감성기반 서비스를 위한 통화 음성 감정인식 기법

Call Speech Emotion Recognition for Emotion based Services

저자
(Authors)

방재훈, 이승룡
Jae Hun Bang, Sungyoung Lee

출처
(Source)

정보과학회논문지 : 소프트웨어 및 응용 41(3), 2014.3, 208-213 (6 pages)
Journal of KISS : Software and Applications 41(3), 2014.3, 208-213 (6 pages)

발행처
(Publisher)

한국정보과학회
KOREA INFORMATION SCIENCE SOCIETY

URL

http://www.dbpia.co.kr/Article/NODE02373829

APA Style

방재훈, 이승룡 (2014). 감성기반 서비스를 위한 통화 음성 감정인식 기법. 정보과학회논문지 : 소프트웨어 및 응용, 41(3), 208-213.

감성기반 서비스를 위한 통화 음성 감정인식 기법

(Call Speech Emotion Recognition for Emotion based Services)

방 재 훈 † 이 승 룡 ††
(Jae Hun Bang) (Sungyoung Lee)

요 약 기존의 음성기반 감정인식기술은 콜센터나 메디컬 센터에서 고객이나 환자의 감정을 실시간으로 모니터링 하고 추출된 감정에 적절한 대응을 해주는 서비스 어플리케이션으로 사용되고 있다. 이러한 음성기반 감정인식 기술은 일정 주기 혹은 단위 시간동안의 음성데이터를 분석하여 사용자의 감정을 인식한다. 기존 연구 방법론은 하나의 통화 이벤트 전체에 대한 감정인식이 아닌 통화 중 특정 시간동안의 감정을 인식하는 기술로써, 전체 통화기간동안 감정의 변화를 인식하지 못하여 감정 기복이 있는 통화음성 데이터에서 하나의 감정으로 도출해내는 통화단위 감정인식에는 부적합하다. 본 논문에서는 스마트폰에서 통화 음성을 녹음한 뒤 감정인식 구간을 통화 시작부터 종료시점까지 하나의 Window로 보고, 이를 다수의 Time-Window로 나눈 다음, 통화 종료시점에 가까워지는 Window에 감정생존곡선을 기반한 가중치를 부여하는 기법을 제안한다.

키워드: 통화음성 데이터, 음성기반 감정인식, 틸티드 타임 윈도우, 스마트폰, 감정생존곡선

Abstract Existing speech based emotion recognition is used in call center or in medical center to monitor client's or patient's emotion in real time, and respond in appropriate service. This method analyzes speech data in constant period to recognize user's emotion. Existing researches do not analyze the whole phone call but only specific part, which makes it unable to know the undulation of emotion for the whole period. This is inappropriate to use in the unit of phone call, which should conclude the emotion using whole speech.In this paper, we propose the following method. After recording a phone call, consider the whole phone call as a window and divide it into several Time-Windows, then assign weighted value gradually until the end of the call.

Keywords: call speech data, speech emotion recognition, tilted-time window, smartphone emotion survivor curve

1. 서 론

스마트폰이 보급되면서 사용자 정보를 활용한 다양한 개인화 서비스 연구가 활발히 진행 중이다. 사용자 정보의 예로는 상황 정보 및 감정 정보 등이 있다. 특히 감정 정보는 사용자의 현재 감정 상태를 나타내는 정보로 감정 상태에 따라 달라지는 음악 추천과 같은 문화 콘텐츠 서비스와 콜센터나 메디컬 센터에서 고객 감정 모니터링 등에 매우 유용하다.

음성기반 감정인식이란 사용자의 음성신호를 분석하여 사용자의 감정을 자동으로 인식하는 기술이다. 최근 마이크로 폰 센서가 탑재된 스마트폰에서 사용자의 통화 음성 데이터 수집 및 처리가 용이해 짐에 따라 감정인식 기술 연구가 활발히 수행되고 있다.

· 이 논문은 2013년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2013-067321)

† 비 회 원 : 경희대학교 컴퓨터공학과
jhb@oslab.khu.ac.kr
†† 종신회원 : 경희대학교 컴퓨터공학과 교수
sylee@oslab.khu.ac.kr
(Corresponding author임)

논문접수 : 2012년 12월 7일
심사완료 : 2014년 1월 7일

정보과학회논문지 : 소프트웨어 및 응용 제41권 제3호(2014.3)

스마트폰에서의 감정기반 개인화 서비스를 제공하기 위해서는 통화 종료 후 사용자 감정을 통화 단위로 도출해야한다. 그러나 기존의 음성기반 감정인식 기술은 수 초정도의 작은 크기의 타임 윈도우마다 수집한 음성 데이터로부터 사용자의 감정을 주기적으로 인식한다. 이와 같은 감정인식 기술은 하나의 통화 이벤트 전체에 대한 감정의 인식이 아닌 통화 중 특정 시간 동안의 감정을 인식하는 기술로써, 전체 통화기간동안의 감정을 인식하기 어렵다.

예를 들어, 사용자가 통화의 대부분을 화를 내며 대화를 하고 마지막 30초 동안 차분한 상태로 통화를 종료하면 기존의 감정인식 기술은 사용자의 감정을 “평범”이라고 인식한다. 이 경우 통화의 전체적인 감정은 화남이였고 통화 종료 후의 감정도 화남일 것이다. 따라서 통화단위에서의 감정인식은 통화 시작 시점부터 종료시점까지 전체적인 감정 상태를 점진적으로 고려해야 한다.

본 논문에서는 여러 가지 감정이 혼재되어 있는 통화음성으로부터 효과적으로 감정을 인식하는 기법을 제안한다. 제안하는 기법은 통화의 전체적인 감정 상태를 점진적으로 고려하기 위해 시간의 흐름에 따라 최근의 상태를 많이 반영하는 Tilted-Time Window 모델을 사용하여 통화음성에서 하나의 감정을 추론한다. 통화시작 시점부터 종료시점을 하나의 타임 윈도우로 설정하며 이를 일정 크기의 서브 윈도우로 분할하여 감정인식을 수행한다. 각 서브 윈도우의 인식된 감정 결과 값을 통화 시간의 흐름에 따라 점진적으로 가중치를 부여하여 통화 전체에서 사용자의 전반적인 감정을 인식한다. 화남, 즐거움, 두려움, 평범, 슬픔의 5가지 감정을 고려한 실험을 통하여 제안하는 감정인식 기술이 기존의 기술보다 더 높은 통화 단위의 감정인식 정확도를 보임을 입증하였다.

2. 관련연구

현재 음성기반 감정인식 연구는 작은 타임 윈도우 단위의 실시간 감정인식에 집중되어 있다. 대표적인 음성기반 감정인식 연구의 대표적인 어플리케이션 적용사례는 콜센터 고객의 부정적인 감정을 감지하여 상담원에게 알려주는 모니터링 서비스이다[1]. 이 연구는 통화 중 감정 모니터링이 주목적이므로 실시간으로 감정을 인지하는 것이다.

이러한 실시간 감정인식 연구는 크게 두 가지로 새로운 특징을 추출하거나 분류 방법론을 다르게 적용하여 정확도를 개선하는 연구들이 있다. 새로운 특징 추출 연구로는 개인마다 발성의 특징이 다르기 때문에 이를 반영하는 특징을 사용한 연구로써 Window사이즈를 정하지 않고 순간 감정을 인식하는 기술이다[2]. 분류 방법론을 다르게 적용하는 연구로는 계층적 분류 방법론을 적용한 기술이 있다. 이 연구는 총 3개의 분류기를 사용하여 음성에서 비슷한 감정의 인자를 나누어 분류하는 방법으로 높은 정확도를 보이지만 3초가량의 음성만 인지가 가능하다[3]. 다른 연구로는 남성과 여성의 훈련모델을 각각 생성하고 입력되는 음성을 남성, 여성으로 먼저 분류하여 성별에 맞는 훈련모델과 비교, 인지하는 방법이다[4].

그러나 기존의 연구는 단시간 초 단위 감정인식에 대한 연구로써 짧게는 몇 초 길게는 수십 분이 될 수 있는 통화단위 음성 데이터에 적용하기에는 부적합하다. 통화 음성 데이터에서는 감정의 기복이 나타날 수 있고 대화 형식으로 되어 있어 사용자의 음성이 정기적으로 들어오지 않기 때문에 통화 중 실시간 감정 인식이 아닌 통화 후의 전체 통화에 대한 감정을 인식하는 통화단위의 감정을 인식하는 데 부적합하다.

3. 통화음성 기반 감정인식

본 장에서는 여러 가지 감정이 혼재되어 있는 통화음성에서 하나의 감정을 추출해내는 통화음성 기반 감정인식 기술을 제안한다.

그림 1은 제안하는 통화음성 기반 감정인식의 개념도이다. 통화 녹음단계는 스마트폰에서 통화 시작부터 종료까지의 음성 데이터를 마이크로 폰 센서를 이용하여 녹음하고 녹음된 통화 음성에서 사용자가 말하지 않는

그림 1 통화 음성기반 감정인식 개념도
Fig. 1 Conceptual Diagram of Call Speech based Emotion Recognition

묵음을 제거한다. 묵음이 제거된 통화음성은 순간의 감정들을 파악하기 위해 5초의 작은 크기의 서브 윈도우로 분할하고 MFCC(Mel Frequent Coefficient Cepstral) Filter Bank 알고리즘을 사용하여 특징 벡터 값을 추출한 후 5가지 감정(화남, 즐거움, 두려움, 평범, 슬픔)으로 훈련된 SVM 분류모델에 의해 감정을 인식한다. 인식된 감정들은 시간의 흐름에 따라 점진적인 가중치를 부여하는 Tilted-Time Window 모델을 사용하여 통화단위의 최종 감정을 추정한다.

3.1 감정 모델의 선정

감정인식 분야에서 적절한 감정 상태를 유추하기 위해선 체계적인 감정 모델을 선정해야한다. 인간의 감정은 다양하고 복잡하며 수많은 형용사로 표현된다. 이러한 감정 상태를 정량화 하고 각각의 감정 상태간의 상관관계를 밝히려는 연구가 활발히 진행되고 있다[5].

현재 감정인식 분야에서 많이 사용되고 있는 감정모델은 크게 두 가지로써[6], 인간의 감정을 선호-부정 활동성-비 활동성의 2차원 영역으로 표현한 Valence-arousal Model이 있다.

Valence-arousal Model은 감정인식 연구에서 많이 사용되는 감정 모델로서[7] 다양한 감정을 2차원 공간으로 표현한다. 감정에 대한 성향을 나타내는 Valence축과, 감정의 강도를 나타내는 Arousal축을 통해 다양한 감정 상태를 정의하고 있다.

두 번째 방법으로는 즐거움, 놀라움, 두려움, 공포, 화남, 슬픔 등의 대표 감정을 선정하는 방법이다. Valence-arousal Model의 경우 인간의 감정 상태를 연속적으로 표현하여 다양한 감정을 선정할 수 있는 장점이 있지만 2차원 지표와 다양한 감정 형용사간의 구분짓기 힘든 애매모호한 감정이 존재하는 단점이 있다. 대표 감정을 선정하는 경우는 감정의 표현이 명확하여 감정에 따른 음성을 분류하기 쉬우므로 대부분의 음성기반 감정인식 분야에서 감정표현법으로 사용되고 있다.

따라서 본 논문에서는 감정 표현이 명확한 대표감성 표현법을 사용하여 감정인식 분야에서 일반적으로 많이 사용하는 대표 감정인 화남, 즐거움, 두려움, 평범, 슬픔 5가지의 감정 모델을 사용하였다. 본 논문에서 각 감정에 대한 음성의 정의는 아래의 표 1과 같이 정의한다.

표 1 각 감정에 대한 음성 정의
Table 1 Definition of each emotional speech

Emotion	Definition
Angry	Strongness, High ton, very Rapid Speaking
Joyful	High ton, Rapid Speaking
Nervous	Weakness, Low ton, Shaking Speaking
Natural	Normal ton, Normal Speaking
Sad	Weakness, Low ton, Slow Speaking

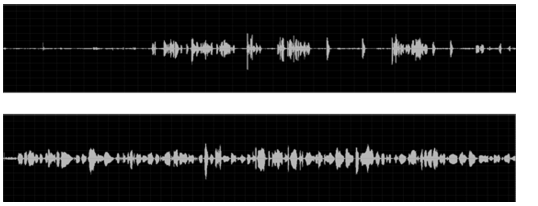

그림 2 묵음 제거 전/후 파형
Fig. 2 Waves of before/after Non-Speech Reduction

3.2 묵음 제거 및 통화음성 분할

통화음성 데이터에는 대화형식으로 이루어지기 때문에 사용자가 말하지 않는 묵음이 발생한다. 음성인식 분야는 다른 타 인식분야와 다르게 말하지 않는 공백데이터는 의미가 없는 데이터로 감정인식에 방해가 되는 요소이다. 따라서 이러한 묵음을 제거하는 것은 긴 단위의 음성기반 감정인식에 있어서 꼭 필요한 기술이다. 본 논문에서 사용하는 음성 묵음 제거 방법으로 소리의 크기에 임계 값을 두어 제거한다. 임계 값은 일반적으로 사람이 소곤대는 소리인 15데시벨(dB)로 설정한다.

그림 2는 통화음성데이터의 묵음을 제거하기 전 파형과 제거한 후의 파형이다. 사용자가 말을 하지 않은 묵음이 효과적으로 제거되었음을 확인할 수 있다.

통화 시간은 가변적이며 매우 길어질 수 있고, 이러한 긴 음성 안에서는 여러 가지 감정이 혼재되어 있다. 따라서 통화의 전반적인 감정을 추론하기 위해서는 통화 기간 내의 순간의 감정들을 파악하는 것이 중요하다.

순간 감정을 인식하기 위하여 묵음이 제거된 통화 음성을 여러 개의 서브 윈도우로 분할한다. 일반적으로 음성기반 감정인식은 3초에서 5초사이의 타임 윈도우에서 높은 정확도를 갖는다[2,3]. 본 논문에서는 순간의 감정을 파악하기 위해 5초 크기의 서브 윈도우를 사용한다.

3.3 단위 윈도우 감정인식

단위 윈도우 감정인식 단계에서 3.2절에서 분할한 5초 단위의 윈도우마다 감정을 인식한다. 단위 윈도우 감정인식은 특징추출 과정과 인식 과정으로 구성된다. 특징추출 과정은 감정인식에 적합한 특징을 추출하기 위해 데이터를 가공하는 필터뱅크 알고리즘을 통해 특징을 추출하는 과정이다. 인식 과정은 추출된 특징들을 기반으로 기계학습 알고리즘을 사용하여 감정을 추론하는 과정이다.

특징추출 과정에서는 필터뱅크 알고리즘으로 13차 MFCC를 사용하였다. MFCC는 인간의 청각 특성을 고려하는 필터뱅크 알고리즘으로 음성 인식분야에서 널리 사용되고 있으며 인식 성능이 우수하다[8]. 13차 MFCC는 64ms를 프레임 단위로 윈도우를 분할하는 해밍 윈도우 기법을 사용한다. 본 논문에서 5초 크기의 타임 윈도우

를 사용하므로 총 78개의 프레임이 생성된다. 그리고 각 프레임마다 13개의 MFCC 값을 추출하여 총 1014개의 특징벡터를 구성한다. 분류과정에서는 추출된 특징벡터로 기계학습 알고리즘인 SVM을 사용하여 감정을 훈련 및 인식한다.

3.4 통화단위 감정인식

통화단위 감정인식 단계에서는 단위 윈도우 감정인식의 결과들을 기반으로 사용자의 통화 단위의 감정을 추정한다. 그림 3에서와 같이 하나의 통화에서는 단위 윈도우마다 서로 다른 감정이 인식된다. 만약 통화 종료 직전의 감정만을 고려하여 통화 단위 감정을 인식하면, 실제 사용자의 감정이 즐거움임에도 불구하고 평범이라는 잘못된 감정을 추론할 것이다.

사람의 감정은 일반적으로 시간의 흐름에 따라 약해진다. 따라서 통화 중에 나타는 감정들 중에 통화 종료 직전의 감정이 가장 큰 영향력이 있다고 고려하여, 시간에 따른 점진적인 가중치를 두는 Tilted-Time Window 기법[9]을 활용한다. Tilted-Time Window는 전체 타임 윈도우를 시간에 따른 서브 윈도우들로 분할하고 각 서브 윈도우들에 시간에 역순으로 큰 가중치를 주어 최근 데이터와 이전 데이터의 중요도를 적절히 반영할 수 있다.

다양한 Tilted-Time Window 기법들 가운데 장기간의 데이터의 분석에 용이한 Logarithmic Tilted-Time Window[10] 기법을 사용한다. 가변적인 통화 길이에 시간에 따른 적절한 가중치를 부여하기 위하여 통화 전체 시간을 하나의 타임 윈도우로 설정한다. 그리고 총 통화 시간을 네 개의 시간 구간으로 분할한다. 그림 4는 전체 통화를 시간을 Tilted-Time Window로 분할한 그림이며 가장 왼쪽은 통화의 시작을 나타내며 통화의 마지막 부분은 오른쪽에 표현하고 있다. 각 시간 구간의

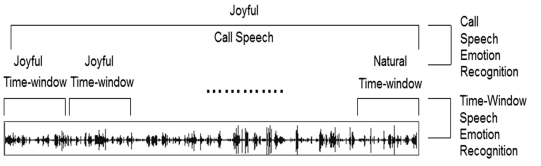

그림 3 통화 단위 감정인식 기법
Fig. 3 Call Speech Emotion Recognition

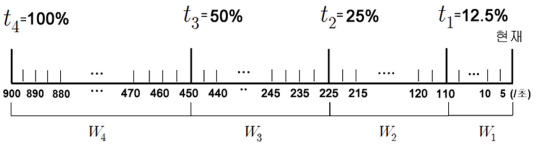

그림 4 Tilted-Time Window의 구성
Fig. 4 Organization of Tilted-Time Window

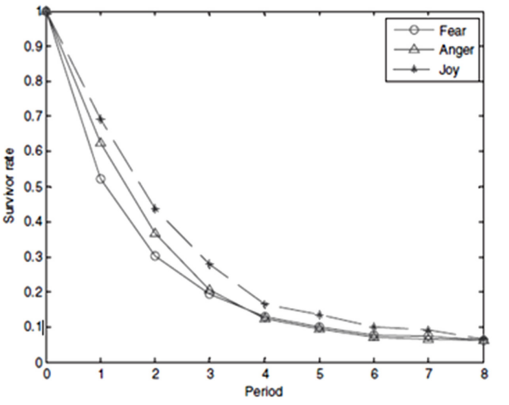

그림 5 감정 생존함수 곡선[11]
Fig. 5 Curve of Emotion Survival Function

비율은 통화 종료 시점을 기준으로 첫 서브 윈도우 t_1은 12.5%, t_2는 25%, t_3는 50%, t_4은 100%로 구성한다.

동일한 서브 윈도우내에 있는 음성 데이터는 동일한 가중치를 부여한다. 가중치는 심리학 연구에 기초하여 감정의 지속시간을 고려한다. 그림 5는 심리학계의 연구로 사람의 감정이 얼마만큼 지속되는지를 60명의 실험자에게 감정을 느꼈던 시점부터 시간의 흐름에 따라 지속적인 설문을 통하여 나온 결과인 감정 생존 곡선[11]이다. 곡선에 대한 결과 값 도출 과정은 실험자에게 해당 감정을 느낄 수 있도록 에피소드를 제공하고 15분마다 설문을 통하여 감정을 계속 느끼고 있는지를 확인하고 이 설문을 바탕으로 생존 함수[12]를 적용하여 도출한다.

감정 생존 곡선은 사람이 느꼈던 감정이 시간이 흐를수록 느끼지 못할 확률을 나타낸다. 식 (1)은 생존 함수를 수식화 해놓은 것으로 시간 t에 따른 생존 확률 R을 구한다. 함수 F(t)는 시간에 따른 해당 감정을 느낀 확률의 누적 함수 분포를 의미한다.

$$R(t) = 1 - F(t). \quad (1)$$

그러나 이 실험에서는 두려움, 화남, 즐거움 3가지 감정에 대한 실험만 하여 다른 평범과 슬픔과 같은 대표적인 감성에 대한 생존 곡선이 존재하지 않는다. 따라서 그림 5에서와 같이 각 감정의 생존 시간 변화는 서로 매우 유사한 추이를 가지고 있어 본 논문에서는 감정들의 생존곡선들의 기울기 평균을 Tilted-Time Window 가중치로 사용한다. 식 (2)는 각 타임 윈도우 t_i에 따른 각 감정에 대한 생존함수 결과 값의 평균 $ES(t)$를 구하는 수식이다.

$$ES(t_i) = avg(Survivor(t_i)) \quad (2)$$

식 (3)은 Tilted-Time Window의 가중치 산출 방법

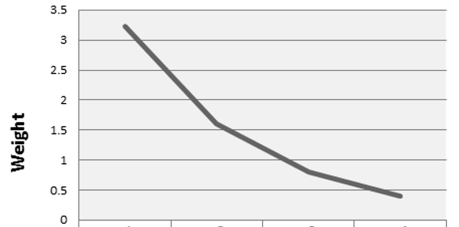

그림 6 시간에 따른 Tilted-Time Window 가중치 그래프
Fig. 6 Time dependent Weighted Graph of Tilted-Time Window

이다. 여기서, t_i는 그림 5와 같이 해당 구간의 시간을 나타내며 각 시간 구간마다의 기울기를 가중치로 사용한다.

$$W(t_i) = \frac{ES(t_{i+1}) - ES(t_i)}{t_i} \quad (3)$$

가중치의 크기는 전체 타임 윈도우의 길이에 따라 변화한다. 그림 6은 Tilted-Time Window 가중치의 예로 통화 시간 15분에 대한 가중치를 곡선과 가중치 값을 표시한 그래프이다.

각 타임 윈도우 구간에서 추출된 Tilted-Time Window 가중치는 해당 타임 윈도우에서 인식된 감정들의 빈도를 곱하여 각 감정에 대한 감정점수를 측정하고 한 통화의 최종적인 감정은 가장 높은 감정 점수를 갖는 감정으로 도출된다. 식 (4)는 감정 e에 대한 감정점수 $Score_e$를 계산하는 식으로 n은 타임 Window의 총 개수를 나타내고 f_e는 타임윈도우 t_i에서 나타난 감정 e의 빈도이다.

$$Score_e = \sum_{i=1}^{n} W(t_i) f_e \quad (4)$$

4. 실험 결과 및 분석

본 장에서는 제안하는 통화 단위 감정인식 기법의 성능을 검증한다. 성능 평가는 단위 윈도우 감정인식의 정확도와 통화단위의 감정인식의 정확도를 각각 측정하였다. 다음절에서는 실험 환경 및 데이터 수집, 성능평가 방법, 감정인식 평가에 대한 내용을 설명한다.

4.1 실험 환경 및 데이터 수집

실험에서 사용된 스마트폰 디바이스는 SHW-M250S 모델을 사용하였으며 음성 데이터는 8kHz, 16bit, 모노(mono)로 설정하였고 잡음이 나지 않는 조용한 환경 녹음하였다. 음성 데이터는 단위 윈도우 감정인식을 위한 음성 녹음과 통화 단위 감정인식을 위한 실제 통화 데이터로 두 가지의 음성데이터를 수집하였다. 단위 윈도우 감정인식을 위한 데이터는 20대의 남자 4명, 여자 4명의 사용자가 직접 5초 단위로 화남, 즐거움, 두려움, 평범, 슬픔에 대한 5가지 감정을 직접 연기를 통해 녹음을 하여 총 1000개로 구축하였다. 또한 모든 상황을 고려하기 위해 훈련데이터는 64ms 간격으로 오버래핑기법을 사용하여 훈련하였다. 통화 단위 감정인식을 위한 통화 음성 데이터는 화남, 즐거움, 두려움, 평범, 슬픔 각 감정마다 85개식 총 425개의 통화음성 데이터를 구축하였다. 구축한 통화음성 데이터 중 데이터 구축에 참여한 사용자가 통화한 후 자기 자신과 익명의 두 명에게 설문을 받아 3명의 평가자가 같은 감정이라고 선택한 각 감정마다 33개의 음성을 선택하여 총 165개만을 평가에 사용하였다. 분류기를 위한 기계학습 알고리즘은 SVM(Support Vector Machine)을 사용하였다.

4.2 성능 평가 방법

실험은 단위 윈도우 감정인식 정확도 측정과 통화 단위 감정인식에 정확도를 각각 측정을 하였다. 단위 윈도우 감정인식 정확도 측정은 10 Fold-Cross Validation 기법을 사용하여 측정하였다. 통화 단위 감정인식에 대한 성능 비교는 기존의 연구와 비교하기 위해 통화의 마지막 부분만을 인지한 방법(방법 1), 본 논문에서 제안하는 Tilted-Time Window기법을 검증하기 위한 Tilted-Time Window를 적용하지 않고 단순히 감정의 빈도수만 반영하여 인지한 방법(방법 2), 제안하는 Tilted-Time Window 기법을 사용한 방법(방법 3)을 사용하여 진행하였다.

4.3 감정인식 정확도

단위 윈도우 감정인식 정확도는 화남은 92.16%, 즐거움은 88.17%, 두려움은 77.30%, 평범은 90.17%, 슬픔은 99.45%, 평균 89.45%의 보였다. 대체로 높은 정확도를 보였으나 가장 분류가 어려운 감정은 두려움이었으며 즐거움과 화남으로 잘못 인지하는 경우가 많았다.

표 2는 각각의 감정에 따른 인식 정확도의 Confusion Matrix 이다. 세로축은 감정의 실제 레이블(true label)이고, 가로축은 각 감정을 인지한 레이블이다. 예를 들어, 화남 감정에 대해서는 92.17%를 정확하게 인지하였고, 화남을 즐거움으로 잘못 인지한 것이 5.8%이고, 두려움으로 잘못 인지한 것이 2.03%이다.

표 2 단위 윈도우 감정인식 Confusion Matrix (단위 %)
Table 2 Confusion Matrix of Time Window Speech Emotion Recognition (unit %)

	Angry	Joyful	Nervous	Natural	Sad
Angry	**92.17**	5.8	2.03	0	0
Joyful	4.61	**88.17**	6.99	0.23	0
Nervous	5.91	12.43	**77.31**	4.18	0.17
Natural	0.09	1.07	8.422	**90.17**	0.25
Sad	0	0.22	0	0.32	**99.46**

표 3 통화단위 감정인식 정확도
Table 3 Accuracy of Speech Emotion Recognition

	Angry	Joyful	Nervous	Natural	Sad	Avg.
Exp.1	43.33	48.57	68.29	39.13	61.53	50.9
Exp.2	46.66	51.42	60.975	52.17	61.53	53.9
Exp.3	**53.33**	**62.857**	**75.60**	**52.17**	**76.92**	**62.4**

표 3은 통화단위 감정인식 정확도를 나타내는 비교표이다. 기존 연구 방법을 사용한 방법 1은 평균 50.9%의 정확도를 보였다. Tilted-Time Window 기법을 사용하지 않고 빈도수만 측정한 방법 2는 평균 53.94%의 정확도를 보였고 제안하는 Tilted-Time Window 기법을 활용한 방법 3은 평균 62.42%의 정확도를 보였다.

세 실험을 비교하였을 때 제안하는 기법이 전체적으로 높은 정확도를 보였으며 기존 연구 대비 즐거움은 17.28%, 두려움은 7.31%, 평범은 13.04%, 화남은 10%, 슬픔은 15.39%, 평균적으로는 11.51%의 높은 정확도 향상을 보였다. Tilted-Time Window 기법을 사용하지 않고 빈도수만 측정한 경우와 비교하였을 때는 즐거움이 11.43%, 두려움이 14.63%, 화남이 6.67%, 슬픔이 15.39% 높은 정확도를 보였고 평범의 경우 인지율이 같았다.

5. 결 론

본 논문은 스마트폰에서 통화 음성기반 감정인식 기법을 제안하였다. 다양한 감정이 혼재되어 있는 통화음성에서 제안하는 기법이 기존의 음성기반 감정인식 기법보다 더 좋은 성능을 보이는 것을 입증하였다. 제안하는 방법은 통화음성에서 5초단위로 감정을 인식하고 인식된 결과를 Tilted-Time Window기법을 사용하여 최종 감정을 추론하였다. 또한 감정의 지속시간을 나타내는 감정생존곡선을 사용하여 Tilted-Time Window의 가중치로 설정함으로써 통화음성 기반 감정인식의 정확도를 높였다.

References

[1] D. Morrison, R. Wang, L. C. De Silva, "Ensemble methods for spoken emotion recognition in call-centres," Speech Communication, vol.49, Issue 2, pp.98-112, 2007

[2] A. B. Kandali, A. Routray, T. K. Basu, "Emotion recognition from Assamese speeches using MFCC features and GMM classifier," *TENCON 2008-2008 IEEE Region 10 Conference*, pp.1-5, 19-21 Nov, 2008.

[3] Z. Xiao, Dellandrea, L. Chen, W. Dou, "Recognition of emotions in speech by a hierarchical approach," *ACII 2009. 3rd International Conference*, 10-12, Sept, pp.401-408, 2009.

[4] Youn-ho Cho, Kyu-Sik Park, "A Study on The Improvement of Emotion Recognition by Gender Discrimination," *Journal of IEEK*, vol.45, pp.401-408, 2008.

[5] Picard, R. W., 1998, Affective Computing, The MIT Press, London, pp.141-192.

[6] Joonyoung Park, Dongsu Park, Jahng-hyon Park, Jihyung Park, "Development of Human Sensibility Recognition Systemusing Hidden Markov Model," *HCI 2004*, pp.605-610, 2004.

[7] J. Posner, J.A. Russell and B.S. Peterson, "The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology," *Development and Psychopathology 2005*, vol.17, pp.715-734, 2005.

[8] A. Klautau, "The MFCC," [Online]. Available: http://www.cic.unb.br/~lamar/te073/Aulas/mfcc.pdf

[9] P. Pitarch, A. Laurent, M. Plantevit, P. Poncelet, "Multidimensional Data Stream Summarization Using Extended Tilted-Time Windows," *2009 International Conference on Advanced Information Networking and Applications Workshops*, 26-29, May, 2009.

[10] P. Pitarch, A. Laurent, M. Plantevit, P. Poncelet, "Multidimensional Data Stream Summarization Using Extended Tilted-Time Windows," *2009 International Conference on Advanced Information Networking and Applications Workshops*, 26-29, May, 2009.

[11] P. Verduyn, E. Delvaux, H. V. Coillie, F. Tuerlinckx, and I. V. Mechelen, "Predicting the Duration of Emotional Experience: Two Experience Sampling Studies," *American Psychological Association*, vol.9(1), pp.83-91, Feb. 2009.

[12] Wekipedia, "Survival Function," URL: "https://en.wikipedia. org/wiki/Survival_function"

방 재 훈

2007년 평택대학교 디지털응용정보학과 학사. 2013년 경희대학교 컴퓨터공학과 석사. 2013년~현재 경희대학교 컴퓨터공학과 박사과정. 현재 경희대학교 동서신의학 u-라이프케어 연구센터 연구원. 관심분야는 유비쿼터스 컴퓨팅, 감정인식, 모바일기반 감정인식

이 승 룡

1978년 고려대학교 재료공학과 공학사 1987년 Illinois Institute of Technology 전산학과 석사. 1991년 Illinois Institute of Technology 전산학과 박사. 1992년~1993년 Governors State University, Illinois 조교수. 1993년~현재 경희대학교 전자정보학부 컴퓨터공학과 교수. 현재 경희대학교 동서신의학 u-라이프케어 연구센터 센터장. 관심분야는 유비쿼터스 컴퓨팅, 상황인지, 인공지능, 실시간 시스템, 미들웨어 시스템, 보안, 클라우드 컴퓨팅