Journal of Korea Multimedia Society Vol. 10, No. 11, November 2007(pp. 1530-1537)

ARM 플랫폼 기반의 음성 감성인식 시스템 구현

오상헌[†], 박규식[††]

요 약

본 논문은 마이크로폰을 통해 실시간으로 습득된 음성으로부터 사람의 음성 감성상태를 평상, 기쁨, 슬픔, 화남 등 4가지로 구별할 수 있는 ARM 플랫폼 기반의 음성 감성인식 시스템 구현에 관한 것이다. 일반적으로 마이크로폰으로 수신된 음성은 화자 주변의 환경 잡음과 마이크로폰의 시스템 특성 때문에 입력 음성 신호가 왜곡되고 이로 인해 시스템의 성능이 저하된다. 본 논문에서는 이러한 잡음 영향을 최소화하기 위해 비교적 단순한 구조와 적은 연산 량을 가진 이동평균(MA, Moving Average) 필터를 입력 음성의 특징벡터 열에 적용하였다. 또한, 효율적으로 감성 특징벡터를 최적화할 수 있는 SFS(Sequential Forward Selection)기법을 적용해 제안 시스템의 성능을 최적화하였으며 감성 패턴 분류기로는 SVM(Support Vector Machine)을 사용하였다. 실험 결과 제안 감성인식 시스템은 모의실험에서 약 65%, ARM 플랫폼에서 약 62%의 인식률을 보였다.

Implementation of the Speech Emotion Recognition System in the ARM Platform

Sang-Heon Oh[†], Kyu-Sik Park[††]

ABSTRACT

In this paper, we implemented a speech emotion recognition system that can distinguish human emotional states from recorded speech captured by a single microphone and classify them into four categories: neutrality, happiness, sadness and anger. In general, a speech recorded with a microphone contains background noises due to the speaker environment and the microphone characteristic, which can result in serious system performance degradation. In order to minimize the effect of these noises and to improve the system performance, a MA(Moving Average) filter with a relatively simple structure and low computational complexity was adopted. Then a SFS(Sequential Forward Selection) feature optimization method was implemented to further improve and stabilize the system performance. For speech emotion classification, a SVM pattern classifier is used. The experimental results indicate the emotional classification performance around 65% in the computer simulation and 62% on the ARM platform.

Key words: Speech Emotion Recognition(음성 감성인식), MA Filtering(MA 필터링), SFS, SVM

1. 서 론

일반적으로 음성 감성인식 시스템은 감성 특징벡터 추출과 감성 패턴인식 2가지 단계로 구성된다. 감성 특징벡터 추출은 음성 신호로부터 감성 상태를 대표할 수 있는 피치(pitch), 포먼트(formant), 에

※ 교신저자(Corresponding Author) : 박규식, 주소 : 경기도 용인시 수지구 죽전동 산44-1 단국대학교 (448-701), 전화 : 031)8005-3252, FAX : 031)8005-3228, E-mail : kspark@dankook.ac.kr
접수일 : 2007년 3월 14일, 완료일 : 2007년 10월 15일
[†] 정회원, 단국대학교 정보 컴퓨터 과학과
(E-mail : taru74@dankook.ac.kr)
[††] 단국대학교 정보 컴퓨터학부
※ 본 논문은 정보통신부 출연금으로 MIC/IITA/ETRI, SoC산업진흥센터에서 수행한 IT SoC 핵심설계인력양성 사업의 연구결과 입니다.

너지(energy) MFCC(Mel Frequency Cepstral Coefficient), LPC(Linear Predictive Coefficient) 등의 특징벡터들을 구하는 과정이다. 한편, 음성의 감성 상태를 분류하는 패턴인식 알고리즘으로는 k-NN (Nearest Neighbor), HMM(Hidden Markov Model), SVM(Support Vector Machine), NN (Neural Network)등 다양한 방법[1]이 사용되고 있다. 일반적으로 음성 감성인식 시스템의 성능은 패턴인식 알고리즘보다 특징벡터에 더 많이 의존하는 경향이 있다.

기존 연구 [2]에서 Dallaert는 음성 신호로부터 피치 변화를 추출하여 음성 감성상태를 기쁨, 슬픔, 화남, 공포 등의 4가지로 분류하였으며 k-NN 패턴 분류기를 사용해서 약 79.5%의 인식률을 달성하였다. Moriyama[3]는 음성 신호의 피치 변화와 파워 포락선(power envelop)을 특징벡터로 사용하여 놀람, 화남, 기쁨, 공포, 슬픔 등의 5가지 음성 감성상태를 분류하였으며 이중에서 놀람, 화남, 슬픔 등의 3가지 감성에서 비교적 높은 인식률을 달성할 수 있음을 보였다. A. Nogueiras는 논문[4]에서 HMM을 이용한 화자종속 방식의 감성인식 시스템을 제안하였고 놀람, 기쁨, 화남, 공포, 혐오, 슬픔, 평상 등 6개 음성 감성상태를 분류하는데 평균 80%의 인식률을 보이고 있다. 또한, N. Amir[5]는 4개 감성상태를 분류하기 위해 피치와 음성 존재 구간 개수(voiced period) 등의 특징벡터를 이용하여 k-NN과 NN 패턴 분류기의 성능을 비교하였다. C. Lee[6]는 음성 신호의 음향학적 특징에 의미론적인 언어 특징 정보를 더하여 콜센터 같은 응용 시스템에서 부정적인 음성과 비-부정적인 음성을 분류할 수 있는 알고리즘을 제안하였다. 실험 결과 언어 정보를 이용해 남자 음성의 경우 약 40%, 여성 음성의 경우 약 36% 향상시킬 수 있음을 밝히고 있다. 반면, G. Zhou는 논문[7]에서 TEO(Teager Energy Operator)라는 새로운 특징벡터를 제안하여 평상 음성과 스트레스 음성을 구분하는 흥미로운 연구 결과를 보고하고 있다. 이외에도 미국의 Microsoft, HP, 일본의 SONY 등의 산업계에서 음성 감성인식 기술을 HCI용 SW나 로봇 등의 응용분야에 적용하기 위한 활발한 연구를 진행하고 있다.

이상에서 살펴본 바와 같이 기존의 음성 감성인식 연구는 주로 PC 기반의 잡음이 없는 깨끗한 환경에서의 연구로서 이를 산업화할 때 필수적으로 고려해야 하는 환경 잡음을 고려한 연구는 부족한 실정이다. 일반적으로 음성 감성인식 시스템은 시스템을 훈련시킬 감성음성 DB와 질의로 입력되어질 음성 데이터 간의 녹음환경 차이(잡음 환경, 마이크로폰 특성 등)로 인해 감성인식 성능이 심각하게 저하된다.

본 연구에서는 이러한 잡음 환경을 고려한 음성 감성인식 알고리즘을 개발하여 ARM 플랫폼으로 구현하였다. 제안된 시스템은 ARM920T(S3C2440A-40) 플랫폼을 이용해 마이크로폰으로부터 음성을 실시간으로 습득하고 감성상태를 평상, 기쁨, 슬픔, 화남 4가지로 구별할 수 있다. 일반적으로 마이크로폰으로 수신된 음성은 화자의 환경 잡음과 마이크로폰 특성 잡음을 포함하고 있어 심각한 성능 저하를 초래한다. 본 논문에서는 이러한 잡음 영향을 최소화하기 위하여 비교적 연산량이 적은 이동평균(MA, Moving Average) 필터를 음성 특징벡터 영역에 적용하였으며 끝점 검출기(End-Point Detector)를 이용해 음성이 존재하는 구간에서만 특징벡터를 추출하였다. 또한, 추출된 특징벡터 중 감성 인식률에 기여가 높은 특징계수들만을 선별해 시스템의 인식 성능을 향상시킬 수 있는 SFS(Sequential Forward Selection)[8] 기법을 적용하였으며 감성 분류를 위한 패턴인식 알고리즘으로는 SVM을 사용하였다.

본 논문의 구성은 다음과 같다. 먼저 2장에서는 제안 감성인식 알고리즘의 전반적인 구성과 음성 전처리, 강인한 감성 특징벡터를 추출하는 기법, 특징벡터 최적화 기법을 설명하고, 3장에서는 ARM 플랫폼 기반의 음성 감성인식 시스템을 소개하였다. 4장에서는 컴퓨터 모의실험과 ARM 플랫폼에서의 성능을 비교하였고, 마지막으로 5장에서는 결론으로 끝을 맺는다.

2. 제안 음성 감성인식 시스템

그림 1은 본 논문에서 구현한 음성 감성인식 시스템의 감성 DB 구축과정과 시스템 동작 과정을 나타낸다.

먼저 감성 DB 구축과정에서는 평상, 기쁨, 슬픔, 화남 각 감성별로 구성된 훈련용 음성 신호를 20ms 프레임 단위로 분할해서 해밍(Hamming) 윈도우를 적용한 후 끝점 검출 등의 음성 전처리 과정과 2차

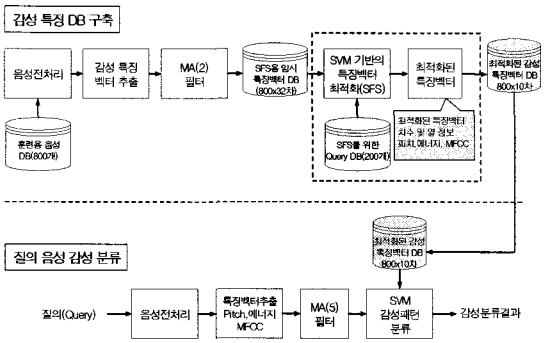

그림 1. 제안 음성 감성인식 시스템

이동평균 필터링을 거쳐 총 32차의 특징벡터를 추출한다. 추출된 특징벡터는 SFS 특징벡터 최적화 과정을 거쳐 감성인식 성공률에 가장 큰 기여를 하는 10개의 특징계수들만을 선별해서 최종 감성 특징벡터를 구성하게 된다.

마이크로폰 질의 음성에 대한 감성 분류는 감성 DB를 구축할 때와 같은 음성 전처리 과정을 거쳐 SFS 특징벡터 최적화 과정으로 선별된 10개의 특징벡터만을 추출한 후, 특징벡터 영역에서 5차 이동평균 필터를 적용해 잡음 영향을 최소화한다. 최종적으로는 SVM 패턴 분류기를 이용해 질의 음성의 감성 상태를 분류한다.

2.1. 음성 전처리

음성 전처리과정은 그림 2와 같이 신뢰성 있는 감성 특징벡터를 추출하기 위한 프레임 단위의 음성 신호 분할, 해밍 윈도우, 끝점 검출기로 구성된다.

2.2. 감성 특징벡터 추출

본 논문에서는 매 20ms 프레임 단위로 운율적 특징을 갖는 피치, 에너지와 음소 특징을 갖는 6차 MFCC 그리고 각 프레임 간 특징계수의 차를 나타내는 델타(Delta) 값을 추출한 다음, 각 특징계수들의 프레임 간 평균과 표준 편차를 구하여 총 32차의 특징벡터를 구성하였다.

그림 2. 음성 전처리 과정

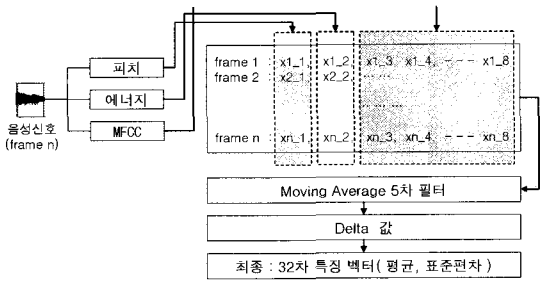

그림 3. 감성 특징벡터 추출 과정

피치(Pitch)는 일반적으로 많이 사용되는 HPS [9], AMDF[10], SHR[11] 방법을 비교 실험하여 잡음 환경에서 가장 높은 감성 인식률을 나타낸 SHR (Subharmonic to harmonic ratio) 알고리즘을 사용하였다. SHR은 음성 신호에 FFT를 취하여 2개의 피크 값 (f_1, f_2)을 피치 후보로 선정하고, 수식 (1)의 SHR 값을 특정 한계 값과 비교하여 SHR이 한계 값보다 작으면 f_2를 최종 피치로 선정하고, 아니면 f_1을 피치로 선정한다. 여기서 DA(·)는 논문[11]에서 주어진 차분 함수이다.

$$SHR = 0.5\frac{DA(\log f_1) - DA(\log f_2)}{DA(\log f_1) + DA(\log f_2)} \quad (1)$$

단-구간 음성 에너지는 수식 (2)를 이용하여 각 프레임에서의 에너지를 계산하였다.

$$E_n = \sum_{m=-\infty}^{\infty} x^2(m) \cdot h(n-m) \quad (2)$$

MFCC는 음성 인식 분야 등에서 널리 사용되는 특징으로 사람의 청각 특성과 유사한 멜-주파수 (Mel-Frequency)상에서 음성 특성을 잘 표현할 수 있다. 본 논문에서는 MFCC 4, 6, 8, 12차에 대한 인식률 비교 실험을 통하여 시스템 연산량과 인식률에서 가장 적합한 6차 MFCC를 사용하였다.

2.3 SFS 특징벡터 최적화

SFS[8]는 전 절에서 추출된 총 32차 특징벡터 간에 중복된 상관성을 제거하는 동시에 시스템 감성 인식률에 가장 기여를 많이 하는 최적의 특징계수들만을 선정하는 방법으로 시스템 성능을 안정화시킬 뿐만 아니라 연산 복잡도를 낮출 수 있는 장점이 있다. SFS는 먼저 각 특징계수들을 개별적으로 사용하

여 감성 분류를 한 후, 가장 좋은 감성 인식률을 나타내는 특징계수부터 순차적으로 하나씩 특징계수를 추가해 나가면서 감성 인식 정확도를 계산한다. 이 과정을 거쳐 최적의 감성 인식률을 얻을 수 있는 처음 몇 개의 특징계수만을 선정해 특징벡터 열을 새롭게 구성하게 된다.

2.4. 잡음 영향 최소화를 위한 이동평균 필터링

마이크로폰을 통해 수신된 음성은 화자의 환경 잡음과 마이크로폰 특성 잡음을 포함하고 있어 감성 특징벡터를 왜곡하게 되고 이로 인해 시스템의 성능이 저하된다. 일반적으로 이러한 복합 잡음들은 예측하기 어려운 패턴을 가지기 때문에 잡음의 통계적 특성을 이용한 기존의 필터링 기법들을 사용하기에는 어려운 점이 있다. 따라서 본 논문에서는 비교적 연산량이 적은 이동평균(MA) 필터를 특징벡터에 적용하여 잡음에 의한 특징벡터 왜곡을 완화시켰다.

이동평균 필터는 기본적으로 저역 통과 필터(Low Pass Filter)이므로 잡음에 의해 급격한 변화를 보이는 특징벡터 열 부분을 부드럽게 하는 역할을 한다. 물론 이동평균 필터 적용으로 인해 원 음성 신호에 포함되어 있는 일부 음성 특징도 왜곡될 수 있으나, 일반적으로 잡음이 섞인 음성신호의 특징벡터 열에서 급격히 변하는 부분은 잡음에 의한 경우가 대부분이기 때문에 이동평균 필터를 사용하는 것이 타당하다 할 수 있다.

이동평균 필터를 특징계수에 적용하는 방법은 다음과 같다. 먼저 음성 신호의 분석시간 동안 매 20ms 단위로 특징 계수를 추출한 후 SFS 특징벡터 최적화 과정으로 선정된 특징계수를 다음과 같이 $T \times D$ 행렬로 표현한다.

$$FM_{td} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,D} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,D} \\ \cdots & \cdots & \cdots & \cdots \\ x_{T,1} & x_{T,2} & \cdots & x_{T,D} \end{bmatrix} \quad (3)$$

여기서 t는 신호 분석시간 동안의 시간 순서별 프레임 번호 $t = 1, 2, \cdots T$ 이고 D는 특징벡터의 차수이다. 따라서 위 수식의 행(row)은 시간 순서별 각 T개 프레임에서 추출된 32개의 특징벡터를, 그리고 열(column)은 각각의 특징계수들이 시간 순서별 프레임에 따른 변화를 나타낸다. 한편, 위 행렬 수식 (3)의 각 열을 평균 0, 표준편차 1이 되도록 정규화해 각 특징계수들 간의 편차로 인한 오 분류 동작을 방지하였다. 여기서 이동평균 필터는 정규화된 각 특징계수 열별 프레임 방향으로 적용되며 본 논문에서는 수식 (4)의 이동평균 필터에 대한 다양한 실험을 거쳐 필터 차수를 M=5로 선정하였다.

$$\widehat{x_{t,f}} = \frac{1}{M}\sum_{i=0}^{M} x_{(t-i),f} \qquad x_{i,f} = \text{특징계수} \quad (4)$$

본 연구에서는 마이크로폰 질의 음성의 특징벡터 열에 5차 이동평균 필터를 적용할 뿐만 아니라, 감성 DB 구축 단계에서도 이동평균 필터를 적용해 감성 DB와 질의 음성 간의 특징벡터 차이를 최소화해 패턴 매칭에 의한 감성 분류 시 오동작을 완화하였다. 감성 DB 구축시 이동평균 필터를 적용하지 않은 경우와 2차, 5차, 7차의 이동평균 필터를 적용한 각 경우에 대한 모의실험 결과 2차 이동평균 필터를 적용하는 것이 가장 좋은 성능을 보이고 있어 본 논문에서는 DB 구축시 2차 이동평균 필터를 적용하였다.

2.5 SVM(Support Vector Machine)

SVM은 기본적으로 두 그룹간의 경계선을 초평면으로 나타낼 수 있는 2진 선형 분류기이지만 커널 함수를 적용해 비선형 형태의 분류기로도 사용될 수 있다. 또한, 이러한 SVM을 여러 개 조합하여 다중 그룹에 대한 분류기로 확장시킬 수 있다. 본 연구에서는 4가지 감성에 대해 Pair wise 기반의 다중 그룹 분류 방식을 적용하였다.

3. ARM 플랫폼 기반의 감성인식 시스템

그림 4는 본 연구에서 구현한 ARM920T (S3C 2440A-40) 플랫폼의 구성을 나타낸다. ARM920T의 CPU는 400MHz로 동작하며 약 440 MIPS의 정수 연산능력과 64Mbyte의 메모리 공간을 갖는다. 그림에서 호스트 PC는 타깃 보드에 음성 감성인식 알고리즘을 로드하는 역할을 한다. 제안 시스템은 마이크로폰으로 입력된 PCM 신호를 1초간 입력 받아 끝점 검출기를 이용해 비-음성 구간에 해당하는 신호 성분을 제거한 다음 음성 존재 구간에서만 특징벡터를 추출한다. 이때 추출된 특징벡터에 5차 이동평균 필

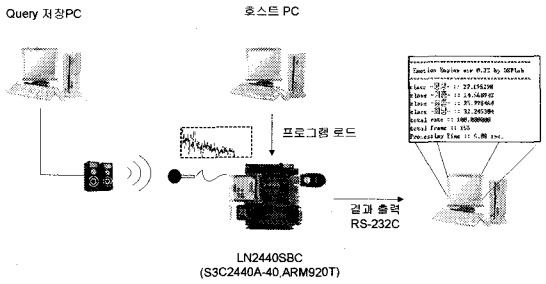

그림 4. ARM 플랫폼 감성인식 시스템의 구성

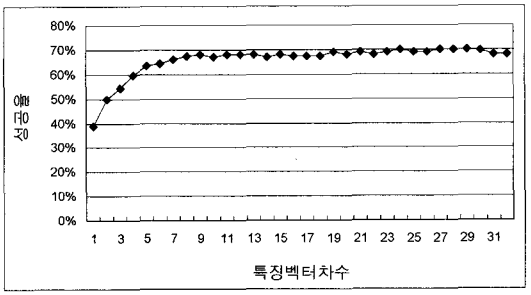

그림 5. SFS 특징벡터 최적화 과정

터를 적용하고 SVM 분류기를 이용해 입력 음성의 감성상태를 평상, 기쁨, 슬픔, 화남 등 4가지로 분류한다. 본 시스템은 3초의 입력 음성에 대해 평균 약 4.8초 정도의 알고리즘 연산 시간을 갖는다.

4. 실험 환경 및 결과

4.1. 실험 환경

본 연구에서는 논문 [12]의 음성 DB로부터 평상, 기쁨, 슬픔, 화남 등 4가지 감성의 음성을 발췌해 실험 DB를 구축하였다. 논문 [12]의 음성 DB는 평소 음성 발성을 훈련하는 아마추어 연극단원 남, 여 각 15명이 45개 문장에 대하여 3회 발성한 음성 16,200개를 8kHz, 16bit로 녹음한 것이다. 본 연구에서는 위 DB에서 비교적 감성이 잘 표현되어 있고 문장 길이가 2초 이상 되는 1,000개 음성만을 선정해 실험에 사용하였다. 1,000개 음성 중 무작위로 800(평상-200, 기쁨-200, 슬픔-200, 화남-200)개 음성을 선정하여 훈련과정을 거쳐 감성 특징벡터 DB를 구축하였다. 제안 시스템의 평가에 사용될 질의 음성은 훈련용 음성 DB와 중복되지 않는 나머지 200(평상-50, 기쁨-50, 슬픔-50, 화남-50)개 음성을 사용하였다.

4.2. 컴퓨터 모의실험 결과

그림 5는 제안 시스템에서 최적화된 특징벡터 계수를 선정하기 위한 SFS 실험 결과를 보이고 있다. 그림은 총 32차 특징벡터 중 인식률에 가장 기여가 큰 특징벡터 열만을 선정하기 위한 것으로 SVM 분류기에 의한 최적화 과정을 나타낸다. 그림에서 보듯이 10차 이후의 결과에서 비교적 높은 인식률을 갖는 것을 확인할 수 있으며, 약 20차 이후에 1~3% 정도 더 증가하는 결과를 볼 수 있다. 본 연구에서는 제안 시스템의 연산량을 고려하여 처음 10차의 특징벡터 열만을 선정해 사용하였다.

표 1은 마이크로폰 잡음 실험 환경에서 이동평균 필터를 적용한 경우와 적용하지 않은 경우의 감성 분류 결과를 비교한 것이다. 다른 방법과의 비교를 위해 전처리 단계에서 대표적으로 많이 사용되고 있는 signal subspace[13] 잡음제거 기술에 의한 결과도 같이 수록하였다.

표 1에서 보듯이 이동평균 필터를 적용한 제안 시스템이 이동평균 필터를 적용하지 않은 시스템에 비해 약 8%~10.5%의 성능 향상을 가져옴을 볼 수 있다. 반면, 제안 시스템은 기존의 signal subspace 잡음 제거 기술에 비해 약 4~5% 정도의 성능 향상을 보이고 있다.

4.3 ARM 플랫폼에서의 감성인식 실험

그림 6은 ARM 플랫폼에 구현한 감성인식 시스템의 동작 순서도를 보여주고 있다. 시스템은 256 샘플 프레임 단위로 동작되며 처음 3초간의 데이터를 입력 받은 후 음성 구간과 비-음성 구간을 계산한다.

표 1. 마이크로폰 잡음 실험환경에서 감성인식 시스템의 성능 비교 (괄호 안의 숫자는 사용된 특징벡터 차수를 나타냄.)

	일반 시스템		signal subspace		이동평균 필터 적용 시스템	
	SVM (32차)	SVM (10차)	SVM (32차)	SVM (10차)	SVM (32차)	SVM (10차)
평균 감성 인식률	58%	54.5%	62%	60%	66%	65%

표 2. ARM 플랫폼에서 제안 감성인식 알고리즘의 시스템성능

항 목	제안 감성인식 알고리즘의 연산 성능					
CPU	ARM920T(S3C2440A-40)					
샘플링 비율	8,000Hz					
샘플당 비트수	16 bits					
평균 입력데이터 길이	약 24,000 sample (3 sec)					
출력시간	평균 4.8 sec (질의 데이터 입력 완료된 시점부터)					
밉스(MIPS)	120 프레임 기준					SVM (10차)
	FFT(256pt)	Pitch(1)	MFCC(6차)	Energy(1)	필터, 및 기타 연산	
	839.87	259.384	653.24	0.3	1,692.244	15.1492
	평균 3445 MIPS					
메모리(KB)	DB 및 상수	감성인식 시스템				
	67.5 KB	268.008 KB				
	335.508 KB					

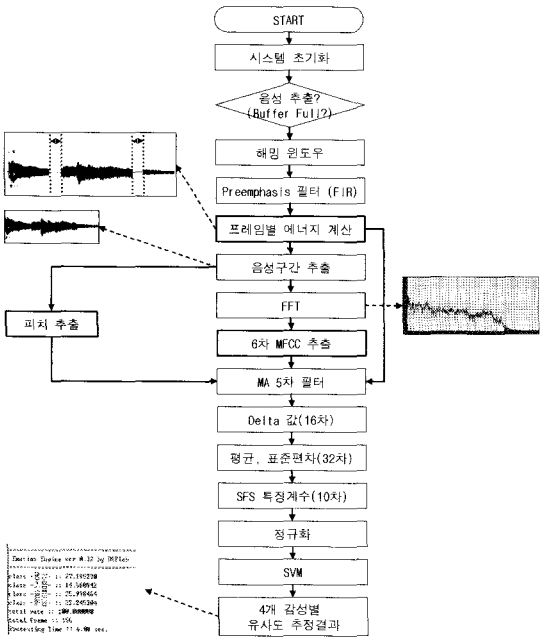

그림 6. ARM 플랫폼 감성인식 시스템의 프로그램 구조

DMA를 통해 주기적으로 계속 입력되는 음성 데이터를 버퍼로 저장하는 동시에 감성인식 알고리즘이 동작한다.

표 2는 ARM 플랫폼에 구현된 감성인식 시스템의 연산 성능을 FFT, 피치, MFCC, 에너지, 이동평균 필터 및 기타 연산 별로 비교한 것이다.

표 2에서 256pt-FFT의 연산량은 1회당 0.4532 MIPS의 비효율적인 연산량을 보이고 있는데 이는 16bit 정수 FFT 연산이 아닌 32bit 정수 연산을 사용하기 때문이다. 실제 ARM 어셈블리의 최적화된 256pt-FFT를 사용할 경우 1/10 이하 수준의 연산량으로 처리가 가능하나 MFCC와 피치에서 연산 오차가 증가하여 감성인식 성능에 심각한 영향을 초래하게 된다.

표 3은 마이크로폰 잡음 환경에서 ARM 플랫폼에 구현된 감성인식 시스템의 결과를 보여준다.

표 3에서 보듯이 주된 오 분류(miss classification)는 평상, 기쁨, 슬픔 감성에서 발생하고 있으며, 화남 감성은 비교적 잡음에 강인한 특성을 가지고 있음을 알 수 있다. 이러한 현상은 논문[3]에서 지적한 바와 같이 화남 감성의 피치, 에너지, MFCC 같은 특징계수들이 비교적 큰 변화를 가지고 있어 다른 감성과

표 3. 마이크로폰 잡음 환경에서 ARM플랫폼으로 구현된 감성인식 실험 결과

	평 상	기 쁨	슬 픔	화 남	소 계
평 상	26	9	10	5	50
기 쁨	3	28	13	6	50
슬 픔	6	8	31	5	50
화 남	4	4	3	39	50
성공률	62.00%				200

구별이 잘되는 반면, 평상 음성은 상대적으로 작은 특징계수 변화를 갖고 있어 잡음 영향을 많이 받기 때문인 것으로 해석할 수 있다. 표 3의 ARM 플랫폼 결과가 컴퓨터 시뮬레이션 결과인 표 1에 비해 약 3% 정도 성능이 떨어지는 것은 ARM 플랫폼의 연산 오차에 기인한 것이다.

5. 결 론

본 연구에서는 마이크로폰을 통해 실시간으로 습득된 음성으로부터 사람의 음성 감성상태를 평상, 기쁨, 슬픔, 화남 등 4개의 감성으로 분류할 수 있는 음성 감성인식 시스템을 ARM 플랫폼에 구현 하였다. 일반적으로 마이크로폰으로 수신된 음성은 화자 주변의 환경 잡음과 마이크로폰의 시스템 특성 때문에 입력 음성 신호가 왜곡되고 이로 인해 시스템의 성능이 저하된다. 본 논문에서는 이러한 잡음 영향을 최소화하고 강인한 감성 특징벡터를 추출하기 위해 비교적 단순한 구조의 이동평균 필터를 적용하였으며, SFS 특징벡터 최적화 기법을 적용하여 시스템 성능을 한층 더 안정화시켰다. 향후 잡음을 더욱 효과적으로 제어할 수 있다면 인공지능 컴퓨터, 로봇, 고객센터, 결혼정보회사, 모바일 콘텐츠 산업 등 다양한 산업 분야에 적용할 수 있을 것으로 기대된다.

참 고 문 헌

[1] Anil Jain and Douglas Zongker, "Feature Selection: Evaluation, Application and Small Sample Performance," *IEEE Pattern Analysis and Machine Intelligence*, Vol.19, No.2, pp. 153-158, February 1997.

[2] Frank Dellaert, Thomas Polzin, and Alex Waibel, "Recognizing Emotion in Speech," *In Proc. International Conf. on Spoken Language Processing*, Vol.3, pp. 1970-1973, 1996.

[3] Tsuyoshi Moriyama and Shinji Oazwa, "Emotion Recognition and Synthesis System on Speech," *IEEE International Conference on Multimedia Computing and Systems*, Vol.1, pp. 840-844, 1999.

[4] Albino Nogueiras, Asunción Moreno, Antonio Bonafonte, and José B. Mariño, "Speech Emotion Recognition Using Hidden Markov Models," *European Conference on Speech Communication and Technology, Eurospeech 2001,* Aalborg, Denmark Vol.1, pp. 2679-2682 2001.

[5] Noam Amir, Ori Kerret, and Dimitry Karlinski, "Classifying emotions in speech: A comparison of methods," *Proceedings of Euro Speech' 2001* Aalborg, Denmark, Vol.1, pp. 127-130, 2001.

[6] Chul Min Lee and Shrikanth S. Narayanan, "Towards Detecting Emotions in Spoken Dialogs," *IEEE Transactions on Speech and Audio Processing*, Vol.13, No.2, pp. 293-303, March 2005.

[7] Guojun Zhou, John H. L. Hansen, and James F. Kaiser, "Nonlinear Feature Based Classification of Speech Under Stress," *IEEE Transactions on Speech and Audio Processing*, Vol.9, No.3, pp. 201-216, March 2001.

[8] Dimitrios Ververidis and Constantine Kotropoulos, "Sequential Forward Feature Selection with Low Computational Cost," *Conf. of 13th EUSIPCO,* Antalya Turkey. Sep. 2005.

[9] de la Cuadra, Patricio, Aaron Master, and Craig Sapp, "Efficient Pitch Detection Techniques for Interactive Music," *International Computer Music Conference,* Havana, Vol.1, pp. 403-406, September 2001.

[10] M. Ross, H. Shaffer, A. Cohen, R. Freudberg, and H. Manley, "Average Magnitude Difference Function Pitch Extractor," *IEEE Transactions on Acoust., Speech, Signal Processing*, Vol.22, No.5, pp. 353-362, Oct. 1974.

[11] Xuejing Sun, "A Pitch Determination Algorithm Based On Subharmonic-to- Harmonic Ratio," *the 6th International Conference on Spoken Language Processing'* 2000, Vol.4, pp. 676-679, 2000.

[12] 강봉석, "음성 신호를 이용한 문장독립 감정 인식 시스템," 석사학위 논문, 연세대학교, 서울,

2001.

[13] Y. Ephraim, "A Signal Subspace Approach for Speech Enhancement," *IEEE Transactions on Speech and Audio Processing*, Vol.3, No.4, pp. 251-266. July 1995.

오 상 헌

1998년 한국해양대학교 전파공학과 학사졸업.
2000년 상명대학교 전자계산학과 석사졸업
2004년 8월 단국대학교 정보컴퓨터과학과 박사수료.
관심분야 : 멀티미디어 신호처리, DSP, ARM 시스템 구현

박 규 식

1986년 Polytechnic University 전자공학과 학사 졸업.
1988년 Polytechnic University 전자공학과 석사 졸업.
1993년 Polytechnic University 전자공학과 박사 졸업.
1994년~1996년 삼성전자 마이크로사업부, 선임 연구원
1996년~2001년 상명대학교 컴퓨터·정보통신공학부 조교수
2001년~현재 단국대학교 정보컴퓨터학부 부교수
주관심분야 : 음성 및 음향신호처리, 멀티미디어 신호처리, DSP 시스템 구현