

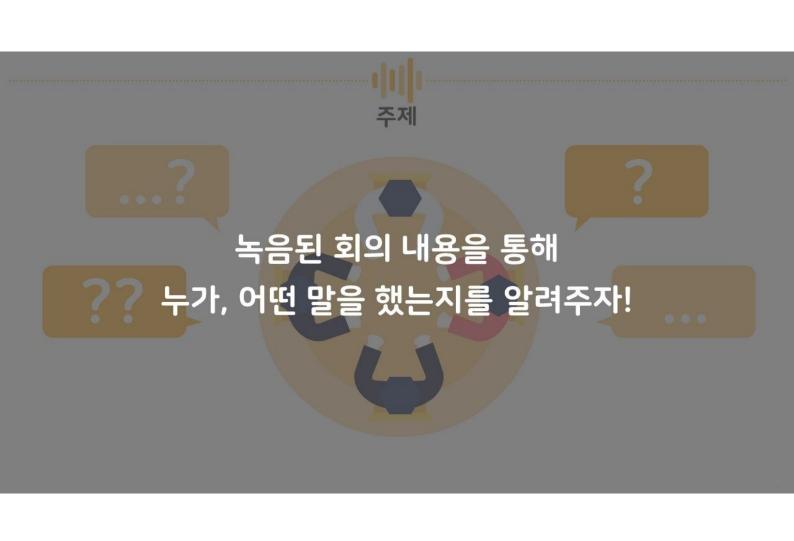
우리 팀 서기는 내 컴퓨터! 화자인식과 음성인식을 이용한 회의 Dialogue

김 강 열 / 박 이 삭 / 이 다 경 / 이 태 권 / 조 혜 진



- 01 주 제
- 02 프로젝트 개발방법
- 03
 구현 및 시연

 04
 팀 소 개







North America Voice Recognition Market - Forecast to Grow at a CAGR of 10.8% During 2017-2023 - Research and Markets

December 06, 2017 05:48 AM Eastern Standard Time

Voice Recognition Technology Could DUBL Resea **Pose New Threat To Advertisers**

st period (20

2023).

reco

iden

Voic

thar

by Laurie Sullivan, Staff Writer @lauriesullivan, December 19, 2017 Cops figuring out criminals by voice recognition

Voice Overview Integra that us to sign Description are "re-invented" with built-in wifi and bright Evaluation Based are voice interfaces. are fur Prizes But, for independent makers and entrepreneu a simple speech detector using free, open dat Timeline voice recognition datasets require preprocessing befo

TensorFlow recently released the Speech Commands

includes 65,000 one-second long utterances of 30 sh

🛗 January 1, 2018 🛔 Kaplan Contributor pers help .icx-toolbar{padding: 0 0 5px 0;} tech Sma equi network model can be built on them. To help with this beco

출처: Search Inside WORDNETDAILY.COM BusinessWire

Tutorials & More Info

Tensorflow Speech Recognition Challenge

The national and international groups that address crime on a global scale say their new strategy to identify active criminals through nothing more than hearing their voice has passed its latest test, explains a new report in While the idea may be scary to privacy advocates, the police forces of the world say they need a way to quickly ident

criminal suspect.

Their new effort is called the Speaker Identification Integrated Project, and authorities say it recently was tested by the Kingdom's Metropolitan Police Service and Portugal's Policia Judiciaria.



" 회의 시작하겠습니다. "



프로젝트 개발방법



1. 데이터 수집



2. 데이터 처리



3. Speaker Recognition

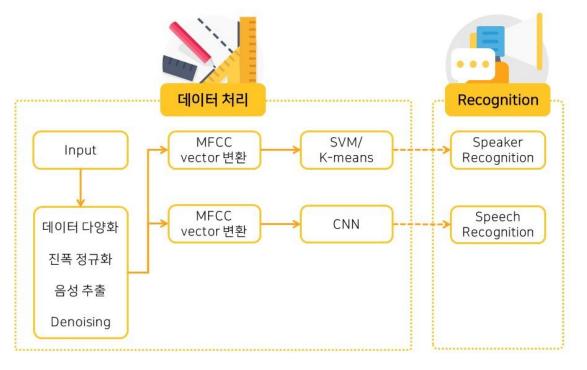


3. Speech Recognition



프로젝트 개발방법



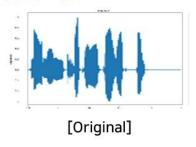


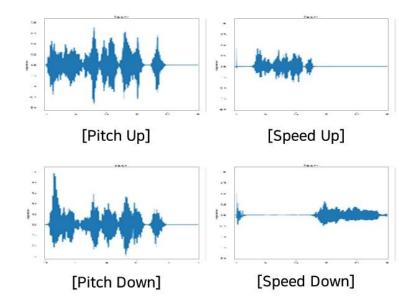
7



분석의 정확도를 높이기 위해 <mark>데이터 개수 확장</mark>

Pitch(높낮이) 조절 Speed(속도) 조절





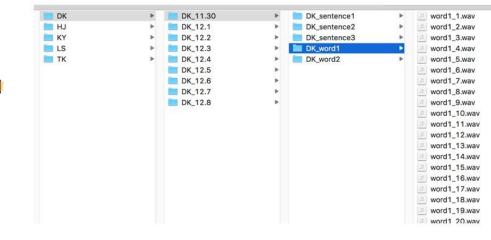


팀원 5명, 녹음

사과, 바나나, 시계, 안경 등 2900개 단어

'나는 밥을 먹었다' 등 1350개 문장

총 4250개음성 데이터







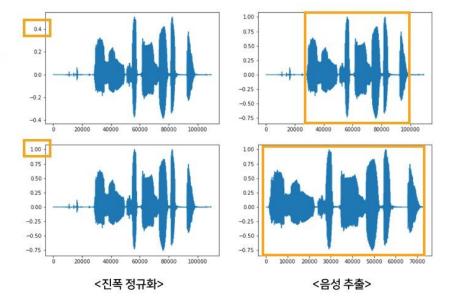


진폭 정규화 / 음성 추출

음성 마다 크기가 다른 진폭을 -1 ~ 1로 정규화

진폭이란? 주기적으로 진동하는 파의 진동

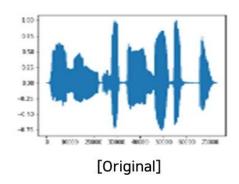
분석의 정확도를 높이기 위해 음성 부분만 추출

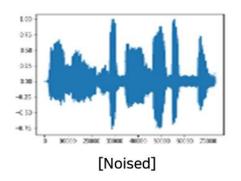




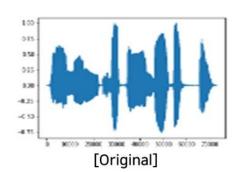
Noise가 합성된 파일을 Denoising 하는 방법을 통해 Denoising의 성능을 평가

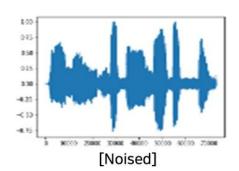
Train Data가 아닌 다른 환경에서의 Data 처리를 위한 단계

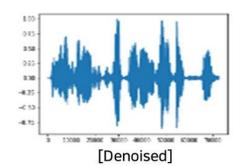






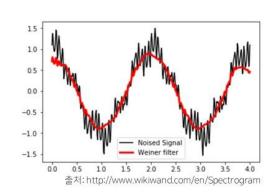








위너 필터(Wiener Filter)란? 불규칙적으로 섞여 있는 noise sound를 smoothing 해주고 기존 신호의 추정치를 추출하는 방법





Denoising_Spectrogram 변환

스펙트로그램(Spectrogram)이란? Time, Frequency, Amplitude를 각 축으로 하는 3-dimensional 음성 표현

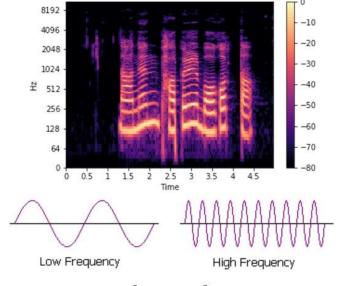
Time: 시간

Amplitude: 진폭

Frequency: 주파수(단위: Hz)



[Amplitude]



[Frequency]

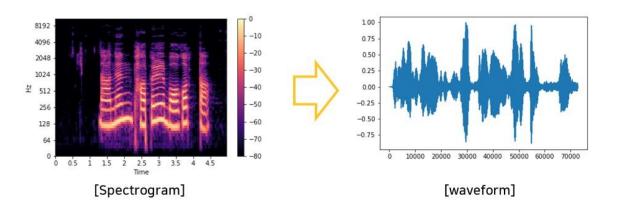
13



Denoising_Time domain waveform

Spectrogram이었던 Audio 다시 waveform으로 변환하는 작업

Speaker recognition을 위한 Data로 쓰기 위해

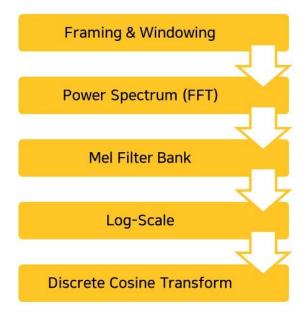


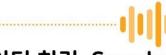




MECC

MFCC (Mel Frequency Cepstral Coefficient) 입력된 신호에서 노이즈 및 배경 소리로부터 인간이 소리를 듣는 방식을 반영해서 추출한 음성 특징 벡터

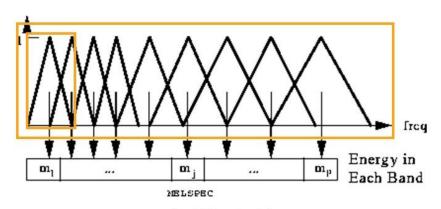




데이터 처리_Speaker Recognition MFCC_Mel Filter Bank

Mel Filter란?

인간이 잘 들을 수 있는 저주파대의 정보를 상대적으로 더 반영하기 위한 필터

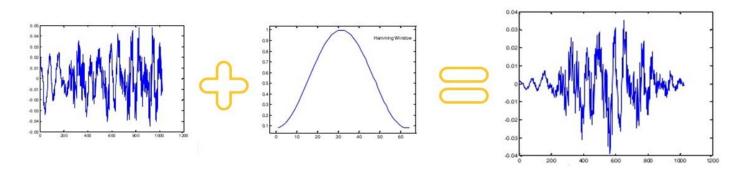


[Mel filter bank]



MFCC_Framing & Windowing

Framing이란? 음성 신호를 일정 크기의 작은 프레임으로 나누는 작업 Windowing이란? Cosine Window를 씌워서 특정 구간에 가중치를 주는 작업

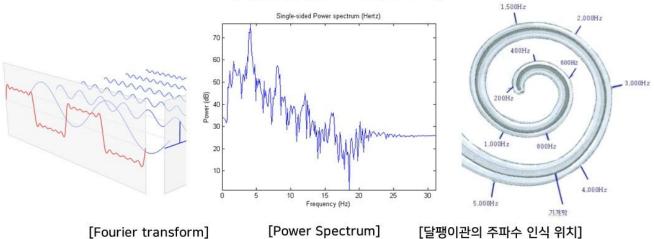




MFCC_Power Spectrum (FFT)

푸리에 변환(Fourier transform)이란? 임의의 입력 신호를 다양한 주파수를 갖는

주기함수들의 합으로 분해하여 표현

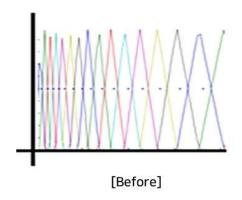


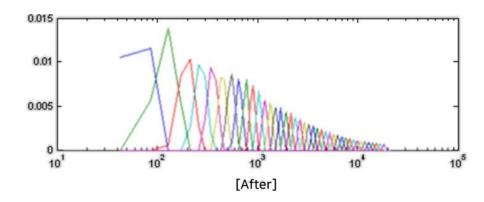


MFCC_Log-Scale&Discrete Cosine Transform

log Scale: 인간이 소리를 감지하는 형태(Non-Linear Scale)와 유사하게 만들기 위한 작업 Discrete Cosine Transform(이진코사인변환): 에너지들 사이에 <mark>상관관계를 감소&분리</mark> 해주는 역할을 한다.

Filter Bank는 모두 Overlapping 되어 있기 때문에

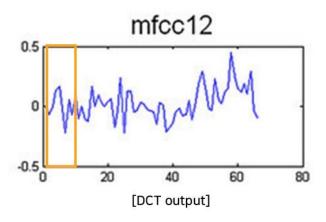






MFCC: DCT까지 완료한 값들(일반적으로 26개) 중에 1~12 or 13번째 까지의 coefficient만 선택

인간이 들을 수 있는 영역인 저주파 대에 집중하기 위함





분류 / 군집화(클러스터링)

군집화

입력 데이터의 분포 특성을 분석하여 임의의 복수 개의 그룹으로 나누는 문제

클래스에 대한 정보 없이 단순히 입력 가능

분류

주어진 데이터 집합을 이미 <mark>정의된 클래스내</mark>에서 구분하는 문제

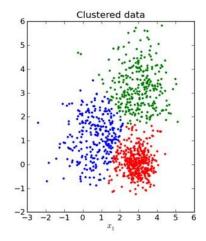
입력 데이터와 각 데이터의 클래스를 <mark>학습하는 과정</mark>이 필요



군집화_K-means

K 평균 알고리즘(K-Means)이란?

- 주어진 데이터를 <mark>k개의 클러스터</mark>로 묶는 알고리즘
- 기계 학습의 분야 중 하나로 <mark>학습하지 않은 데이터에</mark> 자동으로 레이블을 지정



[Accuracy 80%]

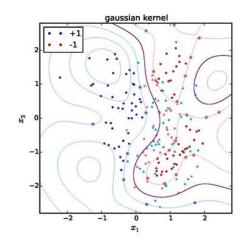


데이터 처리_Speaker Recognition 분류_SVM

SVM(Support Vector Machine)이란?

- 데이터가 사상 된 공간에서 경계로 표현되며 그 중 가장 큰 폭을 가진 경계를 찾는 알고리즘
- 기계 학습의 분야 중 하나로 패턴 인식, 자료 분석을 위한 지도 학습 모델

선형 분류와 더불어 비선형분류에서도 사용가능

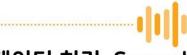


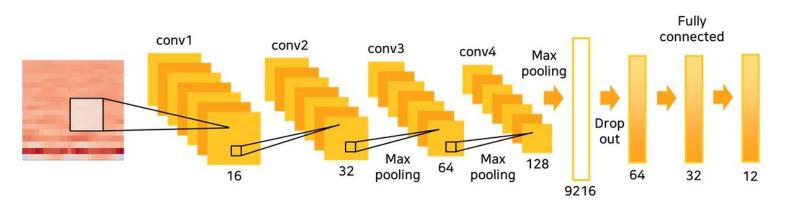
[Accuracy 99.02%]





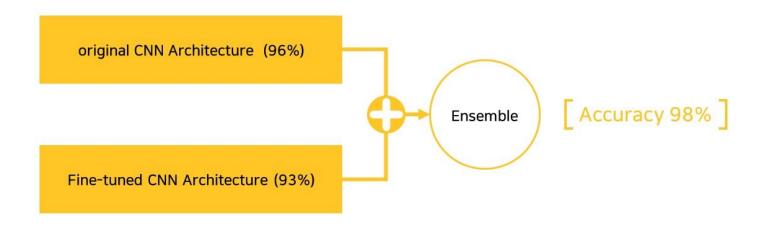






[CNN Architecture]

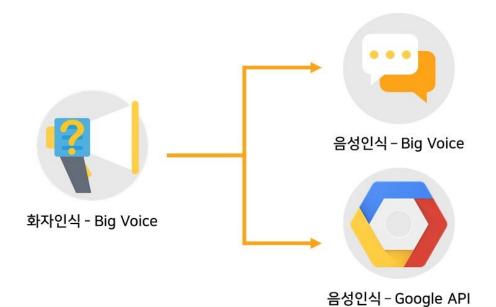






Recognition

Speaker& Speech Recognition



안녕하세요,
Big_Voice 팀 입니다,
발표는 어떠셨나요?
저희 발표는 여기까지 입니다,
지금까지 들어주셔서, 감사합니다





김강열 고려대학교 통계학과 투빅스 8기



박이삭 인하대학교 통계학과 투빅스 5기



이다경 인하대학교 통계학과 투빅스 5기



이태권 인하대학교 컴퓨터공학과 투빅스 8기



조혜진 인하대학교 컴퓨터공학과 투빅스 8기

