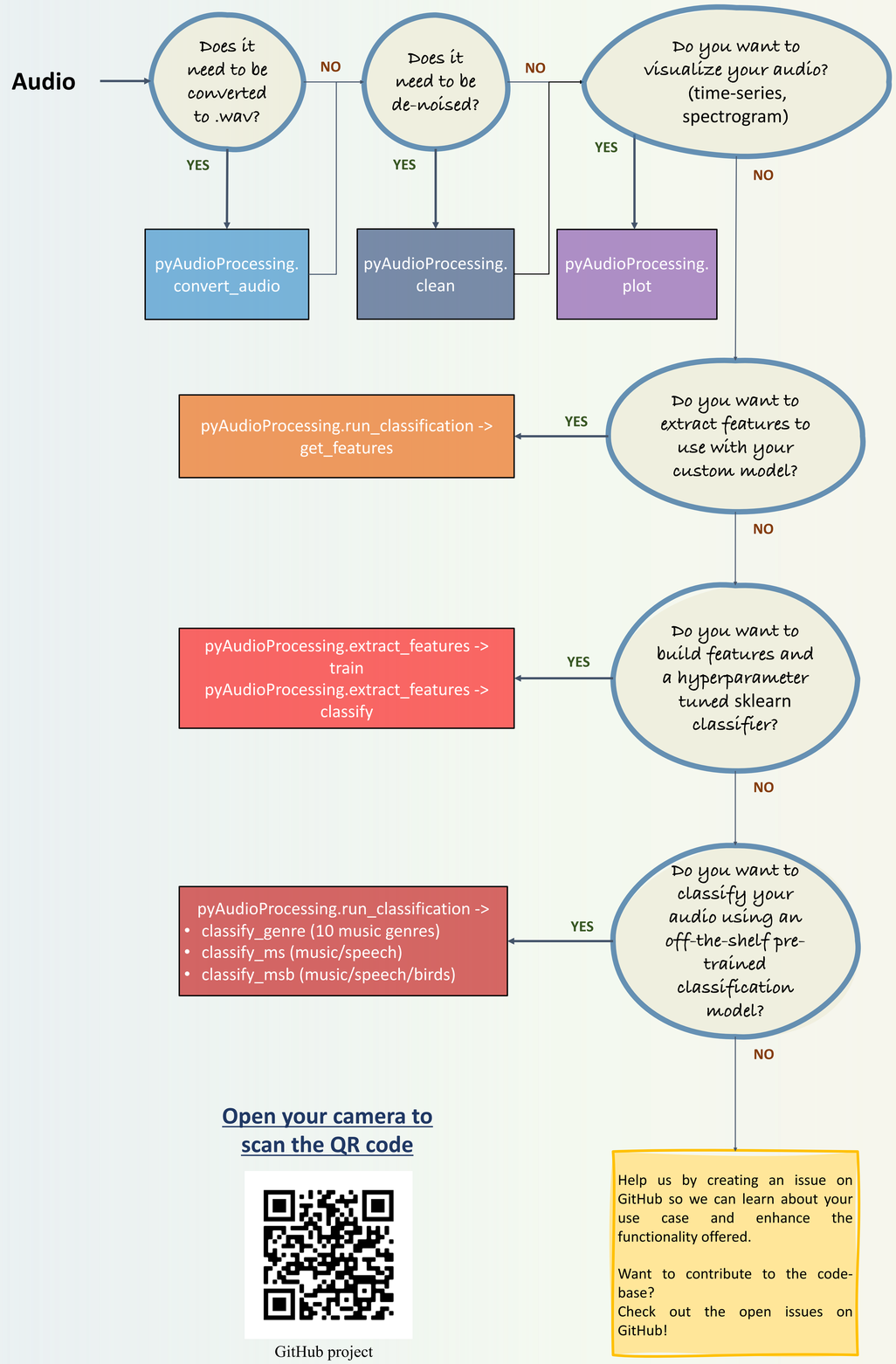
- Plot the audio in time-domain and frequency domain, including spectrograms



Author

Twitter
@jyotikasingh_

LinkedIn
jyotikasingh



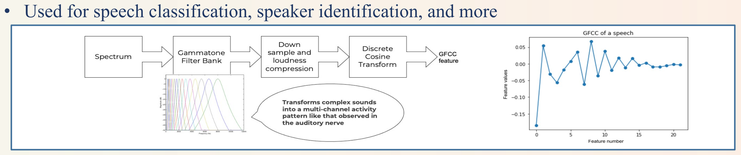
Open your camera to scan the QR code

GitHub project

Numerical feature options for training ML models

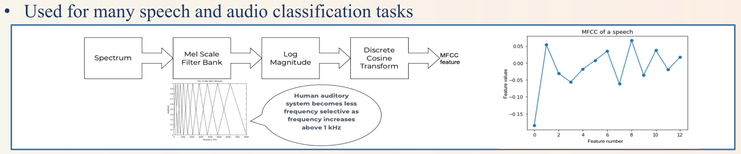
Gammatone Frequency Cepstral Coefficients (GFCC)

- Conceived to be a good approximation to the human auditory filters
- Used as the front-end simulation of the cochlea
- Used for speech classification, speaker identification, and more



Mel Frequency Cepstral Coefficients (MFCC)

- Mel-frequency cepstrum is a representation of the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a nonlinear mel-scale of frequency
- Used for many speech and audio classification tasks



Temporal features

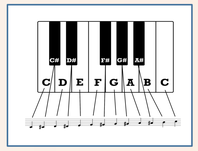
- Signal information in its time domain representation
- Examples include signal energy, entropy, zero crossing rate, etc.

Spectral features

- Signal information contained in the frequency domain representation
- Examples include fundamental frequency, spectral entropy, spectral spread, spectral flux, spectral centroid, spectral roll-off, etc.

Chroma

- Popular for music
- 12 categories: A, A#, B, C, C#, D, D#, E, F, F#, G, G#
- Captures the harmonic and melodic attributes of audio, while being robust to changes in timbre and instrumentation



Train sklearn classifiers

- Options: SVM, SVM RBF, Random forest, Logistic regression, k-NN, Gradient boosting, Extra trees
- Runs automatic hyper-parameter tuning
- Generates evaluation: Cross-validation confusion matrix
- Allows the user to use the trained classifier to classify audio files

Pre-trained models

Music genre classifier

- Support Vector Machine (SVM) classifier
- 10 classes: pop, metal, disco, blues, reggae, classical, rock, hip-hop, country, jazz
- Trained using MFCC, GFCC, spectral features, and chroma features

Class	*Metric*		
	Accuracy	*Precision*	*F1*
pop	72.36%	78.63%	75.36%
met	87.31%	85.52%	86.41%
dis	62.84%	59.45%	61.10%
blu	83.02%	72.96%	77.66%
reg	79.82%	69.72%	74.43%
cla	90.61%	86.38%	88.44%
rock	53.10%	51.50%	52.29%
hip	60.94%	77.22%	68.12%
cou	58.34%	62.53%	60.36%
jazz	78.10%	85.17%	81.48%

Music/Speech classifier

- Support Vector Machine (SVM) classifier
- 2 classes: music, speech
- Trained using MFCC, spectral features, and chroma features

Class	*Metric*		
	Recall	*Precision*	*F1*
Music	97.60%	98.79%	98.19%
Speech	98.80%	97.63%	98.21%

Music/Speech/Bird classifier

- Support Vector Machine (SVM) classifier
- 3 classes: music, speech, bird sound
- Trained using MFCC, spectral features, and chroma features

Class	*Metric*		
	Recall	*Precision*	*F1*
Music	94.60%	96.93%	95.75%
Speech	97.00%	97.79%	97.39%
Birds	100.00%	96.89%	98.42%