

사회맞춤형 산학협력 선도대학 (LINK+) 육성사업 스마트클라우드 인공지능트랙
산학공동기술개발과제 연구발표회 2020.09.22 (화) 오후 1시

제4과제 : 음성 데이터 기반 정보 추출연구 (AI-04)



연구관리자: 강환일 교수
연구총괄: 대표이사: 신대진 이드웨어㈜



목차

1. 연구원 소개

(강환일: 명지대 교수)

2. 연구목표 및 연구내용

3. 연구기간

4. 음성인식 기술현황

5. 연구의 미래

6. 연구의 내용 및 범위

(신대진: 이드웨어 대표이사)

6.1 데이터셋

6.2 음성인식 알고리즘

6.3 IOT 기반 음성인식 장비

7. 연구의 분담 및 계획

8. 결론

9. 참고문헌 및 참고자료

1. 산학 협동 연구원 구성

구분	제1연구모임	제2연구모임	이드웨어 대표이사	명지대학교수
인원(명)	4	5	1	3

▶ 주요연구원 소개

분야	명지대정보통신공학과	명지대정보통신공학과	이드웨어	명지대학교
성명			이드웨어 대표이사	명지대 교수 강환일
성명				
			태풍 진로예측프로그램개발 삼성전자 음성인식 첫 상용화 기계학습/임베디드 전문가	
			미국 현지법인 설립 (SOUNDMIND TECHNOLOGY INC. 19.10.15)	

1. 산학 협동 연구원 구성

이드웨어 (주)

1. 응용 소프트웨어 개발/공급업체
2. 세계최고수준의 임베디드형 음성인식 엔진 VRSoft 개발
 - 세계최고수준의 초소형엔진, 저전력 칩(chip)체 작동
 - IoT 스마트 홈
 - 순수 국내 기술로 자체 개발
 - 고객요청에 따라 다양한 형태로 변환



음성인식노래검색



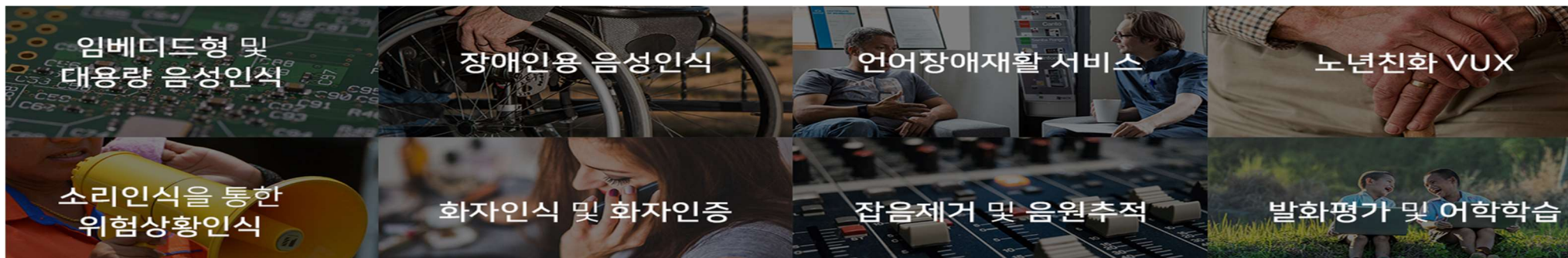
이드웨어 음성인식 모듈



실시간잡음제거이어폰



울음소리 인식 장치



1. 산학 협동 연구원 구성

신대진 대표이사 이드웨어 (주)

VRSoft™

임베디드 음성인식/합성

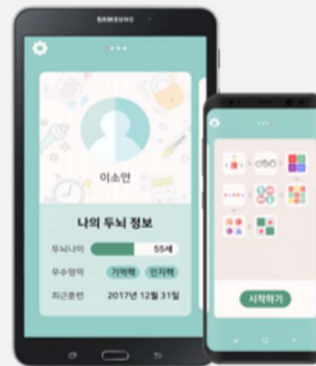
세계최초 칩 탑재 음성합성(TTS) 엔진
세계최소형 (30KB) 임베디드 음성인식 엔진
국내유일 연령에 독립적인 음성인식 엔진
(아동/성인/노년층 3-path 디코더)



Soundmind™

노년층 치매예방/인지재활

- 국내최초 한국형 CCT(디지털 인지재활)
- 세계최초 노년층 음성인식 적용 솔루션
- 인공지능 기반 엑스퍼트 시스템 적용



SeetheSound™

DNN/임베디드 소리인식

- 세계최초 울음소리인식 엔진 상용화
- 세계최초 소리인식전용 칩 개발 예정
(임베디드형 소리인식 개발 국가과제)



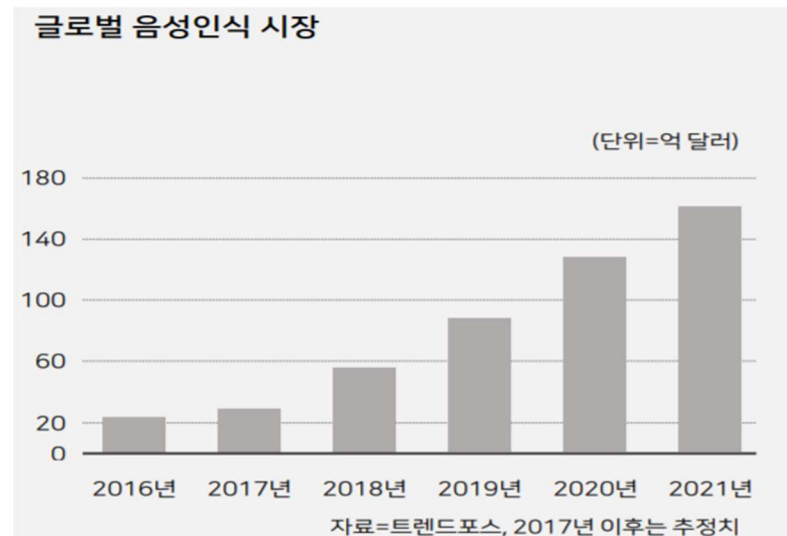
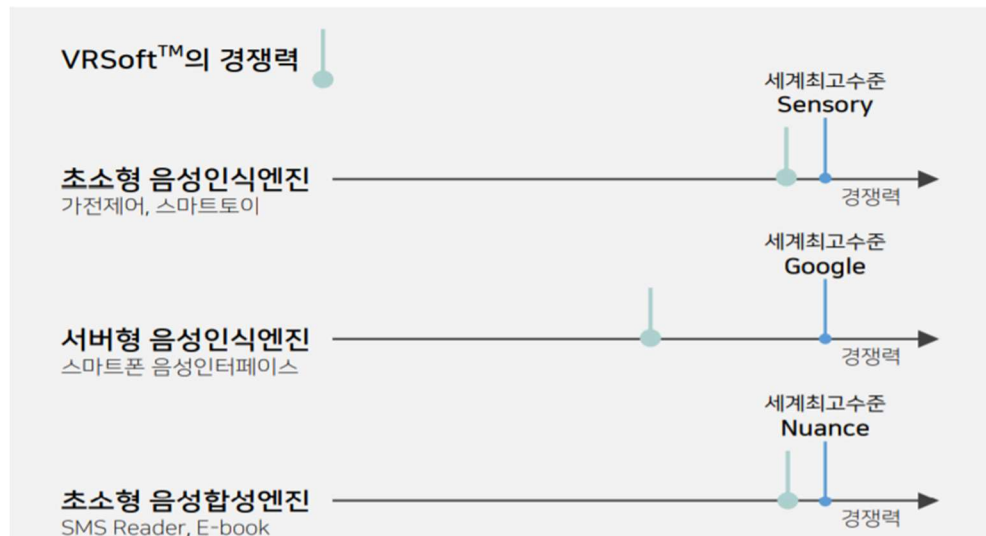
1. 산학 협동 연구원 구성 이드웨어 (주)

1. 지적 재산권:

- 전자책낭독을 통한 발화 정확도 및 표현력 향상을 제공하는 방법 및 시스템(2015)
- 스마트믹싱 모듈을 갖춘 스마트 이어폰, 스마트 믹싱 모듈을 갖춘 기기, 외부음과 기기음을 혼합하는 방법 및 시스템(2016)
- 콥심음 집단을 포함하는 차량 집단 시스템 및 방법(2017)

2. 상용화 내용

- K2전차용 잡음제거 칩 제공 계약
- 쿠키전자용 음성인식 칩 납품 계약
- 음성인식 기반 응급관제 소프트웨어 개발 (발주처: 서울대학교)



2. 연구목표 및 연구내용:

▶ 연구 목표

- 1) 인공지능 기반 음성인식 기술확보

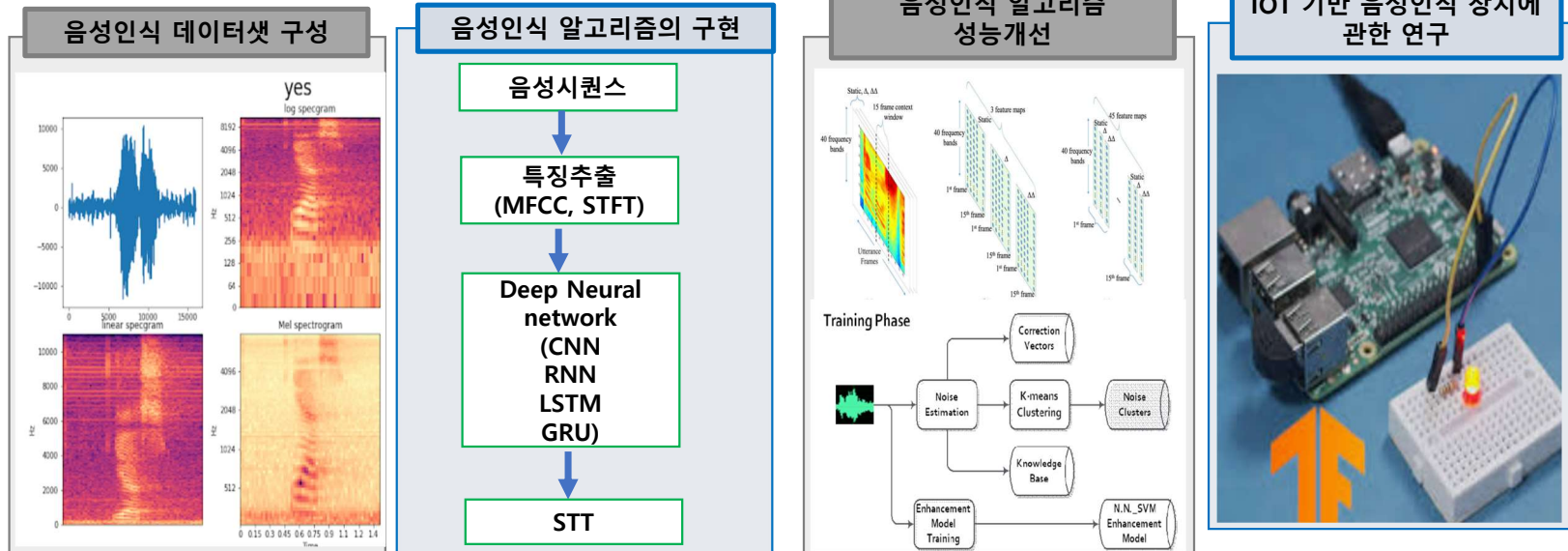
산학협동연구를 통한 취업 및 연구에 활용

- 2) 음성인식 기술의 저변 확대를 통한 신학협력 사업의 자립화

음성관련 데이터 세트의 활용및 IOT를 이용한 음성인식 장치에 관한 연구

▶ 본 과제의 연구내용

- 1) 음성인식 관련 데이터 셋의 구성
- 2) 음성인식 알고리즘의 구현 및 성능개선
- 3) IOT 기반 음성인식 장치에 관한 연구

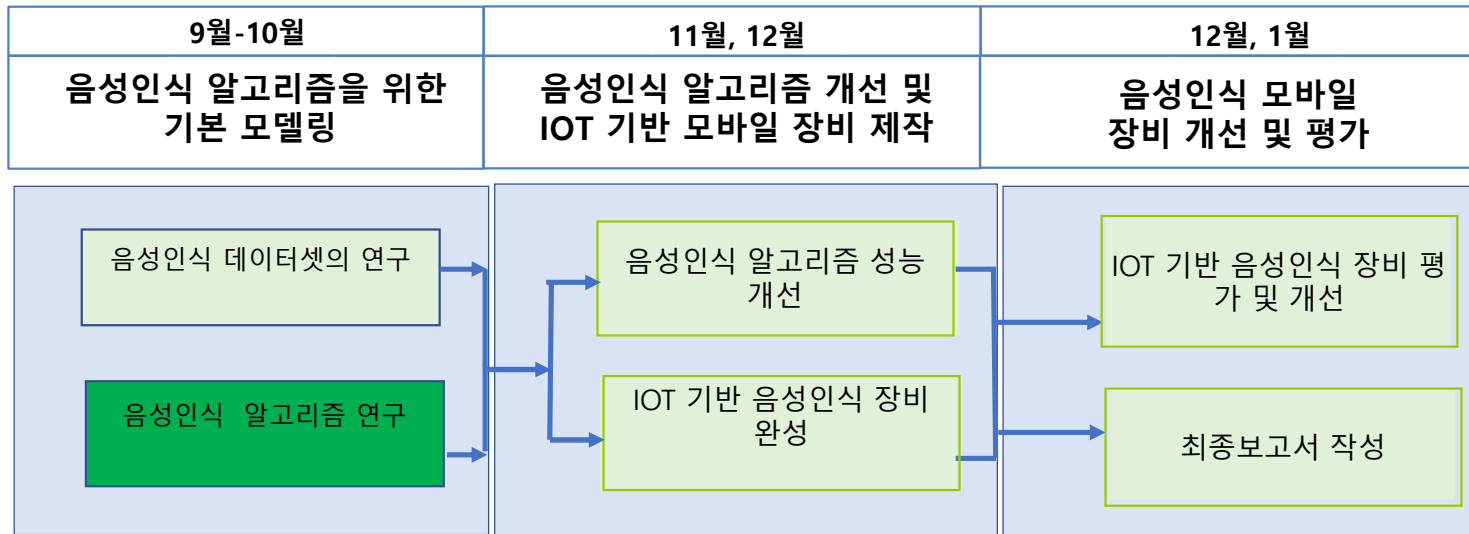


3. 연구기간

▶ 연도별 목표 계획 및 연구내용

• 계획

□ : 계획 ■ : 실현

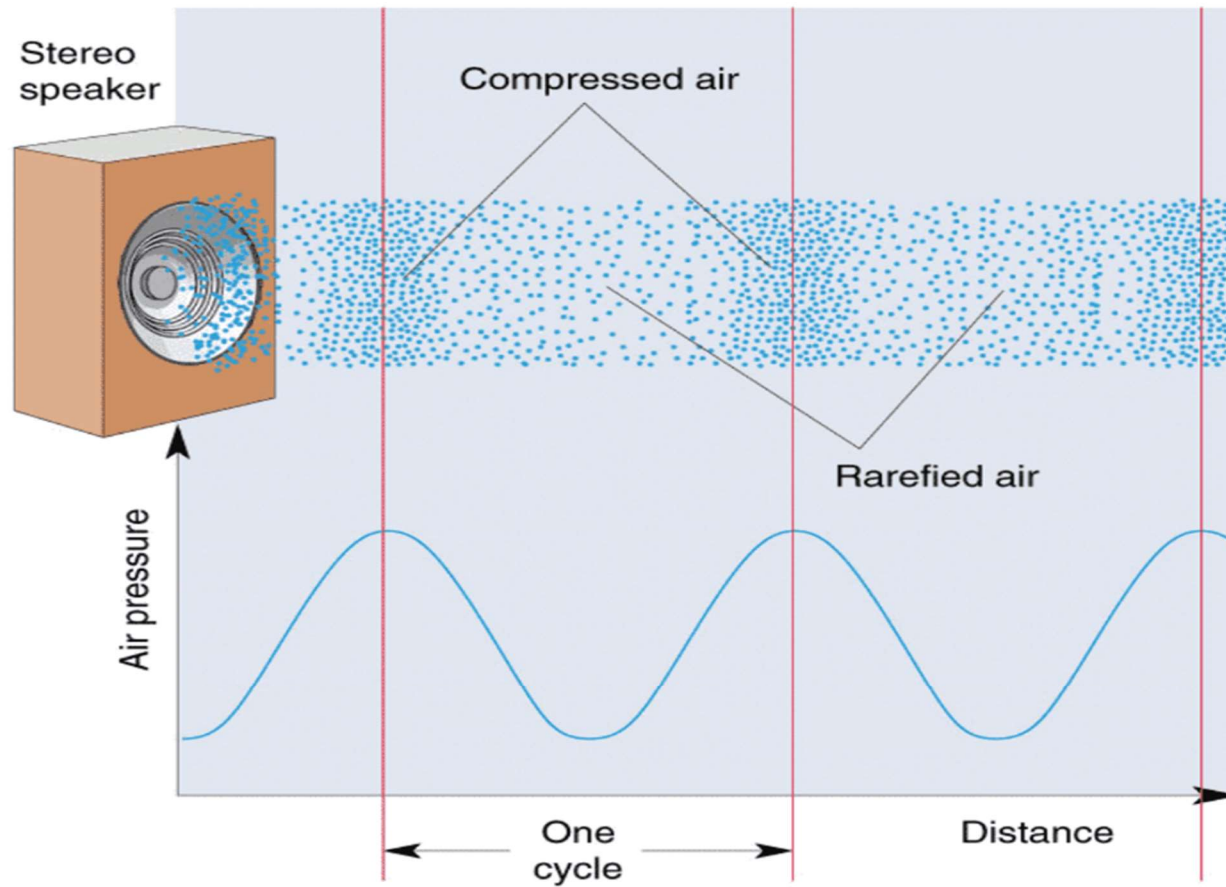


▶ 실적및 계획 요약

- 1) 산학공동기술개발 과제 중간평가 (서면평가 혹은 온라인 평가: 평가위원: 한승수교수)
 - 9월 말-10월초 (발표준비, 프로젝트진행상황, 참여학생 역할 및 참여도)
 - 협약업체 역할
- 2) 논문: 1편 (강환일)
- 3) 주요산출물: 최종보고서: (신대진: 대표이사)
- 4) 산학 EXPO 참석 (12.2—12.09, COEX) (학생, 신대진 대표이사)

4.. 연구의 내용

소리의 정의



4.. 연구의 내용

Frequency?

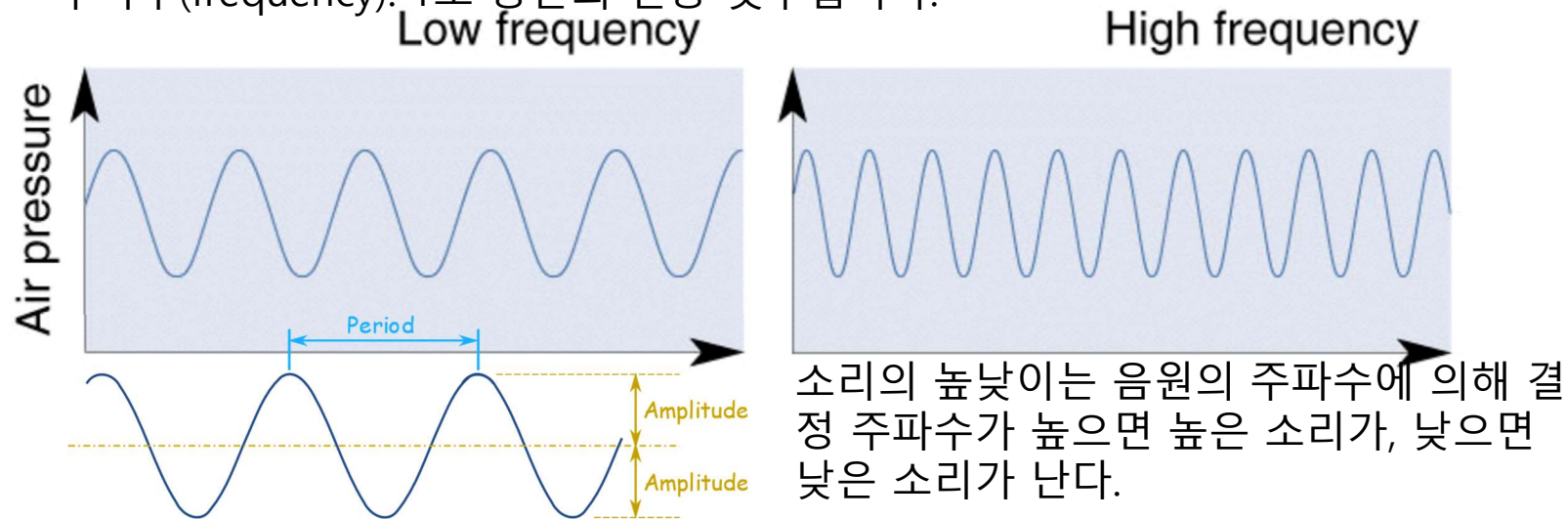
Frequency는 The number of compressed

단위는 Hertz를 사용하며, 1Hertz는 1초에 한번 Vibration을 의미합니다.

- 주기(period): 파동이 한번 진동하는데 걸리는 시간, 또는 그 길이, 일반적으로 \sin 함수의 주기는 $2\pi/w$

입니다.

- 주파수(frequency): 1초 동안의 진동 횟수입니다.

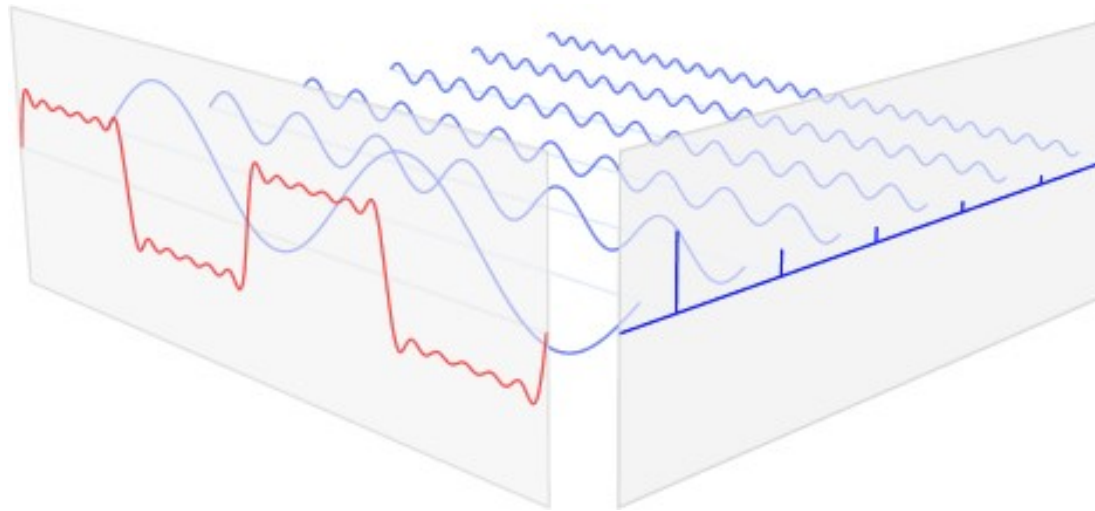


4.. 연구의 내용

푸리에 변환 (Fourier transform)

임의의 입력 신호를 다양한 주파수를 갖는
주기함수(복소 지수함수)들의 합으로 분해하여 표현하는 것

$$A_k = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} f(t) \exp\left(-i \cdot 2\pi \frac{k}{T} t\right) dt$$



4.. 연구의 내용

Python waveform generation

```
import librosa
```

```
audio_path = 'test_voice.wav'
```

```
y, sr = librosa.load(audio_path)
```

```
import matplotlib.pyplot as plt
```

```
plt.title("voice-wave")
```

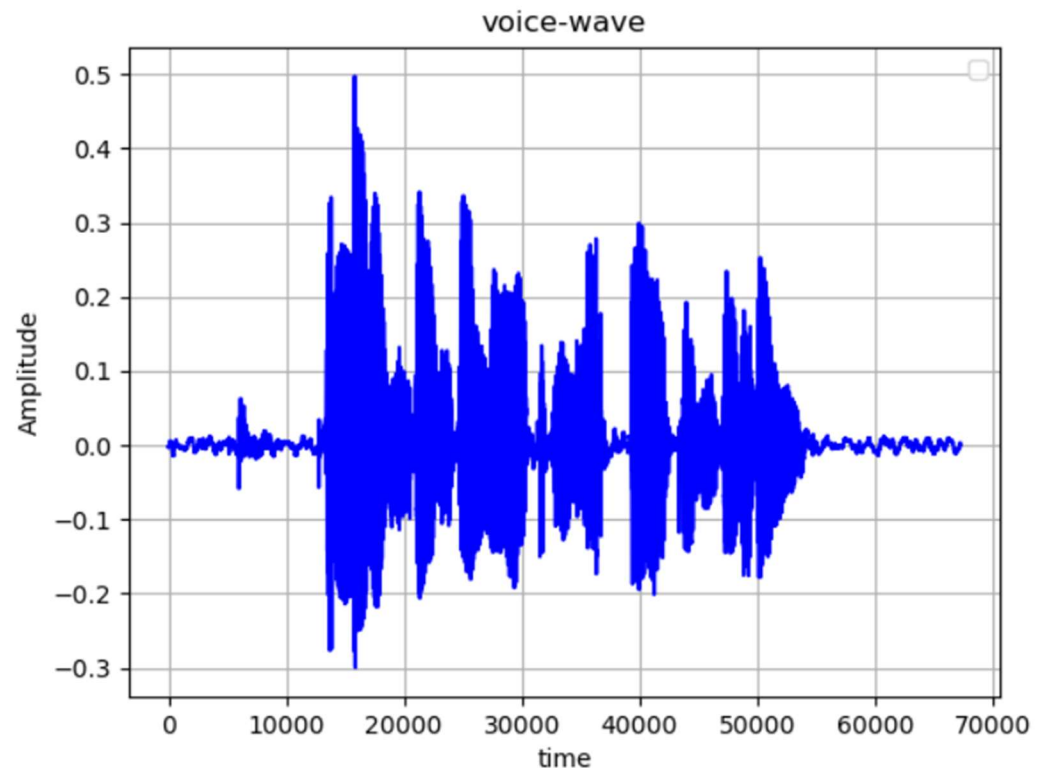
```
plt.xlabel("time")
```

```
plt.ylabel("Amplitude")
```

```
plt.grid()
```

```
plt.plot(y,color='blue')
```

```
plt.show()
```



4.. 연구의 내용

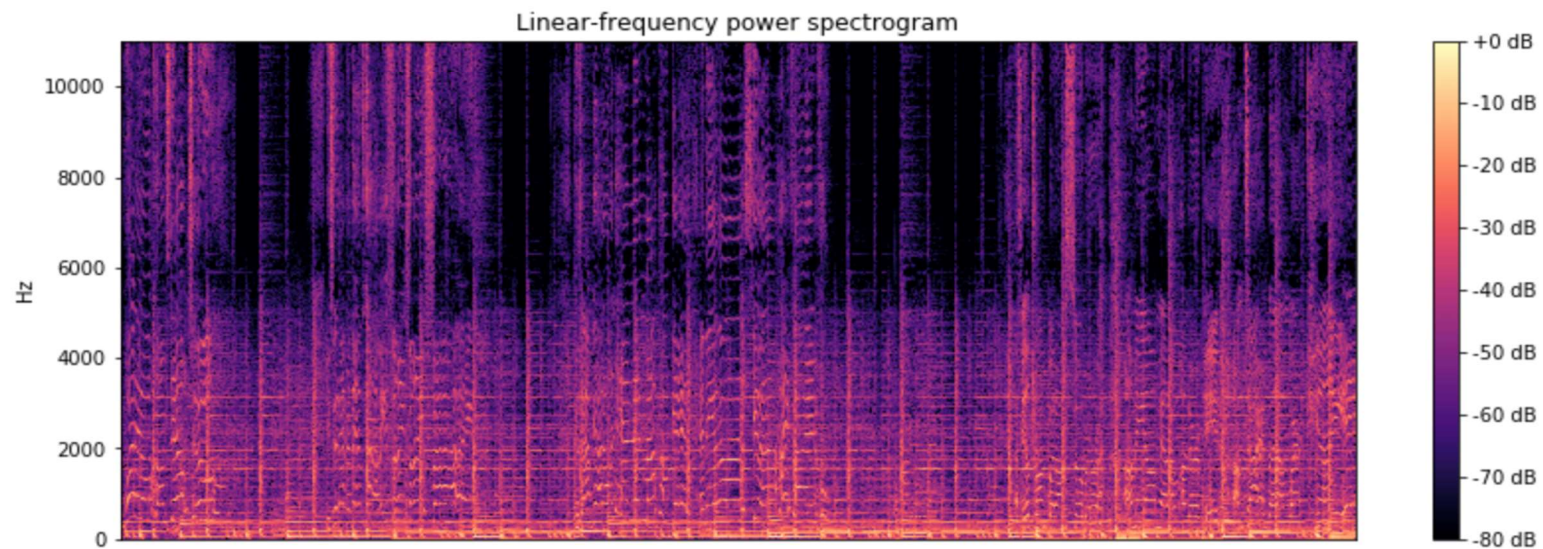
STFT

```
D = librosa.amplitude_to_db(np.abs(librosa.stft(y)), ref=np.max)
D[0:1] ,D.shape

(array([[ -57.238033, -68.030106, -58.63585 , ..., -60.651947, -35.656086,
        -29.75666  ]], dtype=float32), (1025, 1293))
```

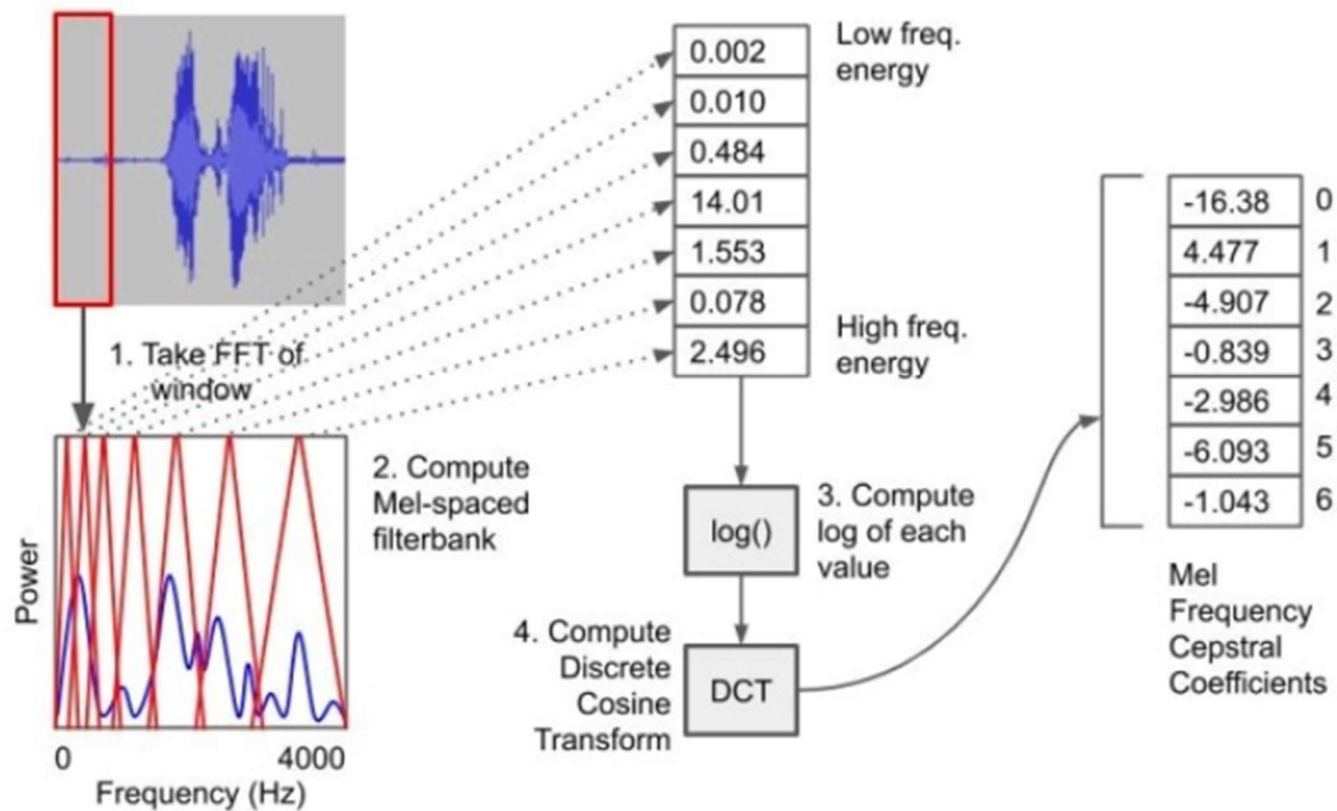
```
fig = plt.figure(figsize = (14,5))
librosa.display.specshow(D, y_axis='linear')
plt.colorbar(format='%+2.0f dB')
plt.title('Linear-frequency power spectrogram')
```

```
Text(0.5, 1.0, 'Linear-frequency power spectrogram')
```



4.. 연구의 내용

멜주파수 (MFCC—Mel Frequency Cepstral Coefficient)



<https://blog.naver.com/damtaja/221998249683>

4.. 연구의 내용

MFCC

0914 MFCC.py - C:/Users/hwank/Documents/python_파일/audio/0914 MFCC.py (3.8.2)

File Edit Format Run Options Window Help

```
import
from pyfeatures import mfcc
```

```
def compute_mfcc(audio_data, sample_rate):
    #audio_data = audio_data - np.mean(audio_data)
    #audio_data = audio_data / np.max(audio_data)
    mfcc_feat = mfcc(audio_data, sample_rate, winlen=0.010, winstep=0.01,
                     numcep=13, nfilt=26, nfft=512, lowfreq=0, highfreq=None,
                     preemph=0.97, ceplifter=22, appendEnergy=True)

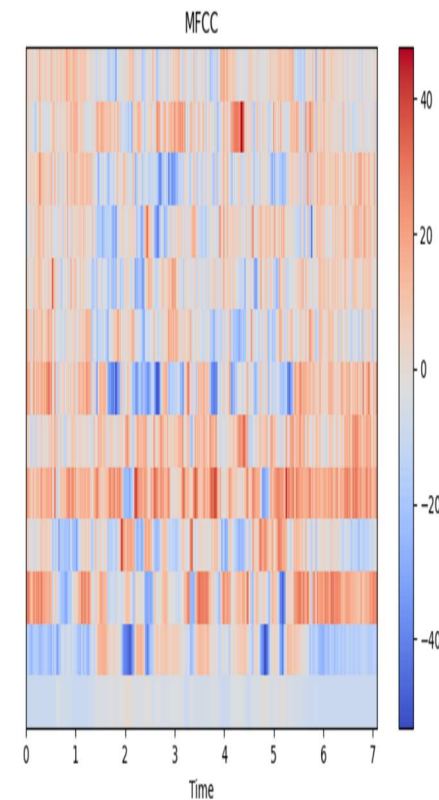
    return mfcc_feat
```

```
audio_sample, sampling_rate = librosa.load("test_voice.wav", sr = None)
mfcc = compute_mfcc(audio_sample, sampling_rate)
```

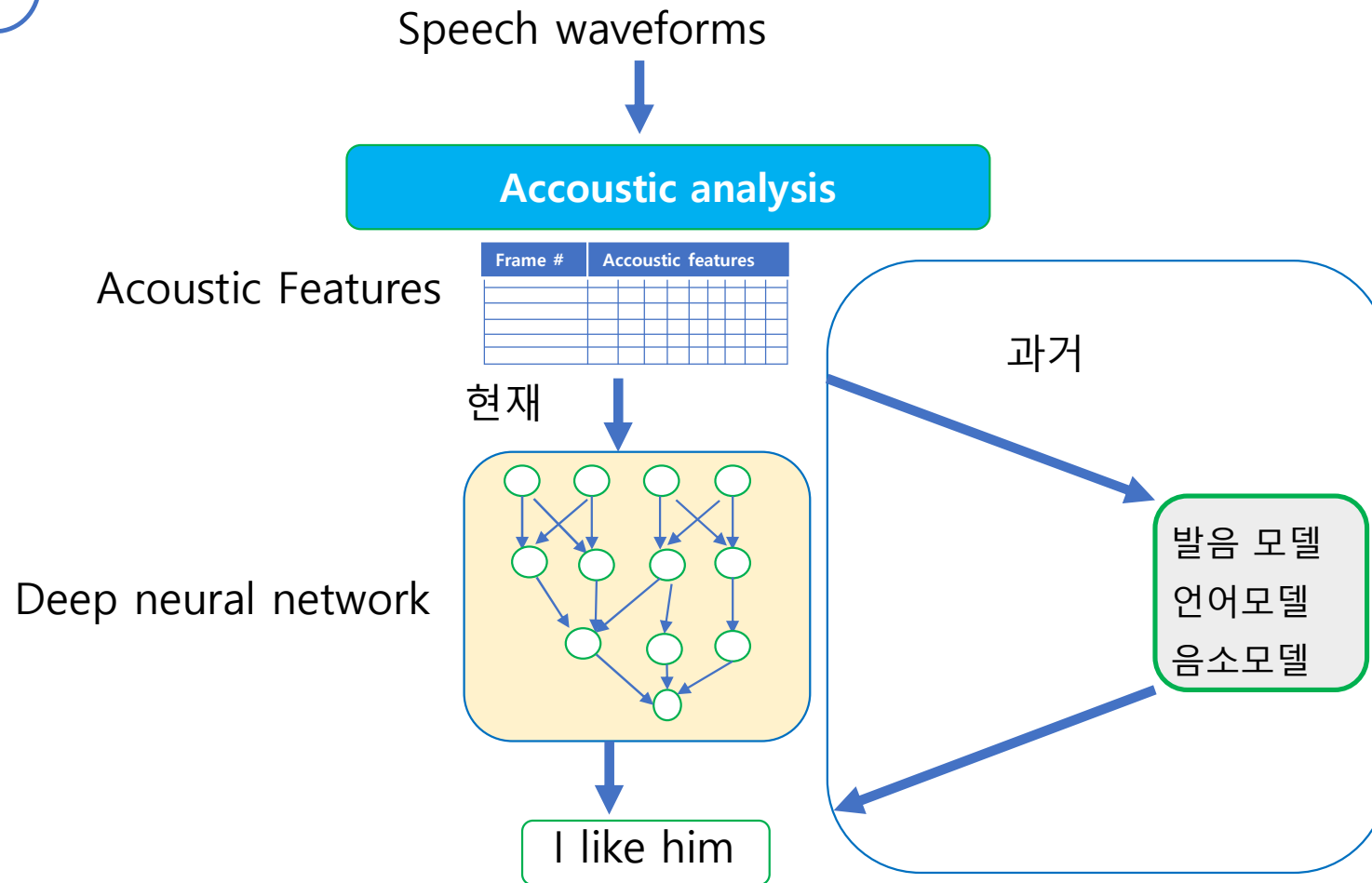
```
import matplotlib.pyplot as plt
import librosa.display
librosa.display.specshow(mfcc.T, x_axis='time')
plt.colorbar()
plt.title('MFCC')
plt.tight_layout()
plt.show()
```

#[출처] [음성/오디오 신호처리] Python을 이용한 오디오 신호 분석 - MFCC 계수 추출

<https://blog.naver.com/daechuni/221652489714>



4.. 연구의 내용



4.. 연구의 내용

입력(input)

Waveforms

STFT(short time Fourier transform)

Mel spectrogram

신경망(framework)

Deep Neural network

CNN(Convolutional neural network : 합성곱 신경망)

Simple CNN

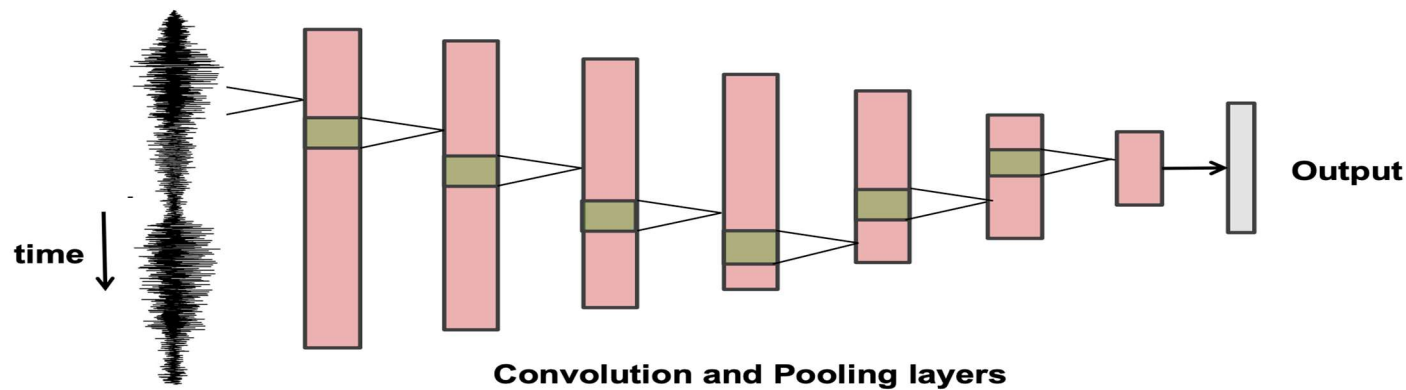
RNN: (Recurrent Neural Network: 회귀 신경망)

LSTM (long short term memory Network)

GRU (Gated Recurrent unit)

4.. 연구의 내용

CNN in Audio!



Sample CNN

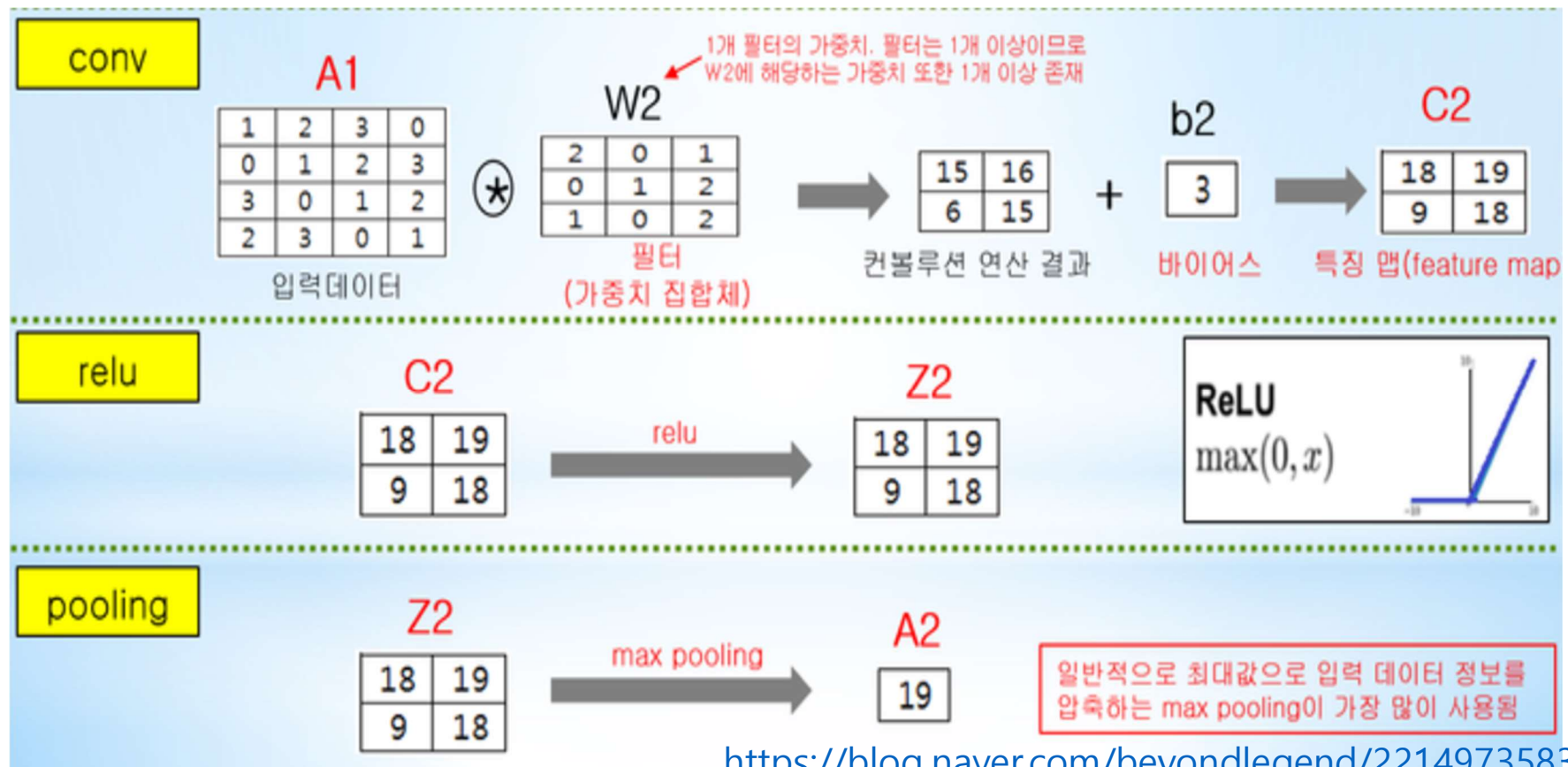
Sample level CNN 의 가장 큰 특징은 바로 input데이터를 waveform 그 자체로 사용할 수 있다는 점입니다.

- Advantage
 - CNN이 "phase-invariant" representation을 반영한다는 점입니다.
 - 커널이 input signal에 대한 spectral bandwidth를 계산해 준다는 점입니다.

- Frequency : F
- Time : T
- Channel : N
- strides : $(s1, s2)$

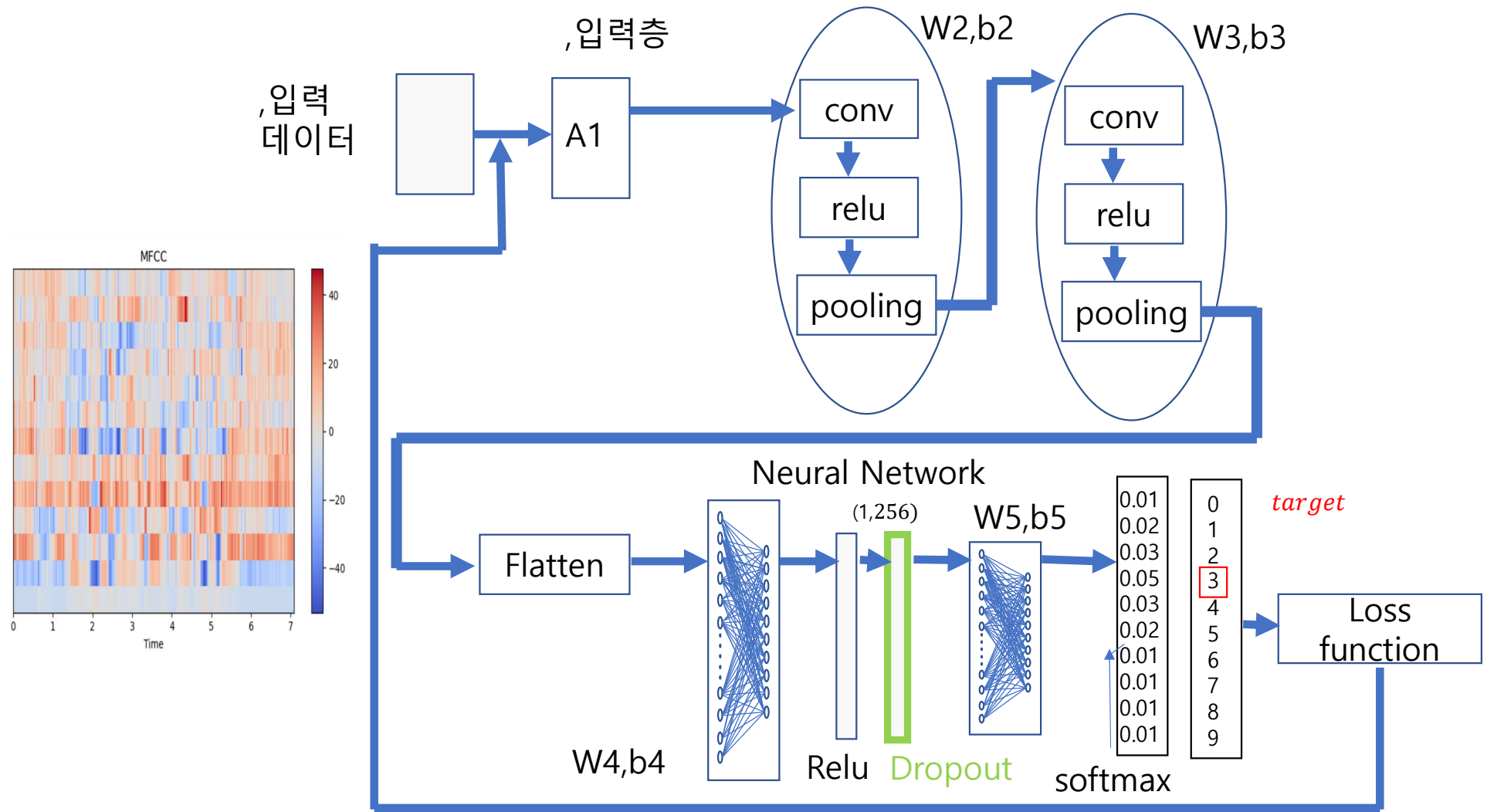
합성곱 층 만들기

Conv2D와 MaxPooling2D 층을 쌓는 일반적인 패턴으로 합성곱 층을 정의합니다.

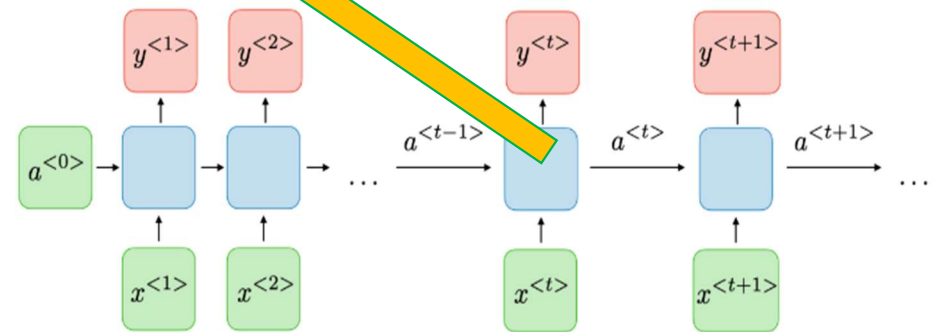
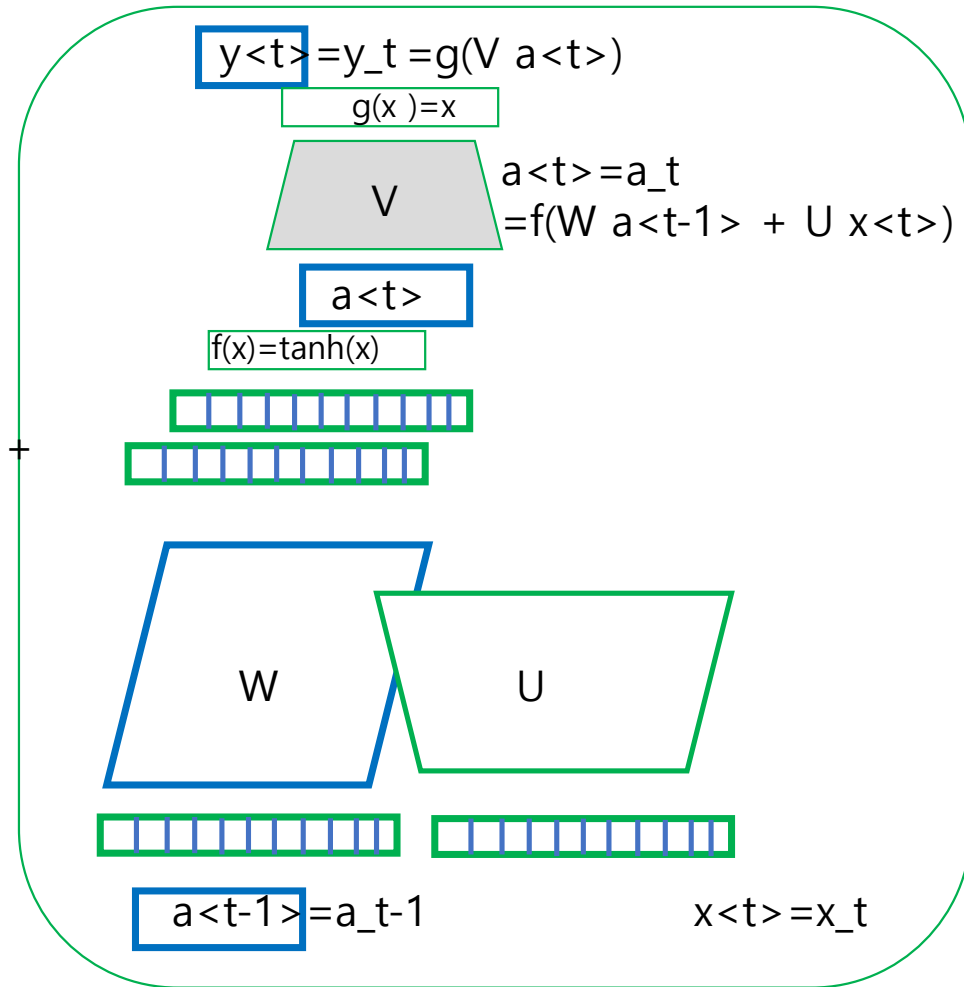


<https://blog.naver.com/beyondlegend/221497358389>

CNN

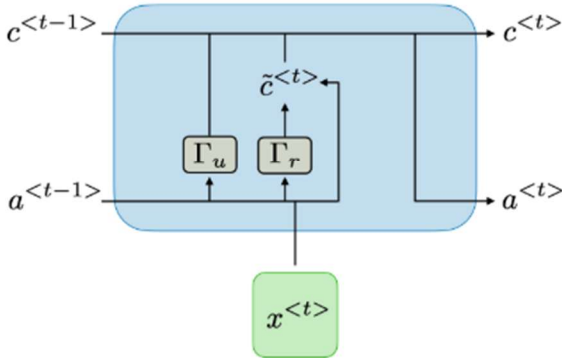
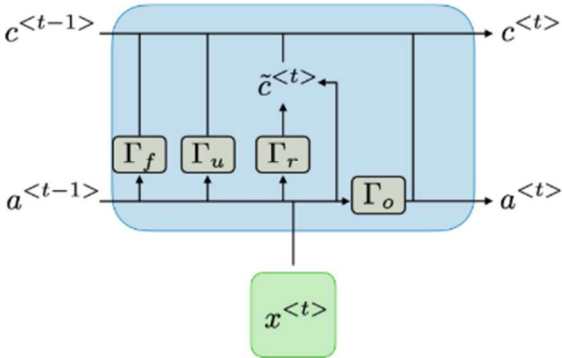


RNN: (Recurrent Neural Network: 회귀 신경망)



4.. 연구의 내용

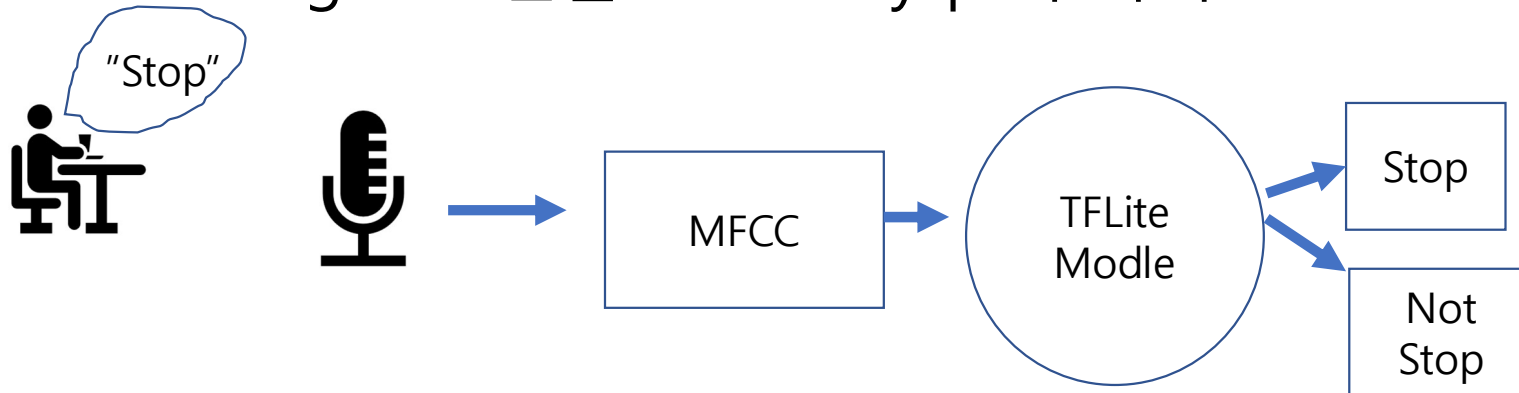
RNN 에 의한 미분값이 사라지는 것을 방지하기 위해 GATED Recurrent Unit(GRU) 와 LSTM(Long short term memory network)의 구조

Characterization	Gated Recurrent Unit (GRU)	Long Short-Term Memory (LSTM)
$\tilde{c}^{<t>}$	$\tanh(W_c[\Gamma_r \star a^{<t-1>}, x^{<t>}] + b_c)$	$\tanh(W_c[\Gamma_r \star a^{<t-1>}, x^{<t>}] + b_c)$
$c^{<t>}$	$\Gamma_u \star \tilde{c}^{<t>} + (1 - \Gamma_u) \star c^{<t-1>}$	$\Gamma_u \star \tilde{c}^{<t>} + \Gamma_f \star c^{<t-1>}$
$a^{<t>}$	$c^{<t>}$	$\Gamma_o \star c^{<t>}$
Dependencies		

4.. 연구의 내용

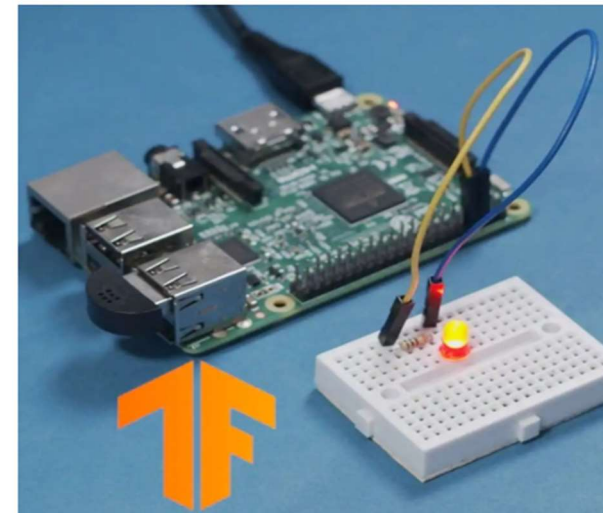
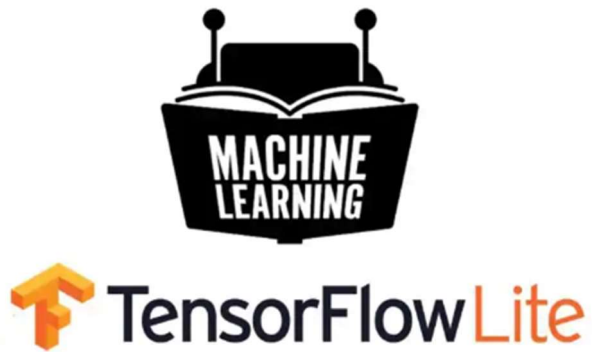
라즈베리파이와 음성인식

- ◆데이터 수집->특징추출->훈련과 테스트 모델->deploy(이용)
 - 특징: MFCC(음성의 스펙트럼을 구하고 스케일을 변경해 얻음)
- ◆코 드 (텐서플로우로 작성)-> 변환기 -> Tensor flow lite model
- ◆Tensorflow light 모델을 Rasbbery pi에 복사



4.. 연구의 내용

라즈베리파이와 사용가능한 기계학습 소프트웨어



<https://blog.naver.com/damtaja/221998249683>

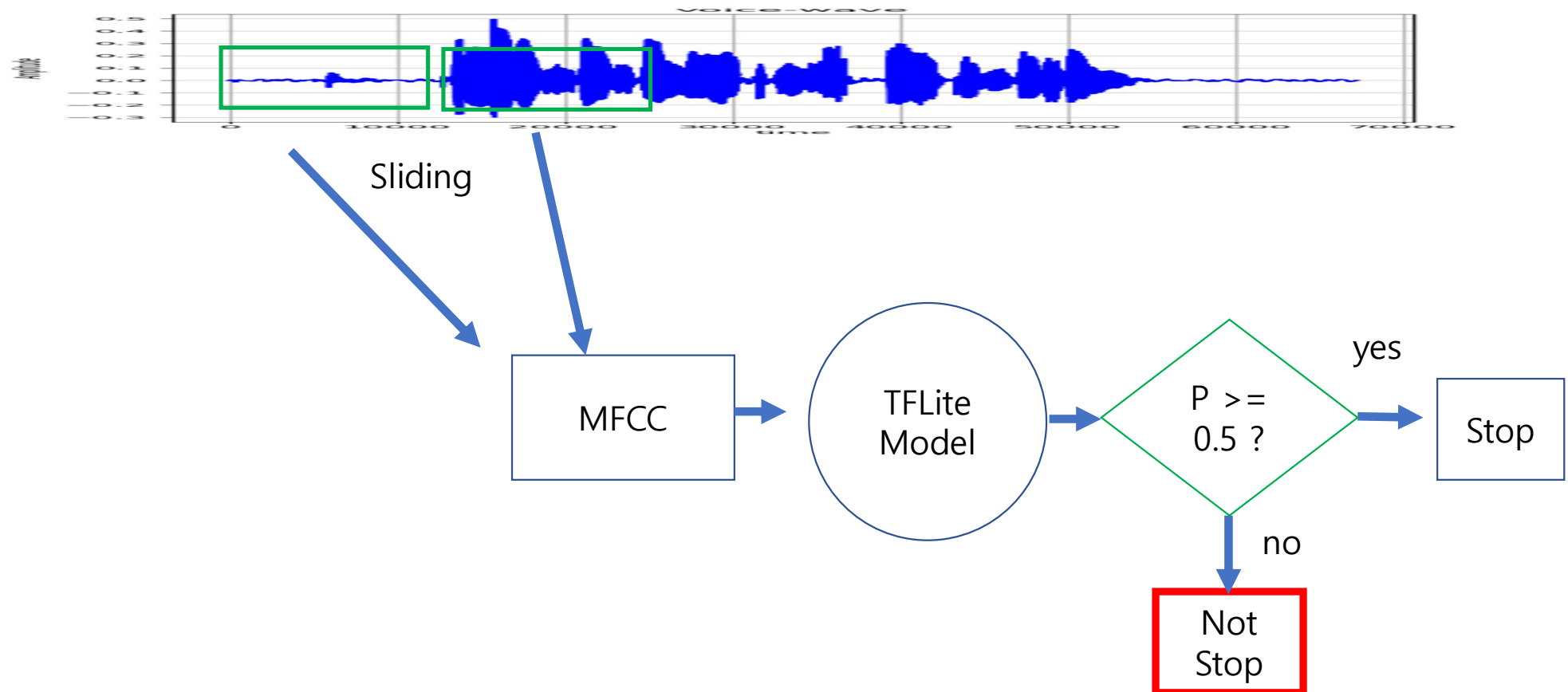
<https://blog.naver.com/damtaja/221997394045>

- 원문 : <https://www.digikey.com/en/maker/projects/how-to-train-new-tensorflow-lite-micro-speech-models/e9480d4a38264604a2bf0336ce11aa9e>

[출처] [번역] 새로운 TensorFlow Lite 마이 <https://www.digikey.com/en/maker/projects/how-to-train-new-tensorflow-lite-micro-speech-models/e9480d4a38264604a2bf0336ce11aa9e>

4.. 연구의 내용

라즈베리파이와 음성인식



<https://bog.naver.com/damtaja/221999558070>



5. 연구의 미래

- 음성합성 (Tacotron)
- 나이, 억양, 언어능력에 변화에 강인한 모델
- 소음이 가득찬 환경에서 음성인식
- 소음하에서 음성과 다른 소리와 구분
- 발음의 변이에 적응
- 새로운 언어 취급

6. 연구의 내용 및 범위

1. 딥러닝을 이용한 영어 단어 인식
2. 딥러닝을 이용한 음성스위치 만들기
 - ① 음성스위치 만들기
 - ② False Alarm 실험
 - ③ 데이터셋 만들기 및 재 훈련
3. 팀별 주제정하기

Ex) Team A : 비명소리인식

Team B : Ambient Noise 인식

<https://www.youtube.com/watch?v=ks9h60WZjhQ>

6. 연구의 내용 및 범위

핵심논문

<https://arxiv.org/pdf/1711.07128.pdf>

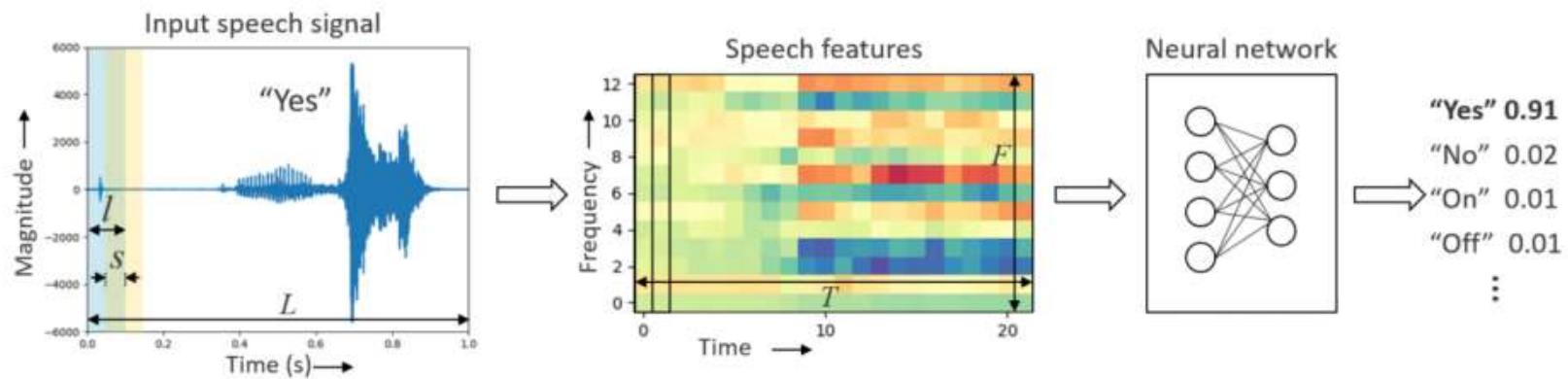


Figure 1: Keyword spotting pipeline.

6. 연구의 내용 및 범위

1. 딥러닝을 이용한 영어 단어 인식

Google Simple Audio Recognition Data Set

https://github.com/tensorflow/docs/blob/master/site/en/r1/tutorials/sequences/audio_recognition.md

- 1초짜리 wav 파일
- 폴더별로 정렬
- 파일이름에 Hash tag
- 16kHz, 16bit, mono, PCM

학습스크립트

<https://github.com/ARM-software/ML-KWS-for-MCU>



6. 연구의 내용 및 범위

2. 딥러닝을 이용한 음성스위치 만들기

- 호출어 선정 : 하이마트 기타 등등
- Google Simple Audio Recognition Data Set + 직접 만든 음성 데이터
- 한국어 말뭉치(A.I. Hub)
- 환경잡음 (TV소리등)

6. 연구의 내용 및 범위

- 음식인식 알고리즘

NN model	S(80KB, 6MOps)			M(200KB, 20MOps)			L(500KB, 80MOps)		
	Acc.	Mem.	Ops	Acc.	Mem.	Ops	Acc.	Mem.	Ops
DNN	84.6%	80.0KB	158.8K	86.4%	199.4KB	397.0K	86.7%	496.6KB	990.2K
CNN	91.6%	79.0KB	5.0M	92.2%	199.4KB	17.3M	92.7%	497.8KB	25.3M
Basic LSTM	92.0%	63.3KB	5.9M	93.0%	196.5KB	18.9M	93.4%	494.5KB	47.9M
LSTM	92.9%	79.5KB	3.9M	93.9%	198.6KB	19.2M	94.8%	498.8KB	48.4M
GRU	93.5%	78.8KB	3.8M	94.2%	200.0KB	19.2M	94.7%	499.7KB	48.4M
CRNN	94.0%	79.7KB	3.0M	94.4%	199.8KB	7.6M	95.0%	499.5KB	19.3M
DS-CNN	94.4%	38.6KB	5.4M	94.9%	189.2KB	19.8M	95.4%	497.6KB	56.9M

Table 5: Summary of best neural networks from the hyperparameter search. The memory required for storing the 8-bit weights and activations is shown in the table.

6. 연구의 내용 및 범위

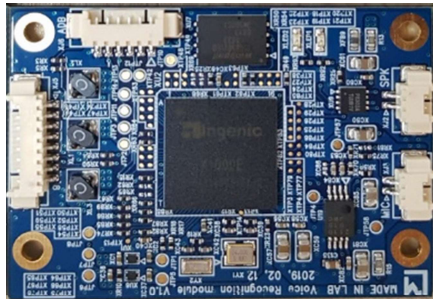
3. 팀별 주제 정하기

- 울음소리인식, 비명소리인식, 음성스위치, 주변상황인식등등
- 주제에 맞는 음성DB구축
- FA를 위한 다양한 DB구축
- 실험 → 재훈련 → 실험 → 재훈련...

6. 연구의 내용 및 범위

- IOT 기반 음식인식 장비

작게만 만들고 싶다면?



VRS M5

다양한 주변 장치와 연동이 필요하다면?



Raspberry Pi 3 or 4



8. 결론

- 열심히 공부하고 제작하여 EXPO에서 우수한 상을 수상하기로 함

9. 참고문헌 및 참고자료

- [TensorFlow Lite Tutorial Part 2: Speech ...digikey.com](https://www.digikey.com/en/maker/projects/how-to-train-new-tensorflow-lite-micro-speech-models/e9480d4a38264604a2bf0336ce11aa9e)
- Convolutional Neural Networks for Speech Recognition Ossama Abdel-Hamid, Abdel-rahman Mohamed, Hui Jiang, Li Deng, Gerald Penn, and Dong Yu(2014, IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 22, NO. 10, OCTOBER 2014)
- Kaggle Tensorflow Speech Recognition Challenge (Feb 23, 2018)
- Convolutional Neural Networks for Raw Speech Recognition
- By Vishal Passricha and Rajesh Kumar Aggarwal
Submitted: April 24th 2018Reviewed: July 6th 2018Published:December 12th 2018
DOI: 10.5772/intechopen.80026
- <https://www.nextpng.com/en/transparent-png-yoydh>
- <https://blog.naver.com/damtaja/221997394045>([출처] [번역] 새로운 TensorFlow Lite 마이크로 음성 모델을 훈련시키는 방법|작성자 해바우)
- <https://www.digikey.com/en/maker/projects/how-to-train-new-tensorflow-lite-micro-speech-models/e9480d4a38264604a2bf0336ce11aa9e>
- <http://www.eidware.com/>

9. 참고문헌 및 참고자료(계속)

- NIPS 2017 Tutorial – Deep Learning: Practice and Trend
<https://drive.google.com/file/d/1SuwilCLERd7SfYo3FiqNG0tCEBUjKcT7/view>
- stanford CS 230 - Deep Learning
<https://stanford.edu/~shervine/teaching/cs-230/cheatsheet-convolutional-neural-networks>
- Audio for Deep Learning (남기현님)
https://tykimos.github.io/2019/07/04/ISS_2nd_Deep_Learning_Conference_All_Together/
- 오디오 전처리 작업을 위한 연습 (박수철님)
<https://github.com/scpark20/audio-preprocessing-practice>
- Musical Applications of Machine Learning
<https://mac.kaist.ac.kr/~juhan/gct634/>
- Awesome audio study materials for Korean (최근우님)
<https://github.com/keunwoochoi/awesome-audio-study-materials-for-korean>
- Graves & Jaitley, "Towards end to end speech recognition with recurrent neural networks", ICML '14
- Maas et al., "Lexicon free conversational speech recognition with neural networks", NAACL '15
- Chan et al., "listen, attend and spell: NN for large vocabulary conversational speech recognition", ICASSP '16.
- Youtube.com, Automatic Speech Recognition-An overview, 2017.9.12.
- 도승헌 딥러닝을 활용한 오디오 분류 및 음성인식. KAIST, Tacademy, <https://tacademy.skplanet.com/>
- <https://stanford.edu/~shervine/teaching/cs-230/cheatsheet-recurrent-neural-networks>