

python_speech_features

This library provides common speech features for ASR including MFCCs and filterbank energies. If you are not sure what MFCCs are, and would like to know more have a look at this [MFCC tutorial](#)

[Project Documentation](#)

Installation

This [project is on pypi](#)

To install from pypi:

```
pip install python_speech_features
```

From this repository:

```
git clone https://github.com/jameslyons/python_speech_features
```

```
python setup.py develop
```

Usage

Supported features:

- Mel Frequency Cepstral Coefficients
- Filterbank Energies
- Log Filterbank Energies
- Spectral Subband Centroids

[Example use](#)

From here you can write the features to a file etc.

MFCC Features

The default parameters should work fairly well for most cases, if you want to change the MFCC parameters, the following parameters are supported:

```
python

def mfcc(signal,samplerate=16000,winlen=0.025,winstep=0.01,numcep=13,
         nfilt=26,nfft=512,lowfreq=0,highfreq=None,preemph=0.97,
         ceplifter=22,appendEnergy=True)
```

Parameter	Description
signal	the audio signal from which to compute features. Should be an N*1 array

Parameter	Description
samplerate	the samplerate of the signal we are working with.
winlen	the length of the analysis window in seconds. Default is 0.025s (25 milliseconds)
winstep	the step between successive windows in seconds. Default is 0.01s (10 milliseconds)
numcep	the number of cepstrum to return, default 13
nfilt	the number of filters in the filterbank, default 26.
nfft	the FFT size. Default is 512
lowfreq	lowest band edge of mel filters. In Hz, default is 0
highfreq	highest band edge of mel filters. In Hz, default is samplerate/2
preemph	apply preemphasis filter with preemph as coefficient. 0 is no filter. Default is 0.97
ceplifter	apply a lifter to final cepstral coefficients. 0 is no lifter. Default is 22
appendEnergy	if this is true, the zeroth cepstral coefficient is replaced with the log of the total frame energy.
returns	A numpy array of size (NUMFRAMES by numcep) containing features. Each row holds 1 feature

Filterbank Features

These filters are raw filterbank energies. For most applications you will want the logarithm of these features. The default parameters should work fairly well for most cases. If you want to change the fbank parameters, the following parameters are supported:

python

```
def fbank(signal,samplerate=16000,winlen=0.025,winstep=0.01,
         nfilt=26,nfft=512,lowfreq=0,highfreq=None,preemph=0.97)
```

Parameter	Description
signal	the audio signal from which to compute features. Should be an N*1 array
samplerate	the samplerate of the signal we are working with

Parameter	Description
winlen	the length of the analysis window in seconds. Default is 0.025s (25 milliseconds)
winstep	the step between seccessive windows in seconds. Default is 0.01s (10 milliseconds)
nfilt	the number of filters in the filterbank, default 26.
nfft	the FFT size. Default is 512.
lowfreq	lowest band edge of mel filters. In Hz, default is 0
highfreq	highest band edge of mel filters. In Hz, default is samplerate/2
preemph	apply preemphasis filter with preemph as coefficient. 0 is no filter. Default is 0.97
returns	A numpy array of size (NUMFRAMES by nfilt) containing features. Each row holds 1 feature vector. value is the energy in each frame (total energy, unwindowed)

Reference

sample english.wav obtained from:

```
wget http://voyager.jpl.nasa.gov/spacecraft/audio/english.au
sox english.au -e signed-integer english.wav
```