2000년도 한국음향학회 학술발표대회 논문집 제19권 제1(s)호

감정 편집기를 이용한 감정 음성 합성

천희진, 이양희
동덕여자대학교 전자계산학과

Emotional Speech Synthesis using the Emotion Editor Program

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

감정 표현 음성을 합성하기 위하여 본 연구에서는 감정 음성 데이터의 피치와 지속시간의 음절 유형별 및 어절 내 음절 위치에 따른 변화를 분석하였고, 스펙트럼 포락이 감정 변화에 어떤 영향을 미치는지를 분석하였다. 그 결과, 피치와 지속시간의 음절 유형별, 어절 내 음절 위치에 따른 변화와, 스펙트럼 포락 등도 감정 변화에 영향을 미치는 것으로 나타났다.

또한, 감정 음성의 음향학적 분석 결과를 적용하여 감정 음성을 합성하고 평가하기 위하여, 평상 음성의 음운 및 운율 파라미터 (피치, 에너지, 지속시간, 스펙트럼 포락)를 조절함으로써 감정 음성을 생성하는 감정 편집기를 구현하였다.

I. 서 론

감정 음성을 합성하기 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 수행되어왔다. 반면, 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, 기존 연구 결과를 토대로 평상 음성의 피치, 에너지, 지속시간 및 스펙트럼 포락 등의 음운 및 운율 요소를 변화 시키기 위한 감정 편집기를 구현하여 감정 음성을 합성한다.

본 논문에서는 1장에서는 서론을, 2장에서는 감정 음성 데이터베이스에 대해 간략히 기술하고, 3장에서는 감정 음성의 음절 유형별 및 어절 내 음절 위치에 따른 피치 및 지속시간의 변화와 스펙트럼 포락을 분석하며, 4장에서는 감정 편집기를 이용한 감정 음성 합성에 대해 기술하고 마지막으로 5장에서 결론을 맺는다.

II. 감정 음성 데이터베이스

본 논문에서는 감정에 영향을 받지 않는 평상 문장 80문장을 수집하여, 대학 아마추어 연극 배우인 여성 화자 2인이 서로 다른 네 가지 감정(화남, 기쁨, 슬픔, 평상)으로 각각 녹음하고, 녹음된 감정 음성 데이터에 시간 정보를 나타내기 위해 음절 레벨의 세그먼트를 행하고, 음운 기호 및 품사를 표기하기 위해 각각 음절 라벨링, 음절별 품사 및 문법 정보를 태깅하여 구축된 감정 음성 데이터베이스를 토대로 한다.

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

III. 감정 음성 데이터베이스 분석

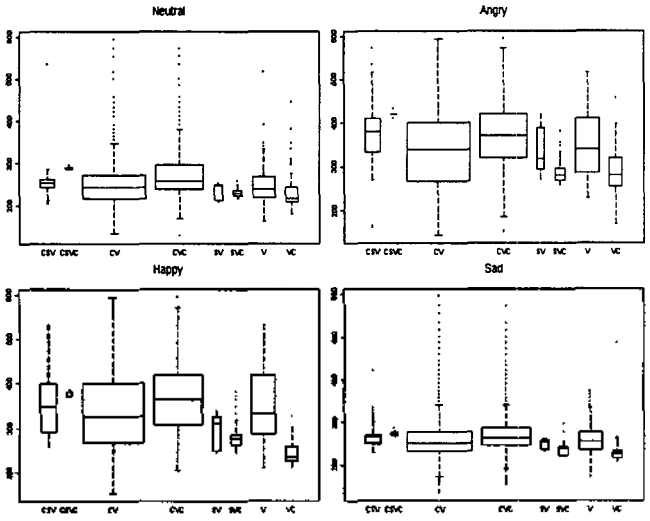
감정을 표현하는 음성 합성 시스템을 구현하기 위하여, 이전 연구에서는 감정 음성의 음운 및 운율 요소인 피치, 에너지, 지속시간에 대한 분석을 행하였다. 또한 통계적인 방법을 이용하여 감정별 피치를 정규화 하였으며, 감정 음성 데이터베이스 내의 문장별 피치 변화에 대해 분석하였다[4]-[6].

본 연구에서는 감정 음성의 음절 유형별, 어절 내 음절 위치에 따른 피치 및 지속시간의 변화와 스펙트럼 포락의 변화에 대해 분석하였다.

1. 음절 유형

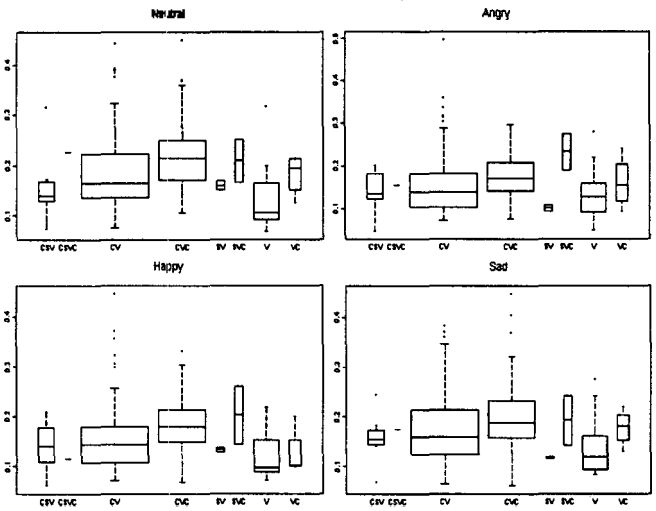
음절 유형은 8 분류(CV, CVC, V, VC, SV, SVC, CSV, CSVC)로 나뉜다. V 앞의 C 는 초성이고, V 뒤의 C 는 종성이며, S 는 반모음이고, V 는 모음이다. 아래 그림의 상단 왼쪽부터 평상, 화남, 기쁨, 슬픔의 음절 유형에 따른 피치 및 지속시간의 변화를 나타낸다.

음절 유형에 따라 감정 음성의 피치를 분석한 결과, 화남과 기쁨의 감정에서는 그림과 같이 모음으로 끝나는 CSV, CV, SV, V 등의 음절 유형이 다른 음절 유형 보다 피치가 넓게 분포하고, 큰 변화를 보이는 것을 알 수 있으며, 평상과 슬픔의 피치 분포가 좁게 나타나는 반면, 화남과 기쁨의 감정에서는 피치가 넓게 분포하여 피치의 동적인 범위가 크게 나타나는 것을 알 수 있다.



[그림 1] 음절 유형에 따른 각 감정별 피치 분포

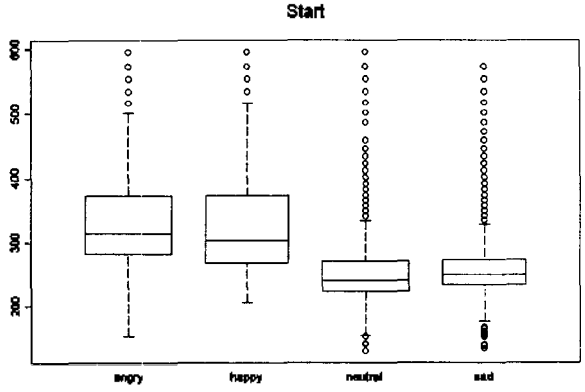
음절 유형에 따라 음절 지속시간을 분석한 결과, 화남과 기쁨의 감정이 평상과 슬픔의 감정에 비해 지속시간이 짧게 나타났으며, 모음으로 끝나는 CSV, CV, SV, V 등의 음절 유형이 다른 음절 유형에서보다 지속시간이 비교적 길게 나타나며, 동적인 변화도 크게 나타나는 것을 알 수 있다.



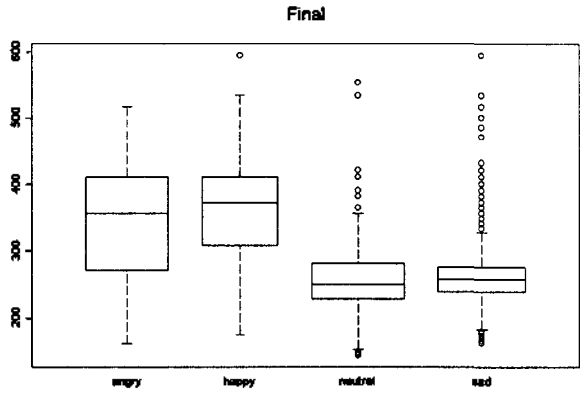
[그림 2] 음절 유형에 따른 각 감정별 지속시간 분포

2. 어절 내 음절 위치

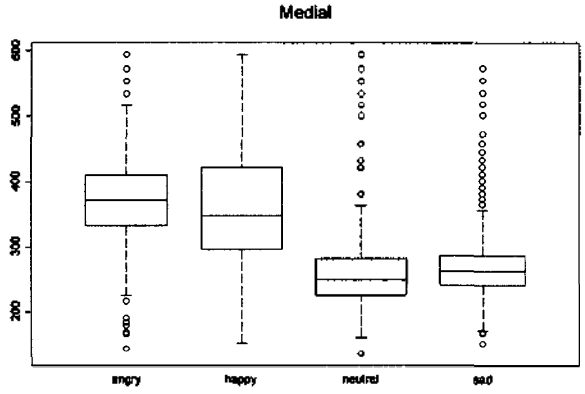
어절 내의 음절 위치를 3 분류(Start, Medial, Final)로 나누었다. 어절 내의 음절을 첫 음절(S), 마지막 음절(F), 그리고 이 두 음절을 제외한 중간 음절(M)의 3 가지로 분류하여 피치를 분석한 결과, 마지막 음절에서 화난 감정의 피치가 넓게 분포하고, 변화가 크게 나타나며, 기쁜 감정의 평균 피치가 마지막 음절에서 크게 나타나는 것을 알 수 있다. 또한 중간 음절에서 기쁜 감정의 피치가 다른 감정에 비해 크게 변화함을 알 수 있다.



[그림 3] 어절 내 첫 음절의 피치 분포



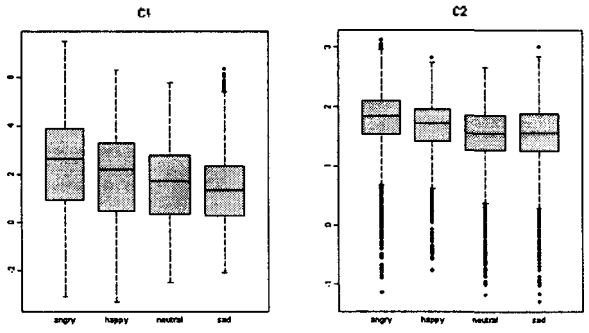
[그림 4] 어절 내 마지막 음절의 피치 분포

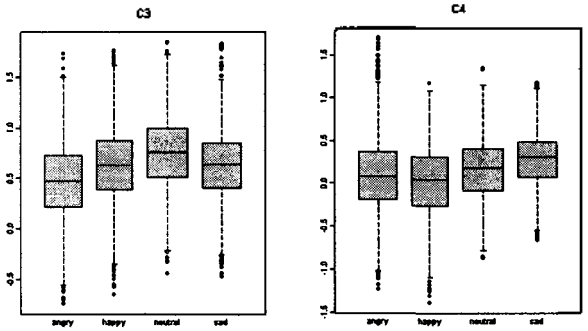


[그림 5] 어절 내 중간 음절의 피치 분포

3. 스펙트럼 포락

감정에 의해 음성의 스펙트럼 포락이 어떻게 변화하는가를 분석하기 위해, 켑스트럼 계수에서 저차항인 에너지(C0)를 제외하고 1 차부터 24 차까지의 값(C1~C24)의 분포를 각 감정별로 분석하였다.





[그림 6] 감정별 스펙트럼 포락의 분포 (1 차~4 차)

[그림 6]의 상단 왼쪽부터 C1, C2, C3, C4 의 감정별 스펙트럼 포락의 분포를 나타낸다.

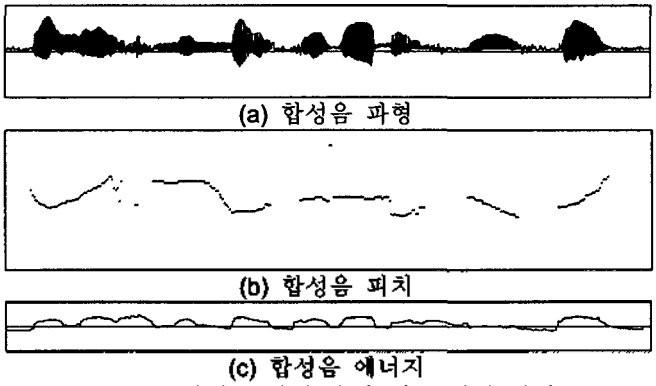
스펙트럼 포락의 분석 결과, 감정에 따라서 가장 많은 변화를 보이는 켑스트럼 차수는 1 차부터 8 차(C1~C8)까지이고, 11 차부터 15 차(C11~C15)까지는 작은 변화를 보이며, 나머지 켑스트럼 차수에서는 감정에 따른 변화가 거의 나타나지 않는 것을 알 수 있다.

IV. 감정 음성 합성 및 평가

본 연구에서는 분석합성계[4]를 이용하여, 평상 음성의 음운 및 운율 파라미터를 조절함으로써 감정 음성을 생성하는 감정 편집기를 구현하였다.

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. 감정 합성음의 파형, 에너지, 피치는 그림과 같다.

(예문) 원하시는 날짜가 맞습니까



(a) 합성음 파형

(b) 합성음 피치

(c) 합성음 에너지

[그림7] 합성된 감정 음성 예 – 화난 감정

V. 결론

본 논문에서는 감정 표현 음성을 합성하기 위해서 피치 및 지속시간의 음절 유형별 및 어절 내 음절 위치에 따른 변화를 분석하고, 스펙트럼 포락의 변화도 분석하였다.

그 결과, 음절 유형에 따른 피치 변화는, 화남과 기쁨의 감정에서는 모음으로 끝나는 CSV, CV, SV, V 등의 음절 유형에서 피치 변화가 크게 나타나고, 평상과 슬픔의 피치 분포가 좁게 나타나는 반면, 화남과 기쁨의 감정에서는 피치가 넓게 분포하여 피치의 동적인 범위가 크게 나타나는 것을 알 수 있으며, 음절 지속시간은, 화남과 기쁨의 감정이 평상과 슬픔의 감정에 비해 지속시간이 짧게 나타나고, 모음으로 끝나는 CSV, CV, SV, V 등의 음절 유형에서 지속시간이 비교적 길게 나타나며, 동적인 범위도 크게 나타나는 것을 알 수 있다.

어절 내 음절 위치에 따른 피치 변화는, 마지막 음절에서 화난 감정의 피치 변화가 크게 나타나며, 기쁜 감정의 평균 피치가 마지막 음절에서 크게 나타난다. 또한 중간 음절에서 기쁜 감정의 피치가 크게 변화함을 알 수 있다.

음성의 스펙트럼 포락의 변화는, 감정에 따라서 가장 많은 변화를 보이는 켑스트럼 차수는 1차부터 8차(C1~C8)까지이고, 11차부터 15차(C11~ C15)까지는 작은 변화를 보이며, 나머지 켑스트럼 차수에서는 감정에 따른 변화가 거의 나타나지 않는 것을 알 수 있다.

본 논문에서 감정 음성을 생성하는 감정 편집기를 구현하여 평상 음성의 음운 및 운율 파라미터 즉, 피치, 에너지, 지속시간, 스펙트럼 포락 등을 조절함으로써 감정 음성을 합성하였다.

[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", KSCSP '98, Vol. 15, No. 1, 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. Elissaveta Abadjieva, Murray I.R. and Arnott J. L. "Applying Analysis of Human Emotional Speech To Enhance Synthetic Speech", Eurospeech '93, 1993.