

감정변화에 따른 음성정보 분석에 관한 연구

장인창*, 박미경, 김태수, 박면웅(KIST CAD/CAM 연구센터)

Study of Emotion in Speech

I. C. Jang, M. K. Park, T. S. Kim, M. W. Park (CAD/CAM Research Center, KIST)

ABSTRACT

Recognizing emotion in speech is required lots of spoken language corpus not only at the different emotional statues, but also in individual languages. In this paper, we focused on the changes speech signals in different emotions. We compared the features of speech information like formant and pitch according to the 4 emotions (normal, happiness, sadness, anger). In Korean, pitch data on monophthongs changed in each emotion. Therefore we suggested the suitable analysis techniques using these features to recognize emotions in Korean.

Key Words : Speech signal (음성 신호), formant (포먼트), Pitch (피치), Recognizing Emotion (감정 인식)

1. 서론

인간의 감정을 인지하고, 그에 정서적인 반응을 하는 시스템의 개발은 보다 고차원적인 HCI(Human Computer Interface) 제품의 개발을 가능하게 한다.

인간의 감정정보는 얼굴표정, 음성, 몸동작, 심장 박동수, 체온, 혈압 등 다양한 방법으로 얻을 수 있고, 응용분야에 따라 감정 정보 취득 방법이 달라진다.

음성을 이용한 시스템의 경우 센서가 신체부위에 직접 접촉되지 않거나, 전화와 같이 반드시 음성을 이용하여야 하는 시스템의 적용에 유리하다. 미국 MIT대 미디어 랩의 Rosalind W. Picard 교수를 중심으로 진행되고 있는 감정적 착용 컴퓨터의 개발은 감정과 관련된 연구의 실용화 가능성을 보여준다. 또한 1999년 일본 SONY사가 개발하여 시판한 오락용 애완로봇 AIBO는 6가지 감정을 포함하는 감정 모델을 적용하여 주인과의 관계에 의해서 감정상태가 변화하고 반응하도록 개발되었다.

그러나 산업용 및 연구용 기준이 될 만한 감정 인식용 음성 DB 조차 구하기 힘든 실정이라서 영어로 된 DB와 연구 결과물이 있다고 하더라도, 우리 말은 영어와는 다른 음향, 음성학적 특징을 가지고 있으므로 국외에서 연구된 기법이 한국에서도 동일하게 적용 될 경우 이에 대한 성능 검증이 필요하

며, 한국어에 동일하게 적용되기 위해서는 적합한 독자적인 기술 개발이 시급히 요구된다.

본 논문에서는 한국어의 특성을 분석하기 위하여 모음 부분(7 단모음)에 레이블링(Labelling) 절차를 거쳐 감정의 변화와 관련된 모음부분에서 음성정보를 분석하였다. 모음 부분에서의 음성정보(Pitch, Formant 등) 분석 방법은 감정인식 시스템에서의 효과적인 특징추출이 이뤄질 수 있도록 시도되었다.

2. 음성 DB

2.1 음성 DB의 구성

음성 DB를 구성하기 위해서는 DB의 용도를 고려하여 감정 선정, 문장 선정, 녹음 대상 선정, 녹음 환경, DB 규모 등의 결정 작업이 필요하다.

대상 감정은 인간의 주요 감정인 기쁨, 슬픔, 화남의 3가지 감정과 이들의 기준이 되는 평상 감정을 포함한 4가지 감정으로 결정하였다. DB 제작 시 화자독립-문장독립 형 감정인식 시스템개발 및 테스트 용도의 DB를 제작한다는 목적 하에 문장 선정 시 다음과 같은 사항을 고려하였다.

- 1) 3가지 감정(기쁨, 슬픔, 화남) 상태로 발음하기에 용이한 문장.
- 2) 자연스런 감정 표현이 담긴 대화체 문장.

- 3) 전체적으로 우리말의 모든 음소를 고루 포함하도록 구성.
- 4) 법, 불협형 등 다양한 어법을 고려하도록 구성.

2.2 감정의 유도

DB 녹음에 참여할 화자에게 대본을 미리 전달하여 감정에 대한 사전 연습을 시켰고, 녹음 시에는 각 감정 별로 화자가 감정에 몰입 할 수 있는 시간을 할애 하였다.

또 감정에 적절한 상황을 주고 설명할 만한 보조 지문을 대상 문장의 전후에 삽입하여 감정을 유도 하였다. 이 방법은 다양한 문장에 대해서 즉각적인 감정 유발이 어려운 경우에 지문이 감정 연상 작용의 역할을 하였고, 동일 문장에 대해서도 대상 참여자마다 감정 기준이 다를 수 있는 상황을 최소화 시키는데 효과가 있었다.

참여한 화자들의 감정 유도를 보조하기 위하여 대본에 상황 설명을 위한 지문을 삽입 하였다. 지문은 녹음 문장 별로 평상을 제외한 기쁨, 슬픔, 화남 3 종류의 감정에 대하여 각각 작성하였다.

2.3 주관적 평가

구축된 DB 가 화자의 감정을 어느 정도로 정확히 반영하는 지를 판단하기 위해서 각 감정에 대한 주관적 평가를 실시하였다. 감정 별 인식률 및 오인식 패턴을 알 수 있는 주관적 평가에 대한 결과는 Table 1 과 같다.

Table 1 Subjective appraisal confusion matrix

	normal	happiness	sadness	anger
normal	93.9	3.1	8.9	4.1
happiness	26.6	57.8	3.5	12.0
sadness	6.4	0.6	92.2	0.8
anger	15.1	5.4	1.0	78.5
total	78.2			

Table 1 의 결과를 보면 각 감정의 인식률이 균일하지 않음을 보여준다.

슬픔은 인식률이 가장 높았고, 기쁨은 가장 낮았다. 오인식 패턴은 기쁨을 평상이나 화남으로, 화남을 평상으로 판단하는 경향이 두드러졌다. 이와 같은 결과는 평상시 감정 상태에 대한 주관적 평가자들의 정신적인 기준치가 다르기 때문에 나타났다고 볼 수 있다.

또한 각 감정 별 문장 인식률 실험을 통한 결과 인식률은 최저 71.1%, 최고 82.3%로 10%가 넘는 편차가 있었다. 이를 살펴보면 감정 중립적인 문장

이 있는 반면, 어느 감정으로 치우쳐 해석되기 쉬운 문장이 있음을 알 수 있다.

3. 음성의 특성 분석

음성을 통한 감정 상태의 파악을 위해서는 각각의 감정이 음성에 어떠한 변화를 만들어 내는가를 정확히 규명하여야 한다. 감정 상태를 나타내는 음성정보에는 피치(Pitch), 포먼트(Formant), 발음속도(Speaking rate) 등이 있다.

우리말에서의 자음과 모음의 특성을 잘 고려하여 감정의 변화에 의한 음성정보의 변화에 주목할 필요가 있다.

감정의 인지를 위해서는 음성에서 이러한 음성정보를 잘 반영하는 특징을 제대로 포착해내야 한다. 그러므로 본 연구에서 대상 감정인 평상, 기쁨, 슬픔, 화남을 인식하기 위한 parameter 로 피치, 포먼트를 분석해야 한다고 판단하였다.

음성신호 분석에 앞서 실험이 올바르게 진행되기 위해 감정변화에 따른 모음부분에서의 피치의 유의성을 평가하기 위해 실험 계획을 세워 ANOVA 분석을 실시하였다. 그 결과 유의수준 0.05 에서 유의한 결과를 나타냈다

- 귀무가설 : 성별 감정변화에 따른 발화자의 모음부분의 음성정보의 차이가 없다.
 - 대립가설 : 성별 감정변화에 따른 발화자의 모음 부분의 음성정보의 차이가 있다.
- 즉, 감정변화에 따른 발화자의 모음부분에서 피치 정보가 유의하게 변화된다는 결과를 얻었다.

4. 음성 DB 의 음향적 특징 분석 결과

음성 DB 의 피치변화, 피치범위, 피치최대값에 대한 분석 결론은 Table 2 와 같다.

Table 2 Characteristics of the different emotions

	pitch. var	pitch. ran	Pitch. Max
normal	slow	narrow	low
happiness	normal	wide	high
sadness	fast	narrow	low
anger	fast	very wide	very high

위의 특징 비교 실험 결과의 가장 큰 특징은 기쁘거나 화난 음성인 경우에는 전체적으로 피치와 에너지가 높고, 발음 속도가 빠른 반면, 슬픔과 평상 감정 상태의 음성은 전체적으로 피치와 에너지가 낮고 발음 속도가 느리다는 것을 알 수 있다.

본 논문에서는 각 감정의 변화에 따른 음성정보 값이 모음 부분에서의 유의한 변화를 바탕으로 각 감정 별 피치 값의 패턴과 분포를 비교하였다.

Fig. 1, Fig. 2 에서는 “좋아요 함께 가시죠” 와 “여기 앉아봐” 문장에 대한 각 감정 별, 성별 피치 값의 차이를 나타내고 있다.

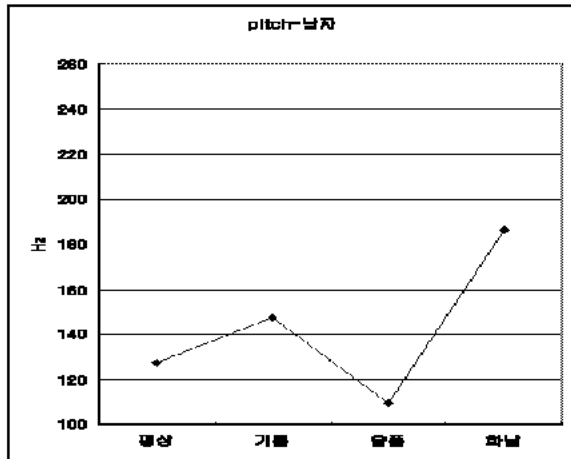


Fig. 1 Comparison of measured pitch (man)

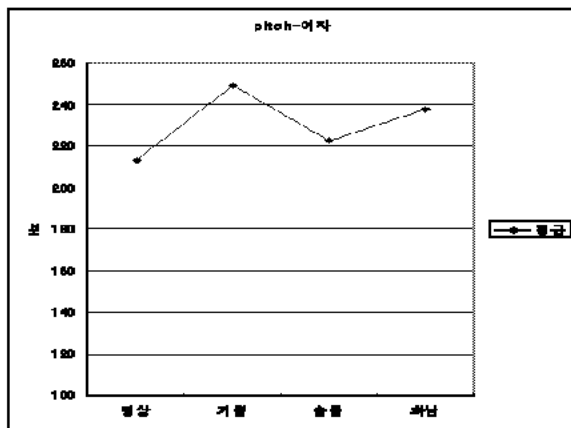


Fig. 2 Comparison of measured pitch (woman)

Fig. 1, Fig. 2 의 피치 값의 변화는 각 감정 별, 성별 확연하게 차이가 나는 것을 알 수 있다. 그러나 기쁨과 화남, 평상과 슬픔 감정상태의 피치는 유사한 패턴을 지내고 있다는 것을 확인 할 수 있었다. 이 결과는 감정인식을 시도할 경우 오인식 패턴이 발생하게 만드는 문제인 것이다.

Fig. 3, Fig. 4 에 나타난 피치의 분포에서는 감정 별 고유의 대역을 가지고 있다고 확인되었다. 앞의 Table 2 의 피치 범위와 거의 일치하는 분포를 나타내고 있지만, 기쁨과 화남 감정에서는 분포대역의 차이가 확연하지 않음을 확인 할 수 있었다

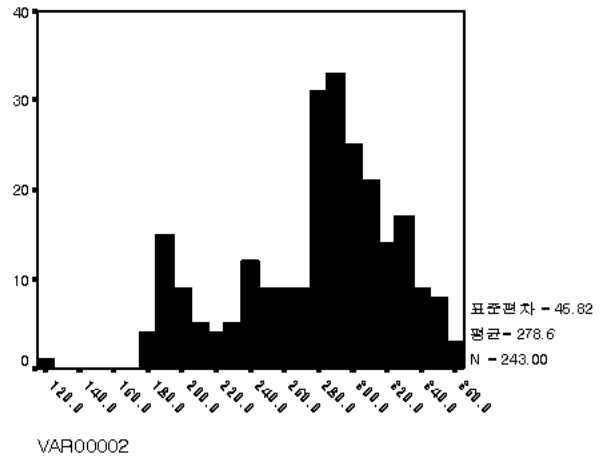


Fig. 3 Histogram of pitch distribution (Happiness)

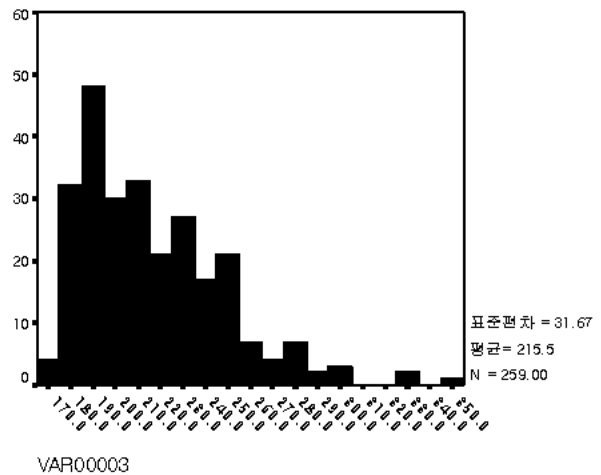


Fig. 4 Histogram of pitch distribution (Sadness)

본 연구에서는 피치 이외의 Parameter 인 포먼트 정보를 모음 부분에서 분석해 보았다. 그러나 모든 모음들을 분석해야 하는 어려움이 있기 때문에 감정을 가장 잘 나타내는 단모음을 선정하였다. 한국어의 단모음(7 단모음)은 /아, 어, 에, 이, 우, 오, 으/로 구성되며 이 중 고모음에 속하는 /우, 오, 으/ 는 모음의 길이가 짧고, F2(제 2 포먼트)의 변동이 심하기 때문에 연구대상에서 제외시켰다. 또한 고모음이지만 비교적 관찰이 용이한 /이/ 모음의 경우 생략된 경우가 많아 데이터가 충분하지 않으므로 연구 대상에서 제외 되었다.

포먼트는 /아, 어, 에/ 모음 부분에서 각 감정 별 분석을 실시 하였다. 포먼트가 나타내는 지표중 F3, F4 는 화자의 특성을 잘 나타내는 포먼트 이기 때문에 F1, F2 에 비해서 특히 관심을 가지고 분석 할 필요가 있다.

Fig. 5, Fig. 6 에서는 /아/, /어/ 모음에서의 각 감정 별 포먼트 정보를 분석한 결과이다.

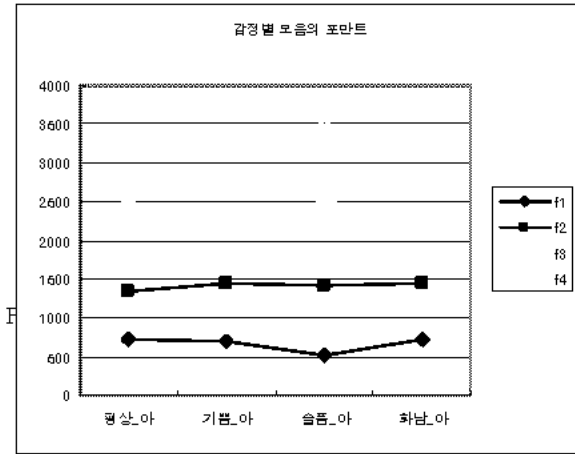


Fig. 5 Formant frequencies of /a/ in 4 different emotions.

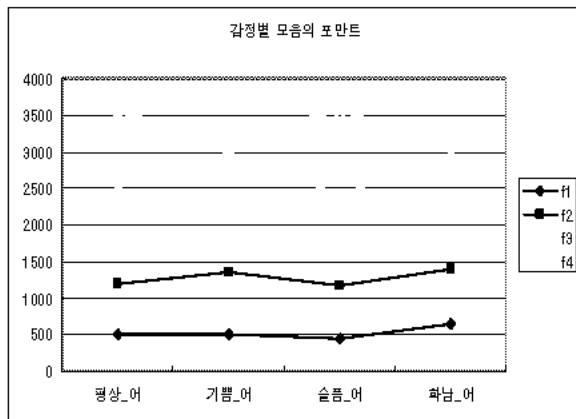


Fig. 6 Formant frequencies of /e/ in 4 different emotions

Fig. 5 의 분석 결과에서는 감정의 변화에 따라 나타나는 F1 과 F2 의 변화에 비해 F3 와 F4 의 변화 패턴이 유사하게 나타나고 있다.

Fig. 6 의 분석 결과에서는 모든 포먼트가 감정의 변화에 따라 일관성 있게 변화하고 있으며, 기쁨, 화남의 감정 상태에서 포먼트가 가파르게 올라가며 슬픔감정에서 가파르게 내려감을 확인 할 수 있다.

즉, 앞에서 설명한 감정을 가장 잘 나타내는 모음에 대한 분석결과를 직접 확인 할 수 있었으며, 그 결과는 모음 별, 감정 별 F3, F4 의 변화 패턴의 유사함과 고유의 주파수대역을 가진다는 것을 알 수 있었다. 그러나 기쁨과 화남감정의 결과 값은 유사하게 나타남을 재 확인 할 수 있었다.

5. 결론

본 연구에서는 4 가지 감정 상태가 표현되는 음성 DB 를 분석하여 표준어를 사용하는 20 대 남자, 여자의 각 감정 별 나타나는 특징들을 비교하였다.

감정의 변화에 따라 모음 부분에서의 피치, 포먼트 등의 음성정보 값의 변화를 확인 하였으며, 감정을 가장 잘 반영하는 모음 또한 확인 할 수 있었다.

본 연구 결과는 감정인식의 시스템에 있어서 음성신호 parameter 의 특징을 추출하는 session 에 있어서 분석방법론을 제안 하였다. 또한 기준 모델이 될 수 있는 codebook 과 입력 음성과의 유사도 측정에 있어서 모음 부분의 유의한 차이를 이용하여 인식률을 높일 수 있는 분석 방법론을 제안하였다.

참고문헌

1. 강봉석, 연세대학교 전기전자공학과 석사 학위 논문 "음성신호를 이용한 문장 독립 감정인식 시스템", 2001
2. Frank Dellaert, Thomas Polzin and Alex Waibel. "Recognizing Emotion in Speech", ICSLP '96, pp. 1970-1973, Philadelphia, Pa, USA, 1996
3. Roddy Cowie & Ellen Douglas-Cowie "Automatic Statistical Analysis of the Signal and Prosodic Signs of Emotion in Speech", ICSLP '96, vol ,3, pp. 1989-1992, Philadelphia, Pa, USA, 1996
4. Klaus R. Scherer, D. Robert Ladd, and Kim E. A. Silverman. "Vocal cues to speaker affect: Testing Two Model", Journal of the Acoustic Society of America, Vol. 76, No. 5, 1984