ISSN 1598-0170 (Print)
ISSN 2287-1136 (Online)
http://www.jksii.or.kr

감성인식과 핵심어인식 기술을 이용한 고객센터 자동 모니터링 시스템에 대한 연구

A Study on the Automatic Monitoring System for the Contact Center Using Emotion Recognition and Keyword Spotting Method

윤 원 중* 김 태 홍** 박 규 식***
Won-Jung Yoon Tae-Hong Kim Kyu-Sik Park

요 약

본 논문에서는 고객의 불만관리 및 상담원의 상담품질 관리를 위한 고객센터 자동 모니터링 시스템에 대한 연구를 진행하였다. 제안된 시스템에서는 평상/화남의 2가지 감성에 대한 음성 감성인식 기술과 핵심어인식 기술을 사용하여 상담내역에 대한 보다 정확한 모니터링이 가능하고, 욕설, 성희롱 등의 언어폭력을 일삼는 고객에 대한 전문상담 및 관리가 가능하다. 서로 다른 환경에서 구축된 이종 음성 DB를 이용하여 불특정 고객들의 질의 음성에 안정적으로 동작할 수 있는 알고리즘을 개발하였으며, 실제 고객센터 상담내역 데이터를 이용하여 성능을 검증하였다.

ABSTRACT

In this paper, we proposed an automatic monitoring system for contact center in order to manage customer's complaint and agent's quality. The proposed system allows more accurate monitoring using emotion recognition and keyword spotting method for neutral/anger voice emotion. The system can provide professional consultation and management for the customer with language violence, such as abuse and sexual harassment. We developed a method of building robust algorithm on heterogeneous speech DB of many unspecified customers. Experimental results confirm the stable and improved performance using real contact center speech data.

☞ keyword : 감성인식(Emotion Recognition), 핵심어인식(Keyword Spotting), 고객센터(Contact Center), 자동 모니터링 시스템(Automatic Monitoring System)

1. 서 론

최근 고객들에게 보다 개선된 서비스를 제공하고 기업의 경쟁력을 강화시키는 마케팅 수단의 일환으로 고객센터를 운영하는 기업들이 증가하고 있다. 이러한 고객센터의 상담원들은 고객들에게 기업이 제공하는 다양한 서비스와 정보를 알려주는 역할을 하는데, 대부분의 고객들은 상담원의 신속하고 친절한 응대와 만족스러운 답변을 얻을 수 있을 때 해당기업에 대한 신뢰도가 높아진다고 한다. 그러나 불친절한 응대와 성의 없는 답변으로 고객의 요구를 충족시키지 못했을 때는 기업 이미지에 큰 타격을 줄 수도 있다. 이렇게 기업의 이미지와 평가가 상담원의 상담결과에 따라 영향을 받을 수 있기 때문에, 기업은 상담원의 교육과 관리 차원에서 통화품질관리(Quality Management) 시스템을 운영하고 있다. 현재 시스템에서는 QA(Quality Assurance) Analyst가 상담원들의 상담 내용을 직접 청취하면서 상담원들이 처리한 내역을 평가하는 방법을 사용하고 있다. 그러나 이러한 방법은 상담원이 하루에 처리해야하는 통화 양이 너무 많기 때문에 비용이 너무 많이 발생하고, 현재는 임의로 선택된 통화에 대해서만 평가를 하고 있어 제대로 된 평가가 불가능하다는 현장의 목소리가 높은 실정이다. 또한, 현 시스템 체계는 상담원들에게 업무 평가에 대한 커다란 심리적 압박을 주어, 고객의 성희롱 및 언어폭력 등에 적극적으로 대처하지 못하게 되는 커다란 원인으로 작용하고 있다. 실례로 상담원의 93.2%가 언어폭력을 경험한 적이 있으며, 40% 이상이 성희롱적 언사를 경험하고 있다고

* 정 회 원 : 단국대학교 소프트웨어학과 연구교수
helloril@dankook.ac.kr
** 정 회 원 : 단국대학교 소프트웨어학과 석사과정
peppy0510@dankook.ac.kr
*** 종신회원 : 단국대학교 소프트웨어학과 교수(교신저자)
kspark@dankook.ac.kr
[2012/04/09 투고 - 2012/04/17 심사 - 2012/05/18 심사완료]

http://dx.doi.org/10.7472/jksii.2012.13.3.107

한다. 그러나 이러한 불쾌한 상황에서의 대처방안으로는 고객이 그만둘 때까지 기다리거나(32.7%), 전화를 먼저 끊을 수 없기에 계속 당할 수밖에 없다(56.5%) 등의 소극적 대처방식을 취하는 것으로 나타났다[1]. 이는 언어폭력시의 시스템적인 응대방안이 없거나, 전화를 먼저 끊었을 때의 불만제기 및 민원접수 등의 피해를 받을 수 있기 때문으로 상당한 스트레스를 받는 것으로 파악된다.

이에 본 논문에서는 상담내용에 대한 보다 정확한 모니터링이 가능하고, 욕설, 성희롱 등의 언어폭력을 일삼는 고객에 대한 전문상담 및 관리가 가능하도록 음성 감성 인식기술과 핵심어 인식기술을 이용한 자동 모니터링 시스템에 대한 연구를 진행하였다.

본 고객센터 자동 모니터링 시스템은 고객의 음성을 평상과 화남으로 분석하여 화남 감성으로 인식된 상담내역과 고객이 상담하면서 사용한 욕설 등의 단어가 인식되었을 경우에 대해서만 QA Analyst가 모니터링을 할 수 있도록 하여, QA Analyst의 업무 효율성을 높이고 불만이 있는 고객에 대하여 적절한 상담이 이루어졌는지 확인할 수 있게 도와준다. 또한, 실시간 모니터링시 욕설 및 성희롱에 해당하는 단어가 인식되었을 경우 전담 상담원에게 연결할 수 있도록하여 상담원들의 고충을 덜어줄 수도 있다. 본 논문의 구성은 다음과 같다. 2장에서는 음성 감성인식 기술과 핵심어인식 기술에 대한 기존 연구에 대하여 살펴보고, 3장에서는 제안된 자동 모니터링 시스템의 전반적인 구조와 기술에 대하여 설명한다. 4장에서는 실험을 통해 제안 시스템의 성능을 분석하고, 마지막으로 5장에서는 결론으로 끝을 맺는다.

2. 기존 연구동향

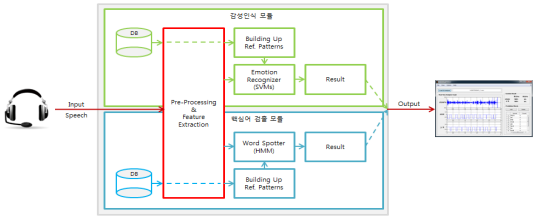
기존의 음성 감성인식에 대한 연구들로, S. Pathak[2]는 LPC 12차 계수와 NN(Neural Network)을 이용하여 슬픔, 화남, 기쁨, 공포 그리고 지루함에 대하여 분류하였고, C. Busso[3]는 화자간의 차이를 최소화하기 위해 평상 음성에서 정규화 매개변수를 추출하여, 다른 감성 음성들에 적용할 수 있는 iterative feature normalization 방법을 제시하였다. 이외에도 음성 감성인식 기술을 HCI용 SW나 로봇 등의 응용분야에 적용하기 위한 활발한 연구가 진행되고 있다[4-8]. 그러나 기존 대부분의 연구가 화자종속 시스템에 대한 연구들이며, 화자독립 시스템에 대한 연구도 동일 환경에서 녹취된 감성음성 DB를 이용하고 있기 때문에, 고객센터와 같이 불특정 다수 화자들의 감성을 인식해야하는 시스템에의 적용이 쉽지 않은 실정이다.

반면 음성인식 기술은 불특정화자에 대한 대어휘 연속 음성인식 시스템이 궁극적 목표이지만, 응용분야에 따라서는 연속 음성내에서 미리 지정된 몇 개의 핵심어의 위치를 찾아내는 핵심어인식 기술이 유용하게 사용된다. 핵심어 인식기술에 대한 연구로는 시간축에서 왜곡되어 있는 두 음성 데이터를 정합시켜 주는 DTW 방법이 있다[9]. DTW는 많은 계산량으로 인해서 인식시간이 많이 소요되고 화자독립 환경에서 성능이 저하되는 단점이 있으나, 적은 양의 음성 데이터로도 동작 가능한 인식시스템의 개발이 가능하다는 점에서 여전히 강점을 가지고 있다[10]. 그러나 최근에는 hidden markov model(HMM)을 이용하여 핵심어와 비핵심어들을 각각 모델링한 후 이 모델들을 연결한 형태로 네트워크를 구성하고 인식의 결과를 나타내는 방법이 주로 이용된다. 이러한 시스템에서는 비핵심어에 대한 모델을 얼마나 효과적으로 모델링하느냐와 핵심어로 오인식된 단어를 어떻게 효과적으로 거절할 것인지에 대한 연구가 진행되고 있다[11, 12]. 각 모델들을 훈련시키기 위해서는 많은 양의 음성 데이터를 필요로 하지만 충분한 훈련용 데이터가 준비된다면 DTW를 이용한 방법보다 우수한 성능을 나타낸다.

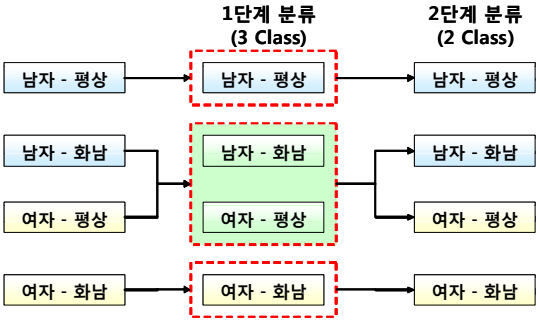
또한, 실제 고객센터의 상담내역에 대한 연구로는 고객의 음성을 평상, 긴장, 투덜거림, 따로 분류할 수 없음 등의 4가지로 분류하여 54%, 평상과 불만(긴장, 투덜거림, 따로 분류할 수 없음을 통합)에 대하여 73%의 성능을 보고한 연구[13]와, 감성인식 결과(평상 50%, 기쁨 57.1%, 화남 60%)와 speech-to-text(STT) 후 텍스트 감성분석 결과(평상 66.7%, 기쁨 71.4%, 화남 60%)를 반영하여 감성인식 성능을 향상(평상 83.3%, 기쁨 71.4%, 화남 80%)시키는 방법을 제시한 연구[14]가 보고되고 있다. 그러나 이러한 연구들도 인위적으로 구축된 음성 데이터가 아닌 불특정 다수의 실제 상담내역을 사용하여 고객의 상태만을 모니터링 하였을 뿐, 고객과 상담원 그리고 기업 모두를 위한 진정한 의미의 고객센터 모니터링 시스템에 대한 연구라고 보기는 힘든 실정이다.

3. 제안 시스템의 구조

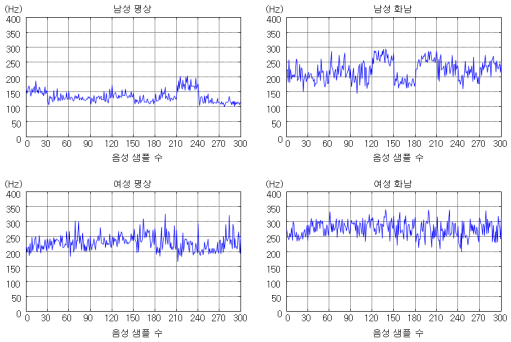
(그림 1)은 본 논문에서 제안한 고객센터 자동 모니터링 시스템의 구조도로써, 고객의 음성이 입력되면 감성인식 모듈과 핵심어인식 모듈을 나란히 거쳐서 현재 상담중인 고객의 상태 및 언어폭력 상황 등을 파악할 수 있다.



(그림 1) 제안된 시스템의 구조



(그림 2) 평상/화남 감성의 2단계 분류 체계



(그림 3) 남·녀 성별에 따른 Pitch mean의 감성별 특성 변화 비교

3.1 음성 감성인식

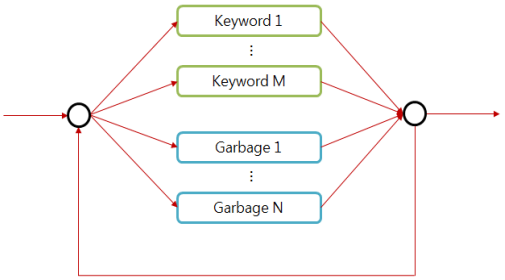
본 논문에서는 평상/화남의 2가지 감성의 분류에 대하여 (그림 2)와 같이 남·녀 성별에 따른 감성별 특성 변화를 적용한 2단계 분류 체계를 이용한다[15].

이는 (그림 3)에서와 같이 감성인식에서 중요한 특징으로 알려져 있는 pitch mean 값의 범위가 남성 평상에 대해 가장 낮고(약 160Hz이하), 남성 화남과 여성 평상의 범위가 유사한 영역에 존재(약 160Hz이상, 250Hz이하)하며, 여성 화남이 가장 높은 범위에 존재(약 250Hz이상)하는 결과를 반영한 것이다.

먼저 1단계 분류에서는 pitch mean 특징을 기준으로 질의 음성을 (남성 화남-여성 평상), (남성 평상), (여성 화남) 등 3개의 class로 분류하고, 2단계에서는 Mel-frequency cepstral coefficients(MFCCs)[16]와 에너지 특징을 이용해 남·녀 성별을 구분해 남성 화남과 여성 평상 음성을 분류한다.

3.2 핵심어인식

본 논문에서는 핵심어인식을 위한 음성 특징벡터로 12차 MFCC와 로그 에너지 그리고 이들의 1·2차 미분 계수를 포함한 39차 벡터를 사용하였고, HMM을 기반으로 하여 핵심어 모델과 비핵심어(garbage) 모델로 구성된 핵심어인식 시스템을 사용하였다. 각 모델은 triphone을 사용하여 sub-word 단위를 기반으로 모델링 하고, 각 triphone은 3개의 state를 가지는 left-to-right HMM 방법을 사용하였으며 각 state는 7개 Gaussian의 mixture로 모델링 하였다.



(그림 4) 핵심어인식 시스템의 구조

4. 실험환경 및 결과

본 논문에서는 감성인식 및 음성인식 실험을 위한 데이터베이스로 (표 1)과 같이 서로 다른 음성 녹취 환경에서 구축된 4가지의 DB를 이용하여 다음과 같은 2가지 주안점을 두고 실험을 진행하였다.

첫째, 화자독립-문장독립 환경에서 시스템 평가에 사용되는 감성 훈련용 음성과 질의 음성이 동일한 환경에서 녹취된 단일 DB인 실험과 서로 다른 녹취 환경에서 구축된 이종 DB간의 실험을 통하여 기존 연구들의 문제점을 파악하고 화자독립-문장독립 환경에서 강인한 성능을 나타내는 감성인식 시스템을 구축한다.

(표 1) 실험에 이용된 DB에 대한 간략한 설명

DB 구분	DB1	DB2	DB3	DB4
DB에 대한 설명	논문[17] DB	TV 드라마 DB	본 연구팀 DB	고객센터 상담 DB
녹음 인원	아마추어 연극단원 남·녀 각15명	전문 연기자 남·녀 각 30명	아마추어 연극단원 남·녀 각 20명	실제 고객센터 상담원
발성 문장	모든 감성에 동일한 45 문장	각기 다른 2,000 문장	각 감성별 100 문장	실제 상담내용
음성 파일수	2,700개	2,000개	8,000개	17,252개
샘플 링주파수	8kHz	8kHz	8kHz	8kHz
특징	화자 독립 문장 종속	화자 독립 문장 독립	화자 독립 문장 독립	화자 독립 문장 독립

둘째, 핵심어인식 기술을 사용하여 감성인식기의 성능개선 효과를 파악하고 실제 고객센터 상담내역을 이용한 실험으로 성능을 검증한다.

본 논문에서는 실험을 위해 Support Vector Machines (SVMs)[18] 분류기를 사용하였으며, 실험결과는 (표 2~5)의 1열에 있는 감성을 갖는 질의 음성들에 대한 인식결과를 2~5열에 표시하였다. 또한, 제안된 방법과의 비교를 위한 실험에서는 특징추출 단계에서 추출한 모든 특징들(pitch, MFCC 그리고 에너지)을 사용하여 단번에 분류하는 전통적인 방법을 사용하였다.

4.1 실험 1 – 이종 훈련용 음성 DB와 질의 음성의 영향 비교〔15〕

실제 기업환경에 적용될 수 있는 화자독립-문장독립 감성인식 시스템의 성능평가를 위해 훈련용 음성과 질의 음성이 동일한 녹취 환경인 경우와 서로 다른 경우에 대한 실험을 진행하였다. 이를 위해 TV 드라마 대사를 녹음한 DB2와 본 연구팀이 준비한 DB3를 사용하였다.

TV 드라마 대사를 녹음한 DB2는 남·녀 각 30명의 연기자가 TV 드라마에서 발성한 대사를 녹음한 것으로 총 2,000개의 음성 파일로 구성되며, 각 화자별로 중복되는 문장이 없다는 것이 특징이다.

(표 2) 훈련용 음성(D)과 질의 음성(Q)의 녹취 환경이 다른 경우의 실험 결과

DB1(D) vs DB2(Q)		
	평상	화남
평상	368	132
화남	87	413
		78.1%

DB1(D) vs DB3(Q)		
	평상	화남
평상	343	157
화남	7	493
		83.6 %

DB2(D) vs DB1(Q)		
	평상	화남
평상	424	76
화남	217	283
		70.7 %

DB2(D) vs DB3(Q)		
	평상	화남
평상	420	80
화남	70	430
		85.0 %

DB3(D) vs DB1(Q)		
	평상	화남
평상	500	0
화남	490	10
		51.0 %

DB3(D) vs DB2(Q)		
	평상	화남
평상	490	10
화남	385	115
		60.5 %

본 연구팀의 DB3는 TV 드라마 대사들을 참고하여 비교적 감성이 잘 표현될 수 있는 2~3초 길이의 문장으로 평상/화남 각 감성별 100문장으로 구성하였으며, 아마추어 연극단원 남·녀 각 20명이 발성한 총 8,000개의 음성 파일로 구성된다. 또한, 실험결과를 쉽게 비교할 수 있도록 이후의 실험에서 질의 음성 파일의 수는 1,000개씩으로 제한하였다.

다음의 (표 2)는 화자독립-문장독립 환경에서 감성 훈련용 음성과 질의 음성의 녹취 환경이 서로 다른 경우에 대한 실험 결과로, 이는 실제 고객센터에 적용되는 감성인식 시스템의 경우 감성 훈련용 음성과 고객의 질의 음성이 서로 다른 녹취 환경(특성)을 갖기 때문에 반드시 고려되어야 할 사항이다. (표 2)에서 괄호 안의 (D)는 감성 훈련용 음성의 종류를 말하며 (Q)는 시스템 평가에 사용된 질의 음성의 종류를 말한다.

(표 2)에서 보듯이 실험 결과는 50%~85%까지 다양한 성능을 나타내었으며, DB1과 DB2를 훈련하여 실험한 결과는 DB3이 서로 특성이 다른 DB1과 DB2의 훈련결과 모두와 높은 인식률을 나타냄을 확인할 수 있었다. 그러나 이러한 DB3를 훈련용 음성으로 사용했을 때는 한쪽 감성, 즉, 평상 감성으로의 쏠림현상이 나타나며 전혀 예측할 수 없는 현상을 나타낸다. 이는 아무리 잘 구축된 DB를 훈련용 음성으로 사용한다고 해도 어떤 환경에서 녹취된(어떤 특성을 가진) 음성이 들어오는가에 따라 시스템의 성능이 불안정해질 수 있다는 것을 의미한다. 이러한 감성 쏠림 현상은 실용적인 감성인식 시스템의 구축에 가장 큰 성능 저하 요소로 작용할 수밖에 없다.

(표 3) 제안된 방법에 의한 감성 분류 결과

DB3(D) vs DB1(Q)				
	남성 평상	남성 화남	여성 평상	여성 화남
남성 평상	195	55		0
남성 화남	0	205	29	45
여성 평상		1	220	
여성 화남	0	7		243
				86.3 %

DB3(D) vs DB2(Q)				
	남성 평상	남성 화남	여성 평상	여성 화남
남성 평상	198	52		0
남성 화남	30	165	29	63
여성 평상		25	188	
여성 화남	0	50		200
				75.1 %

사용된 감성음성 DB에 따라 약간씩의 차이는 있지만, (그림 3)에서 보듯이 pitch mean 값의 범위는 남성 평상에 대해 가장 낮고(약 160Hz이하), 남성 화남과 여성 평상의 범위가 유사한 영역에 존재(약 160Hz이상, 250Hz이하)하며, 여성 화남이 가장 높은 범위에 존재(약 250Hz이상)한다. 만약 훈련용 음성과 질의 음성의 pitch mean 값의 범위가 유사하다면 큰 문제가 없겠지만, 녹취 환경 및 화자의 차이로 인해 값의 범위에 현격한 차이가 있다면 이는 (표 3)의 하단에 보이는 2개의 실험결과처럼 감성인식 결과가 한쪽으로 치우치게 나타날 것이다. 또한, 질의 음성의 남·녀 성별을 고려하지 않고 평상/화남의 감성분류를 할 경우, 남성 화남과 여성 평상의 특징이 상당히 유사한 값의 영역에 존재하고 있기 때문에 상당한 오 분류를 야기하기도 한다. 이러한 문제를 해결하기 위해 본 연구에서는 (그림 2)와 같이 남·녀 성별에 따른 감성별 특성 변화를 감안한 2단계 평상/화남 감성 분류법을 사용하였다.

먼저 1단계 분류에서는 pitch mean을 기준으로 질의 음성을 (남성 화남-여성 평상), (남성 평상), (여성 화남) 등 3개의 class로 분류하고, 2단계에서는 MFCC와 에너지를 이용해 남·녀 성별을 구분해 남성 화남과 여성 평상 음성을 분류한다. (표 3)는 본 연구에서 제안한 2단계 감성인식 방법으로 실험한 경우의 평상/화남 감성 분류 결과를 나타낸다.

(표 2)와 비교해 볼 때 인식 감성이 한쪽으로 쏠리는 현상이 없음을 볼 수 있을 뿐만 아니라, 실험 결과 역시 51%→86.3%, 60.5%→75.1%로 평균 25%의 성능 개선이 있음을 알 수 있다. 즉, 본 연구의 남·녀 성별 특징 변화를 감안한 2단계 감성 분류를 통해 감성 훈련용 음성과 질의 음성의 녹취 환경이 서로 다른 경우에 발생하던 편향된 인식 결과 문제를 해결할 수 있으며, 이와 더불어 상당한 인식 성능 개선 또한 가져올 수 있음을 알 수 있다.

4.2 실험 2 – 핵심어인식을 이용한 감성인식 및 실제 상담내용에 대한 검증

본 논문에서는 일반적인 핵심어 인식이 아닌 욕설 및 성희롱에 특화된 핵심어를 인식하는 것을 목적으로 하기 때문에 DB1~DB4의 화남 감성에 대한 음성 데이터들을 이용하였다. 참고로 DB4는 연구목적으로만 사용한다는 양해를 얻어 획득한 데이터로 고객센터에서 6일간의 inbound(고객이 고객센터로 연락한 경우) 및 outbound(고객센터에서 고객에게 연락한 경우) 상담내역 700개의 데이터를 상담원 part와 고객 part로 나누어 분석한 결과로 3분~30분 정도의 상담내역을 의미있는 내용을 포함하는 1초~5초 정도의 음성 clip으로 수작업 분할하여 17,252개의 데이터를 얻은 것이다. 그러나 이 DB는 특정목적을 위하여 인위적으로 작성한 데이터가 아니기 때문에 평상과 화남의 비율이 일정하지는 않아, 상담원의 경우 평상 9,627개(99.6%)와 화남 38개(0.4%), 고객의 경우 평상 7,529개(99.2%)와 화남 58개(0.8%)로 평상의 비율이 월등히 높은 특징을 가지고 있다.

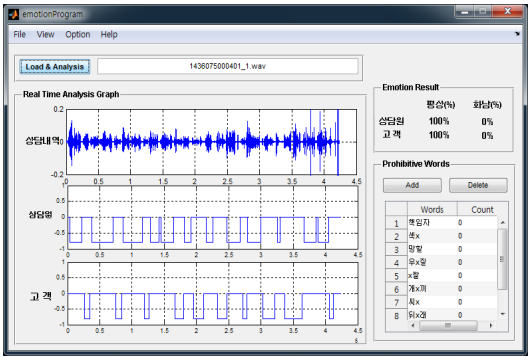
핵심어인식기의 훈련을 위해서는 전체 데이터에서 핵심어에 해당하는 부분을 수작업으로 분할한 음성 clip만을 사용하여 훈련하였다. 다음 (표 4)는 핵심어인식 결과를 포함한 감성인식 결과이다. 감성인식 결과는 질의 음성의 SVM 분류 결과인 화남에 대한 가능성 여부(probability)로 감성으로 판정을 하며, 등록된 핵심어가 인식되었을 경우에는 횟수에 따라 화남에 대한 weight를 20%씩 높여 최종 감성인식 결과에 반영하였다. 예를 들어 질의 음성의 SVM 결과 화남일 확률: 40%에 핵심어가 2회 인식되었을 경우 실험 1에서는 평상으로 판정했으나, 실험 2에서는 화남일 확률: 40%×1.2×1.2=57.6%로 조정되어 화남으로 판정하게 된다.

(표 3)과 비교하여 DB3을 훈련하여 DB1으로 질의한 실험에서는 남성 평상의 분류 결과에서는 변화가 없었으나 남성 화남이 여성 평상으로 오분류되던 문제 중 14건, 여성 화남의 오분류 문제 중 4건이 개선되어 약 2%의 성능이 개선되었고, DB3을 훈련하여 DB2로 질의한 실험에서도 남성 평상의 분류 결과에서는 변화가 없었으나 남

(표 4) 핵심어인식 결과를 포함한 감성인식 결과

DB3(D) vs DB1(Q)				
	남성 평상	남성 화남	여성 평상	여성 화남
남성 평상	195	55		0
남성 화남	0	219	15	45
여성 평상		1	220	
여성 화남	0	3		247
				88.2 %

DB3(D) vs DB2(Q)				
	남성 평상	남성 화남	여성 평상	여성 화남
남성 평상	198	52		0
남성 화남	30	182	12	63
여성 평상		25	188	
여성 화남	0	25		225
				79.3 %



(그림 5) 자동 모니터링 시스템의 UI

성 화남이 여성 평상으로 오분류되던 문제 중 17건, 여성 화남의 오분류 문제 중 25건이 개선되어 약 4%의 성능이 개선되었음을 확인할 수 있었다. 반면, 남성 평상을 남성 화남-여성 평상 그룹으로 오분류하는 문제(55건, 52건)와 남성 화남, 여성 평상을 여성 화남으로 오분류하는 문제(45건, 63건)는 핵심어 검출 방법을 적용하였음에도 여전히 상당한 오류를 보이고 있기 때문에 적절한 후속 연구가 필요할 것으로 사료된다.

또한, 본 논문에서는 (표 4)의 실험 결과를 바탕으로 실제 고객센터의 상담내역으로 구축된 DB4를 이용하여 성능을 검증하였다. DB4의 원 상담내역 700개 중에는 고객이 화(또는 짜증)를 내는 경우가 11건(1.6%)이고 상담원이 화(또는 짜증)를 내는 경우가 1건(0.15%)으로 대부분의 데이터가 평상이며, 화남 12건을 정확하게 검출해냈다. 분석 결과는 (그림 5)와 같이 실시간 상담 또는 고객과 상담원의 음성이 개별적으로 저장되어 있는 경우에는 고객과 상담원 각각의 감성인식 결과를 그래프와 수치로 표현하고 상담내역의 핵심어 인식 결과를 나타내고, 이미 녹취 DB에 상담원과 고객의 음성이 단일 파일로 취합되어 batch로 분석하는 모드에서는 고객과 상담원을 구분하지 않고 결과를 나타낸다. 감성인식 그래프의 하단은 평상 상태를 나타내고 중단(0)은 idle 상태를 나타내며 상단은 화남 상태를 표현한다.

5. 결 론

본 논문에서는 평상/화남의 2가지 감성에 대한 강인한 음성 감성분류 시스템을 이용하여 상담내용에 대한 보다 정확한 모니터링이 가능하고, 욕설, 성희롱 등의 언어폭력을 일삼는 고객에 대한 전문상담 및 관리가 가능하도록 핵심어 검출 기술을 이용한 자동 모니터링 시스템에 대한 연구를 진행하였다. 기존의 감성분류 시스템에 핵심어 검출 기술을 접목시켜 성능을 향상시켰으나 초기 분류시의 오분류로 인한 성능저하 문제는 여전히 해결해야할 과제로 남아있다.

본 콜센터 자동 모니터링 시스템은 고객의 음성을 평상과 화남으로 분석하여 화남 감성으로 인식된 상담내역과 고객이 상담하면서 사용한 욕설 등의 단어가 인식되었을 경우에 대해서만 QA Analyst가 모니터링을 할 수 있도록 하여, QA Analyst의 업무 효율성을 높이고 불만이 있는 고객에 대하여 적절한 상담이 이루어졌는지 확인할 수 있게 도와준다. 또한, 실시간 모니터링시 욕설 및 성희롱에 해당하는 단어가 인식되었을 경우 전담상담원에게 연결할 수 있도록하여 상담원들의 고충을 덜어줄 수도 있다. 제안 시스템은 고객센터, 소방서, 경찰서 등과 같이 불특정 다수로부터 불만이나 공포의 감성을 인식하고 빠르게 대응해야하는 곳에서 유용하게 활용될 수 있을 것으로 예상된다.

감사의 글

이 논문은 2008년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업 연구임(No. KRF-2008-521-H00006)

참 고 문 헌

[1] 유기종, “금융기관 콜센터 근로자의 고충처리에 관한 실증 연구”, 석사학위 논문, 고려대학교, 2010.

[2] S. Pathak and A. Kulkarni, “Recognizing emotions from speech”, International Conference on Electronics Computer Technology (ICECT), pp. 107-109, 2011.

[3] C. Busso, A. Metallinou, and S.S. Narayanan, “Iterative feature normalization for emotional speech detection”, International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 5692-5695, 2011.

[4] C. Busso, S.B. Lee and S. Narayanan, “Analysis of Emotionally Salient Aspects of Fundamental Frequency for Emotion Detection”, IEEE Transactions on Audio, Speech, and Language Processing, vol. 17, no. 4, pp. 582-596, May 2009.

[5] B. Vlasenko, B. Schuller, A. Wendemuth, and G. Rigoll, “Combining Frame and Turn-Level Information for Robust Recognition of Emotions within Speech,” Proc. International Conference Spoken Language Processing, pp. 2225-2228, 2007.

[6] T. Kostoulas, I. Mporas , T. Ganchev , N. Fakotakis, “The Effect of Emotional Speech on a Smart-Home Application”, Lecture Notes in Computer Science Vol. 5027, pp. 305-310, 2008.

[7] J.S. Park, J.H. Kim and Y.H. Oh, “Feature Vector Classification based Speech Emotion Recognition for Service Robots”, IEEE Transactions on Consumer Electronics, Vol. 55, No. 3, pp. 1590-1596, 2009.

[8] Hua A. and D. Litman, “Using System and User Performance Features to Improve Emotion Detection in Spoken Tutoring Dialogs”, ICSLP2006, Pittsburgh, Pennsylvania, USA, 2006.

[9] D. Vicard, “Transient part recognition for continuous speech using transition spotting”, International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 461-464, 1988.

[10] Q. Hong, C. Zhang, X. Chen and Y. Chen, “Embedded speech recognition system for intelligent robot”, International Conference on Mechatronics and Machine Vision in Practice, pp. 35-38, 2007.

[11] R.C. Rose and D.B. Paul, “A hidden Markov model based keyword recognition system”, International Conference on Acoustics, Speech and Signal Processing (ICASSP), Vol. 1, pp. 129-132, 1990.

[12] R.A. Sukkar and C. H. Lee, “Vocabulary independent discriminative utterance verification for nonkeyword rejection in subword based speech recognition”, IEEE Transactions on Speech and Audio Processing, Vol. 4, pp. 420–429, 1996.

[13] K. Vicsi and D. Sztahó, “Emotional state recognition in customer service dialogues through telephone line”, International Conference on Cognitive Infocommunications (CogInfoCom), pp. 1-4, 2011.

[14] P. Gupta and N. Rajput, “Two-Stream Emotion Recognition For Call Center Monitoring”, Interspeech-2007, 2007.

[15] W.J. Yoon and K.S. Park, “Building Robust Emotion Recognition System on Heterogeneous Speech Databases”, IEEE Transactions on Consumer Electronics, Vol. 57, No. 2, pp. 747-750, 2011.

[16] S.J. Young and P.C. Woodland, “state clustering in hidden markov model-based continuous speech recognition”, Computer Speech & Language, Vol. 8, pp. 369-383, 1994.

[17] 강봉석, “음성 신호를 이용한 문장독립 감정 인식 시스템,” 석사학위 논문, 연세대학교, 2001.

[18] C. Cortes and V.N. Vapnik, “Support Vector Network”, Machine Learning, Vol. 20, pp. 1-25, 1995.Barber, R. Flickner, M. Hafner, J. Lee, D. Niblack, W. Petkovie, D. and Ashlen, ’Query By Image Content and Its Applications.’, Proceedings of COMPCON94, pp.424-429, 1994.

◐ 저 자 소 개 ◑

윤 원 중

2003년 상명대학교 정보통신학과(이학사)
2005년 단국대학교 대학원 컴퓨터과학 및 통계학과(이학석사)
2010년 단국대학교 대학원 컴퓨터과학 및 통계학과(이학박사)
2011～현재 단국대학교 소프트웨어학과 연구교수
관심분야 : 음성 및 오디오 신호처리, 감성인식, etc.
E-mail : helloril@dankook.ac.kr

김 태 홍

2010년 단국대학교 컴퓨터공학과(공학사)
2011～현재 단국대학교 대학원 컴퓨터학과(공학석사)
관심분야 : 음악 및 오디오 신호처리, 감성인식, etc.
E-mail : peppy0510@dankook.ac.kr

박 규 식

1986년 NYU-PolytechnicUniversity, EE(공학사)
1988년 NYU-PolytechnicUniversity, EE(공학석사)
1994년 NYU-PolytechnicUniversity, EE(공학박사)
1994～1996 삼성전자 마이크로사업부 선임연구원
1996～2001 상명대학교 컴퓨터·정보통신공학부 조교수
2001～현재 단국대학교 소프트웨어학과 교수
관심분야 : 음성 및 오디오 신호처리, 감성인식, 디지털 통신, etc.
E-mail : kspark@dankook.ac.kr