음성의 피치 파라메터를 사용한 감정 인식

Emotion Recognition using Pitch Parameters of Speech

저자 (Authors)
이규현, 김원구
Guehyun Lee, Weon-Goo Kim

출처 (Source)
한국지능시스템학회 논문지 25(3), 2015.6, 272-278 (7 pages)
Journal of Korean Institute of Intelligent Systems 25(3), 2015.6, 272-278 (7 pages)

발행처 (Publisher)
한국지능시스템학회
Korean Institute of Intelligent Systems

URL
http://www.dbpia.co.kr/Article/NODE06362938

APA Style
이규현, 김원구 (2015). 음성의 피치 파라메터를 사용한 감정 인식. 한국지능시스템학회 논문지, 25(3), 272-278.

Journal of Korean Institute of Intelligent Systems
Vol. 25, No. 3, June 2015, pp. 272-278
http://dx.doi.org/10.5391/JKIIS.2015.25.3.272

ISSN(Print) 1976-9172
ISSN(Online) 2288-2324

JKIIS

음성의 피치 파라메터를 사용한 감정 인식

Emotion Recognition using Pitch Parameters of Speech

이규현† · 김원구
Guehyun Lee† and Weon-Goo Kim

군산대학교 전기공학과
Department of Electrical Engineering, Kunsan National University

요 약

본 논문에서는 음성신호 피치 정보를 이용한 감정 인식 시스템 개발을 목표로 피치 정보로부터 다양한 파라메터 추출방법을 연구하였다. 이를 위하여 다양한 감정이 포함된 한국어 음성 데이터베이스를 이용하여 피치의 통계적인 정보와 수치해석 기법을 사용한 피치 파라메터를 생성하였다. 이러한 파라메터들은 GMM(Gaussian Mixture Model) 기반의 감정 인식 시스템을 구현하여 각 파라메터의 성능을 비교되었다. 또한 순차특징선택 방법을 사용하여 최고의 감정 인식 성능을 나타내는 피치 파라메터들을 선정하였다. 4개의 감정을 구별하는 실험결과에서 총 56개의 파라메터중에서 15개를 조합하였을 때 63.5%의 인식 성능을 나타내었다. 또한 감정 검출 여부를 나타내는 실험에서는 14개의 파라메터를 조합하였을 때 80.3%의 인식 성능을 나타내었다.

키워드 : 감정 인식, 음성 파라메터, 피치

Abstract

This paper studied various parameter extraction methods using pitch information of speech for the development of the emotion recognition system. For this purpose, pitch parameters were extracted from korean speech database containing various emotions using stochastical information and numerical analysis techniques. GMM based emotion recognition system were used to compare the performance of pitch parameters. Sequential feature selection method were used to select the parameters showing the best emotion recognition performance. Experimental results of recognizing four emotions showed 63.5% recognition rate using the combination of 15 parameters out of 56 pitch parameters. Experimental results of detecting the presence of emotion showed 80.3% recognition rate using the combination of 14 parameters.

Key Words : Emotion Recognition, Speech Parameter, Pitch

Received: Nov. 19, 2013
Revised : Apr. 2, 2015
Accepted: Apr. 3, 2015
† Corresponding author
lovely_guehyun@hanmail.net

1. 서 론

음성인식과 인증기술이 안정화 되어가는 시점에서, 현재 많은 관심과 미래기술로 주목받고 있는 분야는 감정인식 또는 감정이해 분야이다. 이는 IT 연구의 전체적인 방향이 PC 중심에서, 네트워크 중심을 거쳐서 고객중심으로 가고 있는 전체적인 흐름과도 관계가 있다. 고객중심의 서비스를 제공하기 위해서는 고객 행동은 물론 감정, 기호, 습관 등을 종합적으로 파악하여 맞춤형 서비스를 제공하는 것이 중요하기 때문이다. 이에 발맞추어 국외의 경우 미국, 영국, 일본 등에서 감정인식 분야가 중요한 연구 주제로 대두되고 있다.

인간의 감정에 대한 정보는 얼굴표정, 음성, 몸동작, 심장 박동 수, 체온, 혈압 등의 다양한 방법으로 얻을 수 있고, 응용 분야에 따라 감정 정보 취득 방법 또한 달라진다. 특히, 센서가 신체부위에 직접 닫지 않거나, 전화와 관련된 응용 분야의 경우, 음성을 이용한 시스템의 응용은 더 많은 이점을 가지고 있다.

음성에는 화자의 감정뿐만 아니라 전달하고자 하는 내용의 단어나 문법에서의 강세 부분, 지역적인 특성이 가미된 억양 등 정서이외의 것들이 많이 담겨져 있기 때문에, 음성에서 감정만을 따로 떼어서 분석하는데 어려움이 있다. 음성을 통한 감정 인식을 위해서는 각각의 감정이 음성에 어떠한 변화를 만들어내는가를 정확히 규명하여야 하는데, 이러한 음성과 감정과의 상관관계에 대한 연구는 서구의 음향학자들과 심리학자들에 의해 먼저 이루어졌다[1-3]. 이런 심리학자들의 연구결과를 바탕으로 공학자들이 다양한 응용분야를

이 논문은 2010학년도 군산대학교 대학자체 학술공모과제 연구비 지원에 의하여 연구되었음.

개발하려는 시도를 하고 있으며 이와 관련하여 미국 MIT대 Media Lab의 Affective Computing Group을 중심으로 인간으로부터 얻을 수 있다는 다양한 정보를 이용한 감정적 착용 컴퓨터의 개발은 감정과 관련된 연구의 실용화 가능성을 밝게 하고 있다. 한편, 감정 인식 기술은 이미 일부 분야에서 상용화가 이루어지고 있다. 이러한 것으로 일본 소니사가 개발하여 시판한 애완로봇 AIBO는 6가지 감정을 포함하는 감정 모델을 적용하여 주인과의 관계에 의해서 감정상태가 변화하고 반응하도록 만들어졌다. IBM에서는 'Blue Eyes 프로젝트'를 통하여 차세대 감정 인식 제품을 개발하고 있으며, 표정을 이용한 감정 인식 시스템, 생체 신호를 이용한 감정 인식 마우스 등을 상용화하는 단계에 이르렀다[4].

음성은 청각에 기반을 둔 방법으로 여기에 내포된 감정을 추출하려는 연구가 활발히 행해지고 있다. Fukuda는 음성신호의 템포와 에너지를 가지고 여섯 개의 기본 감정에 대한 분류를 시도 했으며[5], Moriyama는 음성신호의 피치와 전력의 포락선 검출을 통하여 20개의 일본어 샘플에 대한 실험을 했다[6]. 또한 Silva는 음성 신호의 피치와 HMM(Hidden Markov Model)을 이용하여 영어와 스페인어에 대한 감정인식을 실험하였다[7].

한편 국내에서도 음성 및 얼굴표정을 이용한 감정인식 연구가 활발하게 진행되고 있다. 우리나라 국악의 창에서 인간의 희로애락을 표현하는 음의 고저와 장단을 기본으로 하여 분석하는 연구가 행해졌으며[8], 대화의 내용에 사용된 단어나, 톤, 말의 빠르기, 음질 등을 이용하여 화난 감정의 특성을 찾아내는 연구도 실행되었다[9].

본 논문에서는 음성신호의 특징 중 하나인 피치 정보를 이용한 감정 인식 시스템 개발을 목표로 피치 정보로부터 다양한 파라메터 추출방법을 연구하였다. 이를 위하여 다양한 감정이 포함된 한국어 음성 데이터베이스를 이용하여 피치의 통계적인 정보와 수치해석 기법을 사용한 피치 파라메터를 생성하였고 GMM(Gaussian Mixture Model) 기반의 감정 인식 시스템을 구현하여 각 파라메터의 성능을 비교하였다. 또한 SFS(Sequential Feature Selection)[10] 방법을 사용하여 최고의 감정 인식 성능을 나타내는 피치 파라메터를 선정하였다.

본 논문의 구성은 다음과 같다. 2장에서는 피치에 관한 정보가 사용된 파라메터에 대해 설명하였고 3장에서는 전체적인 시스템 구성 및 실험 과정과 결과를 다루었고 4장에서는 본 논문의 결론을 맺었다.

2. 피치 파라메터

유성음은 성대가 진동하여 발생하며 그 진동의 주기를 피치 혹은 기본 주파수(F0)라고 한다. 피치는 여러 음성 신호 분야에서 이용되고 있는 중요한 음성 변수로 자기상관함수나 AMDF(Average Magnitude Difference Function) 방법을 사용하여 구할 수 있지만 음성 신호의 비 정상성, 성대 진동의 불규칙성, 잡음환경에서의 신호 왜곡 등 때문에 정확하게 구하기는 어렵다.

본 연구에서는 피치를 구하기 위하여 Praat 알고리즘[11]을 이용하였다. 이렇게 구한 피치신호를 이용하여 여러가지 파라메터를 생성하였다. 기본적으로 최대값, 최소값, 평균값 등을 생성하였고 수치해석 기법을 이용하여 표 1과 같은 다양한 파라메터들을 생성하였다. 각 파라메터는 문장 또는 유성음 단위로 생성되었다,

표 1에서 rising과 falling은 각각 피치신호가 시간에 따라 상승 또는 하강 하강하는 특성을 나타내는 파라메터이고 문장단위 또는 유성음 단위로 구하여진다. interquartile은 사분위수라고도 불리며, 자료를 오름차순으로 정렬해서 네 개의 구간으로 나누었을 때, 그 중 몇 번째에 위치한 데이터인지를 알아보는 기법이다. 일반적으로 사분위수는 3/4지점에서 1/4지점의 값을 빼서 interquartile range로 사용한다. Plateaux는 피치신호를 1차미분과 2차미분을 한 후 두 미분값의 상관관계에 따라 정해진다. Plateaux는 크게 maxima와 minima 값으로 나누어진다. 즉, 1차 미분한 값에서 0에 가까운 값을 가지고 2차 미분한 값에서 양수이면 그 위치의 피치값을 minima로 간주하였다. 반대로 1차 미분한 값에서 0에 가까운 값을 가지고 2차 미분한 값에서 음수이면 그 위치의 피치값을 maxma로 간주하였다. 피치의 연속적인 점들을 1차 또는 2차, 3차 함수로 나타내는 것은 점들의 기울기와 곡률, 선형·비선형 정도 등을 알 수 있는 유용한 요소이다. 본 논문에서는 slope, curvature, inflexion로 정의하여 1차, 2차, 3차 함수로 점들을 유사화 하여 계수를 이용해 파라메터를 작성하였다. 연속된 피치들은 식(1)과 같이 함수로 나타내어질 수 있다.

$$\begin{aligned} y &= a_1 x + a_0 \\ y &= b_2 x^2 + b_1 x + b_0 \\ y &= c_3 x^3 + c_2 x^2 + c_1 x + c_0 \end{aligned} \tag{1}$$

n차식으로 수치화하여 가장 높은 차수의 계수를 이용하여 파라메터를 작성하였다.

$$\begin{aligned} a_1 &= slope \\ b_2 &= curvature \\ c_3 &= infexion \end{aligned}$$

표 1. 피치 파라메터와 단위
Table 1. Pitch parameters and its unit

parameter	unit	
	sentence	voiced
rising	O	O
falling	O	O
interquartile	O	O
Plateaux maxima	O	O
Plateaux minima	O	O
slope	X	O
curvature	X	O
inflexion	X	O
skewness	X	O
kurtosis	X	O

skewness는 비대칭도 또는 왜도라고도 불리며 어떠한 집단의 도수분포에서 평균값에 관한 비대칭의 방향과 그 정도를 나타내는 특성 값이다. 도수분포가 대칭일 때는 산술평균, 중앙값, 최빈값은 모두 일치하지만, 비대칭일 때의 도수는 최빈값의 좌우에 균등히 분포하지 않는다. skewness가 음수일 경우에는 분포는 왼쪽으로 치우치고 양수일 경우에는 오른쪽으로 치우친다. kurtosis는 첨도라고도 불리며 유성음 구간의 피치가 얼마나 뾰족하고, 또한 얼마나 평평한지를 나타내는 척도로 사용된다. 얻어진 값이 0보다 크지만 3보다 작으면 분포는 완첨(platykurtic, 평평함)이라고 하고, 3과 같으면 중첨(mesokurtic, 높이가 정상적임)이라고 한다. 또한 3보다 크면 급첨(leptokurtic, 뾰족함)이라고 한다.

3. 실험 및 결과

3.1 데이터베이스

감정인식 시스템의 성능을 평가하기 위해서 데이터 베이스는 다음과 같은 과정으로 구성되었다[12]. 본 연구에서는 인간의 주요 감정인 기쁨, 슬픔, 화남의 3가지 감정과 이들의 기준이 되는 평상 감정을 포함한 4가지 감정을 인식 대상 감정으로 결정하였다. 음성의 녹음은 평소 감정 표현을 훈련하는 아마추어 연극단원 남/녀 각 15명을 대상으로 하였다. 녹음작업은 조용한 사무실 환경에서 이루어졌고 각 화자는 45개의 문장을 4가지 감정으로 녹음하였다. 녹음 동안에 감정 표현이 미흡하다고 판단된 경우에는 다시 녹음을 하였다. 구축된 데이터베이스에 대하여 주관적 평가를 수행하여 구축된 데이터 베이스 중에서 감정이 적절히 반영되었다고 판단되는 문장을 선택하여 감정 인식용 데이터베이스로 구축하였다. 이는 화자의 감정이 어느 정도 정확히 반영된 음성 데이터만을 선별하는 과정으로 평소 음성 신호 처리 실험에 숙련된 연구원들을 대상으로 주관적 평가를 실시하였다. 주관적 평가는 전체 데이터베이스 중에서 5400문장을 문장 당 10명이 청취한 후 감정이 적절히 반영되었다고 판단되는 문장을 선택하였다. 본 논문에서는 총 5400개의 음성 데이터 중에서 2237개의 음성 데이터를 선별하여 데이터베이스로 구성하였다.

3.2 감정 인식 시스템의 구성

본 논문에서는 감정 인식을 위한 피치 파라메터의 성능을 평가하기 위하여 GMM(Gaussian Mixture Model) 기반의 화자 및 문장 독립 감정 인식 시스템을 구현하였다. 감정 인식 시스템은 크게 학습과정과 인식과정으로 나눌 수 있다.

학습 과정에서는 학습데이터를 기반으로 특징을 추출하여 기준 모델을 생성하고 인식 과정에서는 인식 데이터를 사용하여 최대 확률을 갖는 감정 모델을 입력 음성의 감정으로 결정한다. 모델의 학습에는 총 45개의 문장 중에서 35개의 문장을 20명(남성 10명과 여성 10명)이 녹음한 음성이 사용되었고 인식에는 학습에 참여하지 않은 10명(남성 5명과 여성 5명)을 학습에 사용되지 않은 나머지 10개의 문장을 녹음한 음성이 사용되었다.

3.3 피치 파라메터 추출

감정이 포함된 음성 신호의 피치를 추정할 때 Praat 알고리즘[11]을 이용하였다. Praat 프로그램에서 분석창은 40ms 크기로 30ms씩 중첩되어 이동하였고 초당 100개의 분석 구간에서 피치를 계산하였다. 구하여진 피치 결과에 포함될 수 있는 갑작스런 변화를 제거하기 위하여 Praat 알고리즘에 포함된 스무딩(smoothing)이 사용되었다.

표 2에 본 실험에서 사용되는 파라메터를 정의한 하였다[13][14]. 문장단위 파라메터는 따로 표기하지 않았고, 유성음 구간 단위의 파라메터는 파라메터명에 Voiced 기호를 포함하였다.

표 2에서 13번 파라메터인 'infcountnum'은 본 연구에서 제안한 피치가 급격히 변화한 개수를 나타내는 파라메터로 이전의 피치보다 7 이상의 큰 차이를 보이면 급격히 변화하였다고 간주하였을 때 가장 우수한 성능을 나타내었다.

표 2. 실험에 사용된 파라메터
Table 2. Parameter used in the experiment

id	parameter definition(parameter name)
1	maximum of pitch (max)
2	minimum of pitch (min)
3	mean of pitch (mean)
4	standard deviation of pitch (std)
5	maximum minus mean of pitch (mean_max)
6	mean minus minimum of pitch (mean_min)
7	maximum - minimum (max_min)
8	mean of rising slope (risingmean)
9	maximum of rising slope (risingmax)
10	mean of falling slope (fallingmean)
11	maximum of falling slope (fallingmax)
12	number of inflection (infcount)
13	number of rapid inflection (infcountnum)
14	mean of above mean (summaxima)
15	mean of below mean (summinima)
16	mean of above mean of max and mean (maxmean)
17	mean of below mean of max and mean (minmean)
18	mean of higher rank 10% (meanupper)
19	interquartile range (interrange)
20	interquartile range of rising (interrising)
21	interquartile range of falling (interfalling)
22	mean value of plateaux maxima (meanmaxima)
23	mean value of plateaux minima (meanminima)
24	median value of plateaux maxima (medianplamaxima)
25	median value of plateaux minima (medianplaminima)
26	interquartile range of plateaux maxima (interplamaxima)
27	median of rising slope (risingmedian)
28	median of falling slope (fallingmedian)
29	median duration of rising slope (mediandrising)

30	median duration of falling slope (mediandfalling)
31	maximum duration of rising slope (maxdrising)
32	maximum duration of falling slope (maxdfalling)
33	interquartile duration range of rising slope (interdrising)
34	interquartile duration range of falling slope (interdfalling)
35	maximum of plateaux at minima (maxiplaminima)
36	minimum of plateaux at maxima (maxiplamaxima)
37	mean of the voiced segment maximums (meanvoicedmax)
38	mean of the voiced segment minimums (meanvoicedmin)
39	mean of the voiced segment inflexion (meanvoicedinf)
40	mean of the voiced segment interquartile ranges (meanvoicedinter)
41	max of the voiced segment inflexion (maxvoicedinf)
42	max of the voiced segment mean (maxvoicedmean)
43	max of the voiced segment mean (meanvoicedlowquartile)
44	mean of the voiced segment upper quartile (meanvoicedupperquartile)
45	skewness (skewness)
46	kurtosis (kurtosis)
47	least square (leastsquare)
48	mean of the voiced segment slopes (meanslope)
49	mean of the voiced segment curvatures (meancurvatures)
50	mean of the voiced segment inflections (meaninflexions)
51	maximum of the voiced segment slopes (maxslope)
52	maximum of the voiced segment curvatures (maxcurvatures)
53	maximum of the voiced segment inflections (maxinflexions)
54	standard deviation of the voiced segment slopes (stdslope)
55	standard deviation of the voiced segment curvature (stdcurvatures)
56	standard deviation of the voiced segment inflections (stdinflexions)

3.4 실험 결과

각 파라메터의 감정 분류 능력은 클래스간 분산(between class variance)와 클래스내 분산(within class variance)를 이용하여 수치화 할 수 있다. 클래스간 분산은 감정간의 차이를 수치화한 것이고, 클래스내 분산은 감정의 분산정도를 수치화한 것이다. 전체 클래스내 분산을 σ_w^2, 클래스간 분산을 σ_b^2라 하면 σ_b^2는 클수록, σ_w^2는 작을수록 감정 간의 구분 성능이 높은 파라메터이다.

그림 1은 56개의 피치 파라메터별로 σ_b^2와 σ_w^2를 구하여 σ_b^2/σ_w^2로 수치화 하여 상위 15개 파라메터를 나타낸 그래프이다. σ_b^2는 감정간의 거리를 나타내며, σ_w^2는 각 감정의 분산정도를 나타내므로 가장 우수한 파라메터는 σ_b^2는 큰 값, σ_w^2는 작은 값이어야 한다. 이를 쉽게 비교하기 위해 σ_b^2/σ_w^2로 나타내어 그래프에 표현하였다. 그림에서 급격한 피치 변화(infcountnum), Plateaux maxima의 평균(meanmaxima), 피치 평균(mean), 피치 하강 구간의 중간값(fallingmedian), Plateaux maxima의 중간값(medianplamaxima)등의 파라메터가 높은 σ_b^2/σ_w^2값을 나타냈다.

다음 실험에서는 GMM 기반의 감정 인식 시스템을 사용하여 각각의 피치 파라메터에 대한 감정 인식 성능을 비교하였다. 그림 2는 56개의 피치 파라메터중에서 상위 15개 파라메터들의 인식 성능을 나타낸 그래프이다. 그림에서 알 수 있듯이 파라메터를 한 개씩만 사용했을 경우 급격한 피치 변화(infcountnum)가 43.7%로 가장 우수한 성능을 나타내었고 그 다음으로 최소자승(leastsquare), 최대값과 최소값 차이(max_min), 피치 상승구간의 중간 값(risingmedian) 등의 파라메터가 높은 인식률을 나타냈다.

그림 2와 그림 3을 비교하여보면 상위 15위에 공동으로 포함되는 파라메터로는 급격한 피치 변화(infcountnum), 평균 이상값의 평균(summaxima), Plateaux maxima의 최대값(maxiplaminima), 평균과 최소값의 차이(mean_min), 피치 상승구간의 중간 값((risingmedian) 등이다.

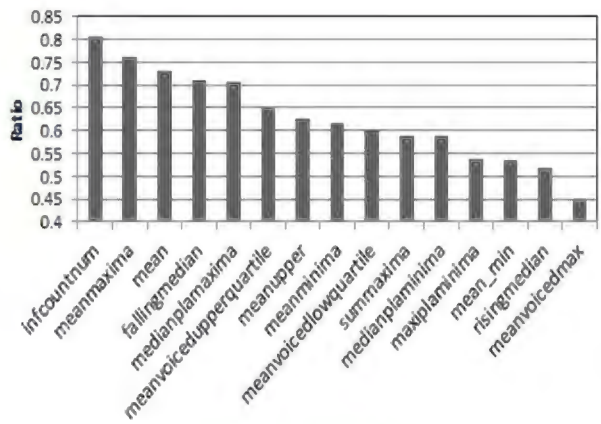

그림 1. 파라메터별 σ_b^2/σ_w^2

Fig. 1. Variance of each parameter

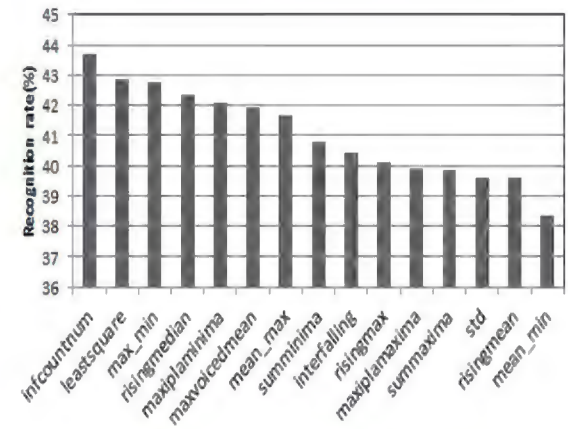

그림 2. 파라메터별 단독 인식률

Fig. 2. Recognition results of single parameter

본 실험에서는 SFS(sequential Feature Selection)[10] 방법을 사용하여 우수한 성능을 나타내는 피치 파라메터의 조합을 선정하였다. SFS 알고리즘은 가장 우수한 성능을 나타내는 한 개의 파라메터를 선정한 후 이와 결합하였을 때 가장 우수한 성능을 나타내는 파라메터의 조합을 찾아 다음 파라메터를 선정하는 방법이다. 따라서 파라메터를 한 개씩 늘려가며 원하는 개수의 파라메터를 선정한다. 그림 3은 최종 인식률이 나온 감정 인식 시스템에 사용된 파라메터의 개수별 인식률을 나타낸 그래프이다. 56개의 파라메터중 15개를 조합했을 때 가장 높은 인식률로 63.5%를 나타냈다. 조합된 15개의 파라메터로를 순서대로 나타내면 급격한 피치 변화율(infcountnum), 최대값과 평균의 차이(mean_max), Plateaux minima의 중간값(medianplaminima), 피치 상승 구간의 최대값(risingmax), Plateaux maxima의 interquartile range (interplamaxima), 평균과 최소값의 차이(mean_min), 피치 상승 구간의 중간값(risingmedian), 최소자승(leastsquare), 피치 최소값(min), 유성음 구간 inflection의 표준편차(stdinflexions), interquartile range(interrange), 유성음 구간 curvature의 최대값(maxcurvatures), 유성음 구간 inflection의 평균(meaninflexions), 유성음구간 기울기의 표준편차(stdslope), Plateaux minima의 평균(meanminima)이다.

그림 1, 그림 2와 그림 3을 비교하여 15개의 파라메터중 공통으로 포함되는 것은 급격한 피치 변화율(infcountnum), 평균과 최소값의 차이(mean_min)와 피치 상승 구간의 중간값(risingmedian)이다.

표 3은 위에서 선별한 실험 요소를 토대로 감정 인식 시스템의 최종 인식률을 나타내는 표이다. 표에서 '평상' 감정의 인식은 65.8%, '기쁨' 감정의 인식은 52%, '슬픔' 감정의 인식은 71.4%, '화남' 감정의 인식은 64.9%로서 최종 인식률은 63.5%이다. '평상' , '슬픔' , '화남' 감정의 인식은 우수한 편이나, '슬픔' 감정의 인식에서는 '화남' 감정과의 구분이 명확하지 않았다.

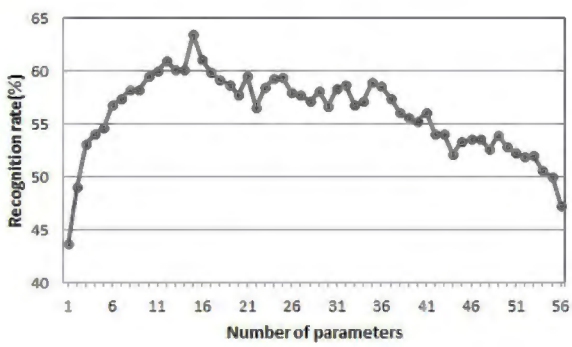

그림 3. 피치 파라메터 개수에 따른 감정 인식 성능
Fig. 3. Emotion recognition performance according to the number of pitch parameters

표 3. 감정 인식 시스템의 인식 성능
Table 3. Recognition performance of emotion recognition system

emotion	recognition rate(%)			
	neutral	happy	sad	angry
neutral	65.8	6.6	22.4	5.3
happy	8.0	52.0	8.0	32.0
sad	16.5	6.6	71.4	5.5
angry	12.3	10.5	12.3	64.9
average	63.5			

본 논문에서는 4가지 감정을 구별하는 감정 인식 외에도 감정이 포함되지 않은 음성(neutral)과 감정이 포함된 음성(emotion)의 인식 실험도 수행하였다. 이를 위하여 감정이 포함된 음성(기쁨, 슬픔, 화남)을 합하여 표 4와 같이 데이터베이스를 재구성하였다.

그림 4는 SFS 알고리즘을 사용하여 감정이 포함되지 않은 음성(neutral)과 감정이 포함된 음성(emotion)에 대한 파라메터 조합을 선정한 결과를 나타낸다. 그림에서 한개의 피치 파라메터를 사용하였을 때 최대 68.5%를 나타내었고 파라메터를 14개 조합하여 사용하였을 때 최대 인식률 80.3%를 나타냈다.

표 4. 감정 검출을 위한 데이터베이스
Table 4. Database for emotion detection

database	number of token
neutral	630
emotion	1607
total	2237

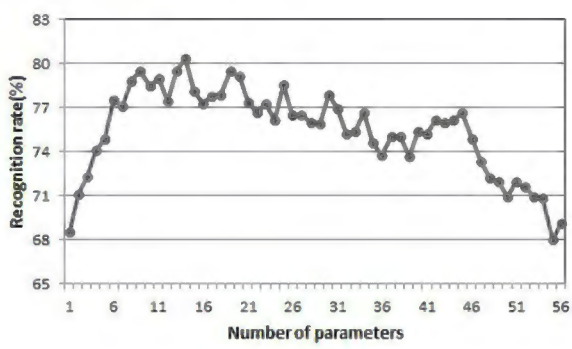

그림 4. 파라메터 개수에 따른 감정 검출 성능
Fig. 4. Emotion detection performance according to the number of pitch parameters

표 5는 감정 검출을 위한 감정 인식 시스템의 최종 인식률을 나타내는 표이다. 표에서 감정이 포함되지 않은 음성(감정

없음)의 인식은 76.3%, 감정이 포함된 음성(감정있음)의 인식은 84.3%로 최종 인식률은 80.3%이다.

표 5. 감정 검출 시스템의 인식 성능
Table 5. Recognition performance of emotion detection system

emotion	recognition rate (%)	
	neutral	emotion
neutral	76.3	23.7
emotion	15.7	84.3
average	80.3	

4. 결 론

본 논문에서는 음성신호의 특징 중 하나인 피치 정보를 이용한 감정 인식 시스템 개발을 목표로 피치 정보로부터 다양한 파라메터 추출방법을 연구하였다. 이를 위하여 다양한 감정이 포함된 한국어 음성 데이터베이스를 이용하여 피치의 통계적인 정보와 수치해석 기법을 사용한 피치 파라메터를 생성하였고 GMM 기반의 감정 인식 시스템을 구현하여 각 파라메터의 성능을 비교하였다. 또한 SFS 방법을 사용하여 최고의 감정 인식 성능을 나타내는 피치 파라메터를 선정하였다.

본 연구에는 다양한 피치 파라메터를 생성하기 위하여 피치의 최대값, 최소값, 평균값, 표준편차 등을 Inflexion, Interquartile, Platueax, Slope, Curvature, Inflexion, Skewness, Kurtosis와 같은 수치해석 기법과 결합하여 파라메터를 작성하였다.

파라메터의 분산을 이용한 방법과 GMM을 사용한 방법으로 감정 인식 성능을 평가하였을 때 급격한 피치 변화를 나타내는 Infcountnum 파라메터가 가장 우수한 성능을 나타내었다. 또한 GMM을 사용한 감정 인식 시스템을 대상으로 SFS 알고리즘을 통하여 56개중 15개의 파라메터 조합을 선택하였을 때 최대 63.5%의 인식률을 얻었다.

감정이 없는 음성과 감정이 포함된 음성과의 1:1 인식에서는 피치 파라메터로 80.3%의 인식률을 얻었으며 문장 독립적 화자 인식 시스템에서뿐만 아니라 화자 및 문장 독립적 감정 인식 시스템에서도 보다 적합한 인식 시스템으로 적용할 수 있다고 판단된다.

References

[1] Janet E. Cahn, "The Generation of Affect in Synthesized Speech", *Journal of the American Voice I/O Society*, Vol. 8, pp.1-19 July 1990.

[2] K. R. Scherer, D. R. Ladd, and K. E. A. Silverman, "Vocal Cues to Speaker Affect: Testing Two Models", *Journal Acoustical Society of America*, Vol. 76, No. 5, pp. 1346-1355, Nov 1984.

[3] Iain R. Murray and John L. Arnott, "Toward the Simulation of Emotion in Synthetic Speech: A Review of the Literature on Human Vocal Emotion", *Journal Acoustical Society of America*, pp.1097-1108, Feb. 1993.

[4] Rosalind W. Picard, "Affective Computing", The MIT Press, 1997.

[5] V. Kostv and S. Fukuda, "Emotion in User Interface, Voice Interaction System," *IEEE International Conference on Systems, Cybernetics Representation*, No.2, pp,798-803, 2000

[6] T. Moriyama and S. Oazwa, "Emotion Recognition and Synthesis System on Speech," *IEEE Intl. Conference on Multimedia Computing and System*, , pp.840-844. 1999

[7] L. C. Siva and P. C. Ng, "Bimodal Emotion Recognition," *in Proceeding of the 4th Intl. Conference on Automatic Face and Gesture Recognition*, pp.332-335. 2000

[8] Y. G. Kim, Y. C. Bae, "Design of Emotion Recognition Model Using fuzzy Logic" *Proceedings of KFIS Spring Conference*, 2000.

[9] K. B. Sim, C. H. Park, "Analyzing the element of emotion recognition from speech", *Journal of Korean Institute of Intelligent Systems*, Vol. 11, no. 6, pp.510-515, 2001.

[10] P. A. Devijver and J. Kitteler, "Pattern Recognition : A Statistical Approach", London: Prentice-Hall International, 1982

[11] P. Boersma and D. Weeninck, "PRAAT, a system for doing phonetics by computer," Inst. Phon. Sci. Univ. of Amsterdam, Amsterdam, Negherlands, Tech. Rep. 132, 1996 [Online]. Available: http://www.praat.org.

[12] B. S. Kang, "text-independent emotion recognition algorithm using speech signal," Master thesis, Yonsei University, 2000

[13] Dimitrios Ververidis, Constantine Kotropoulos, Loannis Pitas, "Automatic Emotional Speech Classification", *in Proceedings of ICASSP '04*, 2004.

[14] Carlos Busso, Sungbok Lee, Shrikanth Narayanan, "Analysis of Emotionally Salient Aspects of Fundamental Frequency for Emotion Detection,", *IEEE Trans. Speech and Audio Processing*, Vol. 17, No 4, pp. 582-596, May 2009

저 자 소 개

이규현(Guehyun Lee)

2011년 : 군산대 전기공학과 공학사
2011년~현재 : 군산대 전자정보공학부 공학석사
2013년~현재 : 에이앤아이(주) 검사장비팀 대리

관심분야 : 음성신호처리, 음성인식, 감정인식
Phone : +82-63-469-4741
E-mail : lovely_guehyun@hanmail.net

김원구(Weon-Goo Kim)

1987년 : 연세대 전자공학과 공학사
1989년 : 연세대 전자공학과 공학석사
1994년 : 연세대 전자공학과 공학박사
1994년~현재 : 군산대 전기공학과 교수
1998년~1999년 : Bell Lab, Lucent Technologies(USA) 객원연구원
2008년~2009년 : Griffith University(Australia) 방문교수

관심분야 : 음성신호처리, 음성인식, 감성인식, 음성변환, 화자 인식
Phone : +82-63-469-4745
E-mail : wgkim@kunsan.ac.kr