

표준어와 방언간의 감정변화에 대한 음성적 연구

이연수*, 박용범**
단국대학교 정보아키텍처 연구실
e-mail: *myth4ys@naver.com
**codon62@gmail.com

An acoustic study of feeling between standard language and dialect

Yeon-Soo Lee*, Young-Beom Park*
*Information Architecture Lab. Dankook Univ

요 약

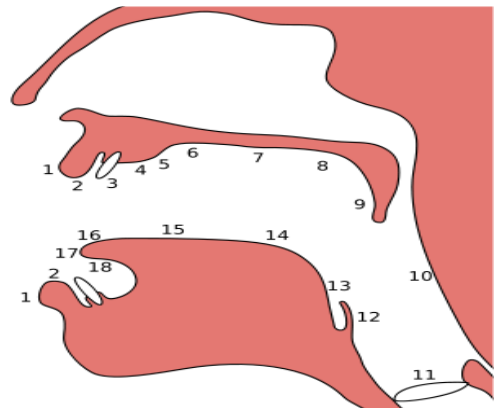
사람의 감정 변화에는 크게 기쁨, 슬픔, 흥분, 보통 4가지 상태로 말할 수 있다. 이 4가지 상태에서 기쁨과 슬픔, 흥분과 기쁨 상태가 음성학적으로 비슷한 형태를 가지고 있다. 흥분과 기쁨의 상태에서 방언의 노말 상태가 표준어의 기쁨, 흥분상태와 비슷한 특징을 가지고 있다. 이와 같은 표준어와 방언간의 특징 때문에 흥분 상태를 인지하는 경우 방언의 보통상태가 흥분상태로 잘못 인식되는 경우가 발생 한다. 본 논문에서는 이와 같은 문제점이 발생하는 음성학적인 차이를 구분 하고자 한다. 이들을 비교하기 위해 Pitch, Formant와 Formant RMS error 3가지 요소를 통하여 표준어와 방언간의 흥분 상태를 연구 하였다.

1. 서론

최근에 화자의 음성을 분석하여 감정의 변화를 인지하는 연구가 증가 하고 있다. 이와 같은 연구가 증가하는 이유는 감정 이라는 부분이 여러 곳에서 사용 될 수 있는 중요한 요소이기 때문이다[1][2]. 감정인지는 대화 중 감정 변화가 발생 될 때 사용 된다. 슬픔이나 흥분 상태 같은 음성은 음성 시스템에서 잘못 인지 할 때가 많기 때문에 시스템이 처리하기 전 사람의 조정을 필요로 한다. 하지만 시스템이 여러 화자간의 대화에서 성능적인 측면을 만족시키기 위해서는 신뢰할 수 있는 음성적 특징들을 알아야만 한다.[5]. 이미 음성의 특징을 이용하여 본래 음성에 감정적인 요소를 추가 하여 변형하기도 한다[6].

실제로 이 논문은 콜센터의 Agent와 Client 사이의 녹음된 음성을 분석하여 감정 변화가 발생한 음성을 찾는 프로젝트가운데 발생되었던 문제점을 연구한 것이다.

감정 변화는 슬픔, 흥분, 기쁨 그리고 보통으로 총 4가지 상태로 구분 할 수 있으며 Hidden Markov models(HMMs)을 이용하여 유사성을 살펴보면 보통과 슬픔, 흥분과 기쁨상태 2가지로 구분 지어진다[5]. 즉 흥분, 기쁨상태는 음성적 특징을 서로 공유하고 있다. 두 상태중에서 기쁨상태를 제외한 방언과 표준어 사이 흥분상태의 특징을 알아보게 된다. 방언과 표준어 사이에서는 보통상태의 방언이 표준어의 흥분상태로 인지된다. 본 논문은 감정 인식에 있어 한국어의 방언과 표준어 사이의 차이를 알아보고자 한다.

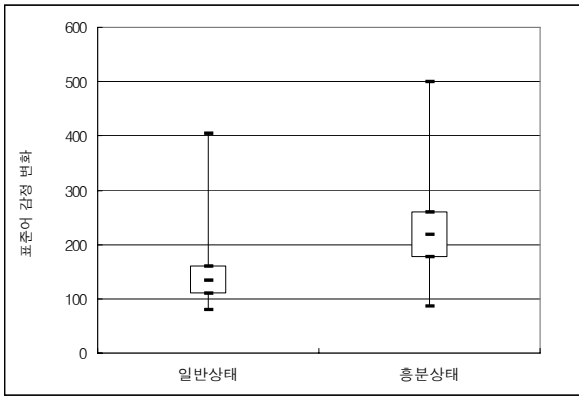


<그림 1> 조음기관

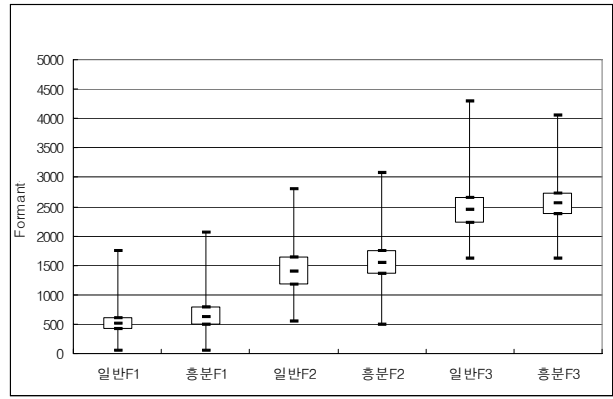
2. 음성 데이터

본 논문에서 사용된 음성 자료는 프로젝트를 진행 하면서 사용된 H사와 N사의 Agent와 Client사이의 녹음된 전화 음성을 사용 하였다. 전화 음성은 8Kz sampling rate, 16bit, Stereo(Agent좌측, Client우측) 형태로 녹음된 음성이다.

Agent는 표준어를 사용 하는 부분이라 제외되고 Client의 음성인 우측 채널만 사용 하였다. 화자의 성별은 남성을 선택 하였다. 방언은 많은 표본을 구할 수 있는 경상도와 전라도로 한정 하였다. 본 논문에서는 사투리와 표준어의 음성적 특징을 알기 위한 것이기 때문에 고객의 음성 데이터 중에서 많은 표본을 구할 수 있는 남성의 음성을 사용 하였다. 일반적으로 남성과 여성의 피치는 100Hz정도 차이가 난다. 언어적인 차이를 연구 하는 것이기 때문



(a)



(b)

<그림2> Box Plot을 통한 표준어 일반 상태와 흥분 상태 비교

에 많은 표본을 추출 할 수 있는 남성의 음성만 사용 하였다. 이는 피치(Pitch)나 포먼트(Formant) 분석에 있어서 보다 정확한 구분을 위해서 이다.

3. 음성의 측정요소

3.1 Pitch

Pitch는 Fundamental frequency라고 부르며 청각적으로 음의 높이의 높낮이로 생각 할 수 있다. Pitch는 사람의 감정과 정서의 변화에 따라 달라진다. 피치 값은 성대의 크기가 클수록 진동하는 속도가 느리기 때문에 여성보다 남성의 피치가 낮다.

3.2 Formant

Formant란 음성 신호를 주파수 영역으로 변환하여 주파수 에너지의 정점을 연결한 선들을 말한다. 낮은 주파수 정점부터 F1, F2.. 순으로 나아간다.

모음이 발생 할때 주파수 영역인 F1 ~ F3에 높은 에너지가 나타난다. 그 특징을 살펴보면 F1은 턱의 열림에 따라 에너지가 증가한다. F2는 <그림1> 14 ~ 18 부분의 혀의 위치에 따라 에너지가 변한다. F3는 입술의 둥근 정도에 따라 에너지가 변하며 <그림1>의 1, 2에 해당 한다.

4. 결과

4.1 표준어의 감정 변화

4.1.1 Pitch

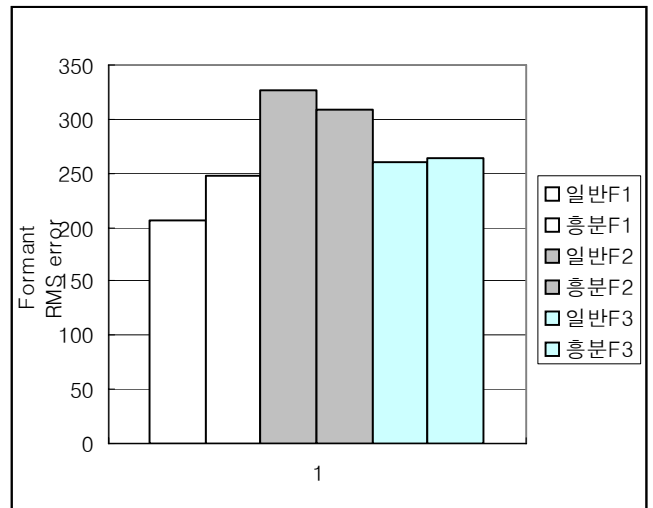
<그림2>는 화자의 일반상태와 흥분상태의 변화를 Median, 1 사분위, 3 사분위로 Box Plot 그림으로 나타낸 것이다. 보통상태는 화자의 감정 변화가 없는 상태를 말하고 흥분상태는 화자가 흥분하여 말하고 있는 상태를 말한다. 보통상태와 흥분상태는 Median값이 각각 134, 218에 있고, Median값 중심으로부터 1, 3사분위 값의 편차가 흥분 상태가 더 크다는 것을 알 수 있다. <그림2>를 통해서 Pitch의 값으로 표준어를 사용할 때 흥분 상태를 인지 할 수 있다.(단, 흥분과 기쁨상태는 서로 비슷한 영역대를

공유한다[5].)

4.1.2 Formant

<그림2>에서 보통상태와 흥분상태의 Formant의 차이는 일반상태 일때 Median값으로부터 편차가 크지 않다. 반면 흥분 상태의 경우 F1의 1 사분위의 값과 3 사분위 값이 넓게 분포 되어 있는 것을 볼 수 있다. 이것은 흥분상태일 때 턱을 크게 벌리는 것을 의미한다. F2, F3 에너지는 흥분상태일 때 높지만 편차는 일반 상태가 좀 더 크다. 흥분 상태일 때 F2,F3는 중용한 요소로 사용되기는 힘들다.

표준어로 대화 중일 때 Pitch의 변화와 F1값을 통해서 일반상태와 흥분상태를 구분 할 수 있다.

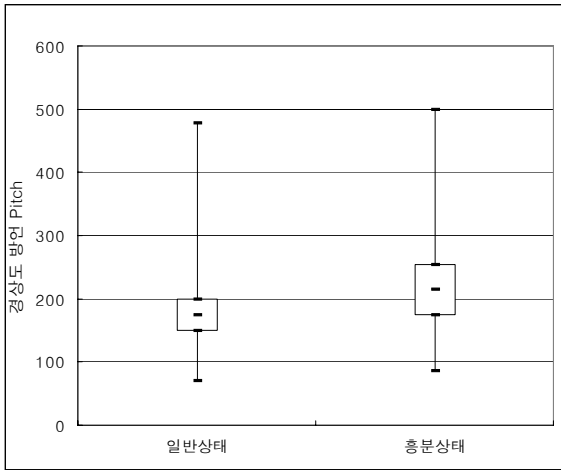


<그림3> 표준어 Formant RMS error

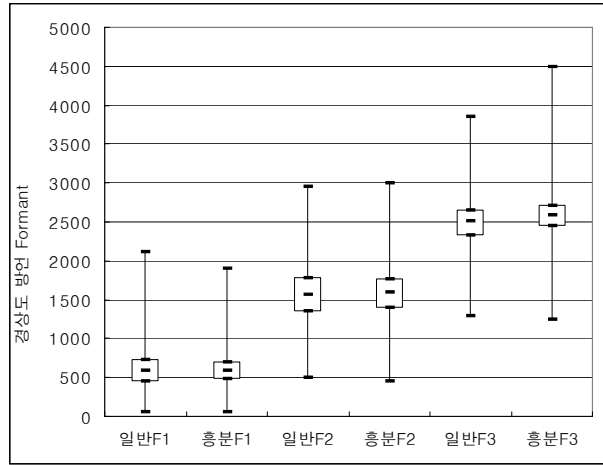
4.2 방언의 감정 변화

4.2.2 Pitch

경상도 방언의 감정 변화는 <그림3>과 같다. (a)의 일반과 흥분의 Median값은 160과 210 이다. 표준어의 흥분상태와 경상도 방언의 흥분 상태는 비슷한 Pitch의 형태를 보여준다. 경상도 방언의 일반 상태는 표준어 Pitch보



(a)



(b)

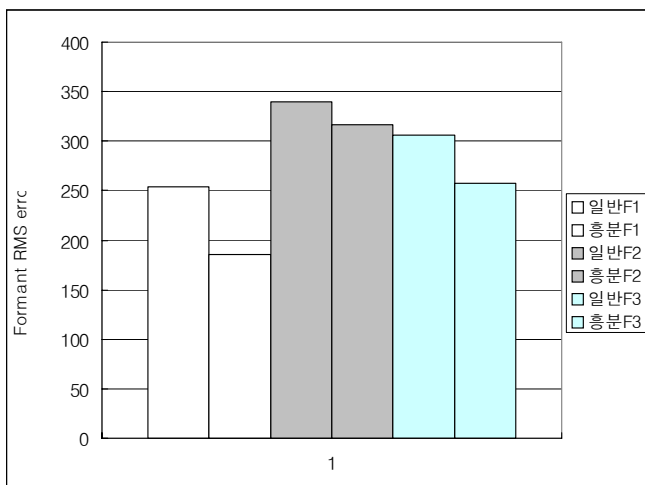
<그림4> 경상도 방언의 Box Plot

다 약 30정도 높은 Mideian값을 보여준다. 표준어와 방언 간의 혼합된 음성에서는 홍분된 음성으로 인식될 수 있다. 전라도의 경우도 마찬가지로 잘못된 인식을 할 수 있다.

4.2.3 Formant

Formant의 표준어 일반상태와 경상도 방언의 일반상태는 비슷한 F1를 보여준다. 반면 홍분상태는 확실하게 표준어와 구분된다. <그림4>의 (b)에서 F1는 700의 차이를 보여준다. 이 데이터 값은 경상도 사투리가 표준어보다 턱의 움직임이 크다는 것을 나타낸다. 결과적으로 F1의 값으로 표준어 홍분과 경상도 방언을 구분 할 수 있게 한다.

F2, F3는 표준어만큼 차이가 크지 않다. 이는 혀의 위치와 입의 둥근 정도의 편차가 적다는 것을 의미 한다. 반면 전라도의 경우는 표준어와 비슷한 형태를 보이고 있다. 전라도 표준의 경우 홍분 상태에 비해 낮은 것을 볼 수 있



<그림5> 경상도 방언 RMS error

다. 이것은 입모양의 움직임이 적은 것을 말해 준다.

4.2.3 RMS error

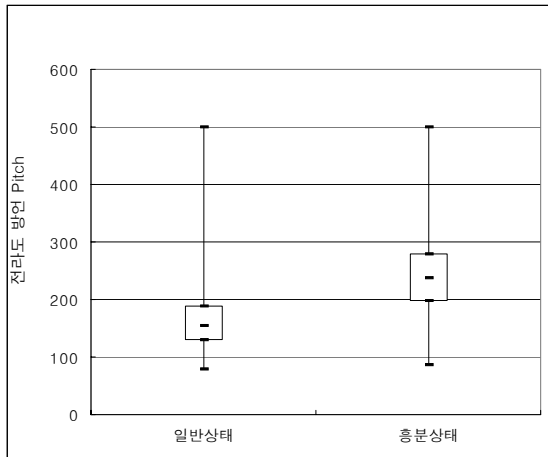
Formant의 RMS error는 <그림5>과 같다. 모든 F1 ~ F3의 Formant가 일반 상태일때 더 높은 결과가 나왔다. 일반 상태때 턱, 혀, 입의 둥근 정도가 좀 더 차이가 나는 것을 볼 수 있다. 반면 홍분 상태일 때는 그 오차가 좀 더 줄어드는 것을 볼 수 있다. 이와 같이 표준어와 대칭되어 나타나는 속성을 이용하여 표준어와 방언의 구분을 지을 수 있게 된다.

5. 결론

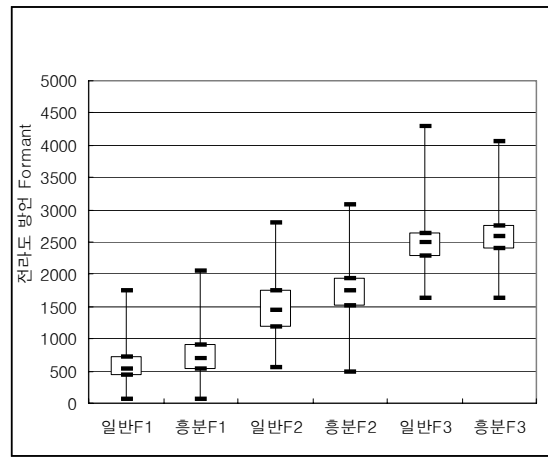
화자의 대화에서 표준어와 방언에 상관없이 Pitch와 F1을 통해서 일반상태와 홍분상태를 구분 할 수 있었다. 표준어와 전라도 방언의 홍분 상태가 나타내는 특징은 홍분할수록 Pitch와 F1의 값이 증가 하는 것을 알 수 있었다. 이는 기본 주파수인 Pitch가 커지고 턱이 많이 움직이는 것을 의미한다. 반면 표준어와 경상도 방언과의 Pitch, Median 값은 사분위 값에서 차이를 보인다. Formant RMS error에서는 일반 상태가 홍분 상태 일때 보다 더 높은 RMS error를 보여준다. 이 결과는 경상도 방언의 일반 상태의 대화에서 턱의 움직임 차이가 심하다는 것을 알 수 있다. 경상도 방언의 홍분 상태일 때는 일반 상태일 때 보다 턱의 움직임의 오차가 더 적은 것을 보여준다.

홍분 상태를 감지할 때 방언의 일반 상태 Pitch가 표준어 홍분 상태로 감지 될 수 있다. 방언의 일반 상태 Pitch Median 값은 표준어에 일반 3사분위 값 상단에 위치하고 홍분 상태는 표준어 홍분 상태 1사분위 값 아래쪽에 위치한다. 방언의 일반 상태 Pitch가 표준어와 홍분 상태와 중첩되어 있기 때문에 홍분 상태를 인식할 때 잘못된 인식을 보인다. 그 결과 분석 하려는 음성의 도메인 영역에 따라 Pitch의 위치를 변경해야 한다.

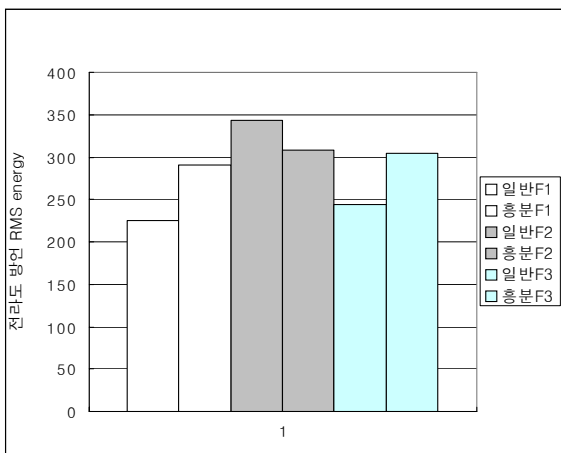
그러므로 홍분된 음성을 표준어와 방언에서 구분하기 위해서는 홍분상태를 구분하기 위한 요소인 Pitch, 사분위 값, Median 값의 편차를 방언 일반상태 값과 표준어 홍분



(a)



(b)



<그림6> 전라도 방언 Box Plot & RMS error

상태 사이의 값으로 조정해야 한다. 경상도 방언의 경우는 F1의 RMS error 형태에 따라 표준어와 경상도 흥분상태를 구분 가능하게 해준다. 하지만 일반상태와 흥분상태를 정확하게 구분하기 위해서는 본 논문의 결과를 바탕으로 시간에 따른 Pitch와 Formant의 변화량에 대해서 더 연구해야 한다.

참고문헌

[1] Lee, C. M., and Narayanan, S., "Towards detecting emotion in spoken dialogs," IEEE Trans. on Speech & Audio Processing, in press
 [2] Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, 문 Taylor, J., "Emotion recognition in human-computer interactions", IEEE Sig. Proc. Mag., vol.18(1), pp. 32-80, Jan 2001
 [3] Patil, V., Preeti Rao., "ACOUSTIC CUES TO MANNER OF ARTICULATION OF OBSTRUENTS IN MARATHI", Proc. of frontiers of research on Speech and Music, February 2008.
 [4] Davis, C., Kim, J., Grauwinkel, K. et al., "Lombard

Proceedings of Speech prosody, Dresden, Germany2006.
 [5] S. Yildirim, M. Bulut, C. Lee, A. Kazemzadeh, C. Busso, Z. Deng, S. Lee, and S. Narayanan, "An acoustic study of emotions expressed in speech," in Proc. 8th Int. Conf. Spoken Lang. Process. (ICSLP 04), Jeju Island, Korea, 2004, pp. 2193-2196
 [6] Cahn, J.E., "Generating Expressions in Synthesized Speech", Master's Thesis MIT, 1989.
 [7] 양병근., "프라트를 이용한 음성분석의 이론과 실제", 만수출판사, 2003