# 음향 파라미터에 의한 정서적 음성의 음질 분석

조철우(창원대), 리타오(창원대)

<차 례>



<Abstract>

## Analysis of the Voice Quality in Emotional Speech Using Acoustical Parameters

Cheolwoo Jo, Tao Li

The aim of this paper is to investigate some acoustical characteristics of the voice quality features from the emotional speech database. Six different parameters are measured and compared for 6 different emotions (normal, happiness, sadness, fear, anger, boredom) and from 6 different speakers. Inter-speaker variability and intra-speaker variability are measured. Some intra-speaker consistency of the parameter change across the emotions are observed, but inter-speaker consistency are not observed.

* Keywords: Voice quality, Emotional speech, Acoustical parameter.

# 1. 서 론

음성의 음질은 화자간의 특성을 구분 짓는 데 관계될 뿐 아니라 동일 화자의 음성 간에도 서로 다른 정서상태의 음성의 경우와 같이 다른 음색을 나타내는 근원이 된다. 합성음성을 다양하게 한다든지 정서상태를 인식하는데 필요한 정보로도 활용될 수 있다. 음질의 변화를 통해서 음성관련 질환을 식별하는 데도 활용될 수 있다. 음질에 관한 연구에는 음성의 발성근원인 음원에 관한 이해부터 시작해서 발성과정에 대한 광범위한 이해와 분석이 필요하다[1]. 음질에 대한 연구는 개념적인 면으로 연구가 많이 되어 왔으나 최근 음성 분석 기법의 발달로 수치적인 파라미터의 분석을 통해 음질을 규정하려는 연구가 많이 수행되고 있다[2][3][4]. 음질에 관한 객관적 연구를 위해서는 개념적인 음성학적 음질을 수치적 파라미터로 변환하는 과정에 대한 연구를 통하여 특정 음성현상을 관찰하고 일반화할 필요가 있다[5]. 음질을 표현하는 수치적 파라미터로서 다양한 특성을 측정하기 위한 수많은 파라미터가 지금까지 제안되어 있다[2][3][4]. 그러나 한국어의 경우 정서적 음성의 특성에 관한 연구결과가 발표된 사례가 많지 않고 특히·정서적 음성의 음질에 관한 연구는 전무하다고 볼 수 있다.

본 연구에서는 음성의 음질을 분석하기 위한 기초 연구의 일환으로서 6가지 다른 정서상태의 음성을 6가지 음질 관련 파라미터로 분석하고 동일 화자 및 화자간의 서로 다른 정서상태별 및 화자간의 음질특성을 비교하였다.

# 2. 연구 방법

본 연구에 사용된 정서음성 데이타베이스는 SiTEC의 정서음성 DB를 사용하였다[7]. 표 1의 화자정보를 가지는 6명의 화자가 발성한 낭독체, 기쁨, 슬픔, 화남, 공포, 지루함의 정서음성이 포함되어 있다. 각 정서음성은 10개의 서로 다른 문장으로 구성되어 있으며 16 kHz, 16 bit로 표본화 되어 있다.

<표 1> 발성화자정보

| 이름 | YJW | JHJ | CWJ | MYS | KKS | PYH |
|---|---|---|---|---|---|---|
| 나이 | 31 | 24 | 29 | 25 | 27 | 28 |
| 성별 | 남 | 여 | 남 | 여 | 여 | 남 |
| 키(cm) | 168 | 166 | 185 | 171 | 162 | 174 |
| 몸무게 | 60 | 53 | 104 | 49 | 48 | 83 |
| 흡연여부 | O | X | X | O | O | O |
| 배우경력 | 10년 | 3년 | 7년 | 6년 | 8년 | 10년 |

녹음된 문장은 다음과 같다.

(1) 난, 가지 말라고 하면서 문을 닫았어.
(2) 이건 내가 원하던 게 아니야.
(3) 야, 이제 그만하자.
(4) 정말 그렇단 말이야.
(5) 예.
(6) 아니오.
(7) 나도 몰라.
(8) 우리가 하는 일이 얼마나 중요한지 너는 모를 거야.
(9) 지금 어디 가는거야.
(10) 바람과 햇님이 서로 힘이 더 세다고 다투고 있을 때 한 나그네가 따뜻한 외투를 입고 걸어 왔습니다.

분석에 사용한 6가지 파라미터는 피치, 피치의 범위, 문장의 발성 길이, Jitter, Shimmer, NHR이다[2][6]. 각 파라미터에 대한 정의는 다음과 같다.

$$Jitter = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|To^{(i)} - To^{(i+1)}|}{\frac{1}{N}\sum_{i=1}^{N}To^{(i)}} \tag{1}$$

여기서 $To^{(i)}, i = 1,2...N$은 피치 주기 값이고, $N$은 추출된 주기의 수를 의미한다.

$$Shimmer = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|Ao^{(i)} - Ao^{(i+1)}|}{\frac{1}{N}\sum_{i=1}^{N}Ao^{(i)}} \tag{2}$$

여기서 $Ao^{(i)}, i = 1,2,..,N$은 각 주기에서의 신호의 진폭크기이고, $N$은 추출된 진폭의 수이다.

NHR은 주파수 범위 1500-4500 Hz 구간의 비하모닉 주파수 성분의 에너지와 70-4500Hz구간의 하모닉 주파수 성분의 에너지와의 비율로 구한다. 자세한 방법은 참고문헌[2][6]에 나타나 있다.

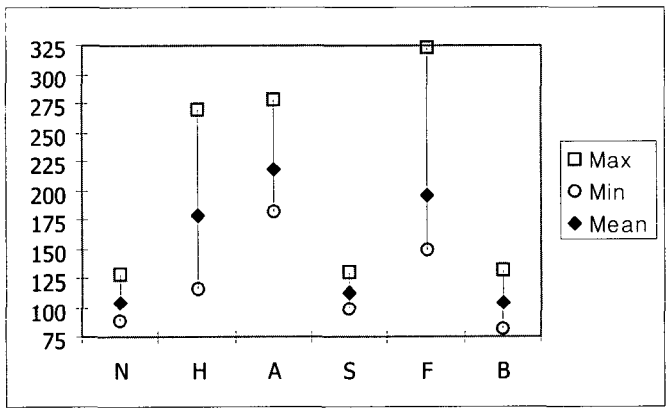
피치는 문장 전체에서 구한 피치 값의 최소, 최대, 평균 피치의 값을 측정하였으며 피치의 범위는 피치의 최대 값에서 최소 값을 뺀 값으로 구하였다. 문장의

발성 길이는 문장의 시작에서 끝날 때까지의 시간을 측정하였다. Jitter, Shimmer, NHR은 10 개의 문장 중에서 추출된 /아/ 음성을 사용하였다. 이렇게 단일 모음만을 추출하여 사용한 이유는 이들 파라미터의 경우 유성음의 경우에만 추출이 가능하기 때문이다.

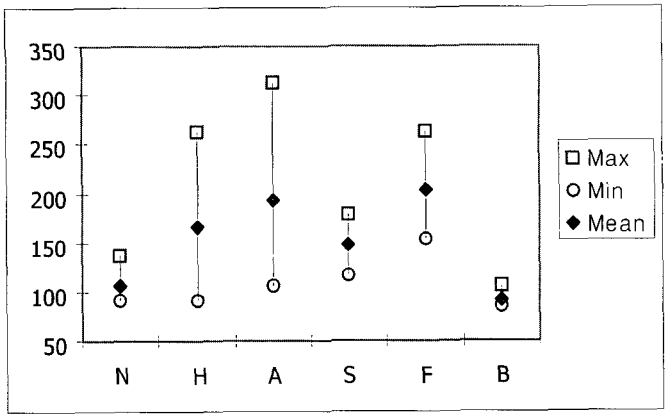
## 3. 분석 결과

### 3.1. 피치 및 피치의 범위

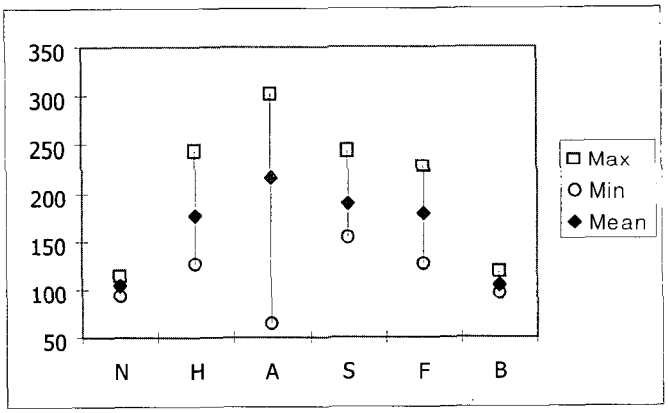
피치자료 분석 결과는 <그림 1>에서 <그림 6>과 같다. <그림 1>에서 <그림 3>은 남성화자의 정서상태별 피치 값의 범위이고, <그림 4>에서 <그림 6>은 여성화자의 정서상태별 피치 값의 범위이다. 이들 그림에서 각 정서상태는 낭독체(N), 기쁨(H), 슬픔(S), 화남(A), 공포(F), 지루함(F)으로 표기되어 있다. 행복, 화냄의 경우는 한명을 제외하고는 일관성 있게 피치 값의 평균이 증가하는 것이 관찰되었다. 화냄의 경우 두 배 가까이 증가하고 있음을 알 수 있다. 슬픔의 경우는 화자간 차이가 있지만 역시 피치의 평균값의 증가를 볼 수 있다. 지루함의 경우는 피치 값이 정상의 경우보다 약간 내려가는 현상을 보였다.



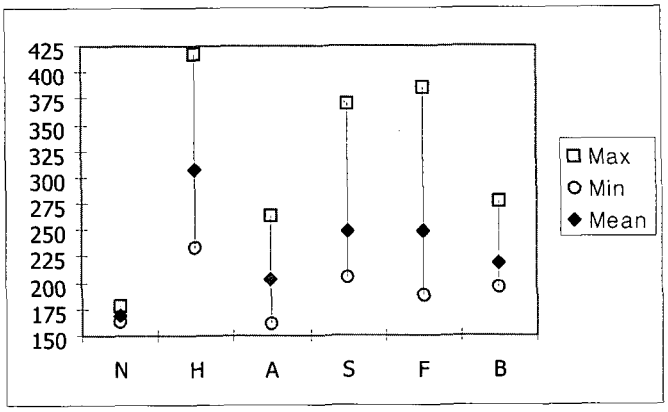

<그림 1> 화자 cwj의 피치 값의 분포



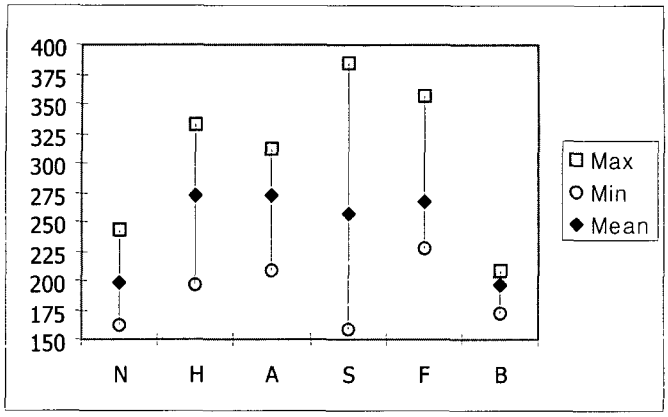

<그림 2> 화자 ysw의 피치 값의 분포



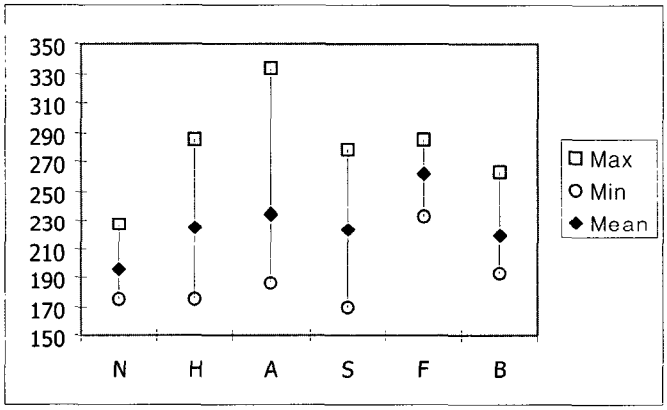

<그림 3> 화자 pith의 피치 값의 분포




<그림 4> 화자kks의 피치 값의 분포

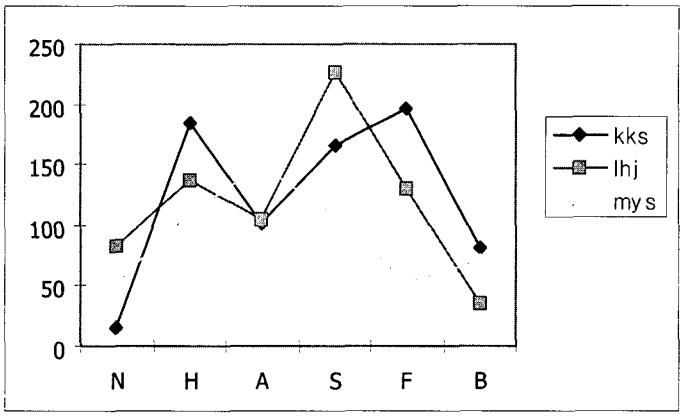



<그림 5> 화자lhj의 피치 값의 분포

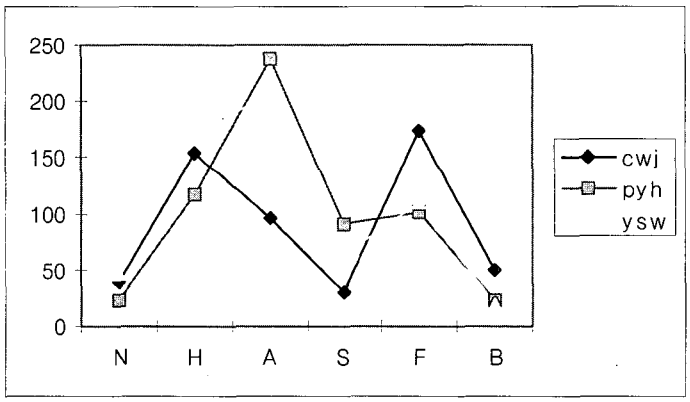



<그림 6> 화자mys의 피치 값의 분포

<그림 7>, <그림 8>은 정서상태에 따른 피치 변동 폭의 변화를 나타낸 것이다. 정상상태에 비해서 정서상태는 모두 변동 폭이 크게 나타나고 있다. 그러나 정서상태에 따른 화자별 변동 폭의 변화는 일관성이 관찰되지 않는다.



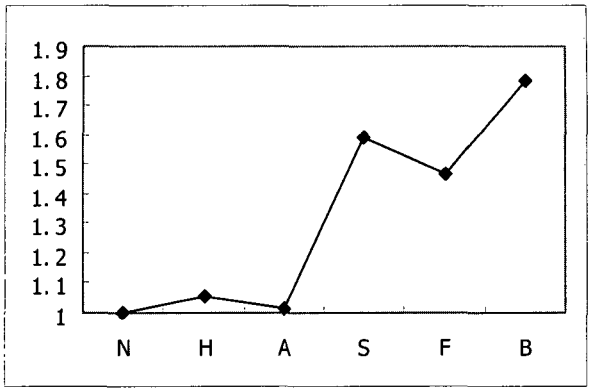

<그림 7> 여성화자의 정서상태별 피치 변동 폭의 변화




<그림 8> 남성화자의 정서상태별 피치 변동 폭의 변화

## 3.2. 발성지속시간

<그림 9>는 화자cwj의 문장별 발성지속시간의 변화를 정상음성을 기준으로 했을 경우 정서 상태에 따른 변화를 비교한 것이다. 지속시간은 낭독체 문장을 기준으로 다른 정서 상태에서의 문장의 길이의 상대적인 길고 짧음을 표시하였다. 각 문장에 대한 상대적인 지속시간의 변화의 평균 값을 나타낸 그래프에서 전반적으로 발성지속시간이 긴 순서는 지루함, 슬픔, 공포, 화냄의 순서로 관찰되었다.

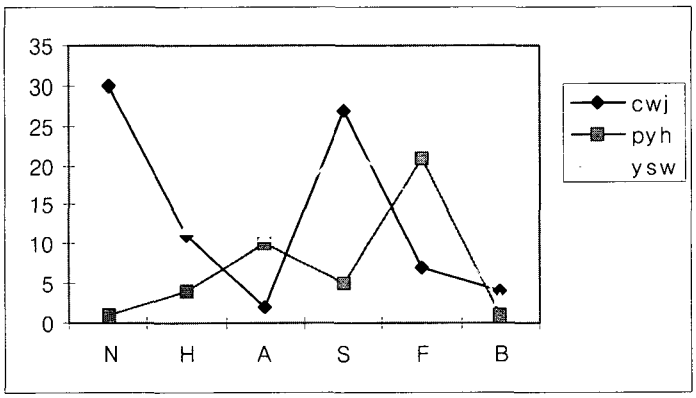



<그림 9> 화자cwj의 정서별 평균 발성지속시간의 변화

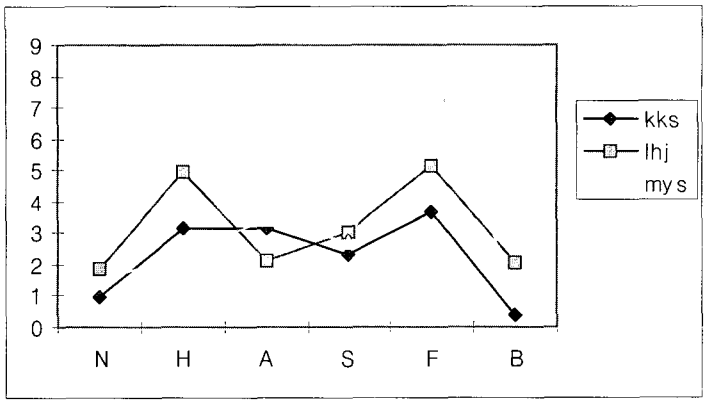
### 3.3. Jitter, Shimmer, NHR

Jitter는 피치 값의 변동률을 말하며 발성과정에 피치 값의 변동이 얼마나 심한지를 나타낸다. <그림 10>, <그림 11>은 Jitter값의 정서의 종류에 따른 평균치의 변화를 보인 것이다. 화자별 대체로 정상음성이 가장 큰 값을 갖고 이에 비해 행복과 화냄의 경우는 작은 값이 관찰되며 슬픔과 공포의 경우 상대적으로 큰 값이, 지루함의 경우 가장 작은 값이 관찰된다.

<그림 10>에서 남성화자의 Jitter값의 경우는 일관성을 찾을 수가 없다. <그림 11>에서 여성화자간 Jitter값의 변화에서는 정상음성의 경우 남성화자와는 다르게 낮은 값을 보였다.



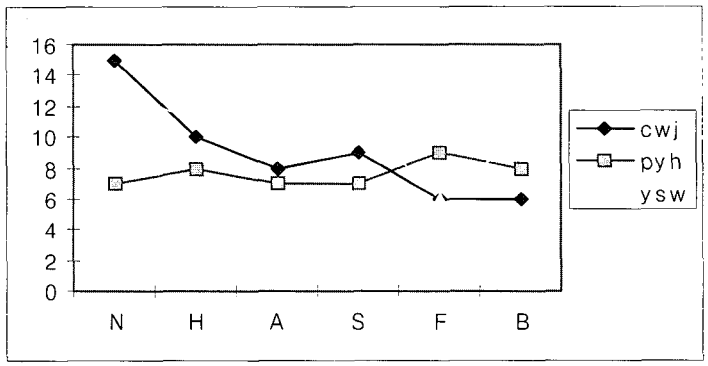

<그림 10> 남성화자의 Jitter값의 변화



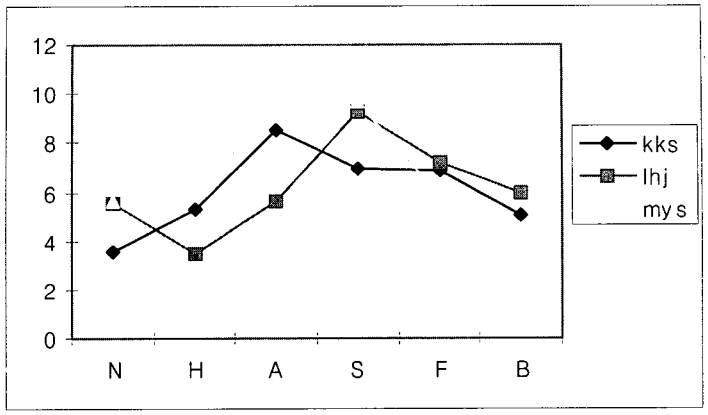

<그림 11> 여성화자간 Jitter값의 변화

Shimmer는 음성의 진폭 변동률을 측정하는 파라미터이다. <그림 12>에서 남성화자간 Shimmer값은 정상음성에서 가장 큰 값을 보이고 있으며 행복, 화냄, 슬픔에서 비슷한 값을 보이고 있으며 공포와 지루함에서 가장 낮은 값을 보였다.

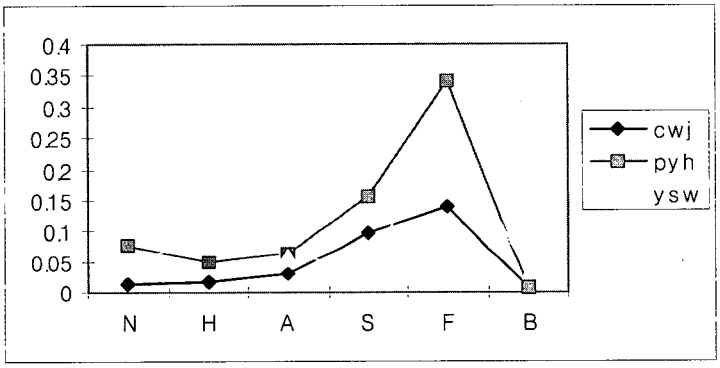
<그림 13>에서 여성화자의 Shimmer값은 화냄과 슬픔의 경우 가장 큰 경향을 보였다. 행복과 공포의 경우는 중간정도의 값의 분포를 보였고 지루함의 경우는 정상음성의 경우와 유사한 분포를 보였다.



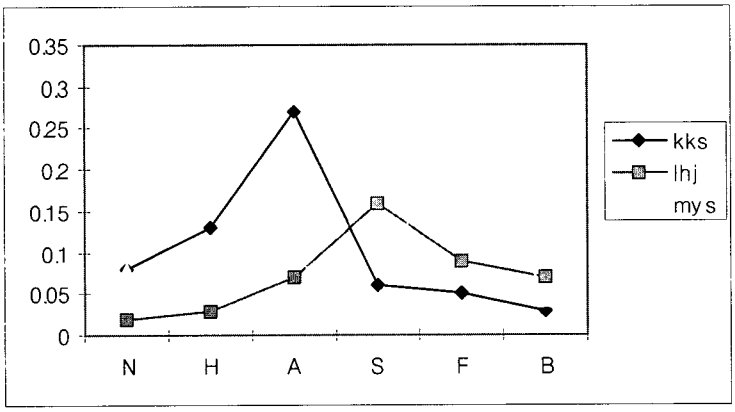

<그림 12> 남성화자의 Shimmer값의 변화




<그림 13> 여성화자의 Shimmer값의 변화




<그림 14> 남성화자의 NHR값의 변화




<그림 15> 여성화자의 NHR의 변화.

NHR은 잡음성분의 비율이 얼마나 많은가를 측정하는 파라미터이다. <그림

14>에서 남성화자간의 NHR값은 공포의 경우 압도적으로 큰 값을 보였는데 이는 음성이 기식 화에 의해 잡음성분이 증가하는데 기인한 것으로 예측할 수 있다. 슬픔의 경우 두 번째로 NHR의 값이 크게 나타났으며 지루함의 경우는 가장 낮은 NHR 값을 보였다. <그림 15>에서 여성화자간의 NHR값은 화자에 따라 일관성 있는 변화를 관찰할 수 없었다. 이는 남성화자의 경우와 대조된다.

## 4. 결 론

본 논문에서는 정서적 음성의 음질에 관한 음향적 특징을 조사하기 위하여 여섯 개의 파라미터에 관하여 분석하고 결과를 조사하였다. 피치, 피치변동 폭, 발성지속시간, Jitter, Shimmer, NHR에 관한 조사에서 동일 화자의 파라미터 변화는 어느 정도 일관성이 있었으나 서로 다른 정서상태에 따른 화자간의 파라미터의 변화는 일관성이 별로 관찰되지 않았다. 이것은 남성과 여성의 경우에도 다른 형태로 나타났다. 이러한 이유로 현재의 데이터를 이용하여 화자에 따른 정서현상의 음향적 표현에서의 일반적인 경향을 언급하기는 어려워 보인다.

이러한 현상의 주된 원인은 정서음성의 경우는 정서상태의 일관성이 결여되었기 때문으로 추정할 수 있는데 정서상태가 유발되는 구체적 상황에 따라 다른 음성 상태로 표현되었기 때문으로 생각된다. 현재의 DB에서는 정서유발상황을 전적으로 배우의 생각에 맡겨 두었으므로 동일한 정서상태에 대하여 세부적인 정서상황이 기록되어 있지 않다. 따라서 앞으로 정서음성 DB의 수집에 있어서 정서적 상태를 세부적으로 정의할 수 있는 방법을 제안할 필요가 있다고 생각된다.

향후의 연구에서는 본 논문에서 제시한 데이터의 다양한 분석을 추가하면서 정서음성 수집에서의 정서상태를 제어하기 위한 방법을 모색할 필요가 있을 것으로 사료된다.

감사의 글

이 논문은 2004년도 창원대학교 연구비에 의하여 연구되었습니다.

## 참 고 문 헌

[1] J. Laver, The Phonetic Description of Voice Quality, *Cambridge University Press*, 1980.

[2] R. D. Kent and M. J. Ball, Voice Quality Measurement, pp.119-244, *Singular Thomson*

*Learning*, 2000.

[3] I. R. Murray, J. L. Arnott, "Toward the Simulation of emotion in Synthetic Speech: A Review of the Literature on Human Vocal Emotion", *Journal of the Acoustical Society of America*, Vol.93, No.2, pp.1097-1108, 1993

[4] G. Klasmeyer, W.F. Sendmeier, "Objective Voice Parameters to Characterize the Emotional Speech", *ICPhS '95-Stockholm*, Vol.1, pp.182-185, 1995.

[5] 조철우 외 2인, "정서정보의 변화에 따른 음성신호의 특성분석에 관한 연구", *한국음향학회지*, 16권 3호, pp.33-37, 1996.

[6] MDVP Manual, Multi-dimensional Voice Program Model 5105, *Kay Elemetrics*, 1993.

[7] 조철우, 박일서, 이용주, 김봉완, "배우에 의한 한국어 정서음성 데이터베이스 수집", *한국음향학회 춘계학술대회논문집*, pp.45-48, 2004.

접수일자 : 2005년 5월 21일
게재결정 : 2005년 9월 22일

▶ 조철우(Cheol-Woo Jo)
주소: 641-773 경상남도 창원시 사림동 9 창원대학교
소속: 메카트로닉스공학부 음성 및 음향신호처리 실험실
전화: 055) 279-7552
E-mail: cwjo@changwon.ac.kr

▶ 리타오(Tao Li)
주소: 641-773 경상남도 창원시 사림동 9 창원대학교
소속: 메카트로닉스공학부 음성 및 음향신호처리 실험실
전화: 055) 279-7550
E-mail: litao2000@hotmail.com