2020 IT 21
Global Conference
Digital New Deal
Technology Essentials
디지털 뉴딜 기술 핵심

Session 4-2

감성미디어(Emotion Recognition based on Speech)

이석필 교수 (상명대학교)

[요약문]

감정미디어란 감정을 포함하고 있는 미디어를 의미한다. 다시 말하면 미디어에 포함되어 있는 사람의 감정을 의미한다. 미디어를 감정에 따라 분류할 수 있다면 또는 사람의 감정상태를 실시간으로 파악할 수 있다면 제공할 수 있는 서비스가 매우 다양 해질 수 있다.

사람의 감정을 인식할 수 있는 큰 두가지 특징은 얼굴 표정과 음성 인데 표정을 읽기 위해서는 카메라가 계속 영상을 촬영하고 있어야 하는 불편함이 있다.

본 발표에서는 순수하게 사람의 음성만을 바탕으로 감정을 인식할 수 있는 기술과 결과를 보여주며 이는 요즘 많이 보급되고 있는 인공지능스피커 등에서 하나의 킬러 앱이 될 수 있다.

[발표자 약력]

1990년 연세대학교 전기공학 학사
1997년 연세대학교 전기전자공학 박사
1997년~2002년 대우전자 중앙연구소 선임연구원
2002년~2012년 KETI 디지털미디어연구센터 센터장
2010년~2011년 Georgia Tech. 방문연구원
2012년~현재 상명대학교 융합전자공학과 교수
관심분야 : 인공지능, 디지털 신호처리, 방송영상시스템, 멀티미디어 콘텐츠 등

Emotion Recognition based on Speech

Seok-Pil Lee
(esprit@smu.ac.kr)

PMP Lab.
Dept. of Media Software
Sangmyung University

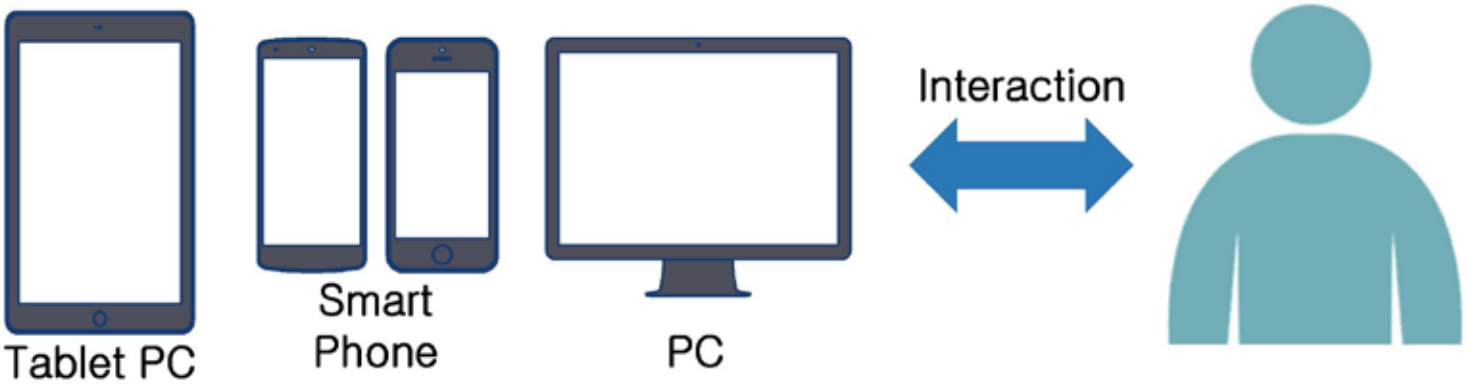
Contents

1. Introduction
2. Database
3. Emotion Recognition
4. Demonstration

Introduction

Contents

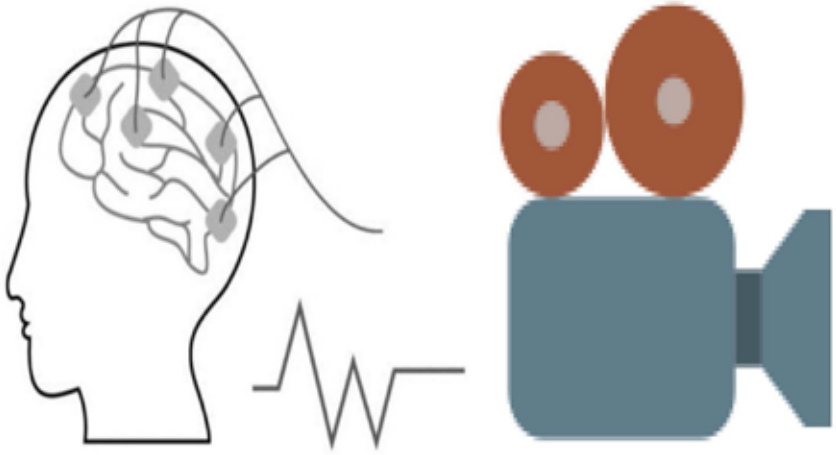
Human–Computer Interaction



4

Emotion Recognition

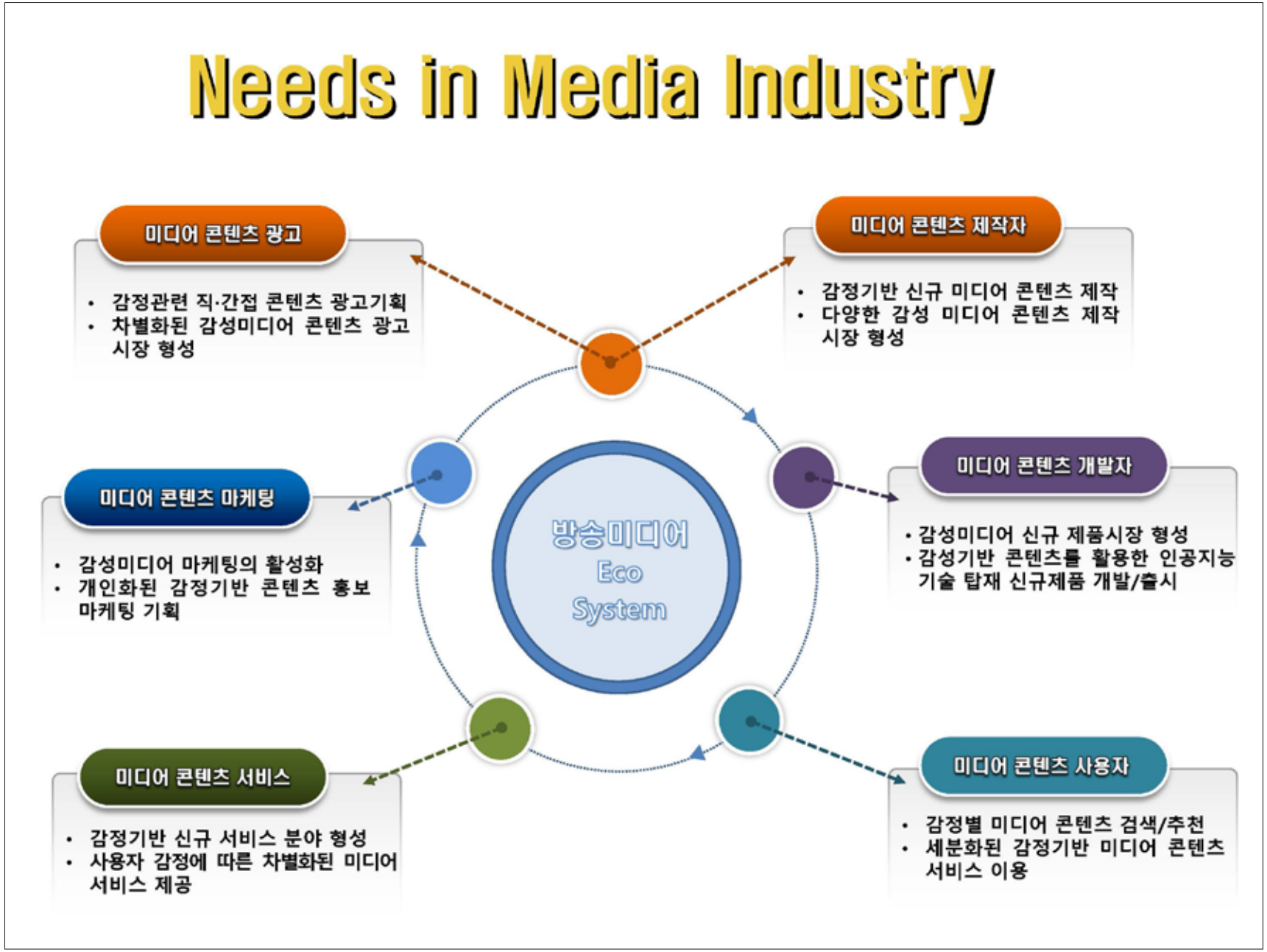
- 음성을 이용하는 방법
- 영상을 이용하는 방법
- 뇌파를 이용하는 방법
- Multi-modal
- ...



Using Speech Signals

- 데이터를 간편하게 취득
- 의사소통의 기본 수단
- 낮은 정확도
  => 양질의 DB 및 인식 알고리즘 필요

아마존 음성을 통한 감정을 읽는 웨어러블 기기 개발 착수


Needs in Media Industry
미디어 콘텐츠 광고
• 감정관련 직·간접 콘텐츠 광고기획
• 차별화된 감성미디어 콘텐츠 광고 시장 형성
미디어 콘텐츠 제작자
• 감정기반 신규 미디어 콘텐츠 제작
• 다양한 감성 미디어 콘텐츠 제작 시장 형성
미디어 콘텐츠 마케팅
• 감성미디어 마케팅의 활성화
• 개인화된 감정기반 콘텐츠 홍보 마케팅 기획
방송미디어
Eco
System
미디어 콘텐츠 개발자
• 감성미디어 신규 제품시장 형성
• 감성기반 콘텐츠를 활용한 인공지능 기술 탑재 신규제품 개발/출시
미디어 콘텐츠 서비스
• 감정기반 신규 서비스 분야 형성
• 사용자 감정에 따른 차별화된 미디어 서비스 제공
미디어 콘텐츠 사용자
• 감정별 미디어 콘텐츠 검색/추천
• 세분화된 감정기반 미디어 콘텐츠 서비스 이용

Conceptual Model

2017 ~ 2019 세종대/건국대/상명대 Co-Project

Database

Emotion Categories

- Anger
- Sadness
- Happiness
- Neutral
- Excitement
- Fear

Acquisition Environment

- The Neumann TLM 103
- Neumann U87 Ai
- Oktava MK-319 microphone

- Universal Audio LA-610 MK1
- WARM AUDIO WA76
- MPAC-01 mixer

녹음 중인
연기자

녹음 환경

Database #1

- 데이터 길이 : 3~5초
- Sampling Rate: 16000Hz
- 남성 11명, 여성 9명 총 20명의 연기자가 녹음
- 감정 유발 문장 20개 X 6감정 = 120문장
- 중의적인 문장 20개 X 6감정 = 120문장
- 한 사람당 240문장 * 20명 => 4800개의 데이터

Examples

감정 유발 문장		중의적 문장	
Anger		Anger	
Excitement		Excitement	
Fear		Fear	
Sadness		Sadness	
Happiness		Happiness	
Neutral		Neutral	

Database #2

- 총 90문장 (평서문 30, 질문 30, 명령 및 요청 30)
- 4가지 감정 (기쁨, 짜증, 우울, 보통) 으로 한 사람 당 총 360문장 녹음
- 10명의 **연기자 및 일반인** 녹음 **(90 * 4 * 10 = 총 3600문장)**

보통 기쁨

우울 짜증

MOS Test

- 감정별로 남성 70개 여성 70개의 데이터를 임의로 선별
- 수집한 4800개의 음성 데이터 평가
- 평균감정점수

- Anger – 4.59
- Excitement - 4.35
- Fear – 4.23
- Sadness – 4.15
- Happiness – 4.33
- Neutral – 4.78

Blind Test

- MOS Test로 선별된 840개의 데이터를 대상으로 진행
- 평가자는 어떤 감정에 해당하는 음성 파일인지 알지 못하는 상태로 진행
- 음성 파일을 듣고 6개의 감정 중 느껴지는 감정으로 분류
- 외국인 포함 총 13명의 인원이 평가에 참여

Test Result

		응 답					
		Anger	Excitement	Fear	Sadness	Happiness	Neutral
정답	Anger	**99.73**	0	0.16	0.05	0.05	0
	Excitement	1.65	**85.**		0.38	12.69	0.05
	Fear	0.05			.58	0.05	0.05
	Sadness	0.11	0.05		**97.42**	0.05	0.27
	Happiness	0.22	7.97	0	0.05	**90.11**	1.65
	Neutral	0	0	0	0.22	0.05	**99.73**

94.89%

Emotion Recognition

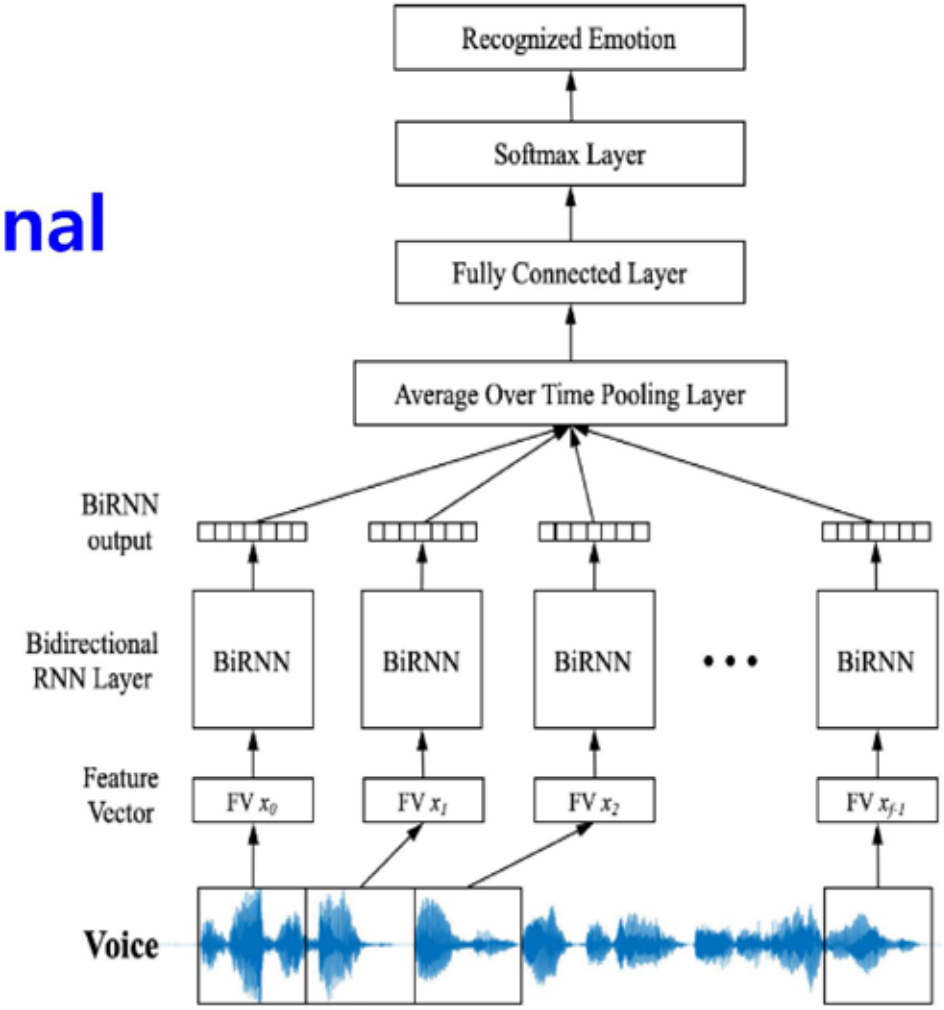
Acoustic Features

	Acoustical characteristics	Acoustic Feature	Expression
Anger	평균 피치가 아주 높으며, 피치의 변화도 크고, 말소리의 속도는 일반적으로 느려지면, 긴장도가 강해짐.	Harmonic Energy	목소리의 톤의 높고, 발화시 매우 얇고 날카로운 목소리가 나타남.
Excitement	운율의 변화가 심함.	Pitch_Medi	소리를 지르는듯한느낌의 높고 큰 목소리가 나 타남
Fear	피치의 평균치가 높아지고 , 피치의 범위는 커짐.	Inharmonicty	목소리의 톤의 변화가 불규칙적이며, 발화의 이어지는 구간에는 묵음이 존재하고, 발화시에는 목소리에 진동과 함께 발생됨
Happiness	피치와 피치의 범위가 증가하며, 크게 증가된 피치의 범위는 아주 서서히 감소됨.	Centroid	호흡은 거의 들리지 않고, 목소리가 크고, 발음이 뚜렷하며, 밝고 활기찬 어조로 말함
Sadness	한번의 발화에서 호흡과 혼재되어 나타남. 억양구의끝에 BPMs(boundary pitch movements) 나타남. 평균 피치 값이가장 낮고 속도도 느리며, 강도가 낮아지고, 불규칙적인 휴지가 발생됨.	NoiseEnergy	단어발화시에 호흡과 같은 잡음과 동시에 발화됨. 이로 인하여 단어가 명확하게 들리지 않음. 발화시에 단어당 음성에서 나오는 잡음(호흡, 울음) 등이 혼재되어 많음. 잦은 쉼이 나타나고 낮다.(Sadness가 낮음)
Neutral	-	Noiseness	한번의 발화 끝에 명확한 쉼이 나타남. 단어가 명확하게 표현이 되고, 목소리 주변의 잡음이 거의 없음. 단어와 단어사이 , 구절과 구절 사이가 쉼이 나타남. 호흡은 규칙적으로 발현됨.

그 외 MFCC, Spectral Roll-off, Chroma 등 이용

Recognition using Deep Learning

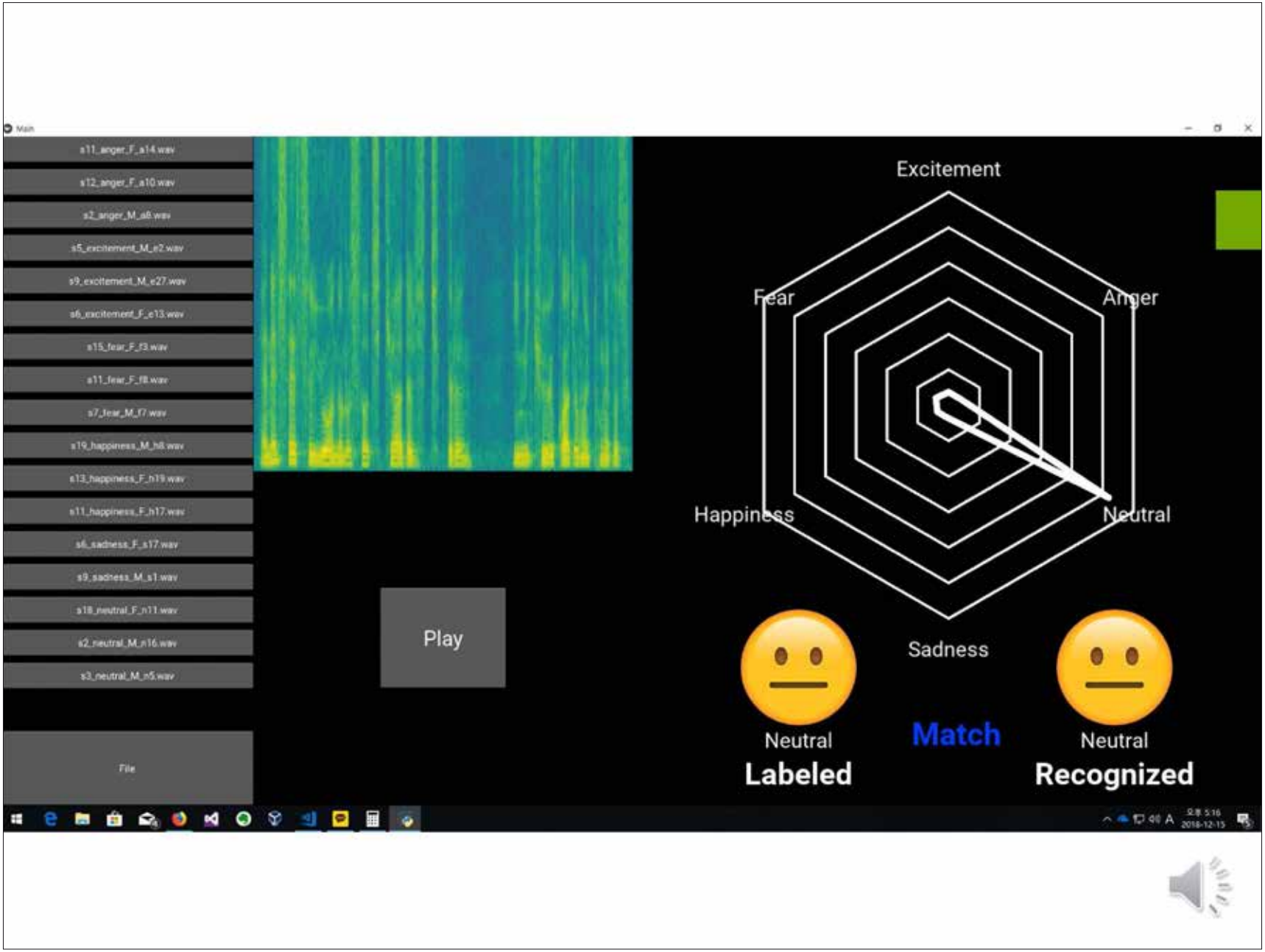
Bi-directional RNN



Recognition Rate

87% for DB #1
94% for DB #2

Demonstration


Main
s11_anger_F_a14.wav
s12_anger_F_a10.wav
s2_anger_M_a8.wav
s5_excitement_M_e2.wav
s9_excitement_M_e27.wav
s6_excitement_F_e13.wav
s15_fear_F_f3.wav
s11_fear_F_f8.wav
s7_fear_M_f7.wav
s19_happiness_M_h8.wav
s13_happiness_F_h19.wav
s11_happiness_F_h17.wav
s6_sadness_F_s17.wav
s9_sadness_M_s1.wav
s18_neutral_F_n11.wav
s2_neutral_M_n16.wav
s3_neutral_M_n5.wav
File
Play
Excitement
Fear
Anger
Happiness
Neutral
Sadness
Neutral
Labeled
Match
Neutral
Recognized

Q&A

Thank you for listening!