

## 음성신호를 이용한 감정인식

### An Emotion Recognition Technique Using Speech Signals

정병욱<sup>1</sup>, 천성표<sup>2</sup>, 김연태<sup>3</sup>, 김성신<sup>4</sup>

<sup>1</sup> 부산시 금정구 부산대학교 전기공학과  
E-mail: wooroogy@pusan.ac.kr

<sup>2</sup> 부산시 금정구 부산대학교 전기공학과  
E-mail: buzz74@pusan.ac.kr

<sup>3</sup> 부산시 금정구 부산대학교 전기공학과  
E-mail: dream0561@pusan.ac.kr

<sup>4</sup> 부산시 금정구 부산대학교 전기공학과  
E-mail: sskim@pusan.ac.kr

#### 요 약

본 논문은 음성신호를 이용한 감정인식에 관한 연구이다. 감정인식에 관한 연구는 휴먼 인터페이스(Human Interface) 기술의 발전에서 인간과 기계의 상호작용을 위한 것이다. 본 연구에서는 음성신호를 이용하여 감정을 분석하고자 한다. 음성신호의 감정인식을 위해서 음성신호의 특징을 추출하여야 한다. 본 논문에서는 개인에 따른 음성신호의 감정인식을 하고자 하였다. 그래서 화자인식에 많이 사용되는 음성신호 분석기법인 Perceptual Linear Prediction(PLP) 분석을 이용하여 음성신호의 특징을 추출하였다. 본 연구에서는 PLP 분석을 통하여 개인화된 감정 패턴을 생성하여 간단하면서도 실시간으로 음성신호로부터 감정을 평가 할 수 있는 알고리즘을 만들었다.

**Key Words** : 음성신호, 감정인식, PLP 분석

#### 1. 서 론

인간의 소리는 의사소통을 하기 위한 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(LPC) 분석을 이용하여 특징 벡터를 생성하고 Neural

Network(NN)을 이용하여 감정을 분류하였다. 감정인식 분류 결과는 8개의 감정에 대하여 약 50%정도로 타났다[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]. 강민구는 화자 및 문장 독립적 감정인식을 위하여 Vector Quantization(VQ)와 Gaussian Mixture Model(GMM)을 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 연구에서 화자에 종속적인 감정인식에서 높은 인식률을 보였고 언어별

감정인식은 비슷한 인식률을 보였다고 결과로 보고하고 있다[7].

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

2장에서 음성신호의 개요 및 특징 분석에 대하여 논하고 3장에서 감정의 개요 및 특징에 대하여 설명한다. 그리고 4장에서 개인화된 음성신호의 감정인식에 대한 알고리즘 및 실험 방법에 대하여 설명하고 5장에서 실험 결과 및 결론을 논하도록 한다.

## 2. 음성신호의 개요 및 특징 분석

인간의 소리에는 많은 정보를 가지고 있다. 소리의 주된 목적은 상대방과의 의사소통을 위한 것이다. 소리는 마이크를 통하여 정량적인 값으로 저장할 수 있다. 이러한 정량적인 소리를 음성신호라고 한다. 그림 1은 음성신호를 나타내는 그래프이다.

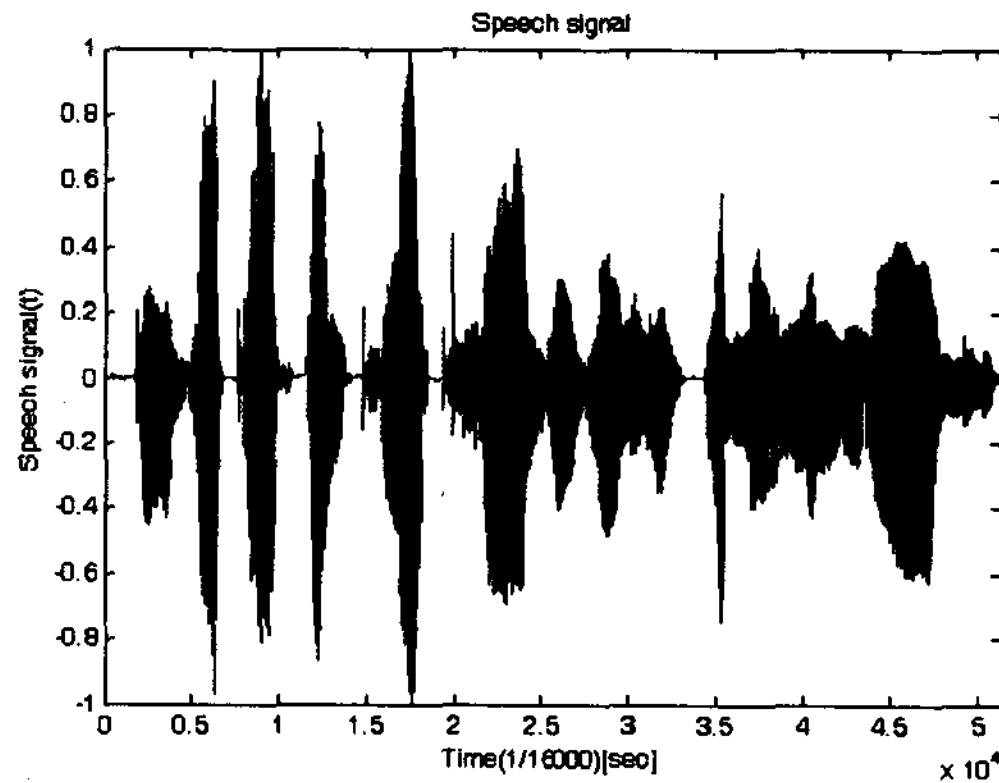


그림 1. 정량화된 음성신호.

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. 본 논문에서는 PLP 분석을 통하여 감정인식을 하고자 한다.

PLP 분석은 인간이 소리에 대한 청감도를 고려하여 음성신호를 분석하는 기법이다. 인간의 청감도는 그림 2와 같이 나타난다. 인간의 청감도는 인간의 귀로 소리를 들을 때 주파수에 따라 다른 감도를 가지고 있음을 나타낸다.

이러한 청감도를 이용하여 음성신호에 대한 주파수의 강도를 조절하고 역 변환을 이용하여 시간에 따른 음성신호의 스펙트럼을 분석한 것이 PLP 분석의 특징이다.

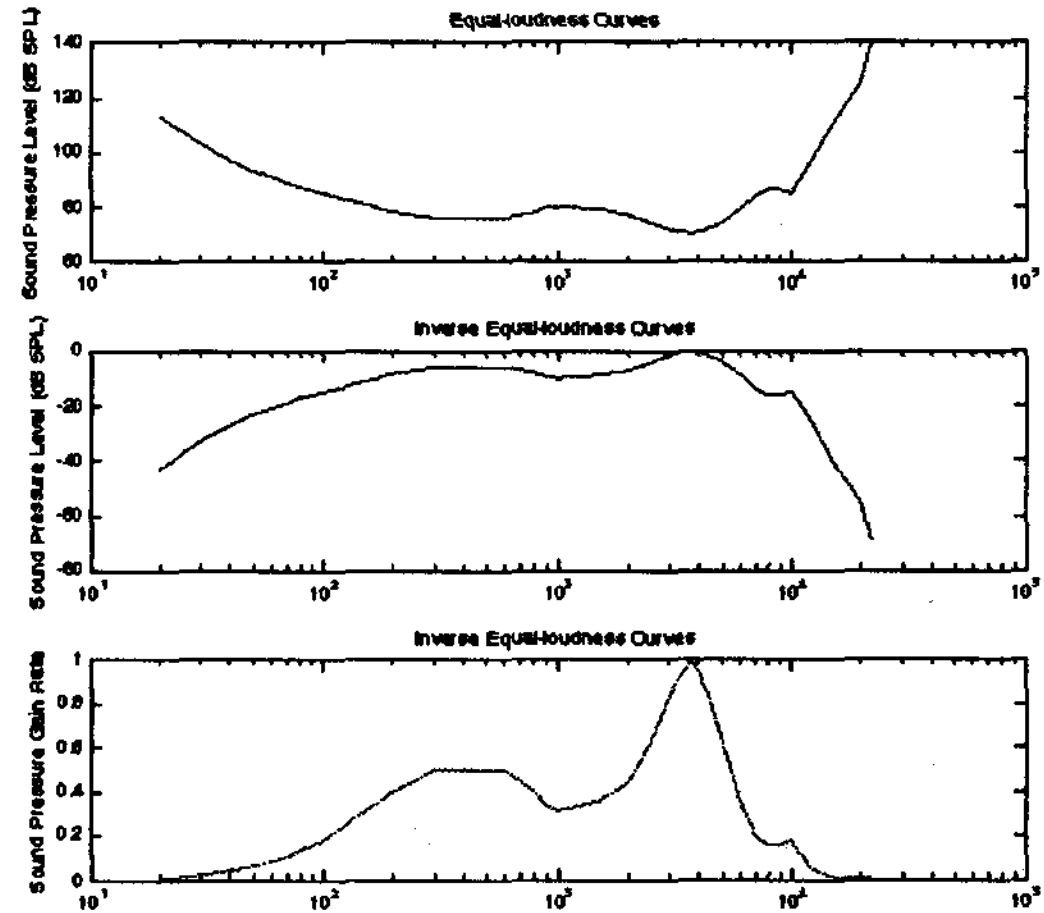


그림 2. 인간의 청감도.

PLP 분석의 알고리즘은 그림 3과 같다. 그림 2.3에서 Critical Band Analysis는 주파수 영역으로 변환된 입력신호의 전력을 구하는 것이다. Equal Loudness는 주파수 변화에 따른 사람이 느끼는 음의 강도를 표현한 곡선이다.

PLP 분석은 음성신호의 개인적인 억양, 음색, 언어 등의 특징을 많이 나타낸다고 할 수 있다.

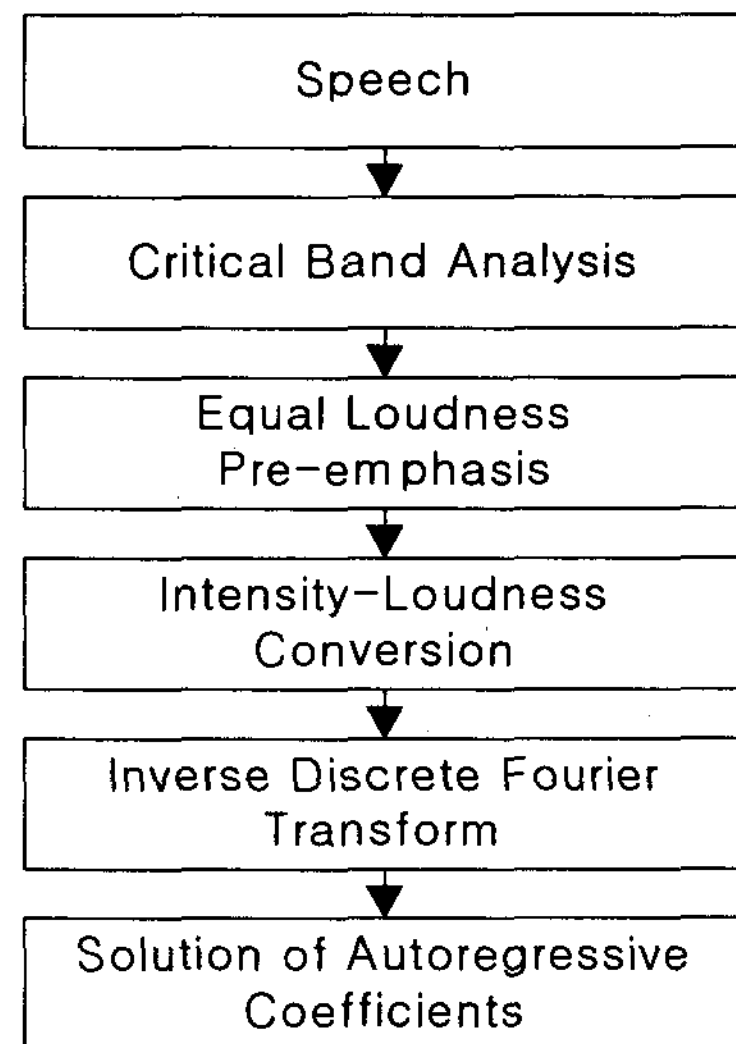


그림 3. PLP 분석 알고리즘.

PLP 분석을 통해 음성신호를 분석하면 그림 4와 같이 스펙트럼 형식으로 나타낼 수 있다. 그림 4는 PLP 분석을 이용하여 시간영역에서의 PLP 스펙트럼을 분석한 그림이다. 그림 4의 세 번째 그림에서 세로축은 스펙트럼의 프레임을 나타낸 것이다. 총 21개의 프레임으로 구성되어있다. 본 논문에서는 각 프레임에 대하여 평균 에너지와 Min-Max 차이를 이용하

여 음성신호의 특징을 추출하였다.

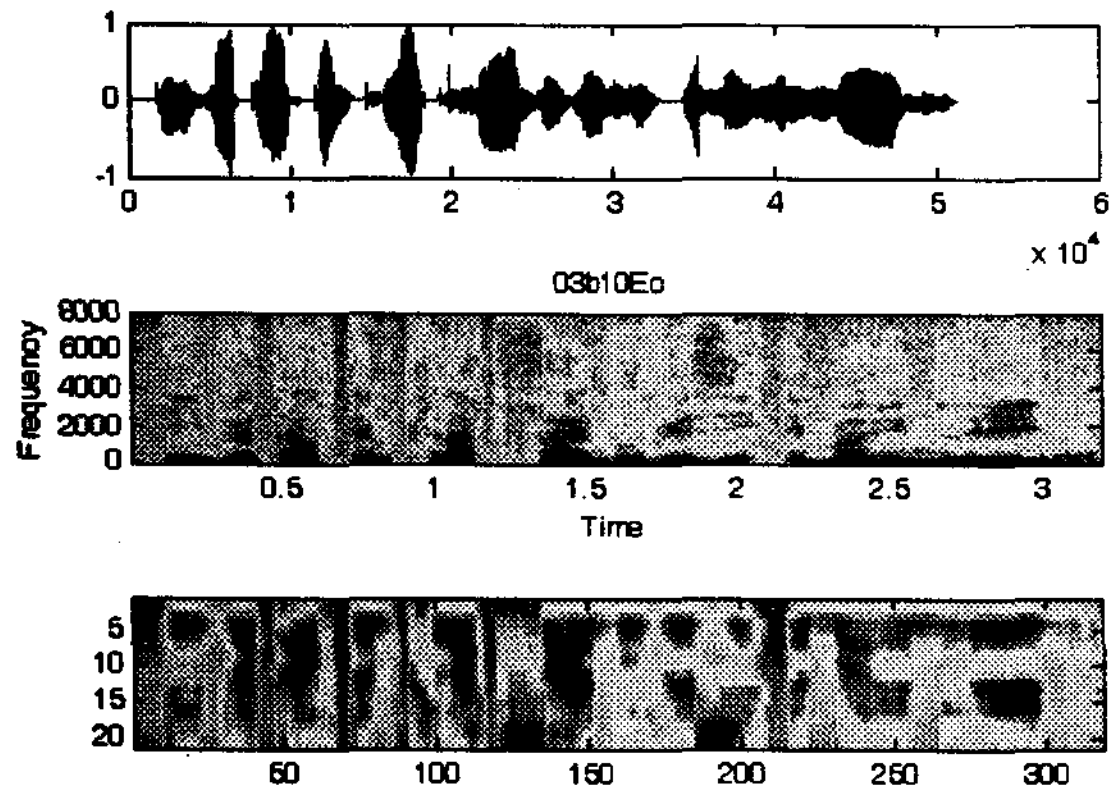


그림 4. 음성신호의 PLP 분석 그래프.

### 3. 감정의 개요 및 특징

#### 3.1 감정의 개요

인간은 자신의 감정을 행동, 표정이나 말로서 표현을 한다. 이러한 인간의 감정은 상대방과의 상호작용을 위해서 필수적인 요소이다. 즉, 상대방에게 자신의 감정을 표현함으로써 자신이 상대에게 전달하고자 하는 바를 쉽게 전달 할 수 있게 된다. 본 논문에서는 인간의 감정을 7가지로 분류하여 사용하였다. 본 논문에서 사용한 감정은 그림 5와 같다.

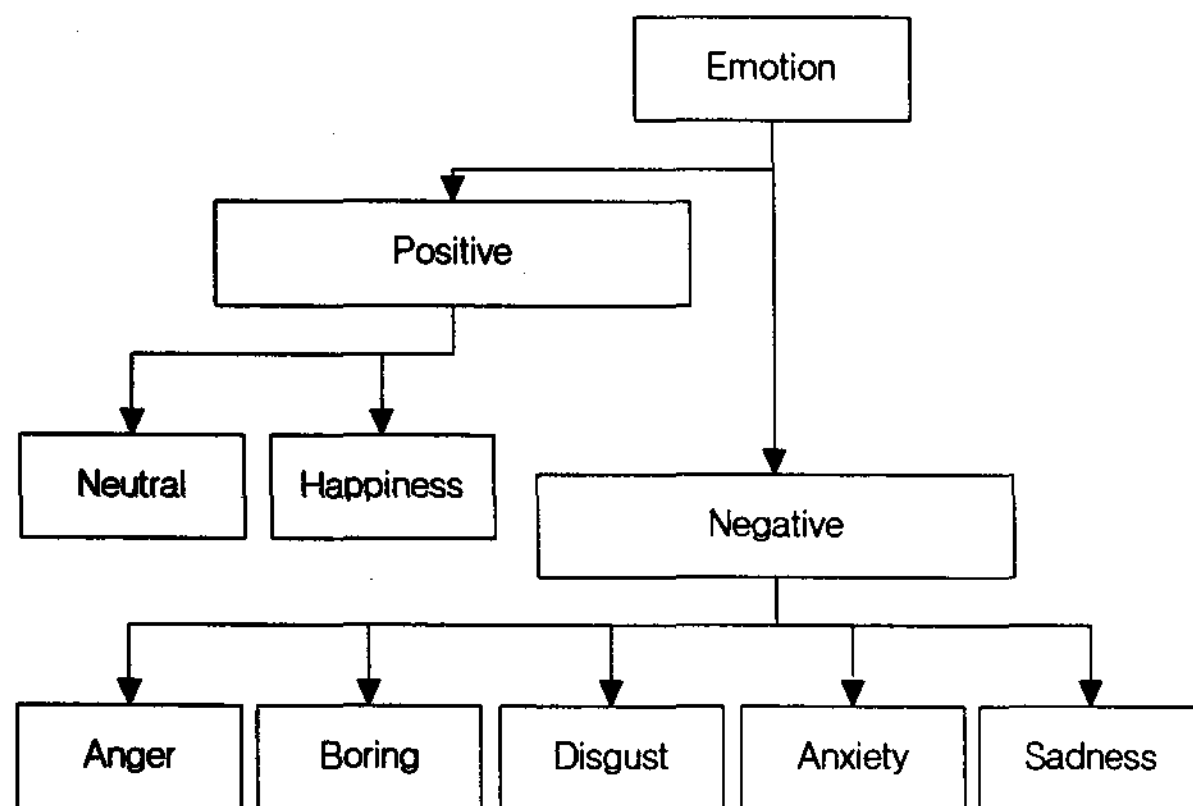


그림 5. 두 가지 감정에 대한 신호 분석.

#### 3.2 음성신호에서 감정별 특징

인간의 말을 마이크를 통하여 음성신호로 정량화함으로써 감정에 따른 음성신호의 특징을 분석할 수 있다. 기본적으로 시간 영역에서의 음성신호 분석이 있을 수 있다. 그리고 주파수 영역에서의 음성신호 분석이 가능하다. 주파수 영역에서 음성신호를 분석하면 감정에 따라 변화되는 부분을 볼 수 있다. 그림 6은 한 명의 화자가 같은 문장으로 기쁨과 슬픔을 표현하였을 때 시간 영역과 주파수 영역에서의 그래프를 나타내었다.

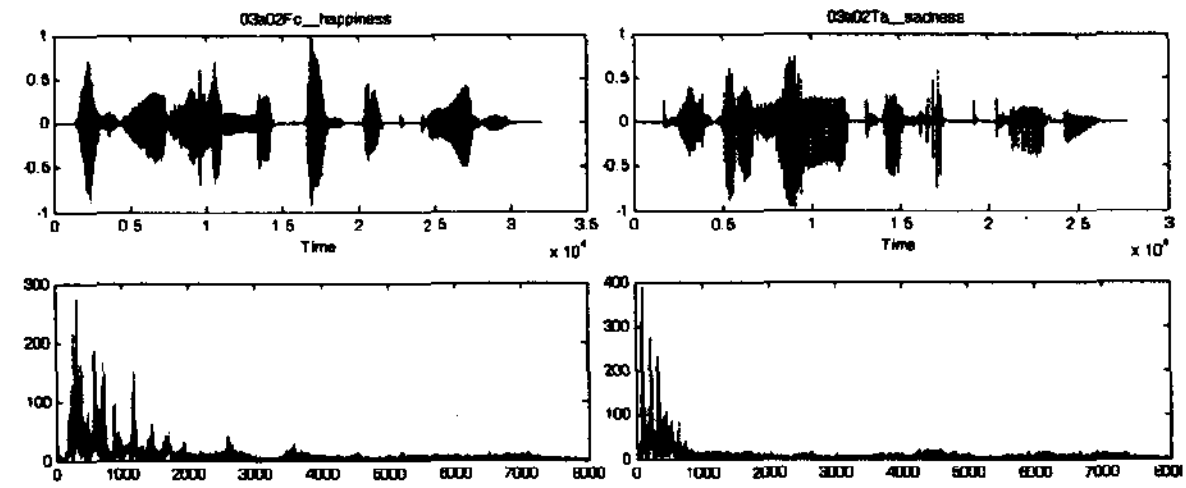


그림 6. 두 가지 감정에 대한 신호 분석.

### 4. 개인화된 음성신호의 감정인식을 위한 알고리즘

본 논문에서 PLP 분석을 이용한 개인화된 감정인식 알고리즘은 크게 두 부분으로 나눈다. 첫 번째 부분은 개인별 감정별 템플릿을 추출하는 것이다. 이때 개인별 감정별 템플릿은 PLP 분석의 에너지에 의한 템플릿과 Min-Max 차이 값에 의한 템플릿으로 만든다. 두 번째 부분은 PLP 분석을 이용하여 음성신호의 특징을 에너지와 Min-Max 차이 값을 이용하여 최종 감정을 판단하는 것이다. 개인별 감정별 템플릿을 추출하는 알고리즘은 그림 7과 같다.

음성신호의 PLP 스펙트럼의 에너지와 Min-Max 차이 값을 이용하여 감정을 판단하는 알고리즘은 그림 8과 같다.

여기서 PLP 스펙트럼의 에너지에 의한 감정인식 결과와 Min-Max 차이 값에 의한 감정인식 결과를 가중치를 이용하여 최종 감정을 판단한다. 가중치를 적용하는 방법으로 본 연구에서는 Genetic Algorithm(GA)을 이용하여 가중치를 최적화 하였다.

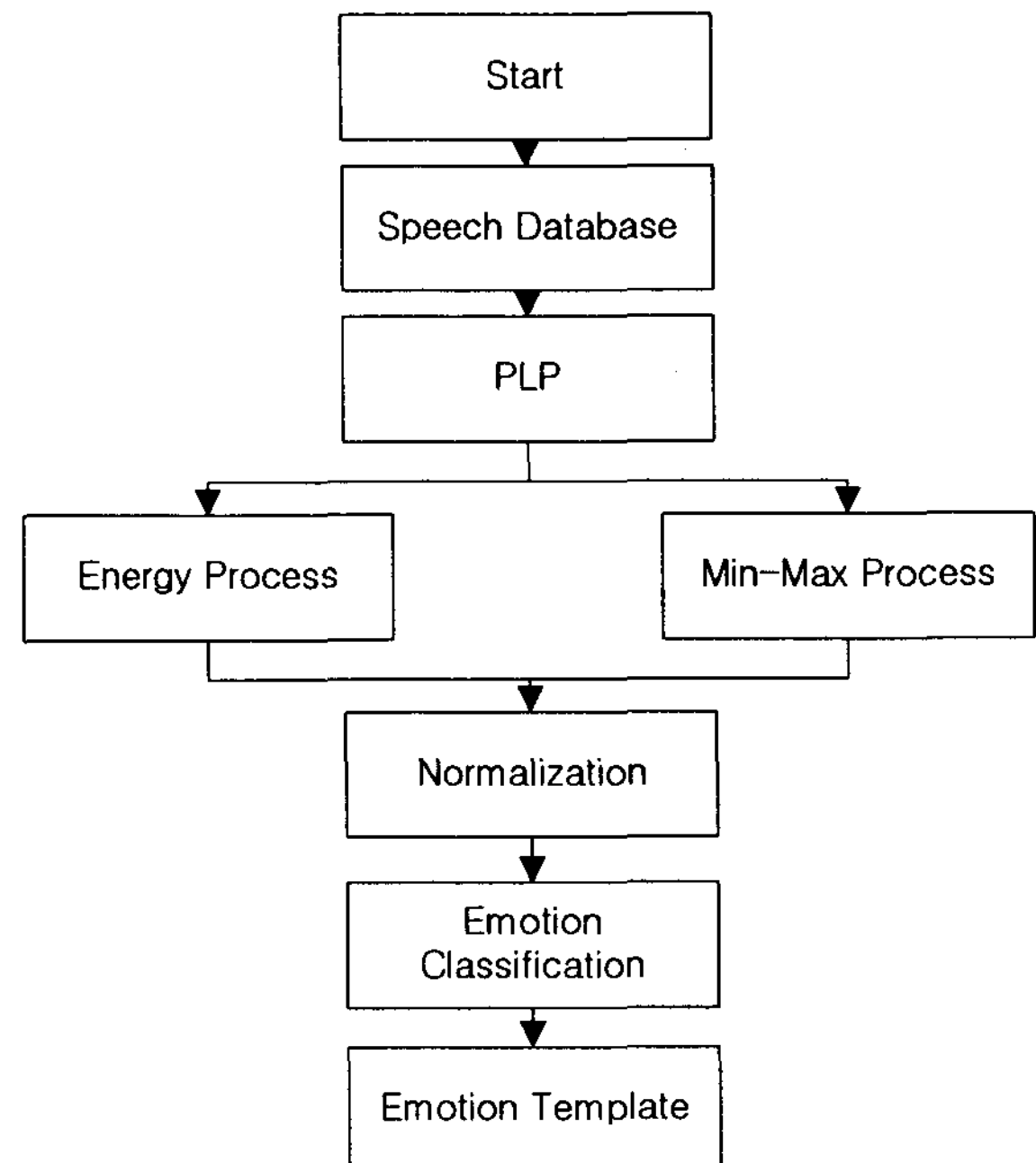


그림 7. 개인화된 감정별 템플릿 생성 알고리즘.

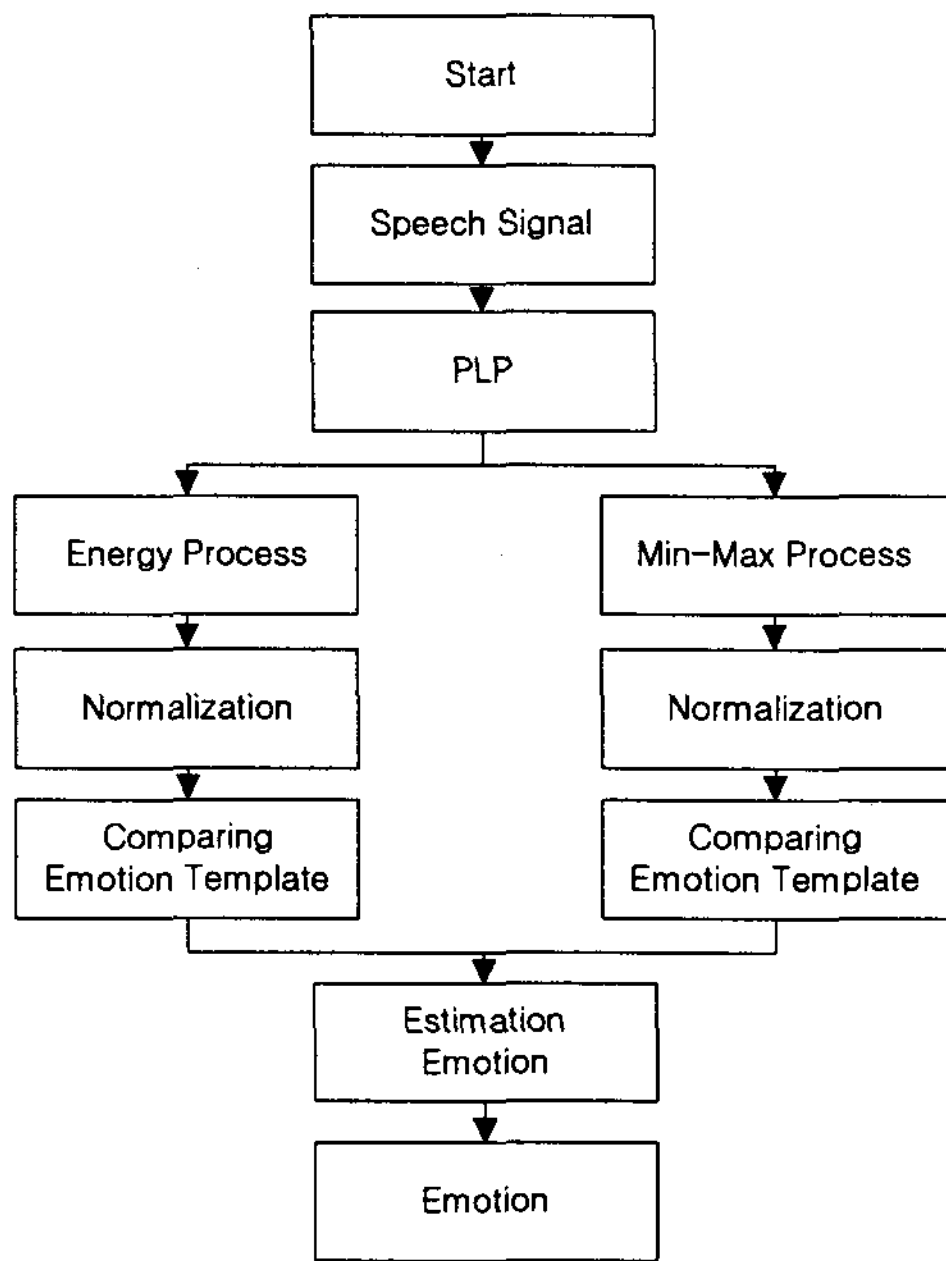


그림 8. 개인화된 감정인식을 위한 알고리즘.

### 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. 4장에서 소개된 알고리즘을 이용하여 감정별 템플릿을 생성한 결과는 그림 9와 같이 나타난다.

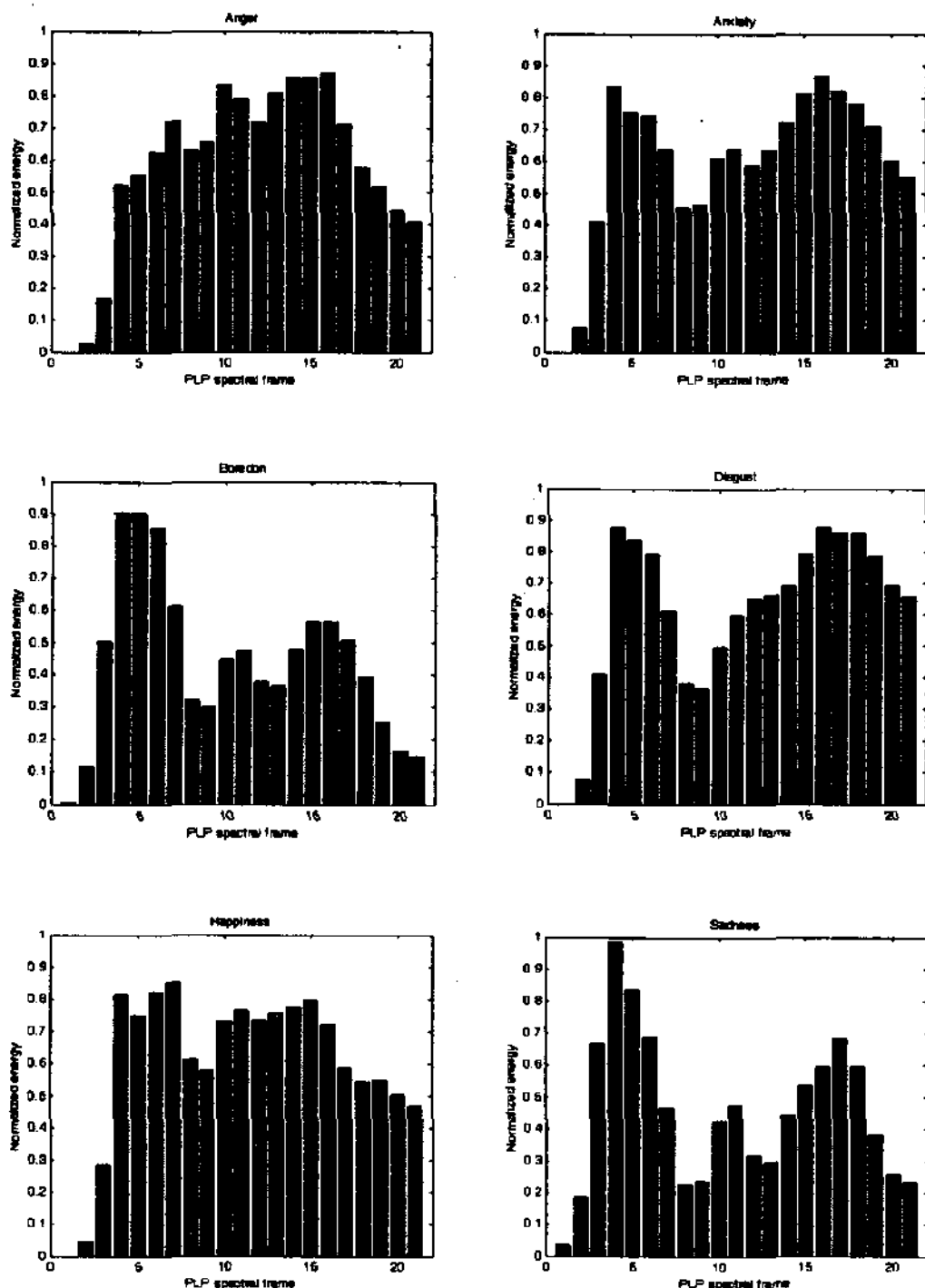


그림 9. 음성신호의 PLP 에너지 템플릿.

그림 9는 PLP 에너지를 이용한 한 개인의 감정별 템플릿이다. 본 실험에서는 이러한 템플릿을 개인화하여 생성하고 감정인식에 적용하였다. 결과는 표 1과 같다. 평균 70%이상의 인식률을 보였다.

표 1. 개인화된 음성신호의 감정인식 결과.

사람	1	2	3	4	5
인식률 (%)	74	77	70	97	76
사람	6	7	8	9	10
인식률 (%)	91	70	72	77	80

### 감사의 글

이 논문은 2007년 한국전자통신연구원(ETRI)의 지원에 의하여 연구되었음.

### 참고 문헌

- [1] Kwang-Dong Jang, Nam Kim, Oh-Wook Kwon, "Speech Emotion Recognition on a Simulated Intelligent Robot," 말소리, 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 IEEE.
- [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, 제25권 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.