姓名：吳政緯

學號：E64062076

多媒體系統與應用 Project1 – Speech Processing，使用python完成

檔案說明：

- Project1\_Speech Process.docx:程式碼說明

- 圖檔，程式生成圖片的截圖:

- Waveform.png

- Zero-crossing-rate.png

- Energy\_contour.png

- Pitch\_contour.png

- End\_point\_detection.png

- sound.m4a:使用的音訊檔案

- Source.py:可執行的python程式碼

--------------------------------------------------------------

import matplotlib.pyplot as plt

import numpy as np

import math

import librosa

import librosa.display

import IPython.display as ipd

引入matplot以及numpy作為繪圖與數學運算，librosa作為音訊處理的工具

y, sr = librosa.load('sound.m4a')

frame\_len = int(0.02 \* sr)

frame\_shift = int(0.01 \* sr)

frameNum = math.ceil(len(y)/256)

讀取檔案並計算所需的變數

--------------------------------------------------------------

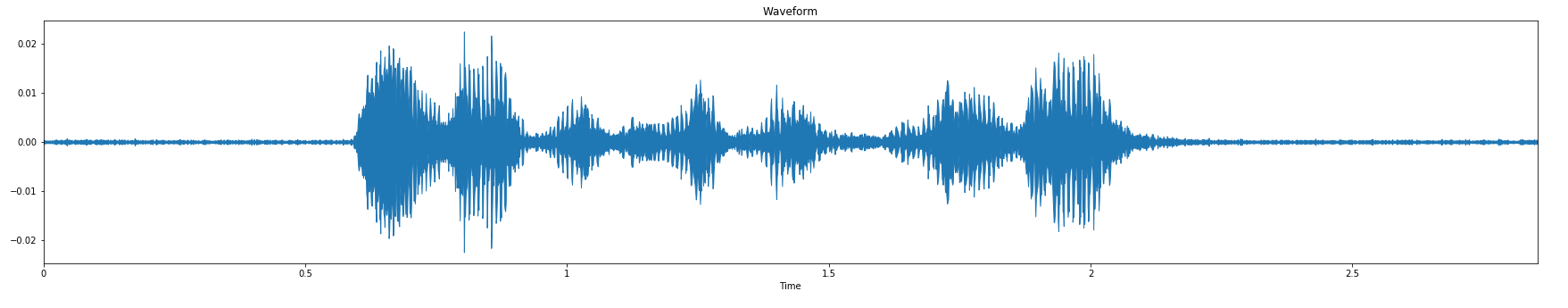
plt.figure(figsize=(30, 5))

librosa.display.waveplot(y, sr=sr)

plt.title('Waveform')

plt.show()

輸出waveform圖形



--------------------------------------------------------------

zcr = np.zeros((frameNum,1))

for i in range(frameNum):

curFrame = y[np.arange(i\*256,min(i\*256+256,len(y)))]

zcr[i] = sum(curFrame[0:-1]\*curFrame[1::]<=0)

zcr\_time = np.arange(0, len(zcr)) \* (len(y)/len(zcr) / sr)

plt.figure(figsize=(30, 5))

plt.plot(zcr\_time, zcr)

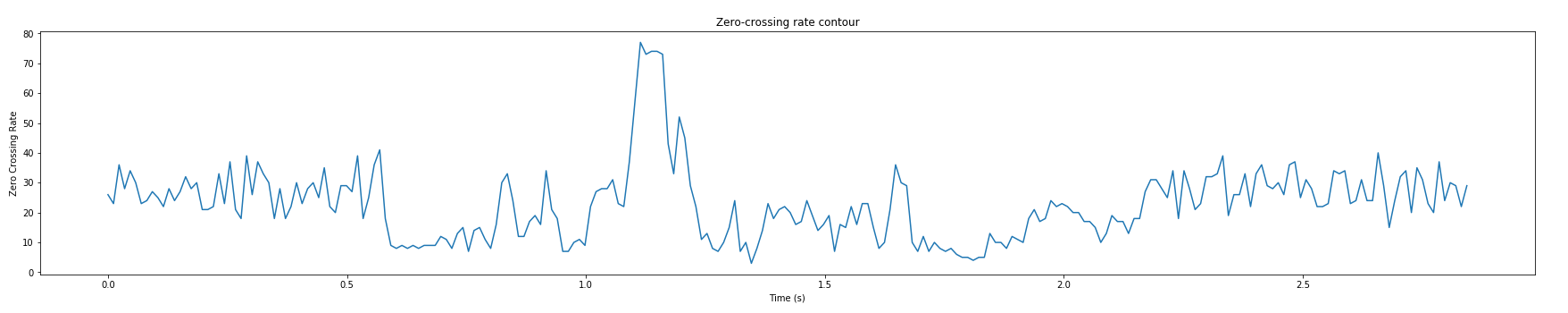
plt.ylabel("Zero Crossing Rate")

plt.xlabel("Time (s)")

plt.title("Zero-crossing rate contour")

plt.show()

定義過零率並輸出圖形



--------------------------------------------------------------

ste = np.zeros((frameNum,1))

for i in range(frameNum):

curFrame = y[np.arange(i\*256,min(i\*256+256,len(y)))]

ste[i] = sum(curFrame[0:-1]\*curFrame[1::])

ste\_time = np.arange(0, len(ste)) \* (len(y)/len(ste) / sr)

plt.figure(figsize=(30, 5))

plt.plot(ste\_time,ste)

plt.xlabel("Time (s)")

plt.ylabel("Energy")

plt.title("Energy Contour")

plt.grid('on')

使用short-time energy作為energy圖表的計算，並輸出圖形

一張含有 螢幕擷取畫面 的圖片

自動產生的描述

--------------------------------------------------------------

frames = librosa.util.frame(y, frame\_length=frame\_len, hop\_length=frame\_shift)

pitches, magnitudes = librosa.core.piptrack(y, sr=sr, hop\_length=frame\_shift, threshold=0.75)

pitch\_track = []

for i in range(0,pitches.shape[1]):

pitch\_track.append(np.max(pitches[:,i]))

x=np.r\_[2\*pitch\_track[0]-pitch\_track[10::-1],pitch\_track,2\*pitch\_track[-1]-pitch\_track[-1:-11:-1]]

temp = np.hanning(11)

temp = np.convolve(temp/temp.sum(),x,mode='same')

pitch\_smoothtrack = temp[11:-10]

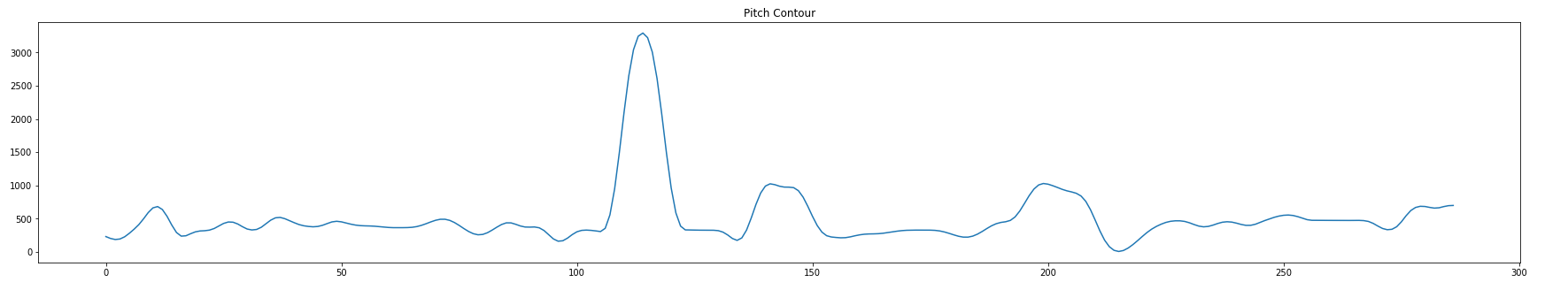
plt.figure(figsize=(30, 5))

plt.plot(pitch\_smoothtrack)

plt.title("Pitch Contour")

plt.show()

使用最大瞬間頻率來推測音高，平滑化圖表後輸出圖形



--------------------------------------------------------------

root\_mean\_square\_energy = librosa.feature.rms(y, frame\_length=frame\_len, hop\_length=frame\_shift)

rms = root\_mean\_square\_energy[0]

rms = librosa.util.normalize(rms, axis=0)

zero\_crossing\_rate = librosa.feature.zero\_crossing\_rate(y, frame\_length=frame\_len, hop\_length=frame\_shift, threshold=0)

zcr = zero\_crossing\_rate[0]

frame\_indexs = np.where( (rms > 0.3) | (zcr > 0.5) )[0]

start\_indexs = [frame\_indexs[0]]

end\_indexs = []

index\_shape = np.shape(frame\_indexs)

for i in range(index\_shape[0]-1):

if (frame\_indexs[i + 1] - frame\_indexs[i]) != 1:

start\_indexs.append(frame\_indexs[i+1])

end\_indexs.append(frame\_indexs[i])

end\_indexs.append(frame\_indexs[-1])

start\_indexs = np.array(start\_indexs)

end\_indexs = np.array(end\_indexs)

start\_temp = start\_indexs \* frame\_shift / sr

end\_temp = end\_indexs \* frame\_shift / sr

plt.figure(figsize=(30, 5))

temp = np.linspace(0, len(y)/sr, len(y))

plt.plot(temp, y)

for start, end in zip(start\_temp, end\_temp):

plt.axvline(x=start, color='#2ca02c')

plt.axvline(x=end, color='#d62728')

plt.title("End point detection")

plt.show()

計算邊界來預測字的起始與結尾，並用紅綠線標示在waveform上

一張含有 水, 船 的圖片

自動產生的描述