배우에 의한 한국어 정서음성 데이터베이스 수집

조철우*, 박일서*, 이용주, 김봉완
창원대학교 메카트로닉스 공학부*, 원광대학교 SiTEC

Collection of Korean Emotional Speech Database from Actors

Cheolwoo Jo*, Il-suh Bak*, Yongju Lee, Bongwan Kim
*School of Mechatronics, Changwon National University, cwjo@sarim.changwon.ac.kr
SiTEC, Wonkwang University, yjlee@wonkwang.ac.kr

요약

본 논문에서는 한국어 정서음성 데이터베이스를 수집하는 과정을 기술하고 및 데이터베이스의 특성에 관해서 논의한다. 데이터베이스는 배우로부터 수집되었으며 주관적 평가에 의해 평가되었다. 배우는 남녀 각 3인씩 총 6인이며, 6가지 정서상태에 의해 10개의 문장을 발성하였고 20명의 평가자가 음성에 포함된 정서상태를 독립적으로 평가하였다. 작성된 데이터베이스는 임의제시 방법에 의한 주관적 평가결과 80%이상의 일치도를 얻었다.

1. 서론

인간은 다양한 정서를 가지고 있고 다양한 방법을 통해 표출된다. 인간의 음성이 인간의 생각을 표출하는 수단으로 사용됨을 생각할 때 음성 신호에는 여러 가지 정서가 담겨 있다. 최근 음성인식, 합성등에 관한 연구가 활기를 띰에 따라 보다 상세한 정서에 관한 정보를 얻고자 하는 연구가 진행되고 있다.[1][2][3][4][5][6] 이러한 정서음성 처리에 관한 연구의 기본은 어떻게 유용하게 활용할 수 있는 정서음성을 얻을 수 있는가 하는 것이다. 인간의 정서는 동일한 정서라고 할지라도 다양한 상황에 따라서 전혀 다른 형태로 나타날 수 있기 때문에 정서음성의 수집에는 많은 어려움이 따른다. 대부분의 정서음성에 관한 연구에 사용된 음성데이타베이스는 배우를 통하여 정서상태를 모사하도록 하여 발성하도록 한 음성을 활용하지만 다큐멘타리에 의한 방법, 게임에 의한 방법, 자전적 회상에 의한 방법, 드라마 또는 영화에 의한 방법등 여러 가지 방법이 시도되고 있다. 그러나 각 수집방법에는 모두 나름대로의 문제점을 안고 있으며 원하는 정서를 자연스런 상태로 얻고자 하는 의도에 전적으로 맞는 경우는 한가지 방법으로는 매우 어렵다. 배우에 의한 방법이 그 중에서도 가장 상황 제어가 잘 되는 방법이기 때문에 현존하는 많은 데이터베이스는 이 방법을 사용하고 있다.

본 논문에서는 배우에 의한 정서유발 방법을 이용하여 정서음성 데이터베이스를 수집하고 수집된 데이터를 제3자의 주관적 평가법에 의해 평가한 과정을 소개하고자 한다.

2. 녹음과정

2.1 화자정보

정서음성을 녹음할 화자로는 연극배우 남,녀 각 3인을 선정하였다. 배우들은 각각 24세에서 31세까지의 연령이며 배우경력이 3년에서 10년까지 있는 직업배우를 택하였다.

2.2 발성내용

정서음성을 발성할 문장으로는 다음과 같이 10개의 중의성 문장을 택하였다.

(1). 난 가지말라고 하면서 문을 닫았어.
(2) 이건 내가 원하던 게 아니야.
(3) 야, 이제 그만하자.
(4) 정말 그렇단 말이야.
(5) 예
(6) 아니오.
(7) 나도 몰라.
(8) 우리가 하는 일이 얼마나 중요한지 너는 모를거야.
(9) 지금 어디가는 거야.

(10) 바람과 햇님이 서로 힘이 더 세다고 다투고 있을 때 한 나그네가 따뜻한 외투를 입고 걸어 왔습니다.

2.3 정서의 종류 및 유발방법

녹음에 적용된 정서는 6가지로 낭독체, 기쁨, 화남, 슬픔, 공포, 지루함 이다. 이 중 낭독체는 특별한 정서상태가 없는 중성 상태를 나타내는 문장으로 택하였다. 그외의 5가지 정서는 기존의 정서음성 관련 연구에서 변별력이 확실하고 기본적인 정서라고 간주되는 것들을 택하였다.

정서를 유발하는 방법은 배우에 의한 방법, 자전적 회상을 통한 방법, 게임에 의한 방법, 드라마 또는 다큐멘타리에 의한 방법등 여러 가지가 있을 수 있으나 각각의 방법들이 갖고 있는 한계성 때문에 가장 실현성이 좋으면서 다양한 정서상태를 효과적으로 유발할 수 있다고 여겨지는 배우에 의한 방법을 택하였다.

배우에 의한 방법에서 정서상태는 배우 개개인의 연기 능력에 의해 결정된다. 각 배우의 경험적인 판단에 의해 개별 정석가 도출되고 음성으로 표현된다.

배우들에게는 주어진 10개의 문장에 대하여 각각 6가지 다른 정서상태로 발성하도록 요구한다. 이 때 한 정서상태에 관해 모든 녹음이 끝난 후 다른 정서상태로 녹음을 시행하였다. 정서의 유발기간은 배우에 따라 상이하였다.

2.4 녹음환경

녹음은 외부대비 10dB정도의 소음감쇠가 있는 간이 방음실에서 행해졌다. 녹음장비에 대한 사양은 다음과 같다.

녹음장비: DAT recorder Sony 59ESJ

마이크로폰: AKG C414B-ULS microphone

표본화: 48000Hz, 16Bit

2.5 수집절차

수집은 다음과 같은 절차로 이루어 졌다.

(1) 화자 명세 기록

(2) 녹음 대본을 주고 화자에게 읽어보게 함

(3) 녹음은 화자가 마이크 앞에 앉자 마자 시작해서 자리를 뜨지 않는 한 계속한다.

(4) 한가지 정서상태에 대한 녹음이 전체 문장에 대해 끝날 때 까지 녹음하고 나서 다른 정서상태로 넘어간다.

(5) 각 정서상태에서 말하는 도중 연습을 위하여 여러번 반복하는 것도 포함하여 기록한다.

3. 데이터베이스의 구성

화자는 남자와 여자로 구분하고 2장의 CD에 남, 녀 각 3인씩의 음성데이타를 저장한다. 데이터는 가공의 정도에 따라 (1) 원녹음자료, (2) 1차편집DB (3) 최종DB의 세 단계로 저장된다. 원녹음자료는 초기 녹음된 자료가 편집 없이 그대로 녹음되어 있다. 이 자료는 향후 별도의 처리가 필요할 경우를 대비해서 원본을 그대로 보존하였다. 정서음성의 경우 아직 분석방법에서도 정해진 기준이 없으므로 향후 다양한 분석을 고려하여 원 녹음을 자료로 저장하였다. 두 번째 1차 편집DB에서는 정해진 문장이 발성된 경우를 모두 저장하였다. 녹음 과정에서 배우는 자연스럽게 연습을 포함해서 여러 번 발성하게 되는데 이들을 모두 편집하여 저장하였다. 최종DB는 1차 편집DB에서 가장 발성상태 및 정서상태가 양호하다고 생각되는 음성을 택하여 저장한 것으로 향후 검증을 거쳐 최종DB로 사용된다.

편집된 각 문장음성은 다음과 같이 이름이 붙여져서 저장된다.

1차편집DB의 경우

정서정보 1자리

녹음문장번호 2자리

녹음횟수 1자리

발성자이름 3자리

+ '.wav'

예) cwj라는 화자가 화남이라는 정서의 3번째 발성한 문장은 다음과 같이 이름이 붙는다.

a032cwj.wav

최종편집DB는 녹음횟수를 제외한 부분으로 이름이 구성된다.

예) a03cwj.wav

4. 검증과정

작성된 정서음성DB의 유효성을 검증하기 위하여 주관적 의견제시 방법에 의한 청취실험을 행하였다. 화자별로 60개의 음성을 임의의 정서와 순서로 각 청취자에게 들려주고 6가지 정서상태중 하나를 택하도록 하여 시행하였다.

실험에 참가한 청취자는 남자18명, 여자2명 총 20명으로 구성되었다. 그림1은 청취실험에 사용된 프로그램

의 메뉴를 보여준다.

검증순서는다음과같다

(1) 본인의이니셜과테스트날짜를기록하고, 테스트를 시작한다

(2) 임의로제시된정서음성을듣고적당한정서버튼을 선택한다

(3) 정서를결정하기전에는여러번의반복청취가 가능하다

(4) 전체데이터를청취할때까지 (2),(3)의과정을 반복한다

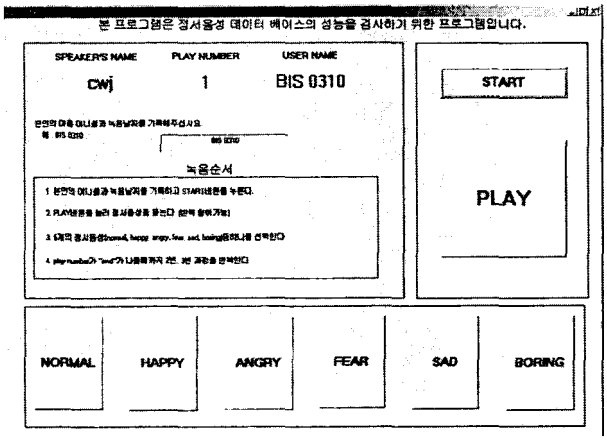

그림1. 검증실험 화면

검증결과는 표1, 표2, 표3과 같이 나타났다.

표1. 화자별 테스트 결과

CWJ						
	낭독체	기쁨	화남	슬픔	공포	지루함
낭독체	179	2	10	0	3	6
기쁨	10	187	3	0	0	0
화남	14	3	177	5	0	1
슬픔	6	1	7	171	14	1
공포	16	0	2	50	118	14
지루함	8	0	2	0	4	186

KKS						
	낭독체	기쁨	화남	슬픔	공포	지루함
낭독체	125	0	20	0	9	46
기쁨	3	181	5	3	7	1
화남	1	1	184	6	6	2
슬픔	3	0	12	161	22	2
공포	2	1	0	25	171	1
지루함	19	0	2	0	15	164

LHJ						
	낭독체	기쁨	화남	슬픔	공포	지루함
낭독체	167	0	12	0	2	19
기쁨	31	135	17	3	13	1
화남	2	0	196	1	1	0
슬픔	7	0	7	169	17	0
공포	6	1	0	8	177	8
지루함	24	0	2	0	6	168

MYS						
	낭독체	기쁨	화남	슬픔	공포	지루함
낭독체	169	0	3	1	0	27
기쁨	7	183	6	0	2	2
화남	15	2	180	2	1	0
슬픔	2	0	3	179	16	0
공포	2	1	0	6	187	4
지루함	19	0	1	1	17	162

PYH						
	낭독체	기쁨	화남	슬픔	공포	지루함
낭독체	170	1	1	0	4	24
기쁨	4	190	0	5	0	1
화남	1	0	198	1	0	0
슬픔	1	0	3	188	8	0
공포	0	0	0	77	123	0
지루함	6	0	0	0	5	189

YSW						
	낭독체	기쁨	화남	슬픔	공포	지루함
낭독체	200	1	2	0	0	7
기쁨	8	184	7	3	7	1
화남	2	0	207	1	0	0
슬픔	0	0	1	186	23	0
공포	13	0	0	51	126	20
지루함	6	0	3	0	3	198

테스트를 수행한 검증인 별, 화자 별 정서를정확히 판정했을 때의 일치도테스트결과와이들값들의평균을 퍼센트(%)로 나타낸정보는다음과같다. 각 배우에 대한 6가지정서의평균일치도는 86.6% 였다. 평균 일치도에 있어서 배우별로는 MYS가 88.3%로 가장 높은 일치도를 보였고, KKS가 82.2%로 가장

낮은 일치도를 보였다.

표5는 각 화자별, 정서별 정서일치도를 보여준다. 낭독체를 제외하고는 화남이 가장 높은 일치도(94.3%)를 보였고 공포가 가장 낮은 일치도(80.3%)를 보였다. 낭독체의 경우는 다른 정서와 혼동된 경우가 많아서 낮은 일치도를 보이고 있는데 지루함과 가장 많은 혼동을 보였다. 정서간의 혼동율이 일어난 경우는 2%이상의 혼동이 발생한 경우를 보면 기쁨은 낭독체 및 공포와, 화남은 낭독체와, 슬픔은 공포 및 화남과, 공포는 낭독체 및 슬픔과, 지루함은 낭독체 및 공포와 가장 높은 혼동율을 보였다.

표2. 검증인별 정서 일치도 테스트 결과

	발성자						
	CWJ	KKS	LHJ	MYS	PYH	YSW	평균
ABK	78.3	85.0	86.7	90.0	86.7	86.7	85.6
BIS	90.0	95.0	91.7	98.3	93.3	93.3	93.6
CDH	75.0	71.7	80.0	90.0	86.7	80.0	80.6
CHW	85.0	90.0	95.0	93.3	91.7	91.7	91.1
HJM	81.7	80.0	86.7	86.7	86.7	85.0	84.4
JKS	86.7	70.0	66.7	81.7	86.7	85.0	79.4
KCM	76.7	81.7	80.0	93.3	91.7	86.7	85.0
KEJ	86.7	75.0	93.3	95.0	91.7	96.7	89.7
KHJ	90.0	71.7	91.7	86.7	88.3	86.7	85.8
KJP	90.0	91.7	85.0	93.3	95.0	95.0	91.7
KMH	86.7	81.7	76.7	80.0	91.7	83.3	83.3
KSC	75.0	81.7	75.0	80.0	91.7	90.0	82.2
KSM	83.3	83.3	83.3	95.0	91.7	86.7	87.2
PHS	85.0	80.0	76.7	78.3	71.7	71.7	77.2
PMC	86.7	85.0	86.7	88.3	88.3	88.3	87.2
PSI	88.3	90.0	88.3	98.3	91.7	83.3	90.0
RJE	98.3	88.3	91.7	95.0	93.3	95.0	93.6
SCS	80.0	76.7	88.3	86.7	85.0	91.7	84.7
YKR	90.0	78.3	86.7	80.0	81.7	90.0	84.4
평균	84.8	82.2	84.3	88.3	83.2	87.6	85.9

표3. 정서별 일치도

	낭독체	기 쁨	화 남	슬 픔	공 포	지루함
CWJ	89.5	93.5	88.5	85.5	59.0	93.0
KKS	62.5	90.5	92.0	80.5	85.5	82.0
LHJ	83.5	67.5	98.0	84.5	88.5	84.0
MYS	84.5	91.5	90.0	89.5	93.5	81.0
PYH	85.0	95.0	99.0	94.0	61.5	94.5
YSW	95.0	89.5	98.5	89.5	93.5	81.0
평균	83.3	87.9	94.3	87.3	80.3	85.9

5. 결론

본 논문에서는 한국어 정서음성의 연구용 데이터베이스를 구축하기 위한 전단계로서 소규모의 정서음성 데이터베이스를 구축하였다. 배우를 이용한 정서음성수집방법이 사용되었으며 각 음성의 정서상태는 주관적 청취실험에 의한 결과 80~94 퍼센트의 일치도를 보였다.

수집에 사용된 배우에 의한 방법은 손쉽게 정서적 음성을 얻을 수 있다는 잇점이 있어서 채용하기는 하였으나 정서음성의 자연성면에서 어떠한지 평가가 이루어지지 않고 있으며, 다양한 정서적 상황의 정의도 이루어져 있지 않고 있기 때문에 이러한 문제들이 향후 보완되어야할 점들이다.

참고문헌

1. 조은경, 조철우, 민경환, "자전적 회상을 통한 자연스런 정서음성정보 수집방법에 관한 연구",한국음향학회지, 제16권 제2호, pp.66-70, 1997

2. 조철우, 조은경, 민경환, "정서정보의 변화에 따른 음성신호의 특성분석에 관한 연구",한국음향학회지, 제16권 제3호, pp.33-37, 1997

3. Cheolwo Jo, 'Aspects on the collection and applications of multimedia emotional speech database', Proceedings of Oriental COCOSDA'98, pp.27-31, 1998

4. I.R.Murray, J.L.Arnott, 'Toward the Simulation of emotion in Synthetic Speech: A Review of the literature on Human Vocal Emotion',JASA, Vol.93, No.2, pp.1097-1108, 1993

5. K.R.Sherer, 'Vocal affective expression: A review and a model for future research', Psychological Buletin, vol.99, pp.143-165, 1986

6. Cheolwoo Jo, 'Comparisons of two mood induction methods for collecting emotional speech signal', Proceedings of ICSP97, pp.65-69, 1997