

3/17/2023 실습 1

audio file recording & 음성인식기 실험.

1-1. 숫자음 10개를 10번씩 발성하여 간단한 DB를 만든다.

① audio recording tools

① Goldwave (아마 shareware, freeware는 아님)

무난하게 사용가능. windows only, ti.

② Adobe Audition (구 CoolEdit)

가장 powerful 하지만 유료 software
windows only

③ Audacity

Open Source freeware

cross platform (windows, linux, etc.)

multi track 지원

* ④ Ocenaudio (추천)

Open source (x) freeware (o) deb/dpkg...

Cross platform (win / linux / etc)

Spectral view 지원.

- 위의 tool들 중 적절히 것을 설치하고 녹음

- 녹음 형식

16 bit PCM / 16 kHz / .wav

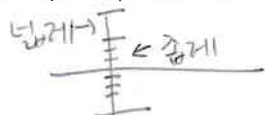
* (raw) PCM: 한 sample을 2 bytes로 직접 저장

lossless {

file size = $2 \times F_s \times T_{(sec)}$

DPCM: delta PCM. raw sample 대신 $x[n] - x[n-1]$ 저장
1~2 bits 줄일 수 있음

ADPCM: adaptive DPCM. #bits를 유동적으로

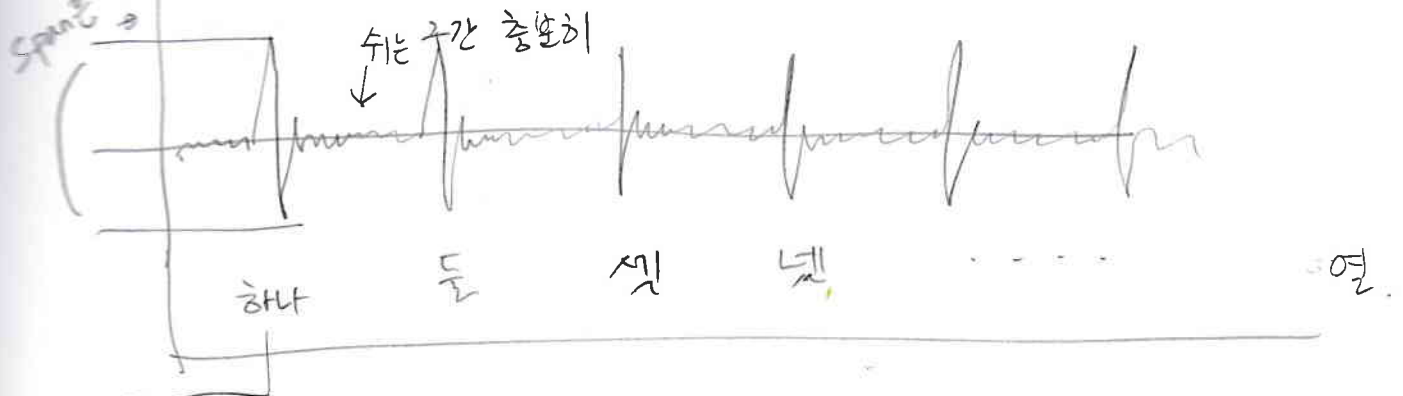
* μ -law, A-law: 8kHz only. 큰 값은 넓게 quantize.

long distance phone에 썼으나 (유선)

digital
지금은 안 씀

- 녹음 방법

$\frac{1}{2} \sim \frac{1}{3}$ (너무 작으면 안 됨)



→ 녹음된 파일을 숫자단위로 분할하여 저장한다.

kdigits0-1.wav

↑ ↑
session 숫자
번호

kdigits 0 - 0 . wav

10번째는 0으로

kdigits 9 - 1 . wav

kdigits 9 - 0 . wav

저장 segmented / 0 /

segmented / 9 /

kdigits 0 - 0 . wav

kdigits 9 - 0 . wav

kdigits 9 - 9 . wav

- 잡음

마이크 quality, PC / notebook 종류, 전원제 따라 다양한 잡음

dc (direct current)



ac (alternating current), 50 ~ 60 Hz

Gaussian white noise



- 음성인식 실험.

python package hmmlearn을 이용하여
간단한 단어인식기 만들 수 있다.

github.com/jayaram125/Single-Word-Speech
사용하여 인식해 본다.

- keras에 업로드된 수록된 코드는 사용

kdigits/ — hmmSr_kdigits.py ~~as~~
└ segmented → 100개.
└ unsegmented → 10개

- file "hmmSr_kdigits.py" 실행

python3 hmmSr_kdigits

(python 3가 default이면 그냥 python으로 실행)

* Windows에서는 Anaconda가 package들이
pre-install되어 있어서 사용하기 편함.