

감정 음성의 피치 변화 분석

천희진, 정지혜, 김병일, 이양희
동덕여자대학교 전자계산학과

An Analysis on the Pitch Variation Of the Emotional Speech

Heejin Chun, Jihye Chung, Byungil Kim, Yanghee Lee
Dept. of Computer Science, Dongduk Women's Univ.
E-mail : heejin@cs4000.dongduk.ac.kr yhlee@www.dongduk.ac.kr

요 약

감정을 표현하는 음성 합성 시스템을 구현하기 위해서 이전 논문에서는 음운 및 운율 요소(피치, 에너지, 지속시간, 스펙트럼 인벨로프)가 각 감정 음성에 미치는 영향에 대한 분석을 수행하였다. 본 논문에서는 네 가지 감정 표현(평상, 화남, 기쁨, 슬픔)을 나타내는 음성 데이터에 대해 음절 세그먼트와 라벨링을 행한 감정 음성 데이터베이스를 토대로 감정 표현에 많은 영향을 미치는 요소인 피치가 어떻게 변화하는지를 분석하였다.

통계적인 방법을 이용하여 감정별 피치를 정규화하였으며, 감정 음성 데이터베이스 내의 문장별 피치 패턴에 대해 분석하였다. 그 결과 감정별 피치의 평균 ZScore는 화남이 가장 작았으며, 기쁨, 평상, 슬픔의 순으로 높았다. 또한 감정별 피치의 범위 변화는 슬픔이 가장 작았으며, 평상, 화남, 기쁨의 순으로 높았다. 문장별 피치의 패턴은 감정 표현에 따라 전체적으로 대부분 유사하게 나타났으며, 문장의 처음 부분은 화남의 경우 다른 감정에 비해 대체로 높게 변화하였고, 화남과 기쁨의 경우 문장의 뒷부분에서 다른 감정에 비해 피치가 상승하는 것을 볼 수 있었다.

I. 서 론

자연스럽고 명료한 감정 표현 음성 합성 시스템은 화상 및 문자 정보 등으로 이루어진 멀티미디어 데이터 처리와 더불어 관심이 고조되고 있으며, 감정 표현이 가능한 합성 시스템은 모든 자동 응답기 및 자동안내, 자동 주문대, 신문 자동 교정기 등 다양한 분야에 응용이 기대된다. 그러나, 감정 음성을 합성하기 위한 음운 및 운율 요소별 패턴 생성이나 모델링에 대한 연구가 미흡한 실정이다. 따라서 본 논문에서는 감정을 표현하는 음성 합성 시스템을 구현하기 위해서, 감정 음성 데이터베이스를 토대로 감정별 피치의 정규화와 문장별 피치 패턴에 대해 분석을 행한다.

본 논문의 1장에서는 연구의 배경 및 목적을 다루는 서론을, 2장에서는 감정 음성 데이터베이스에 대해 기술하고, 3장에서는 감정 표현에 따른 피치 변화를 분석하고 마지막으로 4장에서는 결론을 맺는다.

II. 감정 음성 데이터베이스

감정 표현 음성 합성 시스템에서 합성음에 감정

을 부여하는 규칙을 생성하기 위해서는 대량의 음성 데이터가 필요하다. 음성 합성의 경우는 여러 명의 화자로부터 음성을 수집하기보다는 음성의 특성이 좋은 한 사람의 음성을 철저히 분석하는 것이 중요하다. 따라서 본 논문에서는 화자의 자연스러운 감정을 이끌어 내기 위하여 네 가지 감정(평상, 화남, 기쁨, 슬픔)의 음성 데이터(각 감정별로 80문장)를 대학 아마추어 연극 배우인 여성 화자 2명이 의해 낭독 형태로 녹음하였다. 감정 음성 데이터베이스 내의 모든 문장들은 감정에 영향을 받지 않는 평상 문장으로 구성되었다.

[표1] 감정 음성 데이터베이스 내의 문장 예

문장1 “오늘은 전국이 호린 가운데 비가 조금 내리겠습니다.”
 문장2 “이에 우리의 나아갈 바를 밝혀 교육의 지표로 삼는다.”

구축된 감정 음성 데이터베이스는 각각 시간 정보와 음운 기호를 나타내기 위해서 음절 단위의 세그먼트와 라벨링을 하였다.

III. 감정 표현에 따른 피치 변화 분석

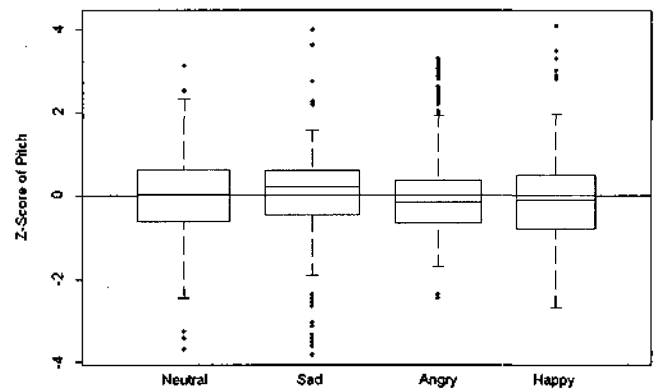
1. 피치 정규화

감정 표현 음성 합성 시스템을 구현하기 위해서는 각 감정에 따라 영향을 미치는 요인이 다양하기 때문에 감정 음성 데이터베이스의 특징을 분석해야 한다. 본 연구에서는 감정 표현에 따른 피치 변화를 분석함에 있어서 화자의 개별성을 제거하기 위해서, 식(1)과 같은 zScore를 사용하여 감정 음성 데이터베이스 내의 세그먼트들의 피치를 정규화 하였다.

$$Z_{if} = (X_{if} - M_f) / S_{Df} \quad \text{----- 식(1)}$$

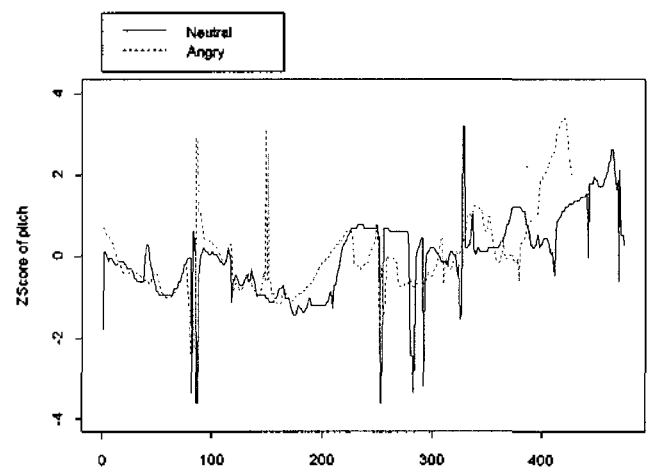
Z_{if} : 감정 f 의 i 번째 관측치에 대한 zScore
 X_{if} : 감정 f 의 i 번째 관측치
 M_f : 감정 f 의 관측치에 대한 평균
 S_{Df} : 감정 f 의 관측치에 대한 표준편차

아래 [그림1]은 식(1)을 이용한 감정별 피치의 정규화의 결과를 나타낸다. 그 결과, 그림에서 보는 바와 같이 감정별 피치의 평균 ZScore는 화남이 가장 작았으며, 기쁨, 평상, 슬픔의 순으로 높았다. 또한 감정별 피치의 범위 변화는 슬픔이 가장 작았으며, 평상, 화남, 기쁨의 순으로 높았다.

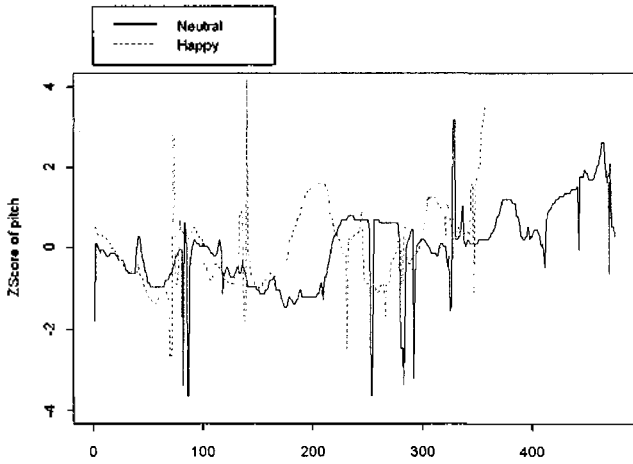


[그림1] ZScore를 이용한 감정별 피치의 정규화

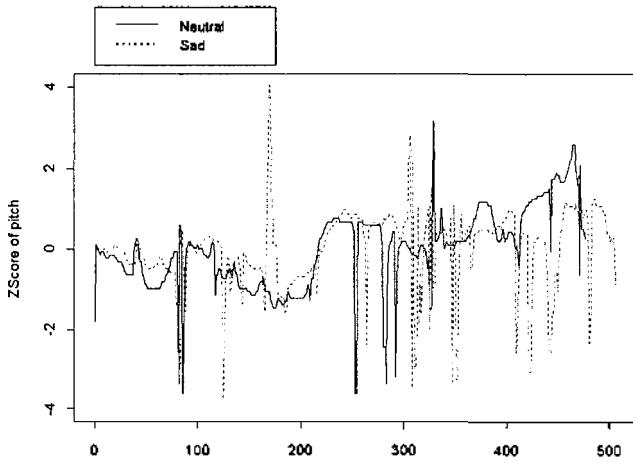
ZScore를 이용하여 정규화한 피치의 평상에 대한 각 감정별 변화를 나타낸 결과는 [그림2]와 같다.



(a) 평상 : 화남



(b) 평상 : 기쁨



(c) 평상 : 슬픔

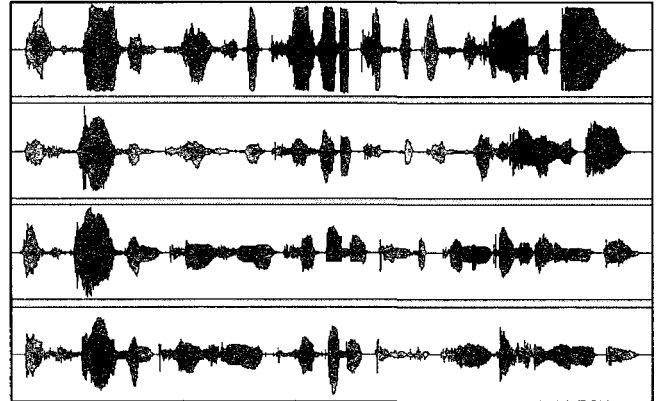
[그림2] 평상에 대한 각 감정별 피치 변화

위의 결과와 같이 화남과 기쁨의 경우 문장의 뒷부분에서 피치가 상승하는 것을 알 수 있다.

2. 피치 패턴

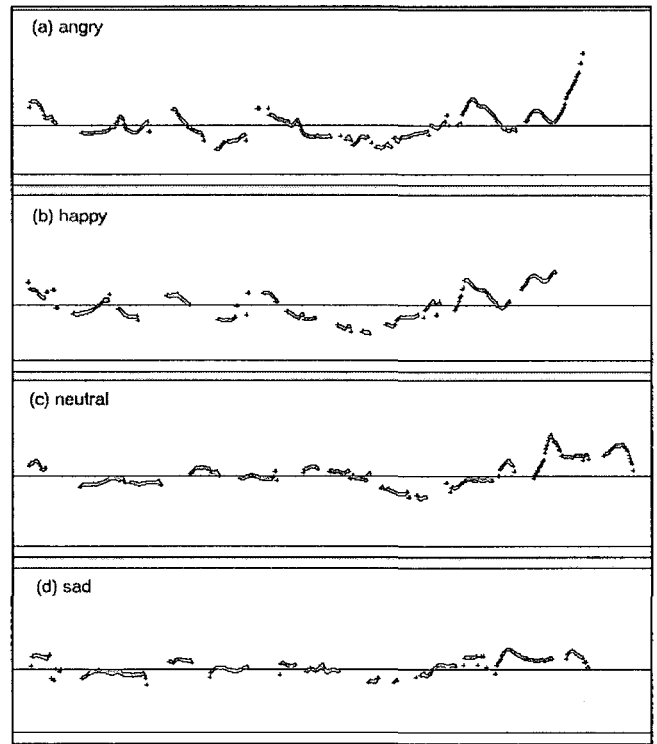
감정 표현에 따라서 피치 패턴의 변화를 분석하기 위해서 감정 음성 데이터베이스 내의 문장별 피치 패턴에 대해 분석하였다. 그 결과 문장별 피치의 패턴은 [그림3]과 같이 감정 표현에 따라 전체적으로는 대부분 유사하게 나타났으며, 화남 감정의 경

우 문장의 처음 부분에서 다른 감정에 비해 대체로 높게 변화함을 알 수 있었다. 또한 화남과 기쁨의 경우 문장의 뒷부분에서 다른 감정에 비해 피치가 상승하는 것을 볼 수 있었다.



(ㄱ) 감정 음성에 따른 파형

“역사학은 단순히 과거를 탐색하는 것이 아니다.”



(ㄴ) 감정 음성에 따른 피치

(a. 화남, b. 기쁨, c. 평상, d. 슬픔)

[그림3] 감정별 피치 패턴의 변화 예

V. 결론

본 논문에서는 감정을 표현하는 음성 합성 시스템을 구현하기 위해서 감정 표현에 많은 영향을 미치는 요소인 피치의 변화에 대해 분석하였다.

화자의 개별성을 없애기 위해서 감정별 피치를 ZScore를 사용하여 정규화 하였으며, 감정 음성 데이터베이스 내의 문장별 피치 패턴에 대해 분석하였다. 그 결과 감정별 피치의 평균 ZScore는 화남이 가장 작았으며, 기쁨, 평상, 슬픔의 순으로 높았다. 또한 감정별 피치의 범위 변화는 슬픔이 가장 작았으며, 평상, 화남, 기쁨의 순으로 높았다. 문장별 피치의 패턴은 감정 표현에 따라 전체적으로 대부분 유사하게 나타났으며, 문장의 처음 부분은 화남의 경우 다른 감정에 비해 대체로 높게 변화하였고, 화남과 기쁨의 경우 문장의 뒷부분에서 다른 감정에 비해 피치가 상승하는 것을 볼 수 있었다. 따라서 감정 표현 음성 합성 시스템을 구현하기 위해서 이러한 피치 분석 결과를 합성에 적용하여야 한다.

[참고문헌]

1. A. Iida, K. Meiseki, Nick Campbell, Designing and Testing a corpus of emotional speech, 일본 음향학회 춘계 연구발표회 강연 논문집 pp311-312, 1997.3.
2. Noad, J.E.H., Whiteside, S.P. and Green, P.D. A macroscopic analysis of an emotional corpus, in Proceedings of Eurospeech 97, pp.517-520, 1997.
3. Engberg, I.S., Hansen, A.V., Andersen, O. and Dalsgard, P. Design, recording and verification of a Danish Emotional Database, in Proceedings of Eurospeech 97, pp.1695-1698, 1997.
4. J.M. Montero, J. Gutierrez-Arriola, S. Palazuelos, E. Enriquez, S. Aguilera, J.M. Pardo, Emotional Speech Synthesis: From Speech Database To TTS, in Proceedings of ICSLP 98, 1998.
5. Akemi Iida, Nick Campbell, Soichiro Iga, Fumito Higuchi, Michiaki Yasumura, Acoustic Nature and Perceptual Testing of Corpora of Emotional Speech, in Proceedings of ICSLP 98, 1998.
6. Heejin Chun, Yanghee Lee, An Analysis on the Emotional Speech for the Speech Synthesis System with Emotion, KSCSP 98, Vol. 15, No. 1, pp350-355, 1998.
7. B. Heuft, T. Portele, and M. Rauth, Emotions in Time Domain Synthesis, in Proceedings of ICSLP 96, (Philadelphia, PA), 1997.
8. Erhard Rank, Hannes Pirker, Generating Emotional Speech with a Concatenative Synthesizer, in Proceedings of ICSLP 98, 1998
9. Heejin Chun, Yanghee Lee, An Analysis on the Acoustic Parameters of Emotional Speech for the Speech Synthesis, in Proceedings of ICSP 99 Vol 1 of 2, 1999