

12.18.2017

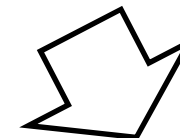
음성파일 위치 추적기

Jinhyun Kim

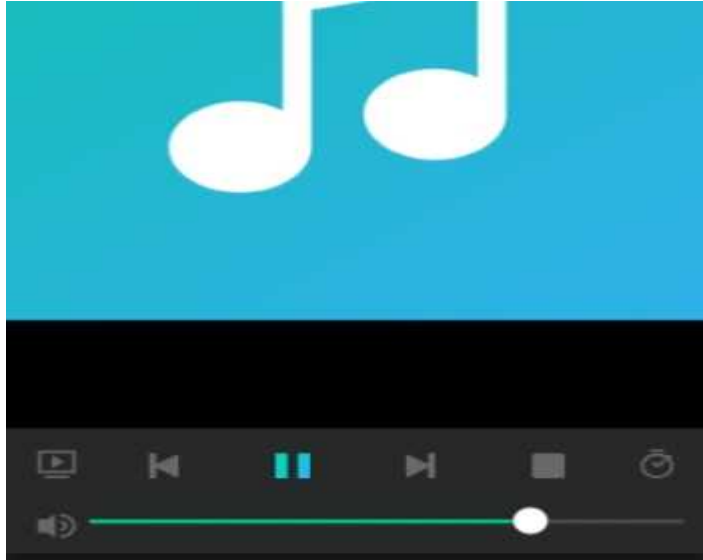
**Signal Processing Systems Laboratory
Department of Electronic Engineering
Sogang University, Seoul, Korea**

Introduction

지금은 회의중

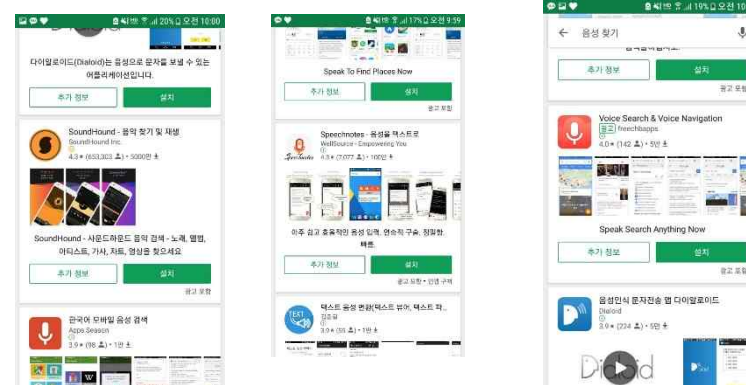


Introduction



플레이 시간 2시간 30분...

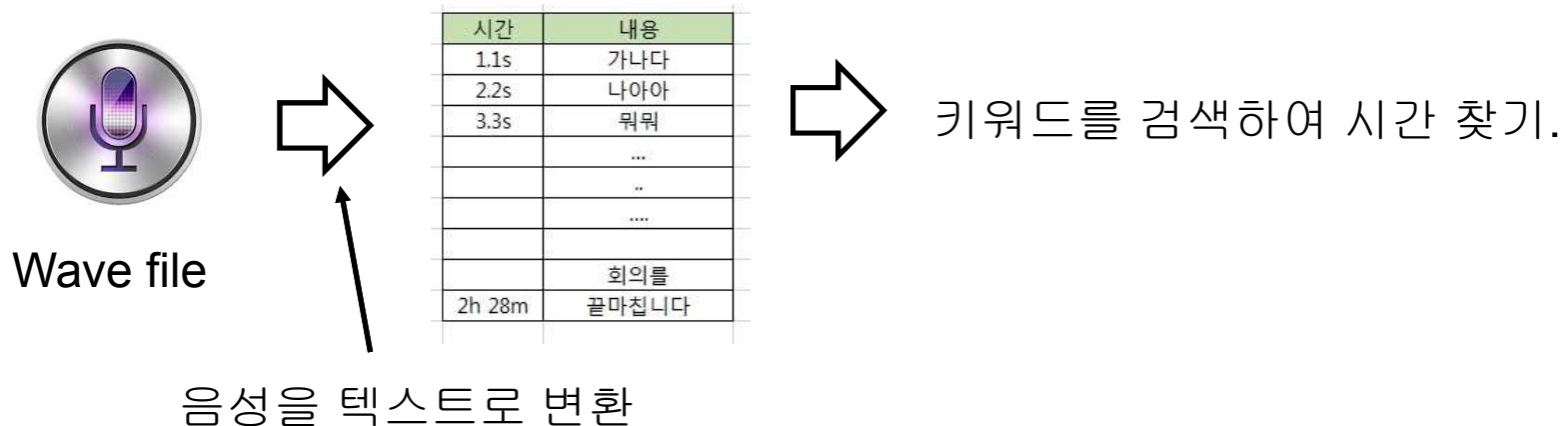
필요한 정보가 있는 위치가 대체 어디지...??



현존하는 어플 없음...

자동으로 찾을 수 있는 방법이 없을까..?

Method1



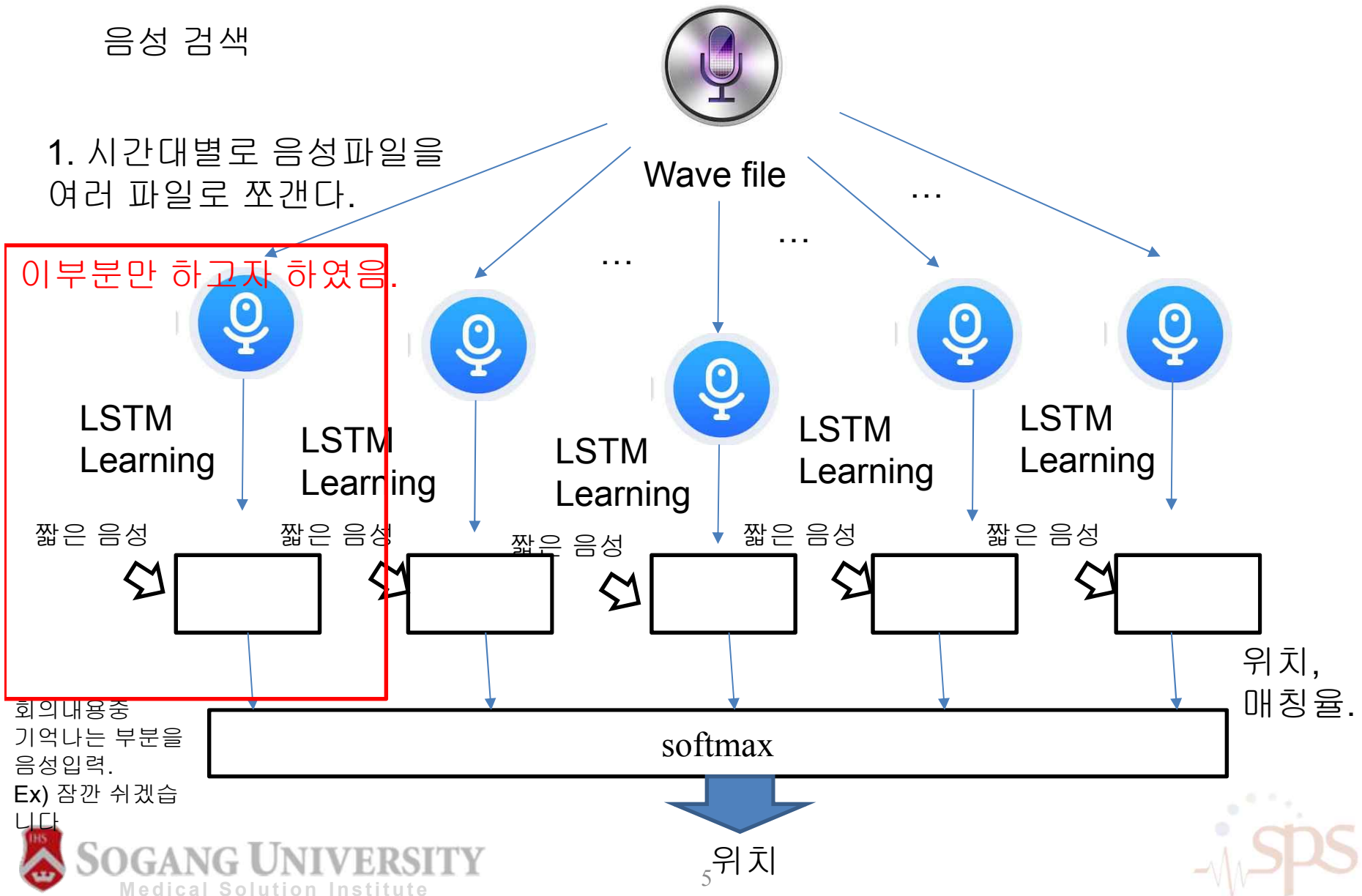
현재 문제점 : 현재 시간정보도 같이 추출할 수 있는 음성→텍스트 변환기가 없다.
텍스트 변환기의 인식률이 그리 좋지 않다.
가장큰 문제점 : 프로젝트로 하기엔 너무 뻥세다.

Method2

음성 검색

1. 시간대별로 음성파일을 여러 파일로 쪼갬다.

이부분만 하고자 하였음.



project



X : size 8000

Wave file

16bit mono input

3분 음성 :
6011537 sample

Sampling rate :
44100

256 LSTM

Dropout(0.2)

256 LSTM

Dropout(0.2)

Dense

Y : 위치



Result

■ ...

```
model.add(Dropout(0.2))
model.add(LSTM(256))
model.add(Dropout(0.2))
model.add(Dense(256, activation='softmax'))
model.compile(loss='categorical_crossentropy', optimizer='adam')
# define the checkpoint
filepath="weights-improvement-{epoch:02d}-{loss:.4f}-bigger.hdf5"
checkpoint = ModelCheckpoint(filepath, monitor='loss', verbose=1, save_best_only=True,
mode='min')
callbacks_list = [checkpoint]
# fit the model
model.fit(X, y, epochs=50, batch_size=64, callbacks=callbacks_list)
```

IndexError Traceback (most recent call last)

<ipython-input-87-e85d5e43ad4f> in <module>()

1 # one hot encode the output variable

→ 2 y = np_utils.to_categorical(dataY)

3

4 # define the LSTM model

5 model = Sequential()

C:\Users\7700JIN\Anaconda3\lib\site-packages\keras\utils\np_utils.py in to_categorical(y, num_classes)

23 n = y.shape[0]

24 categorical = np.zeros((n, num_classes))

→ 25 categorical[np.arange(n), y] = 1

26 return categorical

27

IndexError: index -9888 is out of bounds for axis 1 with size 9214

conclusion

- Preprocessing이 필요 : sampling rate 44100hz로는 input data수가 너무 많아 오래걸린다. Decimation 하여 8000hz, 4000hz로 줄이는 것이 필요.
- 위치정보를 너무 세세하게 하면 one hot encoding 의 수가 너무 많아진다. 세분화된 파일 당 256레벨 정도로 바꾼다.

**THANK YOU FOR
YOUR ATTENTION**