

# 음성으로부터의 감정 인식을 위한 퍼지모델 제안

## Fuzzy Model for Speech Emotion Recognition

문병현, 장인훈, 심귀보

Byung-Hyun Moon, Jang In-Hoon, Kwee-Bo Sim

중앙대학교 전자전기공학부

(E-mail: kbsim@cau.ac.kr)

### 요 약

본 논문에서는 음성으로부터 감정을 인식하고 감성적인 운율로 음성 출력을 산출해 내는 시스템을 제안 한다. 음성적인 운율로부터 감정을 인식하기 위해서 퍼지룰(rule)을 이용한다. 본 논문에서 감정 인식 시스템은 음성 샘플들로 학습 데이터를 구축하고 이를 기반으로 하여 추출된 20개의 특징 집합으로부터 가장 중요한 특징들을 자동적으로 선택한다. 화남, 놀람, 행복, 슬픔, 보통의 5가지 감정 상태를 구분하기 위하여 접근법에 기반한 퍼지를 이용하였다.

**Key Words** : 감정 인식, 퍼지룰(rule), 음성

### 1. 서 론

음성신호는 언어정보, 화자의 개인성 정보, 성별, 국적, 출신지역, 발성 당시 주변환경의 음향정보, 화자의 감정상태, 피로도 등 다양한 정보를 포함하고 있으며, 최근에 인간과 로봇간의 감성 인터페이스에 대한 관심이 고조되면서 화자의 감정상태를 자동 식별하는 방법에 관한 연구가 활발히 이루어지고 있다. 자동 감정식별 분야의 최근 연구들은 대개 네 종류에서 일곱 종류의 감정을 대상으로 하고 있으며, 화남, 슬픔, 지루함, 즐거움, 중립감정 등을 예로 들 수 있다. 이 분야의 주된 연구주제로 감정 분류를 위한 효과적인 특징추출, 감정식별기 설계, 감정식별을 위한 데이터베이스 구축 등을 들 수 있다[1].

감정식별을 위한 특징 값으로는 피치, 에너지, 지체적, 지속길이정보 등의 운율정보 및 스펙트럼정보 등이 이용되고 있다. 이 중 다수의 연구결과에서 피치 및 에너지가 감정식별에 매우 효과적임이 밝혀졌으나, 그 이외의 특징표현에 대한 연구가 지속적으로 이루어지고 있다. Dimitrios와 Constantine은 5개의 감정을

구분하는데 피치, 에너지, 포먼트에서 87개의 특징점을 추출하여 분류하였고[2], Chul Min과 Shrikanth는 남성과 여성을 구분하여 특징점을 추출하였다[3].

인식단계에서는 HMM(hidden Markov model), GMM(Gaussian mixture model), SVM(support vector machine), 신경회로망, 퍼지추론, k-nearest neighbor, LDA(linear discriminant analysis) 등 다양한 방식이 시도되었으며[4], 본 논문에서는 화남, 놀람, 슬픔, 행복, 보통의 감정 상태를 구분하기 위하여 접근법에 기반한 퍼지를 이용한다. 음성 샘플들로부터 학습 데이터를 구축하고 이로부터 자동적으로 퍼지룰(rule) 시스템이 생성된다. 이러한 데이터 베이스를 이용하여 각 감정에 대한 음성 특징들이 분석된 20개의 특징 집합으로부터 자동적으로 결정된다.

본 논문의 구성은 다음과 같다. 먼저 2절에서는 퍼지 감정 인식 시스템의 구조에 대한 설명이다. 퍼지 모델의 자동 생성에 대한 설명은 3절에서 보여준다. 4절 향후 연구 방향과 결론으로 마무리 짓는다.

### 2. 퍼지 감정 인식 시스템 구조

#### 2.1 퍼지 감정 인식 시스템

본 시스템은 하나의 문장을 입력 받아서 그것을 5개의 감정상태(행복, 슬픔, 화남, 놀람,

감사의 글 : 본 연구는 서울시·중소기업청의 연구비 지원에 의한 2007년도 중앙대학교 산학연 컨소시엄 사업에 의해 수행되었습니다. 연구비 지원에 감사드립니다.

보통)으로 분류한다. 그림 1과 같이 학습 모드와 분류(인식) 모드로 구분되며, 이는 자동적으로 감정 인식에 대한 퍼지 모델을 생성한다. 학습하는 동안에는 각 감정들에 대한 퍼지 모델들을 형성하기 위해 이미 알고 있는 감정 샘플들을 이용한다. 이렇게 생성된 퍼지 모델은 임의의 음성 데이터를 분류하기 위해 감정 인식 과정에서 활용된다. 학습 과정에서는 전형적인 패턴 인식 시스템처럼 4개의 모듈을 가지고 있으나, 마지막 단계에서는 분류기 모듈이 아닌 퍼지 생성 모듈이 대신하게 된다. 다음 그림에서 이러한 단계를 확인할 수 있다.

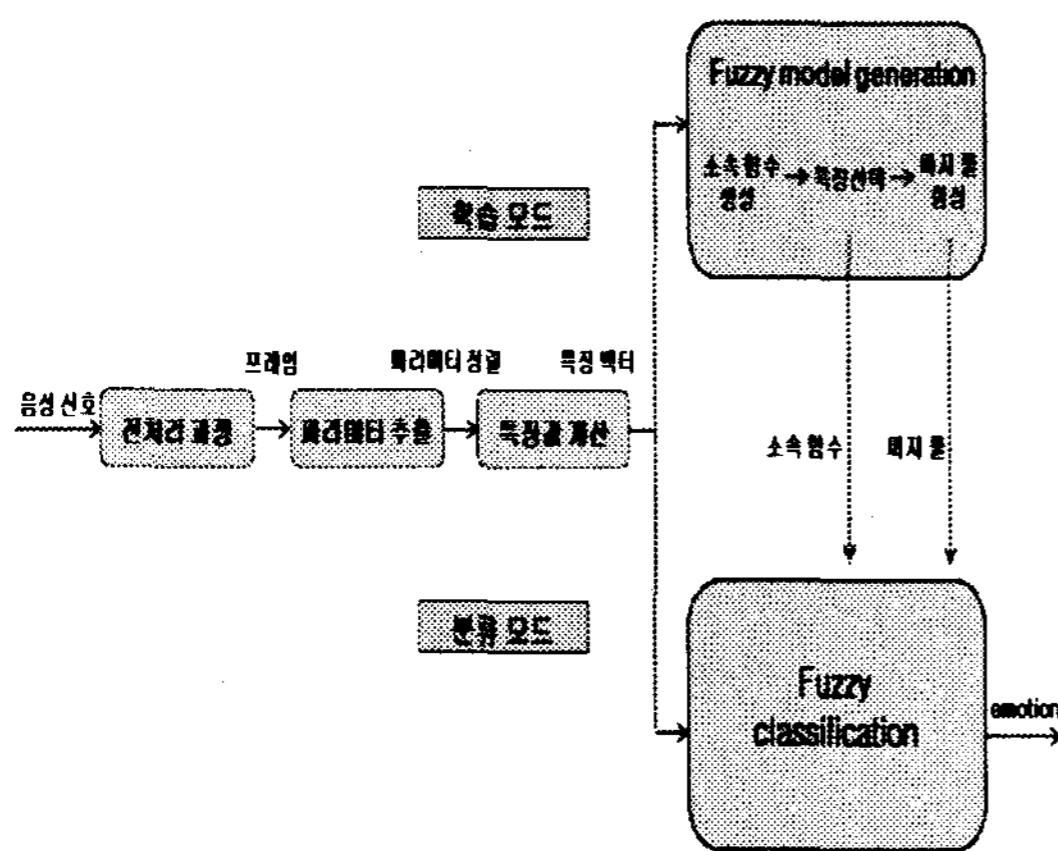


그림 1. 퍼지 감정 인식 시스템 구조

### 2.2 파라미터 추출

각각의 프레임마다 다른 음성 파라미터들을 추출한다. 이후에 이 파라미터들로부터 화자의 감정을 결정하는데 중요한 특징들이 계산된다. 감정을 인식하는데 중요한 정보는 기본 주파수와 음성신호의 시간에 대한 에너지 궤적으로부터 생성된다. 스펙트럼, 피치, 지속길이 등의 운동정보에 기반한 특징 파라미터들도 중요한 정보가 된다. 그러므로 각 프레임에서 음성 구간과 무음 구간이 구해지고 기본적인 주파수, 에너지, 흔들림 등의 값이 결정된다. 파라미터 추출단계를 거쳐 생성된 파라미터들은 일련의 순서를 가지고 다음단계인 특징 계산 단계로 넘어간다.

### 2.3 특징 계산

퍼지 모델은 시간에 대해 비독립적인 데이터를 가지고 동작하기 때문에 음성 신호에서 각각의 파라미터 값들은 직접적으로 퍼지 분류기에 의해 처리될 수 없다. 따라서 특징 계산 단계에서는 통계 분석에 의해 일련의 순서대로 정렬된 파라미터들을 요약한다. 또한 외부 노이즈를 제거하기 위해 기본 주파수에 대해서 계산된 데이터들을 Smoothing 작업이 필요하다. 이는 불충분한 데이터에서 출력 확률 분포

값이 영이나, 거의 영에 가까운 값을 가지게 되어 인식률의 저하를 방지하기 위함이다[5]. 이런 파라미터 값들로부터 기본 주파수와 에너지의 평균값과 편차를 계산한다. 또한 노이즈가 제거된 파라미터들은 통계적인 정보를 결정하는데 이용된다.

### 2.4 퍼지 모델 생성

학습 샘플들로부터 추출된 특징들을 가지고 퍼지 모델 생성기는 각 샘플들과 연관된 감정들을 얻어낸다. 이 데이터로부터 모든 특징들에 대한 소속 함수들을 계산한다. 그리고 각 감정에 대한 n개의 최적 특징들이 선택되어지는데 보통 n은 4에서 6사이의 값으로 정해진다. 마지막 단계로서 각 감정별로 나뉜 퍼지 룰(rule) 시스템이 생성된다.

### 2.5 퍼지 분류기

인식 모드로 동작하면 이미 생성된 5개의 퍼지 룰(rule) 시스템들은 각 5개의 감정들이 속해 있는 실제 음성 샘플들을 분류하게 된다. 이러한 목적으로 각 퍼지 감정 모델은 학습하는 동안에 각각의 감정에 대해 선택된 관련 특징들을 얻는다. 각 감정에 대해 계산된 특징들은 이전 시스템에 의해 비교되고 가장 강한 감정이 최종 인식된 값으로 남게 된다. 그림 2는 퍼지 분류기의 구조를 보여준다.

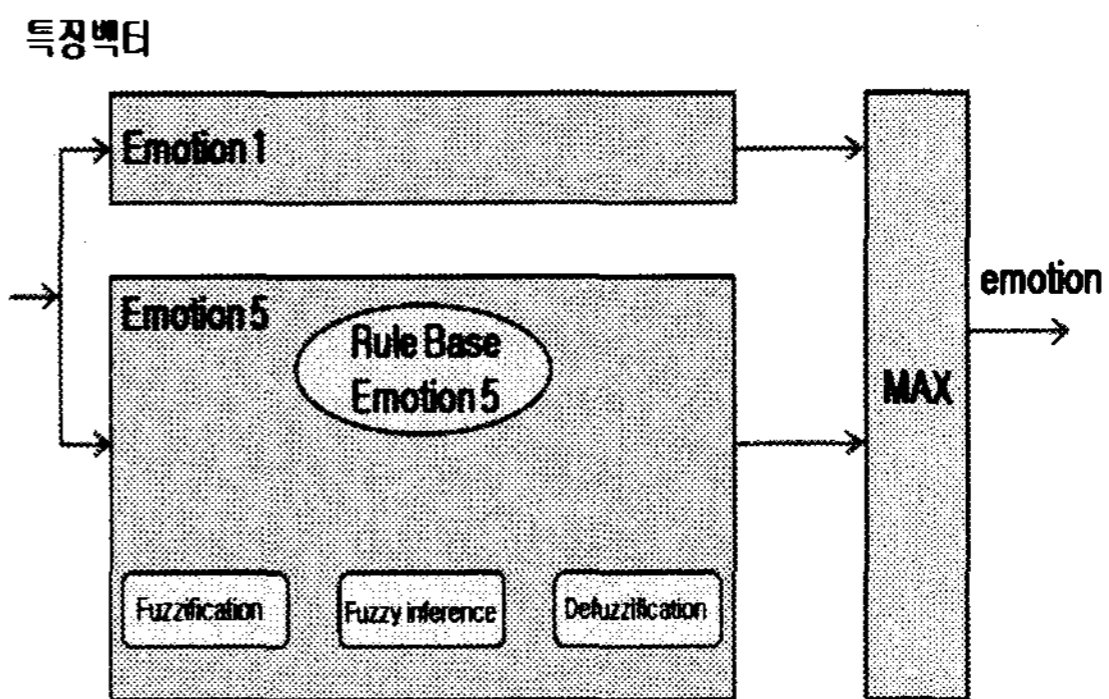


그림 2. 퍼지 분류기의 기본 구조

## 3. 퍼지 모델 자동 생성

퍼지 모델은 입출력 변수들의 퍼지 집합을 나타내는 소속 함수와 이 변수들 사이의 관계를 나타내는 퍼지룰(rule) 시스템으로 구성된다. 입력은 서로 다른 특징들을 의미하고, 출력은 감정들을 나타낸다. 계산량을 줄이기 위해서 간단한 triangular 와 trapezoidal 소속 함수를 선택했고, 특징과 감정을 very-low, low, medium, high, very-high의 5가지 단계로 구분하였다. 그림 3은 소속 함수의 원래 모양을 보여준다. 그것들의 정확한 모양은 학습 샘플

에 따라 다르게 되고 특징들과 감정들에 따라서 자동적으로 결정된다.

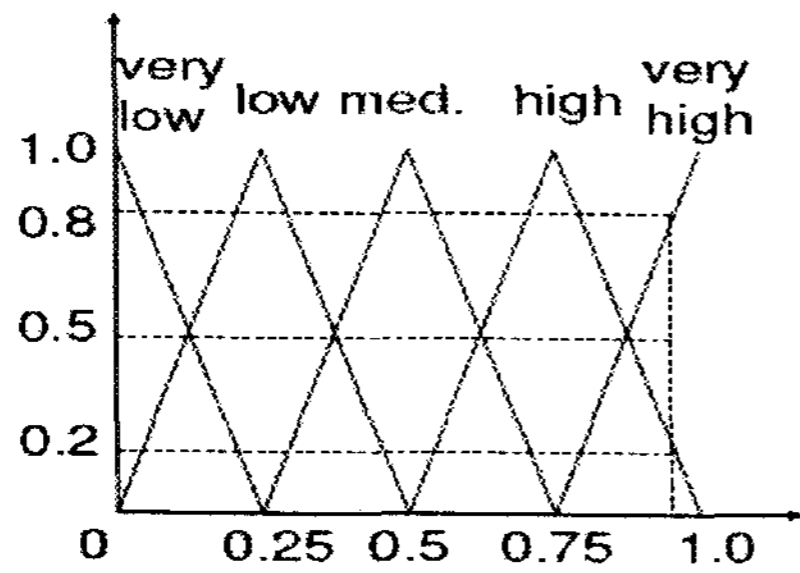


그림 3. 감정별 소속 함수의 기본 모양

학습 알고리즘은 Fuzzy Grid 알고리즘을 택했다. 이 알고리즘은 세 단계로 구성되어 있는데, 첫 번째로 각 감정에 대한 소속 함수가 만들어지고 각 감정에 대한 최적 특징들이 선택되어지고, 마지막으로 알고리즘은 각 감정에 대한 퍼지 룰(rule) 시스템을 생성한다.

### 3.1 소속 함수 생성

모든 소속 함수의 중심을 계산한 이후에는 각 소속 함수의 시작점이 자신의 왼쪽에 이웃한 함수의 중심으로 결정된다. 또한 소속 함수의 끝점은 자신의 오른쪽에 이웃한 함수의 중심으로 결정된다. 이 룰(rule)에는 두 가지의 예외를 두는데 very-low term 소속 함수는 값이 1인 채로 0에서 시작하며 very-high term의 소속 함수는 자신의 중심에서부터 양의 무한대까지 1을 유지한다. 그림 4의 오른쪽에 소속 함수들이 나열되어 있다.

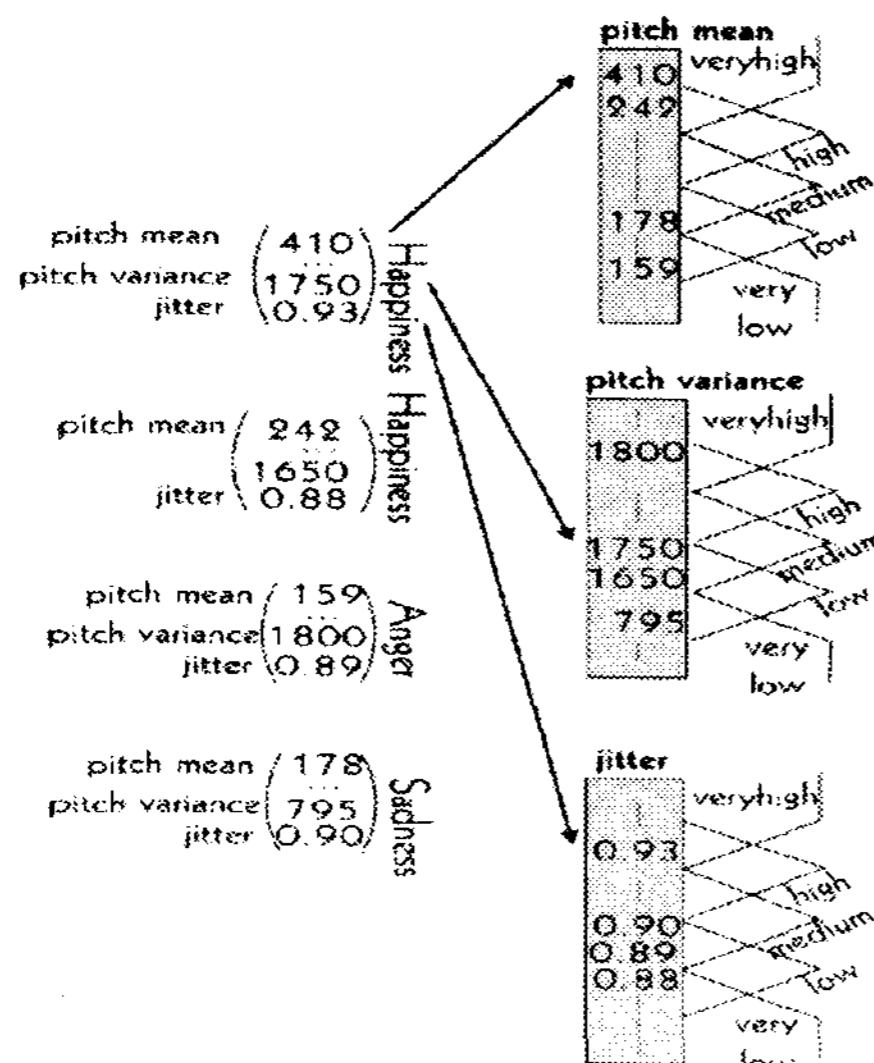


그림 4. 소속 함수 생성

### 3.2 특징 선택

특징 선택은 각 감정에 대해 개별적으로 수행된다. 그 과정은 모든 감정과 특징들에 대한 히스토그램을 생성하면서 시작한다. 이는 하나의 감정 학습 샘플에서 특징 값이 매우 낮음, 낮음 등으로 분류되는 빈도수를 계산하게 된다. 따라서 행복한 감정과 피치 평균에 대한 히스토그램은 very-low, low, medium, high, very-high에 대한 피치 평균을 가진 학습 샘플(행복)들이 얼마나 많은 지에 대한 정보를 포함한다. 히스토그램 상에서 특정 특징에 대한 학습 샘플의 범주는 알고리즘의 첫 번째 단계에서 제공되는 소속 함수를 적용시킴으로써 계산되고 학습 샘플은 소속 정도가 가장 높은 범주로 속하게 된다.

그런데 여기서 가장 높은 소속 값을 가지는 범주만 증가하는 것이 아니라 0보다 더 큰 소속 정도를 가지는 범주에 해당하는 모든 값들도 증가한다. 이렇게 계산되어야 하는 이유는 가장 높은 소속 값을 가지는 범주만 증가시키게 되면 다른 범주에 속하게 되는 명확한 정보들을 버리기 때문이다. 특별히 소속 값이 가장 높은 것과 그 다음으로 높은 것의 차이가 매우 작은 샘플들이 많이 있을 때 이러한 방식은 가장 높은 소속 정도를 가지는 범주만을 선택하는 것보다 더 좋은 결과를 얻게 될 것으로 예상된다. 히스토그램에 모든 학습 샘플들이 입력된 후 하나의 감정에 대해서 가장 중요한 특징이 다음과 같이 선택될 수 있다. 적은 정보만을 포함하는 특징들은 모든 범주에서 학습 샘플과 거의 같은 수만큼 히스토그램들에 의해 표현된다. 감정을 구분하는 좋은 특징들은 뚜렷하게 높은 범주를 나타내는 히스토그램들에 의해서 표현된다. 이것은 그림4에서 보여준다. 왼쪽 히스토그램은 감정 e에 대한 분포가 훨씬 좋은 정보를 포함하는 특징을 보여준다. 오른쪽 히스토그램은 다른 감정들로부터 감정 e를 구분하는데 유용한 정보들을 가지지 못한 특징을 보여준다. 따라서 특징들 특성을 나타내는 방정식  $Q_f$ 는 (1)과 같이 나타낼 수 있다. 여기서  $h_{fev}$ 는 특징 f, 감정 e와 퍼지 성분 v에 대응하는 히스토그램 범주의 값이다.

$$Q_f = \frac{\max h_{fev}}{\sum_{v \in \text{terms}(f)} h_{fev}} \quad (1)$$

### 3.3 퍼지 룰(rule) 구성

특징 선택에 의해 선택된 특징들은 퍼지 룰(rule)을 만드는 세 번째 단계에 이용된다. 이 퍼지 룰(rule)은 선택된 특징들의 소속 함수와 함께 감정의 퍼지 모델을 구성한다. 룰(rule)

생성기는 또한 각각의 감정  $e_i, i=1, \dots, 5$  대해 서로 따로 수행된다. 각 룰(rule)은 퍼지화된 특징들  $f, j=1, \dots, n_i, 4 \leq n_i \leq 6$ , 을 입력으로써 취하고 출력으로 퍼지 감정 값  $e_i$ 를 산출해낸다. 감정  $e_i$ 에 대한 룰(rule)은 다음과 같은 형태를 가진다.

*IF  $f_{1_i}$  IS verylow AND ...  $f_{n_i}$  IS verylow THEN  $e_i$  is veryhigh*

*IF  $f_{1_i}$  IS verylow AND ...  $f_{n_i}$  IS low THEN  $e_i$  is veryhigh*

*IF  $f_{1_i}$  IS verylow AND ...  $f_{n_i}$  IS medium THEN  $e_i$  is medium*

...

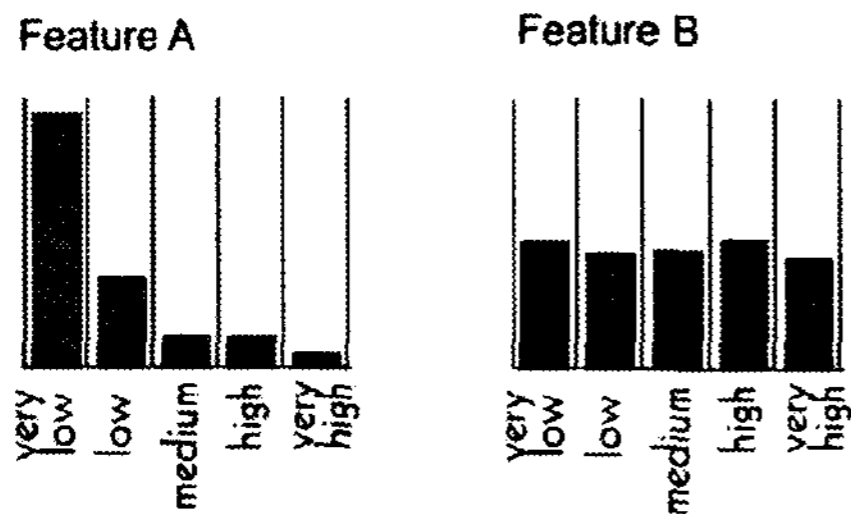


그림 5. 감정에 대한 특징 정보 비교

모든 룰(rule)들의 결론은 알고리즘의 두 번째 단계에 의해 만들어진 히스토그램들을 사용하여 생성된다. 각 룰(rule)에 대해서 룰을 구성하는 다른 성분에 대응하는 룰(rule)의 분포로써 해석될 수 있다. 모든 룰(rule)의 분포를 계산한 후에 "감정은 very-low", "감정은 low" 등의 결론으로 범위를 구분 짓기 위해서 최고 높은 분포 값이 결정되어야만 한다. 그리고 최소 값과 최대 값 사이의 거리를 5개의 동등한 크기로 나누어서 가장 낮은 분포 값에 속하는 룰(rule)은 very-low의 결론으로 정해지고, 가장 높은 분포 값에 속하는 룰(rule)은 very-high의 결론으로 정해진다. 다른 퍼지 출력 값들은 그에 알맞게 정해진다. 룰(rule)의 결론을 계산하는 것은 그림과 같이 도표로 보여준다. 왼쪽에 다른 특징들에 대한 두 개의 히스토그램이 있고 오른쪽에 결합된 히스토그램은 두 특징들의 히스토그램 범주가 모두 결합된 분포 값을 보여준다. 원래의 Fuzzy Grid 알고리즘은 AND 결합에만 이용되기 때문에 감정  $e_i$ 에 대해서  $5^{n_i+1}$ 과 같이 엄청나게 많은 룰(rule)들이 생성된다. 따라서 알고리즘의 마지막 단계로써 퍼지 시스템이 동작 가능하게 하기 위한 룰(rule)의 적당한 수로 줄이기 위해서 OR결합을 이용하게 된다.

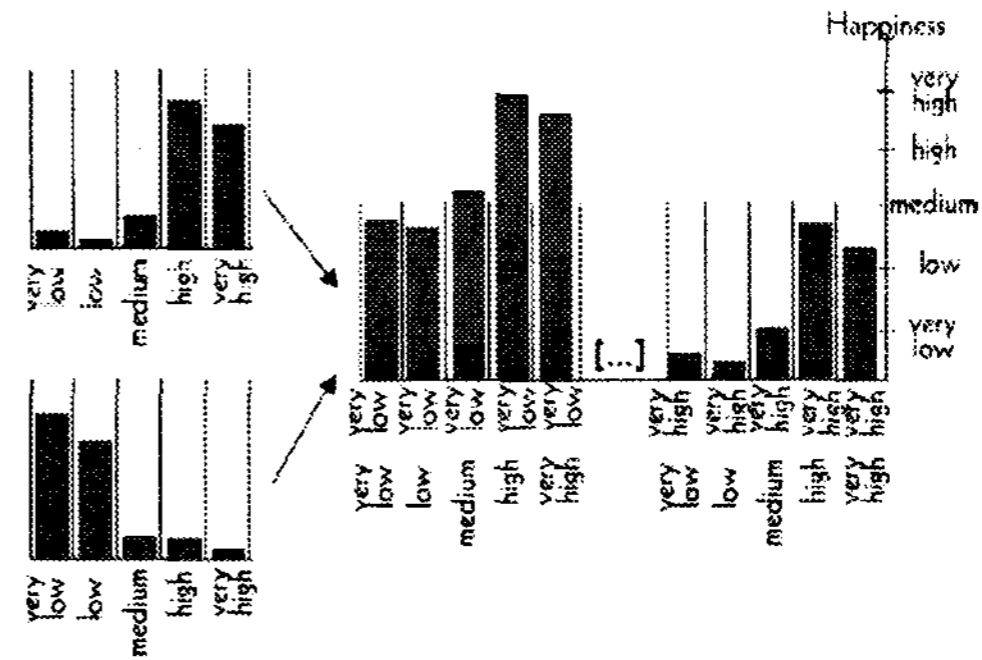


그림 6. 퍼지 룰(rule) 생성

### 4. 결론

본 논문에서는 접근법에 기반한 퍼지 모델을 활용하여 음성으로부터 감정을 인식하고 감성적인 운율로 음성 출력을 산출해 내는 시스템을 제안 하였다. 본 시스템은 20개의 음성 특징 집합들로부터 자동적으로 6개의 가장 중요한 특징을 선택하여 각 감정을 인식하기 위한 룰 시스템을 생성한다. 현재 휴머노이드 로봇에 관한 관심이 크게 늘어가고 있는 시점에서 본 논문에서 제안한 시스템은 로봇과 인간의 상호 감성적 교류에 대한 연구에 조금이나마 이바지 할 수 있을 것으로 생각된다.

### 참고 문헌

- [1] K. Scherer, "Vocal Communication of emotion: A review of research paradigms", *Speech Communications*, 40 227-256, 2003.
- [2] D. Ververidis and C. kotropoulos, "Emotional speech classification using Gaussian mixture models", *Proceedings of ISCAS*, vol. 3, pp. 2871-2874, May 2005.
- [3] C. M. Lee and S. S Narayanan, "Toward detecting emotions in spoken dialogs," *IEEE Transactions on Speech and Audio Processing*, vol. 13, pp. 293-303, March, 2005.
- [4] N. Amir, S. Ziv, and R. Cohen, "Characteristics of authentic anger in Hebrew speech", In *Proc. Eurospeech*, 713-716, 2003.
- [5] 김대수, 김종교, "퍼지 스무딩을 이용한 향상된 음성인식," *한국음향학회 학술대회*, 제 18권 제2(s)호, pp.13-16, 1999.