

음성신호기반의 감정인식의 특징 벡터 비교

A Comparison of Effective Feature Vectors for Speech Emotion Recognition

신 보 라* · 이 석 필†
(Bo-Ra Shin · Soek-Pil Lee)

Abstract - Speech emotion recognition, which aims to classify speaker's emotional states through speech signals, is one of the essential tasks for making Human-machine interaction (HMI) more natural and realistic. Voice expressions are one of the main information channels in interpersonal communication. However, existing speech emotion recognition technology has not achieved satisfactory performances, probably because of the lack of effective emotion-related features. This paper provides a survey on various features used for speech emotional recognition and discusses which features or which combinations of the features are valuable and meaningful for the emotional recognition classification. The main aim of this paper is to discuss and compare various approaches used for feature extraction and to propose a basis for extracting useful features in order to improve SER performance.

Key Words : Speech emotion recognition, Feature extraction, Speech processing, Preprocessing

1. 서 론

최근 급속한 컴퓨터 기술의 발전에 의해 인간과 컴퓨터 사이의 인터페이스가 양방향성으로의 변화가 요구되고 있으며, 사용자 정보를 활용한 다양한 개인화 서비스 연구가 주목받고 있다. 이러한 휴먼 인터페이스 시스템에서는 사용자의 감정 상태를 추출하고 인식하여 그 상태에 맞는 서비스를 제공하는 데에 목표가 있다. 예를 들어, 가정용 로봇은 사람과의 우호적인 상호 작용을 위해 인간의 말로부터 감정을 인식할 수 있다[1]. 고객 관리 인터페이스에서도 감정 인식 시스템을 사용하여 고객 만족도와 서비스 품질을 평가할 수 있다[2]. 이러한 모든 응용 프로그램은 사용자의 감정을 이해하면 도움이 된다. 이 모든 서비스가 감정 정보를 활용한 개인화 서비스이다.

감정이란, 외부의 물리적 자극에 의한 감각, 지각으로부터 인간의 내부에 야기되는 심리적 체험으로 복합적인 감정을 말한다[3]. 이러한 감정 정보는 여러 방법으로 접근할 수 있다. 얼굴 표정, 음성, 심장 박동 수, 혈압, 체온이 대표적인 방법이지만 적용분야마다 유용한 감정 정보를 취득하는 방법은 다르다. 본 연구에서는 음성신호를 통해 감정 상태를 인식하는 접근 방법을 다루고자 한다.

음성기반 감정 인식이란 사용자의 음성신호를 분석하여 사용자의 감정을 자동으로 인식하는 기술이다[4]. 인간의 소리는 의사소통을 하기 위한 수단이자 감정을 전달하는 수단으로 사용된다.

다. 소리에 포함된 감정은 화자의 심리상태를 표현한다. 소리를 이용한 감정인식에는 단어의 의미로부터 감정을 인식하는 방법, 단어의 의미와 상관없이 운율적인 정보만을 이용하는 방법, 그리고 두 가지를 모두 사용하는 방법 등에 대한 많은 연구가 있었다[5]. 하지만 음성기반 감정 인식은 얼굴 표정인식과 같은 다른 감정인식 방법에 비해 다소 낮은 인식률을 보인다. 이는 사람의 감정정보를 효과적으로 포함하는 음성 특징의 부족한 성능 차이라고 볼 수 있다. 인식 시스템 자체의 성능을 높이는 것만큼 음성 감정 인식을 위한 음성신호의 여러 가지 특징을 추출하여 분석하고 어떤 특징이 감정인식에 영향을 주는 유의미한 음성신호의 특징인지를 찾아내는 것도 매우 중요하다.

기존 음성 감정 인식에서 널리 사용되었던 특징에는 음성의 pitch, energy, tempo과 spectral feature인 MFCC, MEDC, LPC, LPCC 등이 있다. 따라서 본 논문은 이러한 음성 특징 중에서 사람의 감정 정보를 효과적으로 포함하는 음성신호의 특징을 찾아 내는 것에 목적을 둔다. 본 연구에서는 IEMOCAP_DB에서 제공하는 감정인식을 위한 데이터 셋을 이용한다. 감정 표현이 명확한 감정인식 분야에서 일반적으로 많이 사용하는 대표 감정인 화남, 즐거움, 두려움, 평범, 슬픔 5가지의 감정을 정의한다. 모든 감정의 음성파일은 남성과 여성의 비율을 1:1로 각 감정당 20개, 총 100개의 데이터를 사용한다. 100개의 데이터에서 감정이 섞여 있는 동일한 구간에서 Mel-Frequency Cepstral Coefficient(MFCC), MEDC(MFCC first derivative), LPC(Linear Predictive Coding), LPCC(Linear Predictive Cepstral Coding), tempo, IAV(Integral Absolute Value), DAMV(Difference Absolute Mean Value), ZCR(Zero Crossing Rate) 특징 값을 각각 추출한다. 각각의 특징벡터들의 분리도를 비교하기 위해 Bhattacharyya거리 측정법을 이용한다[6]. Bhattacharyya 거리

† Corresponding Author : Dept. of Electronic Engineering, SangMyung University, Korea.
E-mail: esprit@smu.ac.kr

* Dept. of Computer Science, SangMyung University, Korea.
Received : July 20, 2018; Accepted : August 31, 2018

값을 비교 분석하여 감정분석에 가장 적합한 feature를 선정한다.

2. 실험 데이터

본 논문에서는 IEMOCAP Dataset을 실험데이터로 선정하였다. 감정 카테고리로는 감정정의가 명확히 분류되면서 감정 인식 분야에서 일반적으로 많이 사용하는 화남(angry), 즐거움(happy), 두려움(fear), 슬픔(sad), 평범(neutral) 5가지로 정의한다. 특징벡터 추출과 분석을 위해 쓰인 데이터는 각 감정 당 20개의 데이터로 총 100개이고 남성 음성과 여성 음성을 동일한 비율로 사용한다. 음성 데이터는 모두 16KHz의 Sampling rate를 갖는다. 감정이 명확하게 섞인 발화구간에서의 특징 추출을 위해 모든 음성데이터에서 1.25~1.5s구간에서의 특징 값들을 사용한다.

3. 특징벡터

3.1 템포

템포는 BPM(beat per minute) 단위를 사용하며 일정한 간격으로 규칙을 띠고 반복되는 소리의 움직임의 단위인 비트(beat)가 1분 내에 들어있는 비트의 수를 의미한다. 하지만 사람의 음성의 경우, 템포는 정량적인 리듬 혹은 사람의 말하는 빠르기를 의미한다. 따라서 음절은 사람의 말 빠르기를 인식하는 척도이며 각각의 음절을 이용하여 음성의 리듬 혹은 빠르기를 측정할 수 있다[7].

3.2 MFCC, MEDC

MFCC는 사람의 청각기관이 저주파수 대역에서 민감한 반면 고주파수 대역에서 상대적으로 둔감한 특성을 표현한 멜 스케일(mel scale)에 기반한 음성 특징으로서 음성인식과 화자 인식 분야에서 모두 널리 사용된다. 멜 스케일은 물리적인 음 높이와 청각 인지적인 음높이의 관계를 표현하는 것으로서, Stevens 등에 의해 명명되었다[8]. 식 (1)은 Hz 단위로 표현되는 물리적인 주파수 f 를 mel 단위의 청각 인지적 음높이 m 으로 변환하는 식이다.

$$m = 1127 \log_c \left(1 + \frac{f}{700} \right) \quad (1)$$

MFCC 추출과정은 <그림 1>과 같다. 우선 음성 신호로부터 매 프레임 단위로 윈도우 함수를 씌운 다음 discrete Fourier transform(DFT) 과정을 통해 시간 영역에서 주파수 영역으로 변환시키며, 실제 DFT 과정은 연산의 효율성을 위해 fast Fourier transform(FFT)의 형태로 구현된다. 그 다음으로 멜 스케일을 가지도록 식 (1)을 사용하여 주파수 축을 워핑한다음 이 스케일에서 동일한 대역폭을 가지는 삼각 필터뱅크를 통해 필터뱅크 별 에너지를 구한다. 여기에 로그 함수를 취한 다음 DCT를 통해 최종적인 MFCC 값들을 구하게 된다[9].



그림 1 MFCC 추출 과정

Fig. 1 The process of MFCC extraction

MFCC 특징 벡터는 단일 프레임의 파워 스펙트럼 값을 기술한다. 매 프레임 단위로 추출되는 음성 특징들의 시간에 따른 동적 변화 특성을 나타내기 위해 DELTA특징이 사용된다. MFCC의 DELTA 특징이 MEDC이다.

3.3 LPC, LPCC

LPC는 선형 결합에 의해 과거의 신호에서 현재의 신호[n]을 예측하는 방법으로 전극(All-pole) 모델을 사용하여 식 (2)과 같이 차분 방정식의 형태로 나타낼 수 있다. 여기서 s_n 은 입력신호, \tilde{s}_n 은 예측신호, a_i 는 선형예측계수이며, p 는 예측계수의 차수이다. 현재 신호와 예측된 신호의 예측오차는 식 (3)과 같다 [10].

$$\tilde{s}_n = -(a_1 s_{n-1} + a_2 s_{n-2} + \dots + a_p s_{n-p}) \quad (2)$$

$$e_n = s_n - \tilde{s}_n \quad (3)$$

식 (4)는 예측오차에 대한 mean square error(MSE) J이며 J를 최소화 하는 선형예측계수를 찾기 위하여 식 (4)에 대하여 편미분하여 0이 되는 p개의 선형 연립방정식 (5)을 얻을 수 있다. 식 (5)는 식 (6)과 같이 나타낼 수 있고, 선형예측계수는 자기상관 행렬의 역행렬을 이용하여 구할 수 있다[10].

$$J = E[e^2(n)] = E[s_n - \tilde{s}_n^2] \quad (4)$$

$$\sum_{j=1}^p a_j E_s(n-i)s(n-j) = E_s(n-i)s(n) \text{ for } i = 1 \dots p \quad (5)$$

$$\begin{bmatrix} R_0 & \dots & R_{p-1} \\ \vdots & \ddots & \vdots \\ R_{p-1} & \dots & R_0 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = - \begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_p \end{bmatrix} \quad (6)$$

LPCC는 C(z)의 inverse z-transform으로 정의되고 식은 다음과 같다.

$$C(z) = \sum_n c(n)z^{-n} \quad (7)$$

3.4 IAV, DAMV

IAV 특징벡터는 신호의 절대 적분치를 나타내는 특징으로, 일정시간 동안의 신호의 절대값을 적분한 것이며 식은 다음 식 (8)과 같다[11].

$$\bar{X} = \sum_{i=1}^N |X(i\Delta t)| \quad (8)$$

DAMV 특징벡터는 각 시계열 신호의 차분 값에 대한 절대 평균치를 나타내는 특징으로 식 (9)와 같다[11].

$$\bar{X} = \frac{\sum_{i=1}^{N-1} |X(i\Delta t) - X((i+1)\Delta t)|}{N-1} \quad (9)$$

3.5 ZCR

Zero Crossing Rate(ZCR, 영 교차율)은 음성 신호 파형이 중심 축(0)을 통과하는 횟수이자, 신호의 부호가 바뀌는 비율이다. 식(10)과 같다.

$$zcr = \frac{1}{T-1} \sum_{i=1}^{T-1} 1R < 0^{(S_i, S_{i-1})} \quad (10)$$

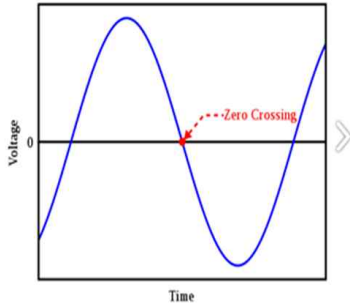


그림 2 Zero Crossing
Fig. 2 Zero Crossing

4. 실험 및 결과

4.1 분리도 비교

특징 추출을 위한 Windowing size는 25ms이며, step size는 10ms이다. 음성 데이터 중 감정이 충분히 담겨있는 발화 구간에 대해 특징 값을 각각 추출한다. 특징벡터 간의 정확한 비교를 위해 모든 특징 값은 동일한 구간에 대한 값이며 구간에 대한 평균값으로 정의한다. MFCC, MEDC, LPC, LPCC계수의 차수는 기존 연구들과 같이 10차로 정의한다[12].

본 논문에서는 특징 벡터들이 감정인식에 적합한지를 평가하기 위해 각 클래스 간 분리도를 계산하는 방법인 Bhattacharyya 거리 측정법을 이용한다. Bhattacharyya 거리의 식은 식 (11)과 같다.

$$\mu(1/2) = \frac{1}{8} (M_2 - M_1)^T \left\{ \frac{\sum_1 + \sum_2}{2} \right\}^{-1} (M_1 - M_2) + \frac{1}{2} \ln \frac{\left| \frac{\sum_1 + \sum_2}{2} \right|}{\sqrt{\left| \sum_1 \right| \left| \sum_2 \right|}} \quad (11)$$

Bhattacharyya거리 값이 크면 각 감정 카테고리 간의 거리가 멀다는 의미로, 해당 특징 벡터가 감정인식에 적합하다는 것을 의미한다.

4.2 특징벡터 분리도 결과

4.2.1 MFCC 분리도

Bhattacharyya 거리 평균값: 2.099877

표 1 MFCC 분리도

Table 1 Separation of MFCC

	Angry	Fear	Happy	Neutral	Sad
Angry	0	2.323838	1.493601	4.002856	4.995931
Fear	2.323838	0	1.6447	1.672888	1.817183
Happy	1.493601	1.6447	0	2.583564	3.37922
Neutral	4.002856	1.672888	2.583564	0	2.334686
Sad	4.995931	1.817183	3.37922	2.334686	0

4.2.2 MEDC 분리도

Bhattacharyya 거리 평균값: 1.259347

표 2 MEDC 분리도

Table 2 Separation of MEDC

	Angry	Fear	Happy	Neutral	Sad
Angry	0	1.619398	1.473508	1.605921	1.725975
Fear	1.619398	0	1.362355	1.436939	1.660916
Happy	1.473508	1.362355	0	1.510874	1.773905
Neutral	1.605921	1.436939	1.510874	0	1.572046
Sad	1.725975	1.660916	1.773905	1.572046	0

4.2.3 LPC 분리도

Bhattacharyya 거리 평균값: 1.787003

표 3 LPC 분리도

Table 3 Separation of LPC

	Angry	Fear	Happy	Neutral	Sad
Angry	0	2.012452	1.156584	2.309482	3.670899
Fear	2.012452	0	1.907243	1.827541	2.369113
Happy	1.156584	1.907243	0	1.888744	3.001996
Neutral	2.309482	1.827541	1.888744	0	2.193484
Sad	3.670899	2.369113	3.001996	2.193484	0

4.2.4 LPCC 분리도

Bhattacharyya 거리 평균값: 1.81059

표 4 LPCC 분리도

Table 4 Separation of LPCC

	Angry	Fear	Happy	Neutral	Sad
Angry	0	2.427193	1.621491	2.760209	3.52193
Fear	2.427193	0	1.84203	1.466132	1.923333
Happy	1.621491	1.84203	0	1.914422	2.680383
Neutral	2.760209	1.466132	1.914422	0	2.47525
Sad	3.52193	1.923333	2.680383	2.47525	0

4.2.5 IAV 분리도

Bhattacharyya 거리 평균값: 0.029627

표 5 IAV 분리도

Table 5 Separation of IAV

	Angry	Fear	Happy	Neutral	Sad
Angry	0	0.024353	0.024552	0.035261	0.007435
Fear	0.024353	0	0.021087	0.051762	0.0278
Happy	0.024552	0.021087	0	0.098548	0.007865
Neutral	0.035261	0.051762	0.098548	0	0.071676
Sad	0.007435	0.0278	0.007865	0.071676	0

4.2.6 DAMV 분리도

Bhattacharyya 거리 평균값: 0.085453

표 6 DAMV 분리도

Table 6 Separation of DAMV

	Angry	Fear	Happy	Neutral	Sad
Angry	0	0.078334	0.092661	0.248974	0.292721
Fear	0.078334	0	0.001661	0.058451	0.099471
Happy	0.092661	0.001661	0	0.050491	0.105606
Neutral	0.248974	0.058451	0.050491	0	0.039795
Sad	0.292721	0.099471	0.105606	0.039795	0

4.2.7 Tempo 분리도

Bhattacharyya 거리 평균값: 0.035547

표 7 Tempo 분리도

Table 7 Separation of Tempo

	Angry	Fear	Happy	Neutral	Sad
Angry	0	0.027412	0.068255	0.013194	0.064162
Fear	0.027412	0	0.108943	0.019765	0.055456
Happy	0.068255	0.108943	0	0.039738	0.027617
Neutral	0.013194	0.019765	0.039738	0	0.019794
Sad	0.064162	0.055456	0.027617	0.019794	0

4.2.7 ZCR 분리도

Bhattacharyya 거리 평균값: 0.086606

표 8 ZCR 분리도

Table 8 Separation of ZCR

	Angry	Fear	Happy	Neutral	Sad
Angry	0	0.058603	0.045496	0.17989	0.189112
Fear	0.058603	0	0.064183	0.061353	0.03649
Happy	0.045496	0.064183	0	0.22584	0.135041
Neutral	0.17989	0.061353	0.22584	0	0.086561
Sad	0.189112	0.03649	0.135041	0.086561	0

4.3 실험 결과

Bhattacharyya 거리 평균값은 MFCC 특징 벡터 값이 2.099877로 가장 높고, IAV가 0.029627로 가장 낮다. 평균값이 큰 특징벡터는 MFCC, LPCC, LPC, MEDC, ZCR, DAMV, Tempo, IAV순이다. 감정 카테고리마다 분리도 비교를 통해 어떤 감정간에 어떤 특징벡터가 분리가 잘 되는지를 확인하고자 표를 정리해보면 다음과 같다.

표 9 감정 카테고리 간의 분리도 비교표

Table 9 Separation comparison of emotion categories

	A&F	A&H	A&N	A&S	F&H
MFCC	2.323838	1.493601	4.002856	4.995931	1.644700
MEDC	1.619398	1.473508	1.605921	1.725975	1.362355
LPC	2.012452	1.156584	2.309482	3.670899	1.907243
LPCC	2.427193	1.621491	2.760209	3.521930	1.842030
IAV	0.024353	0.024552	0.035261	0.007435	0.021087
DAMV	0.078334	0.092661	0.248974	0.292721	0.292721
Tempo	0.027412	0.068255	0.013194	0.064162	0.064162
ZCR	0.058603	0.045496	0.179890	0.189112	0.189112

	F&N	F&S	H&N	H&S	N&S
MFCC	1.672888	1.817183	2.583564	3.379220	2.334686
MEDC	1.436939	1.660916	1.510874	1.773905	1.572046
LPC	1.827541	2.369113	1.888744	3.001996	2.193484
LPCC	1.466132	1.923333	1.914422	2.680383	2.475250
IAV	0.051762	0.027800	0.098548	0.007865	0.071676
DAMV	0.058451	0.099471	0.050491	0.105606	0.039795
Tempo	0.019765	0.055456	0.039738	0.027617	0.019794
ZCR	0.061353	0.036490	0.225840	0.135041	0.086561

카테고리 간의 비교분석을 통해, 감정 인식에 사용되는 특징 벡터들 중 MFCC, LPC, LPCC값이 분리도가 비교적 높게 나타나는 것을 볼 수 있다. MFCC값은 angry&neutral, angry&sad, neutral&happy, happy&sad 분리에서 높은 값을 가지고, LPC는 fear&happy, fear&neutral, fear&sad 분리, LPCC는 angry&fear, angry&happy, neutral&sad에서 높은 값을 갖는다. 그림 3은 감정 카테고리 간의 분리도 비교 분석 그래프이다.

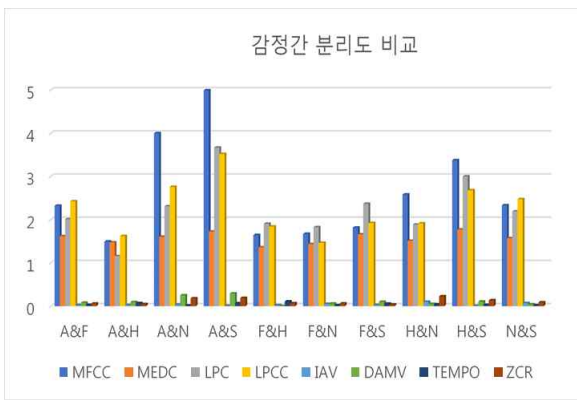


그림 3 감정 카테고리 간의 분리도 비교 분석 그래프
Fig. 3 Chart of separation comparison of emotion categories

3. 결 론

음성신호 기반의 감정인식 정확도를 높이기 위해서는 정확한 분류엔진을 선택하는 것만큼 적절하고 의미있는 특징벡터를 선택하는 것이 매우 중요하다[13]. 본 논문에서는 음성신호의 다양한 음성 특징들을 비교하여 감정 인식에 적절한 특징벡터를 제시한다. 실험 데이터로는 IEMOCAP DB의 음성 파일을 사용하였고, Angry, Fear, Happy, Neutral, Sad 총 5가지 감정 카테고리를 분류한다. 데이터는 각 감정 별로 20개, 총 100개의 데이터를 사용하고 남성 음성과 여성 음성은 일대일로 균일하게 사용하였다. 전체 데이터에서 감정이 잘 담겨있는 약 0.25초의 음성 구간을 추출한다. 추출된 동일한 음성구간에서 기존 연구에서 빈번히 사용되었던 MFCC, MEDC, LPC, LPCC, IAV, DAMV, Tempo, ZCR 총 8가지의 특징 벡터를 추출한다. 추출한 특징벡터 값의 평균값을 구하고 Bhattacharyya 거리 측정법을 사용해 각 특징벡터간

의 분리도를 비교 분석한다. 분석 결과에 따르면, 앞서 정의한 5 가지 감정 카테고리에 대하여 평균거리값이 가장 큰 특징벡터는 MFCC이고, 가장 분리도가 낮은 특징벡터는 IAV이다. 2가지 감정 마다의 분리도 비교에서는 MFCC는 angry&neutral, angry&sad, neutral&happy, happy&sad 분리에서 높은 값을 가지고, LPC는 fear&happy, fear&neutral, fear&sad 분리, LPCC는 angry&fear, angry&happy, neutral&sad에서 높은 값을 갖는다.

향후, 본 논문의 실험 결과를 바탕으로 적절한 특징벡터의 조합과 정확한 분류엔진의 병렬 사용으로 더 높은 음성신호 기반의 감정 인식 성능을 기대할 수 있다.

감사의 글

본 연구는 2018년도 상명대학교 교내연구비를 지원받아 수행하였음.

References

- [1] Xu Huahu, Gao Jue and Yuan Jian, "Application of speech emotion recognition in intelligent household robot", *inproceedings-International Conference on Artificial Intelligence and Computational Intelligence*, AICI 2010, Vol. 1, 2010, pp. 537-541.
- [2] Laurence Vidrascu and Laurence Devillers, "Detection of reallife emotions in call centers", *in Interspeech 2005*, 2005, pp. 1841-1844.
- [3] Yigon Kim and Yong-Chel Bae, 2000, "Design of Emotion Recognition Model Using Fuzzy Logic", *Journal of Korean Institute of Intelligent Systems*, Vol. 10, No. 1, pp. 268~282.
- [4] Jae Hun Bang and Sungyoung Lee, 2014, "Call Speech Emotion Recognition for Emotion based Services", *Journal of KISS : Software and Applications*, Vol. 41, No. 3, pp. 208-213.
- [5] Byungwook Jung, Seungpyo Cheun, Yountae Kim and Sungshin Kim, 2008, "An Emotion Recognition Technique using Speech Signals", *Journal of Korean Institute of Intelligent Systems*, Vol. 18, No. 4, pp. 494~500.
- [6] Seok-Pil Lee, Sang-Hui Park, Jeong-Seop Kim, Ig-Jae Kim "EMG pattern recognition based on evidence accumulation for prosthesis control", *Proc Ann Intl Conf IEEE Eng Med Biol 4*, pp. 1481-1483, 1996.
- [7] Sung-Woo Byun and Seok-Pil Lee, 2016, "Emotion Recognition Using Tone and Tempo Based on Voice for IoT", *The transactions of The Korean Institute of Electrical Engineers*, Vol. 65, No. 1, pp. 116-121.
- [8] T.-L. Pao, Y.-T. Chen, J.-H. Yeh, and P.-J. Li, "Mandarin

emotional speech recognition based on SVM and NN,” *Proc. of the 18th Int’l Conf. on Pattern Recognition (ICPR), Washington, DC*, pp. 1096-1100, Aug 2006.

- [9] Choi, Young Ho, Ban, Sung Min, Kim, Kyung-Wha and Kim, Hyung Soon, 2015, “Evaluation of Frequency Warping Based Features and Spectro-Temporal Features for Speaker Recognition”, *Phonetics and Speech Sciences*, Vol. 7, No. 1, pp. 3~10.
- [9] Choi, Young Ho, Ban, Sung Min, Kim, Kyung-Wha and Kim, Hyung Soon, 2015, “Evaluation of Frequency Warping Based Features and Spectro-Temporal Features for Speaker Recognition”, *Phonetics and Speech Sciences*, Vol. 7, No. 1, pp. 3~10.
- [10] Ha-Na Choi, Sung-Woo Byun and Seok-Pil Lee, 2015, “Discriminative Feature Vector Selection for Emotion Classification Based on Speech”, *The transactions of The Korean Institute of Electrical Engineers*, Vol. 64, No. 9, pp. 1363~1368.
- [11] Sung-Woo Byun and Seok-Pil Lee, Kunyun Kim and Sang-Hyun Han, 2017, “Gesture recognition with wearable device based on deep learning”, *Broadcasting and Media Magazine*, Vol. 22, No. 1, pp. 58~66.
- [12] Kwang-Seung Heo, Chang-Hyun Park, Dong-Wook Lee, and Kwee-Bo Sim “speaker identification using incremental neural network and LPCC”, *Journal of The Korean Institute of Intelligent Systems*, Volume 12 Issue 2, 2002.12. pp. 341-344.
- [13] Hyun Woo Kim, Sung Yong Lee “The Phoneme Kernel Technique based on Support Vector Machine for Emotion Classification of Mobile Texts”, *Journal of KIISE: software and application*, Volume 40 Issue 6 (2013.6) 350-355.



이 석 필(Seok Pil Lee)

1990년 연세대학교 전기공학과 공학사
 1992년 연세대학교 전기공학과 공학석사
 1997년 연세대학교 전기공학과 공학박사
 1997년~2002년 대우전자 영상연구소 선임연구원. 2002년~2012년 KETI 디지털미디어연구센터 센터장. 2010년~2011년 미국 Georgia Tech. 방문연구원. 2012년~현재 상명대학교 융합전자공학과 교수.
 <주관심분야> 멀티미디어 검색, 방송통신시스템, 인공지능

저 자 소 개



신 보 라 (Bo-Ra Shin)

2018년 상명대학교 디지털미디어학과 이학사. 2017년~현재 상명대학교 컴퓨터과학과 학·석사 연계과정.
 <주관심분야> 멀티미디어처리, 인공지능, 음성신호처리