

음성신호를 이용한 감정인식

An Emotion Recognition Technique using Speech Signals

정병욱 · 천성표 · 김연태 · 김성신[†]

Byungwook Jung, Seungpyo Cheun, Yountae Kim, and Sungshin Kim

부산대학교 전기공학과

요 약

휴먼인터페이스 기술의 발달에서 인간과 기계의 상호작용은 중요한 부분이다. 감정인식에 대한 연구는 이러한 상호작용에 도움을 준다. 본 연구는 개인화된 음성신호에 대하여 감정인식 알고리즘을 제안하였다. 감정인식을 위하여 PLP 분석을 이용하여 음성신호의 특징으로 사용하였다. 처음에 PLP 분석은 음성인식에서 음성신호의 화자 종속적인 성분을 제거하기 위하여 사용되었으나 이후 화자인식을 위한 연구에서 PLP 분석이 화자의 특징 추출을 위해 효과적임을 설명하고 있다. 그래서 본 논문은 PLP 분석으로 만들어진 개인화된 감정 패턴을 이용하여 쉽게 실시간으로 음성신호로부터 감정을 평가하는 알고리즘을 제안하였다. 그 결과 최대 90%이상의 인식률과 평균 75%의 인식률을 보였다. 이 시스템은 간단하지만 효율적이다.

키워드 : 감정인식, 음성신호, 지각선형예측분석

Abstract

In the field of development of human interface technology, the interactions between human and machine are important. The research on emotion recognition helps these interactions. This paper presents an algorithm for emotion recognition based on personalized speech signals. The proposed approach is trying to extract the characteristic of speech signal for emotion recognition using PLP (perceptual linear prediction) analysis. The PLP analysis technique was originally designed to suppress speaker dependent components in features used for automatic speech recognition, but later experiments demonstrated the efficiency of their use for speaker recognition tasks. So this paper proposed an algorithm that can easily evaluate the personal emotion from speech signals in real time using personalized emotion patterns that are made by PLP analysis. The experimental results show that the maximum recognition rate for the speaker dependant system is above 90%, whereas the average recognition rate is 75%. The proposed system has a simple structure and but efficient to be used in real time.

Key Words : emotion recognition, speech signals, perceptual linear prediction

1. 서 론

인간의 소리는 의사소통을 하기위한 수단뿐만 아니라, 감정을 전달하는 수단으로 사용된다. 소리에 포함된 감정은 화자의 심리상태를 나타내어 상대방과의 의사소통을 더욱 자연스럽게 한다. 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[1].

앞으로 휴먼 인터페이스(human interface) 기술의 발전에서 인간과의 상호작용은 필수적인 요소이다. 인간과 기계의 상호작용에서 중요한 요소 중의 하나가 감정을 인식하는 기술이다. 현재까지 인간의 감정을 인식하는 방법으로 영상을 이용하는 방법과 음성을 이용하는 감정인식 연구 그리고

인간의 생체신호를 이용하는 연구 방법들이 시도되었다[2]. J. Nicholson은 음성의 파워(power), 피치(pitch), Linear Predictive Coding 분석을 이용하여 특징 벡터를 생성하고 Neural Network을 이용하여 감정을 분류하였다. 감정인식 분류 결과는 8개의 감정에 대하여 약 50%정도로 나타났다 [3]. Aishah Abdul Razak는 감정인식에서 퍼지 모델(fuzzy model)과 NN을 비교 분석하였다. 사용한 음성신호의 특징은 에너지(energy), LPC Coefficients, 주기(duration), 피치, Jitter를 사용하였다. 6가지 감정에 대하여 약 60%의 인식률의 결과를 보였다. 퍼지 모델의 경우 작은 데이터 수에 대하여 이점을 가지는 것으로 보고하고 있다[4]. 강면구는 화자 및 문장 독립적 감정인식을 위하여 Vector Quantization과 Gaussian Mixture Model을 이용한 알고리즘을 사용하였다[5]. 조운호는 이동전화(cellular phone)를 통해 실시간으로 습득된 음성으로부터 사람의 감정 상태를 인식할 수 있는 음성 감정인식 시스템을 제안하였다. 두 가지 감정에 대하여 감정 인식률은 86.5%의 성능을 보였다 [6]. Vladimir Hozjan 과 Zdravko Kacic은 음성신호를 이용한 감정인식 연구에서 문장에 독립적인 다양한 언어를 이용하였다. 이 연구에서는 화자에 종속적인 감정인식에 대하여 높은 인식률을 보였고 언어별 감정인식은 비슷한 인식률

접수일자 : 2007년 10월 30일

완료일자 : 2008년 1월 28일

[†]Corresponding Author : sskim@pusan.ac.kr

이 논문은 한국전자통신연구원(ETRI)의 지원에 의하여 연구되었음.

을 보였음을 결과로 말하고 있다[7].

본 연구에서는 음성신호의 분석 기법 중 하나인 PLP(perceptual linear prediction) 분석을 이용하여 감정인식을 하였다. PLP는 음성신호를 이용한 언어인식이나 화자인식 등에 많이 사용되는 분석기법으로서 본 논문에서는 개인화된 음성신호의 감정인식에 적용하였다.

2. 음성신호의 개요 및 특징 분석

인간의 소리에에는 많은 정보를 가지고 있다. 소리의 주된 목적은 상대방과의 의사소통을 위한 것이다. 소리는 마이크로폰을 통하여 정량적인 값으로 저장할 수 있다. 이러한 정량적인 소리를 음성신호라고 한다. 그림 1은 음성신호를 나타내는 그래프이다. 이러한 음성신호를 분석하면 사람의 목소리가 시간에 따라 크기가 어떻게 변하는지 확인할 수 있다.

2.1 음성신호의 특징 분석

음성신호의 특징을 분석하는 방법에는 여러 가지가 있다. 기본적으로 많이 사용하는 음성신호의 특징 분석은 크게 피치 특징과 에너지 특징으로 나눌 수 있다. 피치 특징 분석은 음성신호의 최대나 최소가 나타나는 부분의 값을 분석하는 방법이다. 그리고 에너지의 특징 분석은 음성신호의 시간 영역이나 주파수 영역에서의 구간 세기를 나타내는 값을 분석하는 방법이다. 두 가지 분석방법이 음성신호에서 감정을 분석하는데 많이 사용되는 분석 방법이다.

그리고 음성신호는 특정 주파수 영역에서 특징을 가지고 있다. 표 1은 주파수 영역별 특징을 나타내고 있다[8]. 주파수 영역에서 이 같은 특징은 화자가 자신의 감정을 음성신호로 표현할 때 사용된다. 이러한 주파수 영역에서의 특징을 바탕으로 본 논문에서는 PLP 분석을 이용하여 감정인식을 하고자 한다.

표 1. 주파수영역에서 음성신호의 특징

Table 1. Features of speech signals in the freq. domain

주파수	청각적 느낌
125Hz~250Hz	증폭하면 목직함이 생기고, 줄이면 평장히 약한 목소리가 된다.
2000Hz~5000Hz	증폭하면 목소리의 명확도가 높아진다.
4000Hz~8000Hz	치찰음, S, SH, CH, C를 말할 때 듣기 싫은 소리가 나옴, 줄이게 되면 목소리의 깨끗함이 감소 됨.
8000Hz 이상	입에서 나오는 공기소리

2.2 음성신호의 PLP 분석

PLP 분석은 인간이 소리에 대한 청감도를 고려하여 음성신호를 분석하는 기법이다[9]. 인간의 청감도는 그림 2와 같이 나타난다. 인간의 청감도는 인간의 귀로 소리를 들을 때 주파수에 따라 다른 감도를 가지고 있음을 나타낸다. 이러한 인간의 청감도를 이용하여 음성신호에 대한 주파수에 따른 강도를 조절하고 다시 역 변환을 이용하여 시간에 따른 음성신호의 스펙트럼을 분석한 것이 PLP 분석이다. PLP 분석 순서도는 그림 3과 같다. 그림 3에서 critical band analysis는 주파수 영역으로 변환된 입력신호의 전력을 구

하는 것이다. 이때 구성되는 필터의 중심 주파수는 음성의 특징과 청각 모델을 기준으로 한다. 사람의 음성은 고주파보다 저주파에 더 많은 정보를 담고 있다. 청각적 특징이 대수화하여 이루어지기 때문에 이 역시도 필터 구성에서 필터의 중심주파수를 결정하는데 사용된다. 사용하는 단위는 bark단위 이다. 식 1과 같이 주파수를 변환하게 되면 필터의 중심 주파수의 범위를 구할 수 있다. 여기서 f_s 는 샘플링 주파수이다.

$$z = 6 \log \left[\frac{f_s}{600} + \sqrt{\left(\frac{f_s}{600} \right)^2 + 1} \right] \quad (1)$$

Bark 단위에서 각 필터마다의 대역폭은 1 bark를 취하게 되며, 필터의 가중치함수는 식 2와 같다. 여기서 z_k 는 중심 bark이다. 식 2.6에서 보면 필터의 가중치 함수는 제한된 범위 즉, 해당 필터의 대역폭 안에서는 가중치를 1로 하여 푸리에 변환된 파워를 그대로 유지하며, 그렇지 않은 곳에서는 그 값을 통과하지 못하도록 하는 필터 특징을 가진다.

$$C_k(w) = \begin{cases} 10^{1.0(z-z_k+0.6)} & , \text{ for } z \leq z_k - 0.5 \\ 1 & , \text{ for } z_k - 0.5 \leq z \leq z_k + 0.5 \\ 10^{-2.5(z-z_k-0.5)} & , \text{ for } z \geq z_k + 0.5 \end{cases} \quad (2)$$

Equal loudness는 주파수 변화에 따른 사람이 느끼는 음의 강도를 표현한 곡선이다. 즉 저주파에서의 주파수 변화는 강도의 변화를 심하게 느끼는 반면에 고주파로 갈수록 주파수 변화에 대한 감도가 떨어지는 특성을 나타낸 것이다. 이런 특징을 가진 equal-loudness함수를 이용해 각 필터에서 발생한 값들을 pre-emphasis한다. 식 3은 equal-loudness함수를 나타낸 것이다.

$$E(w) = 1.151 \sqrt{\frac{(w^2 + 144 \cdot 10^4)w^2}{(w^2 + 16 \cdot 10^4)(w^2 + 961 \cdot 10^4)}} \quad (3)$$

Intensity-loudness conversion은 최종적으로 식 4와 같은 데이터 compression을 통해서 주파수 영역에서의 최종적인 처리를 한다. 여기서 $X(w)$ 는 주파수 영역으로 음성신호를 변환한 것이다.

$$F(w) = |X(w)|^{1/3} \quad (4)$$

위에서 언급한 3단계는 PLP 파라미터를 찾기 위한 주파수 영역에서의 처리과정이다. 이는 식 5와 같이 나타낼 수 있다.

$$F_k = E(w_k) \int_0^{\pi} C_k(w) F(w) dw \quad (5)$$

식 5와 같이 주파수 영역에서 최종 처리된 데이터는 역 푸리에 변환을 통해 최종적인 PLP 스펙트럼(P)을 구하게 된다. PLP 분석을 통해 음성신호를 분석하면 그림 4와 같이 스펙트럼 형식으로 나타낼 수 있다. 그림 4은 PLP 분석을 이용하여 시간영역에서의 PLP 스펙트럼을 분석한 그림이다.

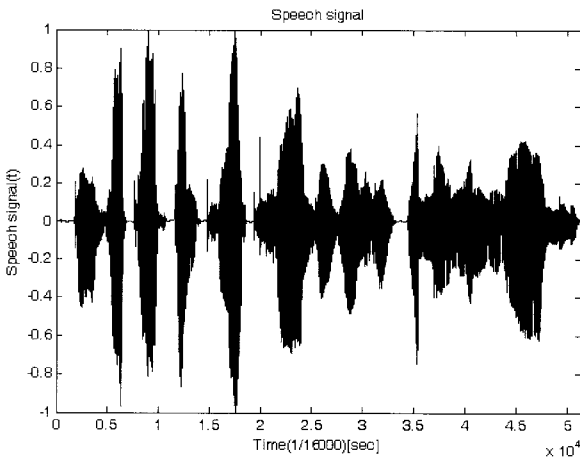


그림 1. 정량화된 음성신호
Fig. 1. Quantitative speech signals

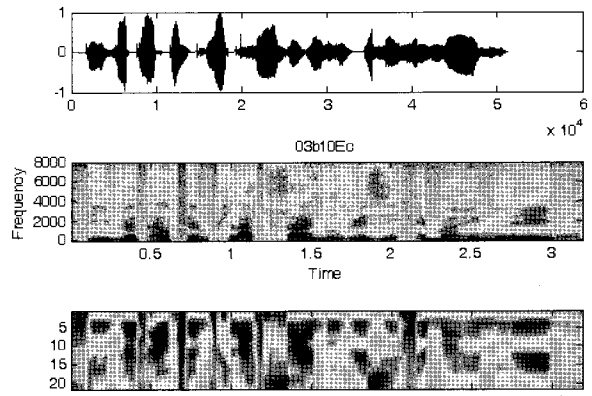


그림 4. 음성신호의 PLP 분석 스펙트럼
Fig. 4. The PLP analysis graph of speech signals

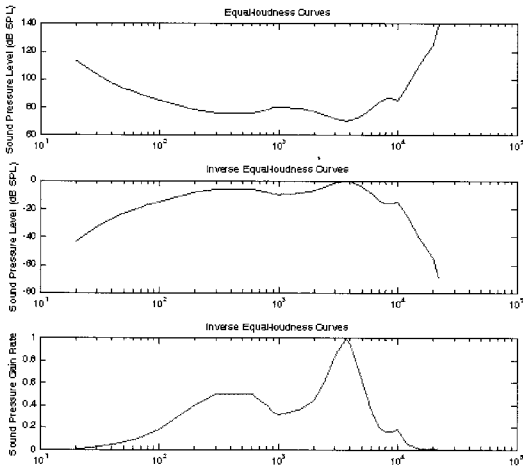


그림 2. 주파수에 따른 인간의 청감도
Fig. 2. Equal-loudness curves

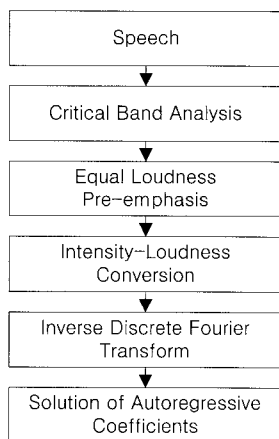


그림 3. PLP 분석 순서도
Fig. 3. PLP analysis algorithm

3. 감정의 개요 및 특징

3.1 감정의 개요

감정은 주관적인 요소이므로 개인에 따라서 표현하는 방법이 조금씩 다르다. 감정은 마음의 움직임이다. 'emotion' (감정)은 라틴어의 'emovere'에서 파생되었는데 'e'는 '밖으로'를 'movere'는 '움직이다'를 의미한다. 따라서 감정은 '밖으로 움직여 나간다'는 뜻을 가진다. 감정은 '개인의 평가에 의해 결정되는 것이며, 육체의 감각 변화에 의해 나타나는 정신 상태로 즉각적이고 직접적으로 사물을 파악하는 기능이다.'로 요약할 수 있다. 그래서 감정은 인간의 생리와 이성적인 판단에 의해 이루어지는 것으로 개인의 욕구를 충족하고 그 충족 지향적인 활동을 자극하는 대상들을 식별하는데 도움이 된다. 본 논문에서는 인간의 감정을 7가지로 분류하여 사용하였다. 본 논문에서 사용한 감정은 그림 5와 같다.

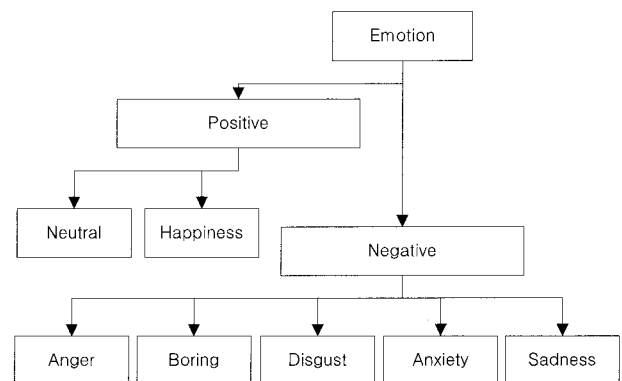


그림 5. 인간의 감정 분류
Fig. 5. Classification of emotions

3.2 감정에 따른 음성신호의 특징

인간의 말속에는 다양한 감정들이 녹아들어 있다. 그림 6은 한 명의 화자가 같은 문장으로 기쁨과 슬픔을 표현하였을 때 시간 영역과 주파수 영역에서의 그래프를 나타내었

다. 그림 6에서 보는 바와 같이 화자가 기쁨을 표현 할 때는 다양한 주파수 영역에서 에너지를 가진다는 것을 볼 수 있다. 그리고 슬픔을 표현 할 때는 특정 주파수 성분들이 많은 에너지를 가진다는 것을 볼 수 있다. 시간영역에서는 슬픔을 표현하는 음성신호에서 거칠게 나타남을 볼 수 있다. 이 처럼 음성신호는 개인의 감정 상태에 따라 시간영역과 주파수 영역에서 다양한 특성이 나타남을 알 수 있다.

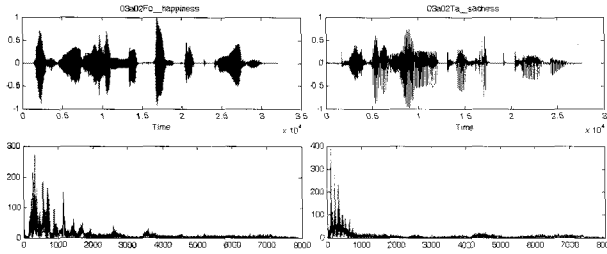


그림 6. 기쁨(happiness)와 슬픔(sadness)에 대한 음성신호

Fig. 6. Speech signals of happiness and sadness

4. 음성신호의 PLP 분석을 이용한 감정인식 시스템 구현

4.1 음성신호의 PLP 분석을 이용한 감정인식의 개요

감정인식 시스템을 구현하기 위해서는, 우선 어떤 분야에 적용될 것인지를 고려하고, 해당 분야에서 발생 가능하거나, 필요로 하는 화자의 대상 감정을 선정하고, 화자 종속 시스템인지 화자 독립 시스템인지의 여부를 결정해야한다. 본 논문에서 감정인식을 위한 알고리즘은 개인과 로봇의 상호 작용을 위한 기술에 초점을 두고 있다. 그래서 감정인식 알고리즘은 개인화된 감정인식 알고리즘으로 만들고자 한다. 음성신호를 이용한 감정인식 방법을 크게 나누면 그림 7과 같다.

우선 speech signal 부분은 소리를 받아들이는 부분이다. 사람의 감정을 인식하기 위해서는 사람의 말을 음성신호로 만드는 부분이라고 할 수 있다. 사람의 말을 마이크를 이용하여 음성신호로 받아들일 때 사람의 말과 함께 외부 노이즈도 함께 들어 올 것이다. 이 노이즈를 speech signal 부분에서 최소화 해주는 것이 필요하다. 본 연구에서는 Emo-DB에서 제공하는 감정인식을 위한 데이터를 이용하였다. 그리고 feature analysis 부분은 들어온 음성신호의 여러 가지 특징을 추출하는 부분이다. 이 부분에서 중요한 점은 음성신호의 특징 분석을 어떻게 하고 특징 값은 무엇으로 할 것 인가이다. 기존의 연구에서는 감정인식을 위해 필요한 음성신호의 특징이 무엇인지 알 수 없었기 때문에 음성신호를 분석하는 여러 가지 방법을 이용하여 많은 음성신호의 특징을 추출하여 감정인식에 영향을 주는 음성신호의 특징을 찾아내는 방법을 feature analysis 부분에 추가하였다. 하지만 본 연구에서는 감정인식에 영향을 주는 음성신호의 특징을 찾아내는 부분을 제거하기위하여 PLP 분석을 이용하여 음성신호를 분석하였다. PLP 분석을 이용하는 이유는 본 연구가 개인화된 감정인식을 목적으로 하였으므로 개인의 음성특징을 잘 분석할 수 있는 PLP 분석이 연구

의 목적에 맞았기 때문이다. 그리고 PLP 분석만을 이용하여 음성신호를 분석하므로 감정인식에 영향을 주는 음성신호의 특징을 찾아내는 부분을 사용할 필요가 없는 장점을 가진다.

다음으로 emotion estimation 부분은 감정의 평가를 위하여 감정의 종류는 무엇으로 할 것 인가를 결정하고 감정의 평가를 하는 부분이다. 기존의 연구에서는 emotion estimation 부분에 다양한 데이터 분류 기법을 이용하였다. 예를 들면 Fuzzy model, K-NN, neural network, HMM 등의 데이터 분류 기법을 이용하여 감정인식에 이용하였다. 본 연구에서는 개인화된 감정인식을 위하여 개인에 따른 감정 템플릿을 이용하여 감정인식을 하였다. 감정 템플릿을 이용한 감정인식 방법은 다음과 같은 두가지 과정으로 이루어진다. 첫 번째 과정은 개인별 감정별 템플릿을 추출하는 것이다. 이때 개인별 감정별 템플릿은 PLP 분석의 에너지에 의한 템플릿과 Min-Max 차이 값에 의한 템플릿으로 만든다. 두 번째 과정은 PLP 분석을 이용하여 음성신호의 특징을 에너지와 Min-Max 차이 값을 이용하여 최종 감정을 판단하는 것이다. 최종 감정을 판단하기 위하여 본 연구에서는 PLP 스펙트럼의 에너지에 의한 감정인식 결과와 Min-Max 차이 값에 의한 감정인식 결과에 가중치를 이용하였다. 가중치를 적용하는 방법으로 Genetic Algorithm(GA)을 이용하여 가중치를 최적화 하였다.

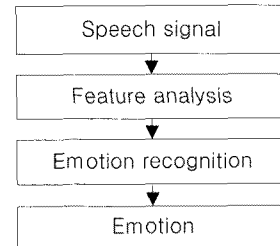


그림 7. 음성신호를 이용한 감정인식 순서도

Fig. 7. Emotion recognition algorithm using speech signals

4.2 PLP 분석을 이용한 음성신호의 감정별 특징 추출

PLP 분석을 이용하여 음성신호를 분석하면 PLP coefficients 값과 시간영역에서의 스펙트럼을 값을 얻을 수 있다. 기존의 논문에서는 PLP coefficients를 이용하여 감정인식에 이용하였다. 그러나 본 논문에서는 개인화에 초점을 두고 음성신호 전체의 특징을 분석하기위하여 PLP의 스펙트럼 값을 이용하여 감정별 특징을 분석하였다. 16KHz로 샘플링한 음성신호를 12차의 PLP 분석을 하게 되면 각 시간대로 21개의 스펙트럼 값이 생성된다. 이 스펙트럼 값을 모든 시간에 대한 평균 에너지를 구하여 하나의 음성신호에 대한 특징을 만들어 낼 수 있다. 식 6은 PLP 분석으로부터 평균 에너지를 구하는 식이다. 여기서 P_i 는 PLP 스펙트럼 값을 나타내고 i 는 PLP 스펙트럼의 프레임 수이고 N 은 PLP 스펙트럼의 총 샘플링 수를 나타낸다.

$$E(i) = \frac{1}{N} \sum_{n=1}^N P_i(n) \quad (6)$$

다음으로 모든 시간에 대하여 각 스펙트럼에 대한 최소와 최대의 차이를 구한 것을 음성신호의 특징으로 만들어 낼 수 있다. 이는 식 7과 같이 나타낼 수 있다.

$$M(i) = \text{Max}(P_i) - \text{Min}(P_i) \quad (7)$$

실험에서는 이 두 가지의 음성신호의 특징을 이용하여 감정인식에 적용하고자 한다.

4.3 음성신호의 특징 데이터 정규화

여기서 중요한 것은 PLP 분석을 통하여 음성신호의 특징 데이터를 일반화 하는 작업이 필요하다. 왜냐하면 모든 음성신호는 서로 다른 크기로 이루어져있기 때문에 그대로 사용한다면 일반적인 특징 분석 데이터로 사용하기에 문제점이 있다. 그러므로 정규화 기법을 통하여 데이터를 전 처리할 필요가 있다. 그래서 본 논문에서는 식 8과같이 Min-Max 정규화 기법을 사용하였다[20]. 여기서 P_i^* 는 Min-Max 정규화한 음성신호의 PLP 스펙트럼 값이고 P 는 한 음성신호의 모든 프레임에 대한 PLP 스펙트럼 값이다.

$$P_i^* = \frac{P_i - \min(P)}{\max(P) - \min(P)} \quad (8)$$

Min-Max 정규화는 하나의 값이 그 영역에서 최소 보다 얼마나 더 큰가를 보고 그 영역의 범위로 스케일링(scaling) 하는 정규화 기법이다. 이 방법을 사용하면 가장 큰 값을 1로 만들고 나머지 값들은 0과 1사이의 값으로 만들 수 있다. 이렇게 정규화를 함으로써 음성신호의 크기에 큰 영향을 받지 않는 음성신호의 특징을 얻을 수 있게 된다. 이러한 데이터 정규화 작업은 일반적인 데이터 분석에서도 많이 사용되는 방법으로 데이터 특징을 일반화 할 수 있는 장점을 가지게 된다.

4.4 GA를 이용한 가중치 설정

최종 감정을 판단하기 위하여 PLP의 에너지 특징으로부터 얻어진 감정결과와 PLP의 Min-Max차이에 의한 결과를 가중치를 이용하여 결합한다. 이를 위해 두 결과에 대한 가중치를 서로 다르게 적용할 필요가 있다. 왜냐하면 에너지에 의한 감정인식 성능과 Min-Max 차이에 의한 감정인식 성능이 다르기 때문이다. 성능이 우수한 결과에 대하여 가중치를 많이 부여할 필요가 있다. 이를 위해 본 연구에서는 GA를 이용하여 최적의 가중치를 구하고자 한다. GA의 목적함수(object function) J 는 학습을 위한 음성신호들의 감정인식률로 사용한다. 목적함수를 감정인식률로 사용한 것은 가중치를 사용하는 목적이 감정인식률을 높이기 위해 사용하기 때문이다. 그리고 GA의 개체 적합도(fitness) 함수 F 는 다음 식 9와 같다.

$$F = \left(\frac{J}{100} \right)^2 \quad (9)$$

그림 8은 전체적인 가중치 결정을 위한 GA의 흐름도이다. 지정한 적합도 값이 나올 때까지 GA를 반복하여 학습을 진행한다.

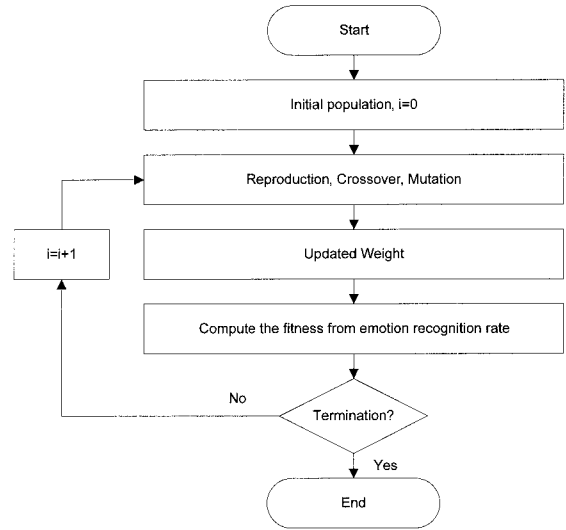


그림 8. 가중치 결정을 위한 GA 흐름도
Fig. 8. Flow chart of the weight tuning algorithm

4.5 PLP 분석을 이용한 개인화된 감정인식 알고리즘

위와 같은 개인화된 감정인식을 위한 방법을 이용하여 본 연구의 실험에서는 그림 9과 같은 개인화된 감정인식 알고리즘을 이용하여 실험을 하였다.

그림 9의 개인화된 감정인식 알고리즘에서 $X(t)$ 는 음성신호를 나타낸다. 이 음성신호 $X(t)$ 를 4.2절에 설명한 PLP 분석을 이용하여 PLP 스펙트럼 값인 $P(n)$ 을 구한다. 계속해서 PLP 스펙트럼 $P(n)$ 을 이용하여 PLP 에너지 데이터 $E(i)$ 와 PLP Min-Max 차이 데이터 $M(i)$ 을 4.2절에 있는 식 6와 식 7을 이용하여 구한다. 다음으로 4.3절의 정규화 과정을 이용하여 PLP 에너지 데이터 $E(i)$ 와 PLP Min-Max 차이 데이터 $M(i)$ 을 정규화한 에너지 데이터 E_i^* 와 Min-Max 차이 데이터 M_i^* 로 만든다. 다음으로 정규화한 E_i^* 와 M_i^* 을 감정별 에너지 템플릿 $T_E(j)$ 와 Min-Max 템플릿 $T_M(j)$ 을 이용하여 유클리디언 거리 계산을 한다. 그리고 계산된 감정별 에너지 템플릿에 대한 유클리디언 거리 $D_E(j)$ 와 Min-Max 차이 템플릿에 대한 유클리디언 거리 $D_M(j)$ 을 낮은 순서대로 정렬시킨다. 그리고 낮은 순서를 가진 감정에 대해서 낮은 가중치를 준다. 그리고 감정별 에너지에 대한 평가 값 $R_E(j)$ 와 Min-Max 차이에 대한 평가 값 $R_M(j)$ 을 더하여 최종 감정을 판단하기 위한 감정별 평가치 $F(j)$ 을 계산한다. 마지막으로 $F(j)$ 가 가장 작은 값을 가진 감정 j 를 최종 감정으로 판단한다.

본 논문에서 제시한 음성신호를 이용한 개인화된 감정인식 알고리즘은 기존의 연구와 달리 개인에 따른 음성신호의 특징변화에 강한 감정인식 방법이라는 점에서 장점을 가진다.

5. 실험 결과 및 결론

본 실험에서 사용된 데이터는 남녀 배우 각각 5명에 의해서 행복(Happiness), 화남(Anger), 걱정(Anxiety), 슬픔(Sadness), 지루함(Bored), 혐오(Disgust) 그리고 일상(Neutral)으로 나누어진 7가지 감정으로 만들어진 약 500개의 음성신호로 이루어져있다. 4절에서 설명한 감정 템플릿 생성 알고리즘을 이용하여 PLP 스펙트럼의 에너지에 대한 감정별 템플릿을 생성한 결과는 그림 10과 같이 나타났다. 그리고 최종적인 감정인식 결과는 표 2와 같이 나타났다. 표 2의 결과를 이용하여 평균 감정인식률을 구하면 77.11%로 나타난다. 이러한 결과는 기존의 연구에서 목표로 하는 음성신호를 이용한 감정인식의 인식률 70%이상과 비교하여 만족할 만한 수치라고 할 수 있다. 그리고 기존의 음성신호를 이용한 감정인식 연구에서 60개 이상의 음성신호의 특징을 이용해 감정인식을 한 것과 비교하여 PLP의 스펙트럼만을 이용하여 감정인식을 하였다는 점에서 장점을 가질 수 있다. 이는 감정인식을 위해 연산시간이 너무 오래 걸리는 단점을 극복 할 수 있다고 하겠다.

앞으로 음성신호를 이용한 감정인식의 발전 방향은 기계와 인간의 상호작용을 위해 필요한 감정의 개요를 보다 적절하게 정의하여야 하겠다. 그리고 음성신호의 특징 추출에 있어서 각 감정에 따른 음성신호의 특징 변화를 명확히 정의할 통한 추출이 필요할 것이다. 이러한 점이 확실하게 정의된다면 음성신호를 이용한 감정인식은 간단하면서도 인식률이 좋아 질 것이다.

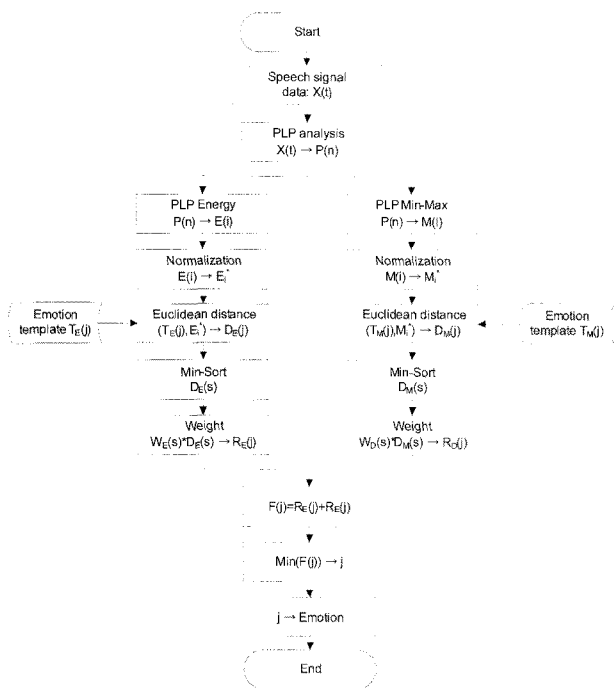


그림 9. PLP 분석을 이용한 개인화된 감정인식 알고리즘
Fig. 9. Personalized emotion recognition algorithm using PLP analysis

표 2. 음성신호의 PLP 분석을 이용한 감정인식 결과
Table 2. Emotion recognition results using PLP energy templates

사람 코드	성별	에너지 템플릿을 이용한 인식률(%)	Min-Max 차이를 이용한 인식률(%)	가중치를 이용한 인식률(%)
1	남	69.4	67.3	73.5
2	여	78.6	71.4	76.8
3	여	62.8	72.1	70.3
4	남	97.4	86.8	97.4
5	남	70.9	58.2	76.4
6	남	91.4	82.8	91.4
7	여	63.9	70.5	70.1
8	여	66.7	66.7	72.5
9	남	71.4	60.7	76.8
10	여	80.3	67.6	80.3

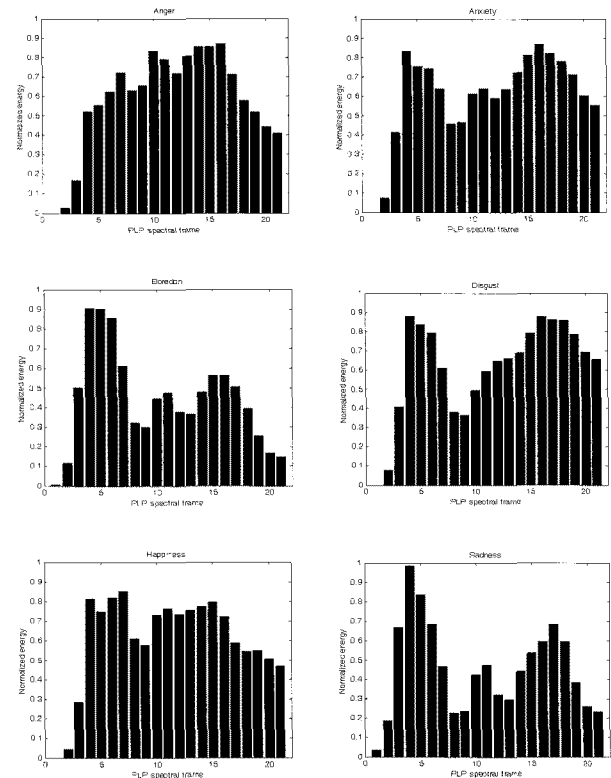


그림 10. 6가지 감정에 대한 음성신호의 PLP 에너지 템플릿

Fig. 10. PLP energy templates of speech signals for six emotions

참 고 문 헌

[1] Kwang-Dong Jang, Nam Kim, Oh-Wook Kwon, "Speech Emotion Recognition on a Simulated Intelligent Robot," *대한음성학회*, vol. 56, pp. 173-183, 2006.

[2] Chang-Huun Park and Kwee-Bo Sim, "Reinforcement Learning Method Based Interactive Feature Selection(IFS) Method for Emotion Recognition," *제어 자동화 시스템공학 논문지*, 제12권 제7호, pp. 666-670, 2006.

[3] J. Nicholson, K. Takahashi and R. Nakatsu, "Emotion Recognition in Speech using Neural Networks," *Neural Comput. & Applic.*, 제9권, pp. 290-296, 2000.

[4] Aishah Abdul Razak, Ryoichi Komiya, Mohamad Izani Zainal Abidin, "Comparison Between Fuzzy and NN Method for Speech Emotion Recognition," *Proceedings of the Third International Conference on Information Technology and Applications*, 2005.

[5] 강면구, 서정태, 김원구, "음성 신호를 사용한 GMM 기반의 감정 인식," *한국음향학회지*, 제23권 제 3호, pp. 235-241, 2004.

[6] 조윤희, 박규식, "A Study on Robust Speech Emotion Feature Extraction Under the Mobile Communication Environment," *한국음향학회지*, 제 25권 제6호, pp. 269-276, 2006.

[7] Vladimir Hozjan and Zdravko, "Context-Independent Multilingual Emotion Recognition from Speech Signals," *International journal of speech technology*, Vol. 6, pp. 311-320, 2003.

[8] 심귀보, 박창현, 주영훈, "음향적 요소분석과 DRNN을 이용한 음성신호의 감정 인식," *퍼지 및 지능시스템학회 논문지*, Vol. 13, No. 1, pp. 45-50, 2003.

[9] 정광우, "화자인식을 위한 음성신호처리," *전자공학회집*, 제26권, 제11호, pp. 135-1145, 1999.



천성표(SeungPyo Cheun)
 1999년 : 부산대학교 전기공학과 (공학사)
 2001년 : 동대학원 전기공학과 (공학석사)
 2008년 : 동대학원 전기공학과 박사과정
 2008년~현재 : Convertteam 연구원

관심분야 : Machine Learning, Bayesian Networks
 E-mail : buzz74@pusan.ac.kr



김연태(Yountae Kim)
 2003년 : 부산대학교 전자전기통신공학부 (공학사)
 2005년 : 동대학원 전기공학과 (공학석사)
 2008년 : 동대학원 전기공학과 박사과정
 2008년~현재 : 대우조선해양 선박해양기술 연구소 연구원

관심분야 : 신호처리, 영상처리, 고장진단, 지능제어
 E-mail : dream0561@pusan.ac.kr



김성신(Sungshin Kim)
 1986년 : 연세대학교 전기공학과 (공학석사)
 1996년 : Georgia Inst. of Technology, 전기및컴퓨터공학과 (공학박사)
 1998년~현재 : 부산대학교 전자전기통신공학부 부교수

관심분야 : 지능 시스템, 데이터 마이닝, 고장예측및진단
 Phone : +82-51-510-2374
 Fax : +82-51-513-0212
 E-mail : sskim@pusan.ac.kr

저 자 소 개



정병욱(Byungwook Jung)
 2006년 : 부산대학교 전자전기정보컴퓨터 공학부 (공학사)
 2008년 : 부산대학교 전기공학과 (공학석사)
 2008년~현재 : LG생산기술원 연구원

관심분야 : 지능 제어, 데이터 마이닝, 고장진단
 E-mail : wooroogy@pusan.ac.kr