

감정 음성의 음향학적 모델링에 관한 연구

천희진, 이양희
동덕여자대학교 전자계산학과

A Study on the Acoustic Modeling of the Emotional Speech

Heejin Chun, Yanghee Lee
Division of Computer and Information Science, Dongduk Women's Univ.
E-mail : heejin@cs4000.dongduk.ac.kr yhlee@dongduk.ac.kr

요 약

본 논문에서는 감정 표현 음성 합성 시스템을 구현하기 위해서, 감정 음성 데이터베이스의 음향학적 특징인 피치, 에너지, 지속시간, 스펙트럼 포락에 대해 분석한 결과와 문법적 요소인 품사에 따른 감정 음성 데이터의 피치 변화를 분석하였다. 분석 결과, 기본 주파수, 에너지, 지속시간, 스펙트럼 포락은 감정 표현에 중요한 영향을 미치는 것으로 나타났으며, 전반적으로 화남과 기쁨의 감정이 평상과 슬픔의 감정 보다 피치 및 에너지의 변화가 크게 나타났으며, 특히 기쁨 감정의 경우 부사, 관형사, 연결어미, 조사, 접미사에서 피치 변화가 많았으며, 화남 감정의 경우, 관형사, 명사, 용언, 접미사에서 피치 변화가 높게 나타났다.

이러한 분석 결과를 적용해 감정 음성을 합성하기 위하여, 평상 음성에 각 감정 음성의 운율 요소를 적용하여 감정 음성을 합성하여 평가한 결과, 기쁨 감정은 기본 주파수의 변화에 의해 86.7%, 화남 감정은 에너지의 변화에 의해 91%, 슬픔 감정은 음절지속시간의 변화에 의해 76.7%가 각각 올바른 감정으로 인지되었다.

1. 서 론

감정은 인간의 의사 소통에 있어서 중요한 역할을 하고 있기 때문에, 자연스럽게 명료한 감정 표현 음성 합성 시스템에 대한 관심이 고조되고 있다. 감정 음성을 합성

하기 위해서, HAMLET (Helpful Automatic Machine for Language and Emotional Talk) [1], ASSESS (Automatic Statistical Summary of Elementary Speech Structures) [2], Affect Editor [3]와 같이 규칙을 기반으로 한 감정 음성 합성기를 구현하기 위한 연구가 수행되어왔지만, 한국어에서 감정 음성을 합성하기 위한 감정 음성 데이터베이스 구축에 대한 연구 및 음향학적 모델링이나 규칙 생성에 대한 연구가 미흡한 실정이다.

따라서 본 연구에서는 감정 표현이 가능한 음성을 평상 음성으로부터 생성하기 위해, 평상 음성의 피치, 에너지, 지속시간 및 스펙트럼 포락 등의 음운 및 운율 요소를 변화 시켜 감정 음성을 합성한다.

본 논문에서는 1장에서는 서론을, 2장에서는 감정 음성 데이터베이스에 대해 간략히 기술하고, 3장에서는 감정 음성 데이터의 문법적 요소에 따른 피치 변화를 분석하고, 4장에서는 분석 결과를 토대로 감정 음성을 합성하여 평가하며, 마지막으로 5장에서 결론을 맺는다.

II. 감정 음성 데이터베이스

본 논문에서는 다양한 감정 표현 중에서, 감정 변화의 구분이 확연한 서로 다른 세 가지 감정 즉, 화남, 기쁨, 슬픔의 감정 음성과 평상 음성에 대해서 분석하고자 한다. 감정 음성 데이터는 화자의 자연스러운 감정을 이끌어 내기 위하여 평상, 화남, 기쁨, 슬픔이 잘 표현될 수 있는 독백 형식의 80문장을 신문이나 뉴스로부터 수집하여 각

감정별로 낭독 형태로 녹음하였다. 화자는 의도적으로 감정을 어느 정도 표현할 수 있는 대학 아마추어 연극 배우인 여성 화자 2인이다.

감정 음성 데이터베이스 내의 모든 문장들은 감정에 영향을 받지 않는 평상 문장으로 구성되었다. 녹음된 감정 음성 데이터는 CSL을 이용하여, 16kHz로 샘플링, 16bit로 양자화 하였고, ESPTS를 이용하여, 각각 시간 정보를 표시하기 위해 음절 레벨 세그먼트를, 음운 기호를 표시하기 위해 음절 라벨링을 행하고, 음절별 품사 및 문법 정보를 태깅하여 구축된 감정 음성 데이터베이스를 토대로 한다.

- 음절 레벨 세그먼트 (시간 정보)
- 음절 라벨링 (음운 기호)
- 어절에 대한 음절 레벨 품사 태깅 (품사)

구축된 감정 음성 데이터베이스는 평상으로 녹음된 음성이 평상 감정으로 인식되는 경우가 89.4%, 화남으로 녹음된 음성이 화남 감정으로 인식되는 경우가 61%, 기쁨으로 녹음된 음성이 기쁨 감정으로 인식되는 경우가 57.6%, 슬픔으로 녹음된 음성이 슬픈 감정으로 인식되는 경우가 70.6%로 평가되었다.

III. 감정 음성 데이터베이스 분석

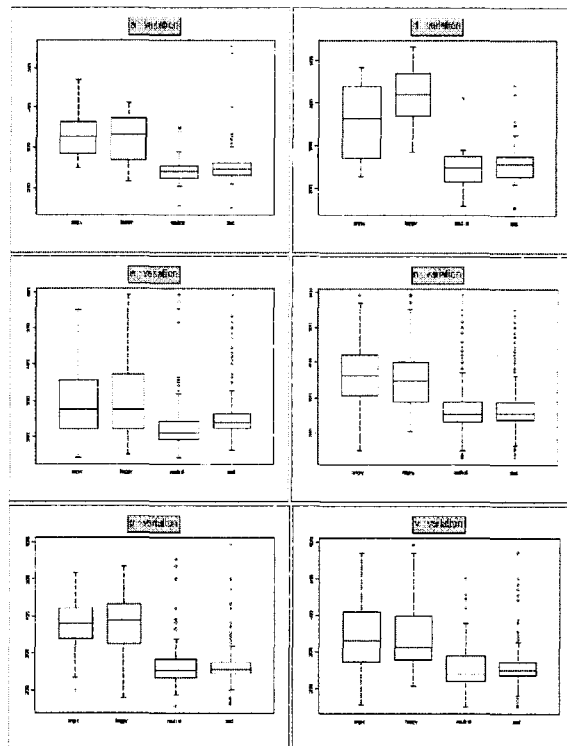
감정을 표현하는 음성 합성 시스템을 구현하기 위하여, 기존 연구에서 감정 음성의 음운 및 운율 요소인 피치, 에너지, 지속시간에 대해 분석하고, 통계적인 방법을 이용하여 감정별 피치를 정규화 하였으며, 감정 음성 데이터베이스 내의 문장별 피치 변화 및 음절 유형별, 어절 내 음절 위치에 따른 피치 및 지속시간의 변화와 스펙트럼 포락의 변화에 대해 분석하였다.[4][7] 본 연구에서는 문법적인 요소에 따른 감정 음성의 피치 및 에너지에 대한 변화에 대해 분석하고자 한다.

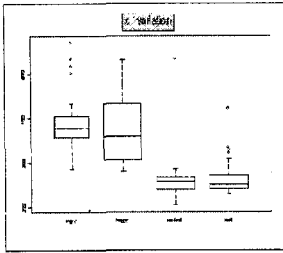
문법적인 요소에 따른 감정 음성의 음향학적 변화를 분석하기 위해서, 본 연구에서 어절에 대한 품사 태깅을 행하기 위해 쓰인 품사표는 [표 1]과 같다.

[표 1] 품사표

ac	접속부사	na	동작성 보통명사
ad	부사	nc	보통명사
dm	지시관형사	nd	의존명사
du	수관형사	np	인칭 대명사
ec	연결어미	nr	고유명사
ed	관형사형 연결어미	ns	상태성 보통명사
ef	어말어미	pa	부사격조사
ep	선어말어미	pd	관형격조사
ex	보조적 연결어미	po	목적격조사
vb	동사	ps	주격조사
vj	형용사	pt	주제격조사
vx	보조용언	xj	형용사 파생접미사
		xv	동사 파생접미사

각 감정별로 품사에 따른 피치 변화를 분석한 결과는 [그림 1]과 같다. 왼쪽 상단부터 부사(a), 관형사(d), 연결어미(e), 명사(n), 조사(p), 용언(v), 접미사(x)의 품사에 대해서 각각 화남, 기쁨, 평상, 슬픔을 나타낸다.





[그림 1] 품사에 따른 감정 음성의 피치 변화

품사에 따른 감정 음성의 피치 변화를 분석한 결과, 전반적으로 화남과 기쁨의 감정이 평상과 슬픔의 감정 보다 피치 및 에너지의 변화가 크게 나타났으며, 특히 기쁨 감정의 경우 부사, 관형사, 연결어미, 조사, 접미사에서 피치 변화가 많았으며, 화남 감정의 경우, 관형사, 명사, 용언, 접미사에서 피치 변화가 높게 나타났다.

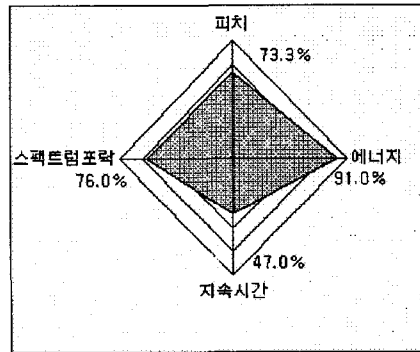
IV. 감정 음성 합성 및 평가

감정 음성을 합성하기 위해서, 개량 켈스트럼 분석법을 이용한 음성 분석계를 사용하여 음성 파형으로부터 피치 및 개량 켈스트럼을 구한다. 이렇게 구해진 감정 음성의 피치 및 켈스트럼 계수를 평상 음성에 적용하기 위해서는 감정 음성의 지속시간이 서로 다르기 때문에 지속시간을 정합하기 위해서 음절 별로 DTW(Dynamic Time Warping) 알고리즘을 사용한다. DTW 알고리즘에서 테스트 패턴으로 감정 음성의 켈스트럼 계수를, 참조 패턴으로 평상 음성의 켈스트럼 계수를 사용하여, 테스트 패턴과 참조 패턴 사이의 거리가 가장 작을 때의 경로를 구하고, 구해진 경로에 따라서 감정 음성의 운율 요소를 평상 음성에 적용한다.

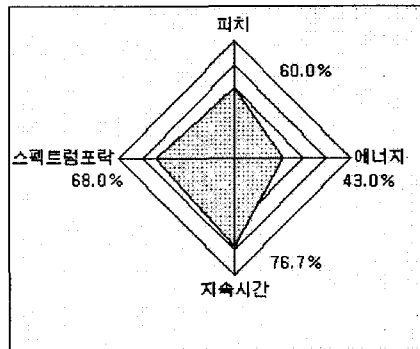
또한, 감정 음성을 합성하기 위한 파라미터로 평상 음성의 평균 피치, 피치 범위, 평균 에너지, 에너지 범위, 지속시간, 첫 음절, 중간 음절과 마지막 음절에서의 피치, 스펙트럼 포락 등을 감정 음성의 파라미터 변화에 따라 조절하여 감정 음성을 합성하였다.

합성된 감정 음성의 감정 표현 인지도를 평가하기 위해서 ABX 청취 테스트를 행하였다[4]. 그 결과 기쁨 감정은 기본 주파수의 변화에 의해 86.7%, 화남 감정은 에너지의 변화에 의해 91%, 그리고 슬픈 감정은 지속시간의 변화에 의해 76.7%가 각각 올바른 감정으로 인지되었다.

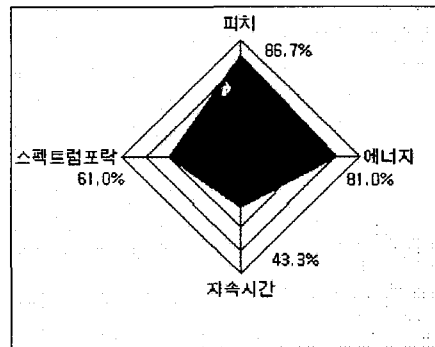
그 결과, 감정 음성과 운율 요소의 상관 관계는 [그림 2]와 같이, 감정 표현에 따라서 피치, 에너지, 지속시간, 스펙트럼 인벨로프는 모두 중요한 영향을 미치는 것으로 나타났다. 특히 화남 감정은 에너지에 의해서, 기쁨 감정은 피치에 의해서, 그리고 슬픈 감정은 지속시간에 의해서 각각 영향을 가장 많이 받는 것으로 나타났다.



(a) 화남 감정



(b) 슬픔 감정



(c) 기쁨 감정

[그림 2] 감정 음성과 운율 요소의 상관 관계

V. 결론

본 논문에서는 감정 표현 음성 합성 시스템을 구현하기 위해서, 감정 음성 데이터베이스의 음향학적 특징 즉 피치, 에너지, 지속시간, 스펙트럼 포락에 대해 분석한 결과와 문법적 요소 즉 품사에 따른 감정 음성 데이터의 피치 변화를 분석하였다. 분석 결과, 기본 주파수, 에너지, 지속시간, 스펙트럼 포락은 감정 표현에 중요한 영향을 미치는 것으로 나타났으며, 전반적으로 화남과 기쁨의 감정이 평상과 슬픔의 감정 보다 피치 및 에너지의 변화가 크게 나타났으며, 특히 기쁨 감정의 경우 부사, 관형사, 연결어미, 조사, 접미사에서 피치 변화가 많았으며, 화남 감정의 경우, 관형사, 명사, 용언, 접미사에서 피치 변화가 높게 나타났다.

이러한 분석 결과를 적용하여 평상 음성에 각 감정 음성의 운율 요소를 적용하여 감정 음성을 합성하여 평가한 결과, 기쁨 감정은 기본 주파수의 변화에 의해 86.7%, 화남 감정은 에너지의 변화에 의해 91%, 슬픈 감정은 음절 지속시간의 변화에 의해 76.7%가 각각 올바른 감정으로 인지되었다. 즉, 화남 감정은 에너지에 의해서, 기쁨 감정은 피치에 의해서, 그리고 슬픈 감정은 지속시간에 의해서 각각 영향을 가장 많이 받는 것으로 나타났다.

[참고문헌]

1. Murray I. R. and Arnott J. L., "Synthesizing Emotions in Speech: Is it Time to Get Excited?", Proc ICSLP '96, 1996
2. Cowie, R., Sawey, M., and Douglas-Cowie, E., "A new speech analysis system: ASSESS (Automatic Statistical Summary of Elementary Speech Structures)", Proc ICPhS '95, pp 278-281, Stockholm, 1995
3. Janet E. Cahn. "The Generating of Affect in Synthesized Speech", MIT Media Technology Laboratory, 1990.
4. 천희진, 이양희, "감정표현 음성합성 시스템을 위한 감정 분석", 음성 통신 및 신호처리 워크샵 논문집, pp 350-355, 1998.
5. 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.

6. 천희진, 이양희, "감정 음성의 피치 변화 분석", 한국음향학회 학술발표논문지 제 18 권 2 호, 1999
7. 천희진, 이양희, "감정 편집기를 이용한 감정 음성 합성", 한국음향학회 하계학술대회 논문집 제 19 권 1 호, 2000.
8. Elissaveta Abadjieva, Murray I.R. and Arnott J. L. "Applying Analysis of Human Emotional Speech To Enhance Synthetic Speech", Eurospeech '93, 1993.