

대본 내용에 의한 정서음성 수집과정의 정규화에 대하여

조 철 우

창원대학교 메카트로닉스 공학부

Normalization in Collection Procedures of Emotional Speech by Scriptual Context

Cheolwoo Jo

School of Mechatronics, Changwon National University

E-mail : cwjo@sarim.changwon.ac.kr

Abstract

One of the biggest problems unsolved in emotional speech acquisition is how to make or find a situation which is close to natural or desired state from humans. We proposed a method to collect emotional speech data by scriptual context. Several contexts from the scripts of drama were chosen by the experts in the area. Context were divided into 6 classes according to the contents. Two actors, one male and one female, read the text after recognizing the emotional situations in the script.

I. 서론

정서음성을 연구하기 위한 시료 수집에 있어서 큰 문제점의 하나는 서로 다른 화자가 동일한 정서상태를 가정하여 발성할 시에 발성하는 상황에 대한 연상이 화자에 따라 모두 다를 수 있다는 것이다. 서로 다른 정서상태는 다른 음성 표현으로 나타나게 되어 분석과정에서 문제가 될 수 있으며 분석을 어렵게 만드는 요인이 되기도 한다. 기존의 정서음성 수집은 배우에 의한 방법[], 다큐멘터리에 의한 방법, 게임[], 자전적 회상에 의한 방법[] 등 여러가지 정서유발 방법을 사용해 오고 있다. 이 중 어느 방법도 특정 정서상태를 일관되게 화자에게 알려줄 수 있지는 못한다. 화자별로 특정 정서상태에 대해 연상하는 상황이 다르기 때문이다.

본 논문에서는 이러한 정서상태의 다양성을 정규화하기 위한 한 방법으로 연극 대본의 내용을 정서상태와 맞도록 설정하고 그 상황을 통해서 정서상태를 유발하는 방법을 제안하고 실제 배우를 대상으로 수집한 과정과 결과에 대해서 서술한다. 대본은 실제 연극의 대본 중 특정 상황을 나타낼만한 부분을 추려서 선택하였으며 '독백' 부분과 '대화' 부분으로 나누어 수집하였다. 대상 정서상태는 기쁨, 슬픔, 화남, 공포, 지루함 의 다섯 가지이며 각 상태의 대본을 낭독형으로 다시 읽도록 하여 비교 대상으로 삼았다.

II. 정서음성 수집에서의 문제들

정서적 음성의 수집은 일반 음성 데이터의 수집 과정과는 여러 가지 면에서 차이가 있으며 어려움이 있다. 일반적인 음성데이터 수집은 다음과 같은 과정을 거친다.

1. 포함 영역 선정 및 용도(숫자음, 대화, 합성, 인식 등등)
2. 텍스트 선정
3. 화자 섭외
4. 발성
5. 녹음

정서적 음성데이터의 수집은 여기에 한두가지 더 고려할 부분이 있다.

1. 정서상태의 선택
2. 포함 영역 선정 (숫자음, 대화, 합성, 인식 등등)
3. 텍스트 선정

4. 화자 섭외
5. 정서상태 유발
6. 발성
7. 녹음

정서음성 수집에 있어서 어려운 부분은 1,2,3,5의 부분이다.

첫 번째, 정서상태의 선택에 있어서의 문제점
대부분의 정서음성 수집에 있어서는 사람의 가장 기본적인 정서라고 하는 기쁨, 슬픔, 화냄, 공포의 네 가지를 택하고 낭독체를 추가하여 다섯 가지를 수집하고 있다. 최근에는 지루함을 더하여 여섯 가지로 수집하기도 한다. 그러나 실제로 정서상태의 수는 매우 다양하므로 어떤 정서상태를 대상으로 하는지를 정하는 것은 간단한 문제가 아닐 수 있다. 대개는 막연히 네가지 상태로 규정하고 있으나 세부적으로 분류하고 규정하기 위해서는 정밀한 연구가 필요하다.

두 번째, 포함할 영역의 선택에 있어서의 문제
어떤 응용을 목적으로 하는지, 어떤 실험을 목적으로 하는지를 규정할 필요가 있다.

세 번째, 텍스트의 선택의 문제
대개의 음성 데이터 수집은 미리 정해진 문장을 주고 읽는 방식으로 수집하게 된다. 자연 발화와 같은 경우는 그렇지 않은 경우도 있지만 목적에 맞는 음성을 수집하도록 제어가 어렵기 때문에 대부분의 경우 대본에 의한 발성을 하게 된다. 통상적인 음성 데이터베이스의 경우는 텍스트 선정에 있어서 음성의 통계적 특성이나 언어적 구성등을 고려하게 되나 내용은 그다지 고려하지 않고 있다. 지금까지 정서음성의 수집을 위해서는 무의미 단어, 유의미 단어, 중의성 문장 등을 발성자로 하여금 다양한 정서상태로 읽도록 한 뒤에 녹음하는 방법을 택하였다. 장시간의 정서음성을 녹음하기 위해서 장문의 문장을 특정 정서상태로 읽도록 하는 방법을 사용한 경우도 있으나 정서적 상태가 어느 정도 시간동안 일관되게 유지될 수 있는가하는 문제도 해결되어야 할 과제이다.

다섯 번째, 정서상태의 유발 문제
이 부분은 정서음성의 수집에서 가장 문제가 되는 부분이다. 어떤 방법을 사용하여야 자연스러운 정서상태가 유발되는가 하는 문제와 서로 다른 화자간에 구현된 정서상태의 일관성을 유지하는 문제가 어렵다. 예를 들어 배우에 의한 정서음성은 정서의 강도가 높은 반면 자연스럽지가 못하다. 자진적 회상에 의한 방법 같은 경우는 비교적 자연스럽지만 정서의 강도가 약해서 응용이 제한적일 수 밖에 없다. 응용 분야에 따라서 정서의 강도가 강한 과장된 경우가 필요한 경우도 있을 수 있고, 약한 경우가 필요한 경우도 있으므로 어느 쪽이 좋다고 말할 수는 없으나 정서의 강도를 어느 정도 규정하고

측정할 수 있는 방안이 필요하다. 기존의 정서음성 수집은 배우에 의한 방법[], 다큐멘터리에 의한 방법, 게임[], 자진적 회상에 의한 방법[] 등 여러가지 정서유발 방법을 사용해 오고 있다.

III. 대본에 의한 수집방법

대본은 실제 연극의 대본 중 특정 상황을 나타낼만한 부분을 추려서 선택하였으며 '독백' 부분과 '대화' 부분으로 나누어 수집하였다. 대상 정서상태는 기쁨, 슬픔, 화남, 공포, 지루함 의 다섯 가지이며 각 상태의 대본을 낭독형으로 다시 읽도록 하여 비교 대상으로 삼았다.

각 정서상태별 대본의 출전

대본의 선택

각 정서별로 독백형, 대화형 원고를 각각 택해서 취함

낭독:

남자독백: 백세개의 모노로그 서문

여자독백: 백세개의 모노로그 엮은이의 말

기쁨:

여자독백: '천의 바보들' 중에서 산드라, 허브 가드너 작, A Thousand Clowns, Herb Gardner

남자독백: '폐츄니아를 짓밟은 거인' 중에서 청년, 테네시 윌리엄즈 작, The Case of the Crushed Petunias, Tennessee Williams

대화: '싸움터에서의 산책', 페르난도 아라발 작

Pique-nique en campagne, Fernando Araval

슬픔

여자독백: '밤마다 실연' 중에서 진자, 이근삼 작,

남자독백: '유리동물원' 중에서 톰, 테네시 윌리엄즈 작, The Glass Menagerie, Tennessee Williams

대화: '그러나 낮설지 않은...', 양지웅 작

화남

여자독백: '겨울새' 중에서 부인, 김수현 작,

남자독백: '택시 택시' 중에서 운전수, 김상수 작,

대화: '덕혜옹주', 정복근 작

공포

여자독백: '심연의 다리' 중에서 무경, 오학영 작

남자독백: '날개' 중에서 나, 이상 원작, 윤고성 각색

대화: '불어를 하세요', 머레이 슈스갈 작

Parlez vous Francais?, Murray Schisgal

지루함:

여자독백: '침입자' 중에서 화분, 조용석 작

남자독백: '침입자' 중에서 작가, 조용석 작

대화: '침입자' 중에서 작가와 시인의 대화, 조용석 작

VI. 녹음과정

각 정서상태별로 작성된 대본에 의해 두 명의 전문 배우가 수집된 정서적 음성은 평균피치범위, 피치 변동율을 측정하여

IV. 수집된 정서음성의 특성

수집된 정서음성으로부터 평균피치값, 상대적 피치값의 범위를 측정하였다.

V. 결론

본 논문에서는 매 통화마다 변화하는 채널의 변이를 4연속 숫자 전화음성에 대해 특징파라미터를 기반으로 기존의 보상기법인 CMN, RTCN, RASTA에 따라 비교 분석하였고, 인식 실험을 통해 인식률도 확인하였다. 변이분석의 결과, RTCN의 경우가 가장 변이가 작았고 또한 인식성능도 가장 높았다. 그러나 CMN이나 RASTA는, 채널의 변이 분석의 결과와 인식성능과의 관계가 일치하지 않음을 확인할 수 있었는데, 이는 화자의 변이를 고려하지 않은 상태에서 변이분석이 이루어졌기 때문이라 사료된다. 따라서 향후, 화자의 변이를 포함한 채널왜곡 특성을 분석하여, 인식성능과의 관계를 확인해야 한다.

참고문헌

- [1] P.J. Moreno, "Speech Recognition in Telephone Environment," MS. Thesis, CMU
- [2] C. Mokbel, J. Monne and D. Juvet, "On -line adaptation of a speech recognizer to variations in telephone line condition," *Proc. Eurospeech*, pp.1247-1250, 1993
- [3] H. Hermansky and N. Morgan, " RASTA Processing of speech," *IEEE Trans. Speech Audio Processing*, Vol.2, No.4, pp.578-589, 1994
- [4] A. Acero, "Environmental Robustness in

Automatic Speech Recognition," *Proc. ICASSP*, pp.849-852

- [5] J. D. Veth and L. Boves, " Comparision of channel normalization technique for automatic speech recognition over the phone," *Proc. ICSLP*, pp.2332-2335, 1996
- [6] J. G. Wilpon, C. H. Lee, and L. R. Rabiner, "Improvements in the Connected Digit Recognition Using Higher Order Spectral and Energy Feature," *Int. Conf. on Acoustics, Seech, and Signal Processing*, vol.1, pp.349-352, 1991
- [7] 김상진, 서영주, 한민수, "LCMS를 이용한 한국어 연속 숫자인식에 관한 연구," *한국음향학회논문집*, Vol.20, pp.43-46, 2001