ISSN 2005-8063
2017. 12. 31.
Vol.9 No.4
pp. 91-97

말소리와 음성과학

Phonetics and Speech Sciences

한국음성학회지

http://dx.doi.org/10.13064/KSSS.2017.9.4.091

목소리 특성과 음성 특징 파라미터의 상관관계와 SVM을 이용한 특성 분류 모델링*

Correlation analysis of voice characteristics and speech feature parameters, and classification modeling using SVM algorithm

박 태 성 · 권 철 홍**
Park, Tae Sung · Kwon, Chul Hong

Abstract

This study categorizes several voice characteristics by subjective listening assessment, and investigates correlation between voice characteristics and speech feature parameters. A model was developed to classify voice characteristics into the defined categories using SVM algorithm. To do this, we extracted various speech feature parameters from speech database for men in their 20s, and derived statistically significant parameters correlated with voice characteristics through ANOVA analysis. Then, these derived parameters were applied to the proposed SVM model. The experimental results showed that it is possible to obtain some speech feature parameters significantly correlated with the voice characteristics, and that the proposed model achieves the classification accuracies of 88.5% on average.

Keywords: voice characteristics, speech feature parameters, correlation, SVM, classification modeling

1. 서론

음성신호를 활용한 감정인식 시스템은 일반적으로 다수의 화자로부터 수집된 음성 데이터를 이용하여 기계학습 방법 등을 통하여 인식 모델을 생성한다[1]. 그동안 감정인식 연구 분야에서는 인식률을 높이기 위하여 감정인식에 적합한 음성 특징 파라미터를 도출하거나, 감정을 분류하는 모델링 방법론에 대한 연구가 진행되어 왔다.

이러한 방식은 다수 화자의 음성 DB를 기반으로 범용 모델을 생성하는 것으로 화자 개개인의 음성 특성은 고려하지 않는다. 이로 인하여 음성 DB 수집 시 포함되지 않은 다른 화자의 음성 데이터가 입력으로 사용될 경우, 모델 생성 단계에서 얻은 인식률에 비하여 성능이 저하된다.

음성 기반 감정인식 기술이 높은 정확성을 보여주지 못하고 있는 주요 원인은, 사람마다 고유한 음성 특성과 감정표현 방식 등이 있어 같은 감정을 표현하는데 차이가 있는데, 감정인식 분류를 위한 학습과정에서 이런 부분이 고려되지 않아, 학습에 사용된 음성 데이터가 아닌 다른 화자의 음성이 입력되는 경우 인식률이 떨어진다는 점이다.

이러한 감정인식 시스템의 문제점을 보완하기 위하여 개인화된 감정인식 시스템을 만드는 자율 특징 학습 방법이 진행되어 왔다[1], [2]. 이 연구는 사용자의 감정과 이에 해당하는 음성

* 이 논문은 과학기술정보통신부 및 정보통신기술진흥센터의 고용계약형 SW석사과정 지원사업(2015-9-00999) 및 2017학년도 대전대학교 교내학술연구비 지원에 의한 연구결과로 수행되었음
** 대전대학교, chkwon@dju.ac.kr, 교신저자
Received 30 October 2017; Revised 15 December 2017; Accepted 15 December 2017

데이터를 피드백 받아 음성 특징 파라미터를 화자에 맞게 정규화 하여 모델 생성 시 사용한 데이터와의 수치 값을 줄이는 방법이다. 이 방식은 범용적인 데이터와 사용자 데이터 값과의 차이를 감소시켜 인식률을 개선시키려는 방법일 뿐, 감정을 표현하는 음성 특성이 사람마다 다르다는 점을 해결해 주는 방식은 아니다.

음성 기반으로 감정인식을 수행하여 사용자의 감정을 정확히 판별하기 위해서는, 입력 음성 데이터를 더욱 세분화하여 특징을 추출하고 다양하게 분류하는 선행 작업이 필요하다. 이를 위해 본 논문에서는 감정인식 모델을 생성하는 학습과정에 들어가기 전, 목소리 특성을 분류하는 모델을 이용해 입력 음성 데이터를 목소리 특성별로 분류하고 학습을 수행하는 방안을 제안한다.

본 연구에서는, 사람의 목소리 특성을 몇 가지로 정의하고 음성 DB를 청취 평가하여 음성 데이터의 목소리 특성을 분류한다. 수집된 음성 DB에서 다양한 음성 특징 파라미터를 추출하고, ANOVA(Analysis of Variance) 분산분석을 통해 목소리 특성의 각 항목과 상관관계가 높은 유의미한 음성 파라미터를 도출한다. Support Vector Machine(SVM) 알고리즘을 통해 목소리 특성을 분류하는 모델을 제안하고, 도출된 유의미한 음성 파라미터를 적용하여 제안한 모델의 성능을 평가한다.

2. 실험 방법

2.1. 피험자 및 음성 DB

피험자는 섭외가 용이한 20대 남성 222명을 대상으로 하였다. 음성 DB 수집 환경은, 목소리 세기를 올바르게 측정하기 위하여 마이크와 피험자 입 사이 거리가 약 4~5cm가 되도록 녹음 주기 동안 일정하게 유지하고, 잡음의 영향을 배제하기 위하여 조용한 공간에서 수집하였다. 피험자는 '아, 에, 이, 오, 우' 등 5개의 모음을 각각 약 2초간 발성하고, '우리는 높은 산에 올라가 맑은 공기를 마시고 왔습니다.' 라는 단문을 두 번 반복 녹음하였다. 음성 데이터는 모노 형식, 16비트, 샘플링 주파수 16kHz로 저장하였다.

2.2. 음성 특징 파라미터 추출

222명으로부터 5개의 모음과 하나의 문장에 대한 음성 데이터를 수집한 후, 끝점 검출을 수행하여 묵음 구간을 제거하고 음성 구간을 검출한 뒤, 모음과 문장에서 다음과 같은 음성 특징 파라미터를 추출한다.

추출한 음성 파라미터는, Praat를 이용하여 F0 평균(F0_mean), F0 최솟값(F0_min), F0 최댓값(F0_max), F0 표준편차(F0_std), 세기(Intensity), 포먼트와 대역폭(F1~F4, BW1~BW4), Jitter, Shimmer, Long Term Average Spectral Slope(LTAS)를, Multi Dimensional Voice Program(MDVP)을 활용하여 Relative Average Perturbation(RAP), Pitch Perturbation Quotient(PPQ), Smoothed PPQ(sPPQ), Variance of F0(vF0), Amplitude Perturbation Quotient(APQ), Smoothed APQ(sAPQ), Variance of Amplitude(vAm), Noise to Harmonics Ratio(NHR), Voice Turbulence Index(VTI), Soft Phonation Index(SPI), F0 Tremor Intensity Index(FTRI), Amplitude Tremor Intensity Index(ATRI)를, VoiceSauce를 통하여 하모닉의 진폭(H1, H2, H4, H2K), 포먼트의 진폭(A1, A2, A3), 두 진폭의 차이(H1-H2, H2-H4, H1-A1, H1-A2, H1-A3), Cepstral Peak Prominence(CPP), Harmonics to Noise Ratio(HNR) 등을 추출한다[3]-[5]. Praat[6]와 MDVP[7]는 음성공학, 음성학 및 음성의학 분야에서 많이 사용되는 음성 분석 도구이고, VoiceSauce[8]는 UCLA Phonetics lab.에서 구현한 음성분석 툴이다.

2.3. 청취 평가에 의한 목소리 특성 분류

222명이 녹음한 문장에 대하여 음성공학을 전공하는 3인이 청취 평가를 수행하여, 각 피험자의 음성에 대해 높낮이, 세기, 빠르기, 청음/탁음, 가늘다/굵다 등으로 목소리 특성을 분류하였다. 사전에 목소리 특성 각 항목을 어떻게 정의할 것인지를 논의하고, 문장을 수차례 들으면서 평가 훈련을 하여 각 항목에 대해 판정 기준을 세웠다. 실제 분류를 할 때는 피험자의 문장을 듣고 각자 판단을 하여 의견을 내어 판정을 하고, 의견이 통일되지 않는 경우에는 여러 차례 청취하면서 합의하는 과정을 거쳤다.

목소리 특성 각 항목에 대해 '작다/보통/크다'로 1점에서 3점까지 점수를 주었으며, 높낮이는 고음일 경우 3점이고, 세기는 클수록 3점이며, 문장을 말하는 속도가 평균적(3.5초~4.3초)일 경우 2점, 빠를수록 3점이다. 청음/탁음에서 청음은 음색이 맑고 명료한 목소리를, 탁음은 거칠고 둔탁하며 막힌 소리를 의미하며[9], 청음일 경우 1점, 탁음에 가까울수록 3점이다. 목소리 굵기는 얇고 가늘며 힘없을 경우 1점, 반대로 굵고 힘 있는 경우 3점을 주었다.

2.4. 목소리 특성과 음성 파라미터의 상관관계 분석을 위한 통계 처리

2.2절에서 기술한 청취 평가 결과를 이용하여 목소리 특성 각 항목에 대하여 ANOVA 분산분석을 수행하여, 각 항목 별로 세 개 그룹의 평균값을 분류하는데 어떤 파라미터가 유의확률 0.05에서 통계적으로 유의미한 값을 갖는지를 도출한다. ANOVA는 세 개 이상의 집단 간 유의미한 평균 차이를 보여 주는 파라미터를 도출해주는 통계 기법[10]으로, SPSS 버전 23[11]을 사용하여 통계처리 했는데, 요인에 목소리 특성 각 항목 중 한 가지를, 종속변수에 음성 파라미터를 설정하면 유의미한 파라미터를 도출해준다.

2.5. SVM 을 이용한 목소리 특성 분류 모델링

2.4절에서 구한 유의미한 음성 파라미터를 이용하여 SVM을 통해 목소리 특성을 분류하는 모델을 제안한다. SVM은 벡터 분류, 회귀 및 추정 기능을 지원해주며 다중 클래스 분류도 가능하다. 본 연구에서는 벡터 분류를 위해 C-Support Vector

Classification(C-SVC) 타입을 사용하고, 커널 함수로는 가우시안 Radial Basis Function 커널을 사용한다. 분류 모델 생성을 위해 SVM 공개 툴인 LIBSVM[12]을 사용했는데, 이 툴은 MATLAB을 이용하여 개발한 소프트웨어로서 SVM을 사용자들이 비교적 쉽게 사용할 수 있도록 만든 툴이다.

LIBSVM을 통해 목소리 특성 각 항목별로 분류한 데이터를 이용하여 SVM 모델을 학습한다. 먼저 LIBSVM 수행을 위해 학습 및 테스트 데이터를 <그림 1>과 같은 양식으로 작성하는데, 여기서 클래스 번호는 목소리 특성 각 항목의 그룹을 뜻하고, 인덱스는 각 음성 파라미터를 의미하며 1부터 차례대로 증가시켜 적어주고, 값에는 각 음성파라미터의 값을 입력해준다.

[클래스 번호] [인덱스 1]:[음성파라미터 값 1] [인덱스 2]:[음성파라미터 값 2] [인덱스 3]:[음성파라미터 값 3] …

그림 1. SVM 입력 데이터 양식
Figure 1. A form of input data for SVM

예를 들어, 청음/탁음의 경우 1점을 부여 받은 피험자의 유의미 음성 파라미터는 청음 클래스의 값이 되고, 3점을 부여받은 피험자의 파라미터는 탁음 클래스의 값이 되며, 클래스 번호는 1점은 +1, 3점은 -1을 부여한다. 이러한 방법을 목소리 특성의 모든 항목에 적용한다.

<그림 2>는 LIBSVM 툴을 이용해 분류 모델을 학습하고 테스트하는 절차를 보여준다[13]. 여기서 train_data는 <그림 1>과 같은 양식으로 만든 학습 데이터이고, 테스트 데이터는 test_data 파일에 저장한다. SVM 학습 및 테스트에 앞서 다양한 크기를 갖는 음성 파라미터 값을 0에서 1까지의 범위로 조정하기 위해 svm-scale을 이용한다. 학습 데이터에 대해 svm_scale을 수행하여 데이터 범위를 조정하고(<그림 2> 1번) 이를 기반으로 테스트 데이터에 대한 범위를 조정한다(<그림 2> 4번).

학습 데이터를 입력하여 svm-train을 통해 모델을 학습하면 생성된 모델이 train_model에 저장되고(<그림 2> 3번), 테이트 데이터를 입력하면 svm-predict를 통해 분류를 수행하여 학습된 모델의 예측 정확도를 산출할 수 있다(<그림 2> 5번). 예측 정확도를 높이기 위해 LIBSVM에서 제공한 라이브러리인 grid.py 스크립트를 실행하여 C-SVC를 위한 코스트(-c)와 가우시안 커널을 위한 감마(-g) 값을 최적으로 찾는 과정을 거친다(<그림 2> 2번). 최적 코스트와 감마 값을 사용하면 최종 예측 정확도를 개선시킬 수 있으므로 grid.py를 통해 산출된 값을 이용하여 모델을 학습한다.

SVM 모델을 학습하고 성능을 평가할 때, 수집된 데이터 중에서 어느 것을 학습 데이터와 테스트 데이터로 선정하는 가에 따라 성능이 달라질 수 있다. 이러한 문제점을 피하기 위하여 10-fold cross validation 실험방법을 사용한다[14]. 전체 데이터를 10개 집단으로 나누어 모델 생성 및 테스트 과정을 10회 반복 시행하고 예측 정확도의 평균값을 산출하여, 알고리즘의 성능을 모든 데이터에 대해 확인한다.

1. svm-scale -l 0 -u 1 –s data_range train_data > train_data.scale
2. python grid.py train_data.scale
3. svm-train –c (코스트) -g (감마) train_data.scale > train_model
4. svm-scale –r data_range test_data > test_data.scale
5. svm-predict test_data.scale train_model test_data.predict

그림 2. LIBSVM을 이용한 학습 및 테스트 알고리즘
Figure 2. Training and test algorithm using LIBSVM

3. 목소리 특성과 음성 파라미터의 상관관계 실험 결과

3.1. ANOVA 분석을 통한 유의미한 음성 파라미터 도출 결과

목소리 특성 각 항목에 대해 ANOVA 분석을 수행하여 각 모음별로 도출한 유의미한 파라미터 중에서, 5개 모음 중 3개 이상에서 동시에 유의미한 파라미터와, 두 번 녹음한 문장 모두에서 유의미한 파라미터를 최종적으로 유의미한 파라미터로 선정하였다. 이렇게 구한 유의미한 음성 파라미터가 <표 1>에 보인다. 빠르기와 관련된 음성 파라미터는 도출되지 못했다. 빠르기는 발화된 음의 길이로 쉽게 알 수 있으므로, 음의 길이 외 다른 음성 특징 파라미터는 필요하지 않을 것으로 판단된다.

표 1. 통계적으로 유의미한 음성 파라미터
Table 1. Statistically significant speech parameters

목소리 특성	모음/문장	유의미한 음성 파라미터
청음/탁음	모음	F0_mean, Jitter, Shimmer, APQ, CPP, HNR05, HNR15, HNR25, HNR35
	문장	BW2, CPP
가늘다/굵다	모음	CPP
	문장	sAPQ, SPI, LTAS, H1-A2
높낮이	모음	F0_mean, CPP, HNR05
	문장	F0_mean, Jitter, LTAS, SPI
세기	모음	Intensity, H1, A1, A2, H2K, HNR15
	문장	Intensity, H1, H2, A1, A2, A3, H2K, HNR05, HNR15, HNR25, HNR35

3.2. 청취평가 판정 점수별 유의미한 음성 파라미터 평균값 비교 분석

목소리 특성 각 항목에 대해 청취평가에서 판정한 점수별로 도출한 유의미한 음성 파라미터의 평균값이, 청음/탁음은 <표 2>, 가늘다/굵다는 <표 3>, 높낮이는 <표 4>, 세기는 <표 5>에 보인다. 각 파라미터의 이름 뒤에 붙은 _v와 _s은 각각 모음과 문장을 뜻하며, 예로 F0_mean_v는 모음에서의 F0 평균값을 말한다.

<표 2>에 보이는 청음/탁음 항목에서, 모음의 기본주파수 평균 F0_mean의 값은 청음이 탁음에 비해 큰 값을 보여준다. Jitter는 피치주기의 변화율을 나타내는데 탁음일수록 값이 커지는 것을 알 수 있다. Shimmer에서는 인접 프레임 간 진폭의 변화율

을 알 수 있는데 음성이 탁음에 가까울수록 값이 커지는 것을 볼 수 있다. APQ는 11개 프레임에 걸친 진폭 변화율의 평균을 나타내는데 이 또한 탁음에 가까울수록 값이 증가하는 것을 알 수 있다. 주기성의 강도를 나타내는 CPP는 음성이 청음에 가까울수록 값이 커지고 탁음일수록 작아지는데, 이는 청음일수록 주기성이 강해지고 탁음일수록 약해진다는 것을 의미한다. 하모닉에너지 대 잡음에너지의 비율을 보여주는 HNR은 HNR05, HNR15, HNR25, HNR35로 구성되어 있고[9], 쉰 목소리가 작은 값을 갖는다는 특성이 있다[15], [16]. 탁음일수록 모든 값이 작아지는 결과를 보여주는데, 탁음일수록 하모닉에너지는 줄어들고 잡음에너지가 증가하여 쉰 목소리라는 것을 알 수 있다. 두 번째 포먼트의 대역폭을 나타내는 BW2는 탁음일수록 값이 커지는 것을 볼 수 있는데, 포먼트 대역폭의 경우 화자의 발음이 명확하지 않을수록 값이 커지는 경향을 가지고 있어, 이러한 경향으로 볼 때 탁음일수록 발화가 불명확하다는 것을 의미한다. 문장에서의 CPP 값도 앞에서 설명한 모음에 대한 CPP 값과 동일하게 탁음일수록 값이 작아지는 것을 보여주고 있다.

청음/탁음 항목에서 나온 상기 파라미터의 결과를 종합해보면, 탁음은 피치주기 변화율과 진폭 변화율이 청음에 비해 크며 대역폭이 넓어지고, 청음은 주기성이 강해지며 하모닉에너지가 증가함을 알 수 있다.

표 2. 청음/탁음에서 청취평가 판정 점수별 유의미한 음성 파라미터의 평균값

Table 2. Mean values of significant speech parameters according to listening evaluation results in clearness of a voice

	1	2	3	평균
F0_mean_v	116	108	108	111
Jitter_v	0.00376	0.00477	0.00553	0.00469
Shimmer_v	3.72	4.27	4.99	4.33
APQ_v	2.91	3.29	3.80	3.33
CPP_v	28.20	27.19	25.39	26.93
HNR05_v	41.39	38.54	36.50	38.81
HNR15_v	43.75	42.05	39.91	41.90
HNR25_v	41.19	39.78	38.12	39.70
HNR35_v	40.69	39.31	37.81	39.27
BW2_s	265	298	333	299
CPP_s	22.46	21.94	21.04	22.81

<표 3>의 가늘다/굵다 항목에서, CPP은 가는 목소리일수록 값이 작아지는데, 기식음이 작은 값을 갖는 특성이 있으므로 [17], 가는 목소리는 기식음이라는 것을 알 수 있다. 반대로 목소리가 굵으면 값이 크므로 목소리의 굵기가 굵을수록 주기성이 강해짐을 알 수 있다. sAPQ는 55개 프레임에 걸친 장 구간 진폭 변화율의 평균을 뜻하는 파라미터인데, 굵은 목소리일수록 값이 커지는 것을 보여주고 있다. SPI는 부드럽게 말하는 정도를 나타내는데, 가는 목소리일수록 값이 크고 이는 가는 목소리는 부드럽게 말하는 특성임을 나타낸다. LTAS는 장 구간 스펙트럼 기울기를 나타내는데, (-) 부호는 음의 기울기를 의미하므로 절댓값이 큰 것이 스펙트럼의 기울기가 큰 것을 의미한다. 기식음은 LTAS 값이 크며, 쥐어짜는 소리(creaky voice)는 작은 값을 갖는 특성이 있다[18], [19]. 이 파라미터는 가는 목소리일수록 절댓값이 커지므로 이 파라미터 역시 가는 목소리는 기식화된 목소리임을 보여준다. H1-A2는 첫 번째 하모닉과 두 번째 포먼트의 진폭 차이이며 LTAS와 유사하게 스펙트럼의 기울기를 보여주는 파라미터로, 가는 목소리일수록 값이 커지는 모습을 보인다.

가늘다/굵다 항목을 종합적으로 살펴보면, 굵은 목소리는 주기성이 강해지며 장 구간 진폭 변화율이 커지고, 가는 목소리는 부드럽게 말하며 기식음이라는 것을 보여준다.

표 3. 가늘다/굵다에서 청취평가 판정 점수별 유의미한 음성 파라미터의 평균값

Table 3. Mean values of significant speech parameters according to listening evaluation results in thickness of a voice

	1	2	3	평균
CPP_v	26.49	27.25	28.50	27.41
sAPQ_s	31.62	32.90	34.92	33.15
SPI_s	17.67	16.53	12.23	15.48
LTAS_s	-17.15	-16.59	-14.74	-16.16
H1-A2_s	29.31	28.40	27.26	28.32

표 4. 높낮이에서 청취평가 판정 점수별 유의미한 음성 파라미터의 평균값

Table 4. Mean values of significant speech parameters according to listening evaluation results in highness of a voice

	1	2	3	평균
F0_mean_v	99	111	126	112
CPP_v	26.65	27.19	28.43	27.42
HNR05_v	37.61	38.84	43.35	39.93
F0_mean_s	102	114	132	116
Jitter_s	0.01752	0.01702	0.01558	0.01671
LTAS_s	-18.20	-16.52	-13.40	-16.04
SPI_s	18.13	16.23	12.60	15.65

<표 4>에 보이는 높낮이 항목에서는, F0_mean에서 목소리가 고음일수록 평균값이 커지는 것을 보여주는데 기본주파수 평균을 보여주는 특성상 당연한 결과이다. CPP는 고음에서 커지며 이는 고음일수록 주기성의 강도가 강해진다는 것을 의미한다. HNR05는 고음에서 커지므로 고음일수록 저주파대역에서 하모닉 에너지가 증가하는 것을 알 수 있다. 문장에서의 F0_mean도 모음과 마찬가지로 고음에서 커지는 결과를 보여주고, 문장에서의 Jitter는 고음에서 작아지므로 고음일수록 피치 주기 변화율이 작다는 것을 알 수 있다. LTAS는 저음에서 값이 커지므로 저음은 기식음에 가까워진다는 것을 말한다. SPI는 저음에서 값이 커지며 이는 저음에서 부드럽게 말하는 정도가 더 크다는 것을 말한다.

높낮이 항목을 종합적으로 살펴보면, 고음일수록 주기성의 강도가 커지며, 저주파대역에서 하모닉에너지가 증가하고 피치주기 변화량은 작아지며, 저음일수록 부드럽게 말하고 기식음화 됨을 알 수 있다.

<표 5>의 세기 항목에서는, Intensity가 세기가 클수록 값이 커지는데 세기를 나타내는 파라미터인 만큼 당연한 결과이다. H1에서는 세기가 클수록 값이 크다. 첫 번째와 두 번째 포먼트 진폭인 A1, A2는 세기가 클수록 값이 커진다. H2K는 2kHz 근처의 스펙트럼의 크기를 나타내는 파라미터이며, 세기가 클수록 값이 크다. HNR15는 세기가 클수록 값이 작아지는데, 이는 세기가 클수록 하모닉 에너지는 줄고 잡음에너지가 증가한다는 것을 나타낸다. 문장에서 Intensity는 모음에서와 동일하게 세기가 클수록 값이 커지며, H1, H2, A1, A2, A3, H2K 등도 세기가 커질수록 값이 커진다. HNR05, HNR15, HNR25, HNR35는 세기가 작을수록 크므로 작은 목소리가 하모닉 에너지가 크다는 것을 보여준다.

종합적으로 세기 항목에서는, 녹음 환경의 영향을 많이 받게 되는 Intensity로 소리의 세기를 알기에는 어려움이 있으므로 소리의 크기를 구분할 수 있는 다른 파라미터가 요구되는데, H1, H2, A1, A2, A3, H2K, HNR05, HNR15, HNR25, HNR35 등과 같은 파라미터들이 그 좋은 예이다.

표 5. 세기에서 청취평가 판정 점수별 유의미한 음성 파라미터의 평균값
Table 5. Mean values of significant speech parameters according to listening evaluation results in loudness of a voice

	1	2	3	평균
Intensity_v	64.63	69.04	74.19	69.29
H1_v	9.49	11.86	15.73	12.36
A1_v	-22.93	-18.48	-14.26	-18.56
A2_v	-24.03	-19.76	-15.14	-19.64
H2K_v	-12.57	-9.47	-4.91	-8.98
HNR15_v	37.98	36.17	32.21	35.45
Intensity_s	60.22	66.82	73.49	66.84
H1_s	2.27	7.66	13.45	7.79
H2_s	1.14	7.18	12.67	7.00
A1_s	-27.86	-22.36	-16.85	-22.36
A2_s	-27.36	-21.06	-15.74	-21.39
A3_s	-19.62	-14.20	-8.96	-14.26
H2K_s	-21.65	-15.52	-10.52	-15.90
HNR05_s	30.65	26.14	20.08	25.62
HNR15_s	36.57	31.50	25.43	31.17
HNR25_s	39.52	34.29	28.24	34.02
HNR35_s	41.35	35.92	29.92	35.73

4. SVM을 통한 목소리 특성 분류 모델링 실험 결과

청취 평가 판정 결과인 1, 2, 3점 중에서 보통인 점수 2점을 부여받은 데이터는 제외하고, 변별력이 큰 1점과 3점 데이터만을 비교 대상으로 SVM을 수행하였다.

SVM을 이용하여 구축한 모델에 대해 10-fold cross validation 실험방법을 사용하여 구한 평균 정확도가 <표 6> ~ <표 9>에 보인다. 청음/탁음 항목에서 83.2%, 가늘다/굵다 항목에서 82.2%, 높낮이 항목에서 94.8%, 세기 항목에서 93.9%의 평균 정확도를 보여주며, 네 가지 항목에 대한 평균 정확도는 88.5%이다.

표 6. 청음/탁음 판별 SVM 모델의 정확도(%)
Table 6. Accuracy of SVM model for clearness of a voice(%)

혼동행렬		판별		평균 정확도
		청음	탁음	
실제	청음	91.1	8.9	83.2
	탁음	24.8	75.2	

표 7. 가늘다/굵다 판별 SVM 모델의 정확도(%)
Table 7. Accuracy of SVM model for thickness of a voice(%)

혼동행렬		판별		평균 정확도
		가늘다	굵다	
실제	가늘다	86.2	13.8	82.2
	굵다	21.9	78.1	

표 8. 높낮이 판별 SVM 모델의 정확도(%)
Table 8. Accuracy of SVM model for highness of a voice(%)

혼동행렬		판별		평균 정확도
		낮다	높다	
실제	낮다	95.8	4.2	94.8
	높다	6.2	93.8	

표 9. 세기 판별 SVM 모델의 정확도(%)
Table 9. Accuracy of SVM model for loudness of a voice(%)

혼동행렬		판별		평균 정확도
		작다	크다	
실제	작다	90.9	9.1	93.9
	크다	3.1	96.9	

이와 같은 실험 결과는 매우 높은 성능을 보여주는 결과로써, 본 연구에서 제안한 방식인 SVM 기반 목소리 특성 분류 모델링 방법이 적절하다는 것을 보여준다.

5. 결론

본 논문에서는 목소리 특성과 상관관계가 높은 음성파라미터를 도출하기 위하여, 20대 남성 222명으로부터 수집한 음성 DB

에 대해 청취 평가를 수행하여, 목소리 특성을 청음/탁음, 가늘다/굵다, 높낮이, 세기, 빠르기 등으로 분류하고, 다양한 음성 분석 툴을 이용하여 음성 파라미터를 추출하고, ANOVA 분석을 수행하여 목소리 특성의 각 항목과 상관관계가 높은 유의미한 음성 파라미터를 도출하였다.

다음에, 기계학습 기법 중 하나인 SVM을 통하여 목소리 특성을 자동으로 판별하는 모델을 구축하기 위해, SVM 툴 중 하나인 LIBSVM을 이용하였으며, C-SVC와 가우시안 RBF 커널함수를 사용하였다. 10-fold cross validation 실험방법을 사용하여 판별 정확도의 평균값을 구하였다.

목소리 특성 판별 모델의 실험 결과, 청음/탁음 항목에서 83.2%, 가늘다/굵다 항목에서 82.2%, 높낮이 항목에서 94.8%, 세기 항목에서 93.9%의 평균 정확도를 얻었으며, 네 가지 항목에 대한 평균 정확도는 88.5%이다. 이는 감정인식 기술에서 감정인식 훈련에 앞서 목소리 특성을 분류하는 모델로써 사용하기에 충분하다고 판단된다.

본 논문은 20대 남성에 국한된 실험으로 여성 음성이나 다른 연령대에 대한 연구로 확장이 필요하다. 또한, SVM 이외 다른 기계학습 알고리즘을 이용하여 목소리 특성 판별 모델을 구현하여 정확도를 개선시키는 연구를 진행할 예정이다.

참고문헌

[1] Bang, J., & Lee, S. (2015). Adaptive speech emotion recognition framework using prompted labeling technique. *KIISE Transaction on Computing Practices*, 21(2), 160-165. (방재훈·이승룡 (2015). 프롬프트 레이블링을 이용한 적응형 음성기반 감정인식 프레임워크. *한국정보과학회 컴퓨팅의 실제 논문집*, 21(2), 160-165.)

[2] Rahman, T., & Busso, C. (2012). A personalized emotion recognition system using an unsupervised feature adaptation scheme. *Proceedings of International Conference on the Acoustics, Speech and Signal Processing* (pp. 5117-5120).

[3] Kwon, C., Song, S., Kim, J., Kim, K., & Jang, J. (2012). Extraction of speech features for emotion recognition. *Phonetics and Speech Sciences*, 4(2), 73-78. (권철홍·송승규·김종열·김근호·장준수 (2012). 감정 인식을 위한 음성 특징 도출. *말소리와 음성과학*, 4(2), 73-78.)

[4] Kim, J., & Kwon, C. (2014). Measuring correlation between mental fatigues and speech features. *Phonetics and Speech Sciences*, 6(2), 3-8. (김정인·권철홍 (2014). 정신피로와 음성특징과의 상관관계 측정. *말소리와 음성과학*, 6(2), 3-8.)

[5] Kim, T., & Kwon, C. (2015). Correlation between physical fatigue and speech signals. *Phonetics and Speech Sciences*, 7(1), 11-17. (김태훈·권철홍 (2015). 육체피로와 음성신호와의 상관관계. *말소리와 음성과학*, 7(1), 11-17.)

[6] Boersma, P., & Weenink, D. (2016). Praat: doing phonetics by computer [computer program]. Retrieved from http://www.praat.org on December, 2016.

[7] MDVP: Multi Dimensional Voice Program, KayPentax. Retrieved from http://www.kayelemetrics.com on January, 2017.

[8] Shue, Y., Keating, P., Vicenik, C., & Yu, K. (2011). VoiceSauce: a program for voice analysis. *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 1846-1849). Retrieved from http://www.seas.ucla.edu/spapl/voicesauce/ on January, 2017.

[9] Han, S., Kim, S., Kim, J., & Kwon, C. (2011). A preliminary study on correlation between voice characteristics and speech features. *Phonetics and Speech Sciences*, 3(4), 85-91. (한성만·김상범·김종열·권철홍 (2011). 목소리 특성의 주관적 평가와 음성 특징과의 상관관계 기초연구. *말소리와 음성과학*, 3(4), 85-91.)

[10] Song, J. (2015). *SPSS/AMOS statistical analysis method required for preparation of thesis*. Seoul: 21segisa. (송지준 (2015). *논문작성에 필요한 SPSS/AMOS 통계분석방법*. 서울: 21 세기사.)

[11] IBM SPSS statistics, IBM Korea. Retrieved from http://www-01.ibm.com/software/kr/analytics/spss/ on January, 2017.

[12] Chang, C., & Lin, C. (2011). LIBSVM: a library for support vector machines. *ACM Transaction on Intelligent Systems and Technology*, 2(3), 1-27. Retrieved from http://www.csie.ntu.edu.tw/~cjlin/libsvm/ on January, 2017.

[13] Kim, T., & Kwon, C. (2016). An SVM-based physical fatigue diagnostic model speech features. *Phonetics and Speech Sciences*, 8(2), 17-22. (김태훈·권철홍 (2016). 음성 특징 파라미터를 이용한 SVM 기반 육체피로도 진단모델. *말소리와 음성과학*, 8(2), 17-22.)

[14] Alippi, C., Roveri, M. (2010). Virtual k-fold cross validation: An effective method for accuracy assessment. *The 2010 International Joint Conference on Neural Networks (IJCNN)*, 18-23.

[15] Ferrand, C. (2002). Harmonics-to-Noise Ratio: an index of vocal aging. *Journal of Voice*, 16(4), 480-487.

[16] Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of Institute of Phonetic Sciences*, 17, 97-110.

[17] Hillenbrand, J., & Houde, R. (1996). Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech. *Journal of Speech and Hearing Research*, 39, 311-321.

[18] Linville, S. (2002). Source characteristics of aged voice assessed from long-term average spectra. *Journal of Voice*, 16(4), 472-479.

[19] Mendoza, E., Valencia, N., Munöz, J., & Trujillo, H. (1996). Differences in voice quality between men and women: use of the long-term average spectrum (LTAS). *Journal of Voice*, 10(1), 59-66.

• **박태성 (Park, Tae Sung)**
(주)케이웍스
대전광역시 유성구 노은로 71
Tel: 042-280-2555
Email: zskyzk@gmail.com
관심분야: 음성기술

• **권철홍 (Kwon, Chu Hhong)** 교신저자
대전대학교 전자·정보통신공학과
대전광역시 동구 대학로 62
Tel: 04-280-2555
Email: chkwon@dju.ac.kr
관심분야: 음성기술, TTS, 음성기술과 의학 분야의 융합연구