Proceedings of KFIS Autumn Conference 2007 Vol. 17, No. 2.

# 엔터테인먼트 로봇을 위한 음성으로부터 감정 인식 및 표현 모듈 개발

# Development of Emotion Recognition and Expression module with Speech Signal for Entertainment Robot

문병현, 양현창, 심귀보

중앙대학교 전자전기공학과
(E-mail: kbsim@cau.ac.kr)

**요 약**

현재 가정을 비롯한 여러 분야에서 서비스 로봇(청소 로봇, 애완용 로봇, 멀티미디어 로봇 등)의 사용이 증가하고 있는 시장상황을 보이고 있다. 개인용 서비스 로봇은 인간 친화적 특성을 가져야 그 선호도가 높아질 수 있는데 이를 위해서 사용자의 감정 인식 및 표현 기술은 필수적인 요소이다. 사람들의 감정 인식을 위해 많은 연구자들은 음성, 사람의 얼굴 표정, 생체신호, 제스쳐를 통해서 사람들의 감정 인식을 하고 있다. 특히, 음성을 인식하고 적용하는 것에 관한 연구가 활발히 진행되고 있다. 본 논문은 감정 인식 시스템을 두 가지 방법으로 제안하였다. 현재 많이 개발되어지고 있는 음성인식 모듈을 사용하여 단어별 감정을 분류하여 감정 표현 시스템에 적용하는 것과 마이크로폰을 통해 습득된 음성신호로부터 특징들을 검출하여 Bayesian Learning(BL)을 적용시켜 normal, happy, sad, surprise, anger 등 5가지의 감정 상태로 패턴 분류를 한 후 이것을 동적 감정 표현 알고리즘의 입력값으로 하여 dynamic emotion space에 사람의 감정을 표현할 수 있는 ARM 플랫폼 기반의 음성 인식 및 감정 표현 시스템 제안한 것이다.

**Key Words** : 음성인식 및 감정표현 모듈, 엔터테인먼트 로봇, ARM플랫폼, Bayesian Learning(BL)

## 1. 서 론

가정이나 사무실 등 비정형화된 공간에서 로봇이 인간에게 와서 내가 원하는 일을 하고, 원하는 정보를 제공하는 인간 중심적인 서비스 환경이 되도록 하려면 수동적인 서비스만을 제공하는 로봇만으로는 한계가 있다. 이러한 한계를 극복하는 방법으로 로봇이 감정을 인식하고 인식된 감정을 일정한 형태로 표현하는 감정표현 시스템을 개발 하는 것이 하나의 해결 방안이 될 것이다. 따라서 공공 서비스, 홈 서비스, 엔터테인먼트, 매개치료 등의 다양한 분야에서 인간과 로봇간의 상호작용을 통한 감성적인 교류에 대한 연구가 활발히 진행되고 있다. 보다 다양한 로봇이 우리 일상생활의 일부로서 그 역할을 수행함에 따라 인간-로봇의 상호작용을 통한 감성적 교감이 가능한 형태의 연구가 앞으로 더욱 필요하며, 사용자에게 지속적인 흥미를 주기 위해 감정 표현에 대한 연구도 병행되어야 한다. Brooks는 CCD 카메라를 통하여 얻은 시각적인 정보를 이용하여 사람과 의사소통할 수 있는 로봇을 개발하였다[1]. 최근에는 입, 눈, 눈썹 등을 지닌 휴머노이드 로봇으로써 인간의 감정 표현에 최대한 가까이 접근해가려는 연구가 많다[2][3]. 음성은 표정과 함께 감성 인식의 가장 중요한 매체이다. 전화의 통신 대역이 좁음에도 불구하고 대부분의 감정을 서로 인식할 수 있다는 점은 음성으로 감정 인식을 할 수 있다는 가능성을 확인 시켜주는 점이다. 즉, 이미지의 경우 여러 환경적인 조건들로 인해 감정 데이터에 잡음이 많이 끼는 상황에서 인식률 저하의 우려가 있지만 음성의 경우는 좁은 대역에서도 감정인식이 잘 되는 것으로 볼 때 잡음에 강인한 특성을 갖고 있다고 볼 수 있다. B. Schuller는 사람의 음성을 기본으로 해서 Hidden Markov model을 연구하였다[4]. 음성의 감정 상태를 분류하는 패턴 인식 알고리즘으로는 k-NN(Nearest Neighbor), HMM(Hidden

감사의 글 : 본 연구는 서울시·중소기업청의 연구비 지원에 의한 2007년도 중앙대학교 산학연 컨소시엄 사업에 의해 수행되었습니다. 연구비 지원에 감사드립니다.

Markov Model), SVM(Support Vector Machine), NN(Neural Network) 등의 방법들이 사용되고 있으며[5], Yi-lin lin는 HMM(Hidden Markov Model)과 SVM(Support Vector Machine)을 이용하여 음성을 화남, 행복, 슬픔, 놀람, 보통 등의 5가지의 감정 상태로 분류하였다[6].

본 연구에서는 첫 번째 방법으로서 현재 많이 개발되어지고 있는 음성인식 모듈을 응용하여 음성을 통한 감정인식 모듈을 제작, 주변 환경으로부터 감지된 소리를 입력받아서 환경 및 감정을 인식하고 그래픽 LCD상에서 아바타형식으로 감정을 표현하는 시스템을 구현하고, 두 번째 방법으로는 감정 상태를 동적 감정 공간에 표현할 수 있는 ARM 플랫폼 기반의 시스템 구현을 목표로 하였다. 제안한 음성인식 및 감정표현 시스템은 기존의 PC기반의 시스템과는 다르게 ARM 플랫폼에서 구현하였으며 ARM 프로세서에 임베디드하였다.

본 논문의 구성은 다음과 같다. 2절에서는 AVR(ATmega-128)과 음성 인식 모듈 그리고 그래픽LCD를 이용하여 구성한 감정표현 시스템을 설명하고, 3절에서는 ARM 플랫폼 기반의 감정 인식 및 표현 시스템의 구성을 보일 것이다. 4절에서는 향후 연구 방향과 결론으로 마무리 짓는다.

## 2. 음성 인식 모듈을 이용한 감정 표현 시스템 개발

### 2.1 음성 인식 모듈을 이용한 시스템 개발

음성 인식 모듈에 연결된 마이크로폰을 통해 사람의 음성이 음성 인식 모듈의 입력이 된다. 음성 인식 모듈을 본 개발 시스템에 적합하도록 디버깅하고 감성 음원 인식 및 표현 모듈 시스템에 부착하여 일정한 단어별로 감정을 분류한 후, 사용자가 정해진 단어 중 어느 한 단어의 음원을 입력받게 되면 그 단어에 포함된 감정을 MCU가 인식한 후 그 결과를 감정표현 시스템의 입력 값으로 제공하여 감정표현 시스템 모듈에서 아래 그림 과 같이 디스플레이하게 된다. 음성 인식 모듈은 총 8개까지의 음성을 인식할 수 있는 3MS사의 제품을 사용하였으며, 학습할 단어의 수를 늘리기 위해 음성 인식 모듈을 추가 부착하였다. 지정된 사용자만이 아닌 다른 사용자들의 음성 또한 인식할 수 있는 화자 독립 방식을 위하여 200명의 음성 샘플을 데이터베이스화 하였다. 감정 표현 모듈로는 현대 LCD사의 HG12605NY-LY (126X64), MCU는 ATMEL사의 ATmega-128을 사용하였다. 다음의 그림 1은 본 연구에서 구현한 음성인식 모듈을 이용한 감정 표현 시스템의 구성도를 나타내고 있다.

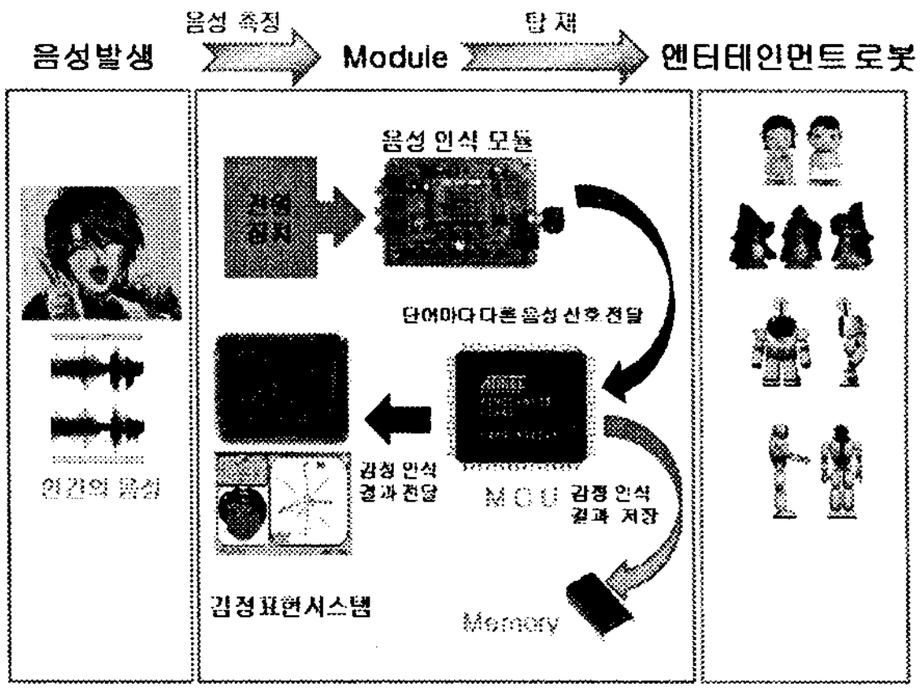


그림 1. 음성인식 모듈을 이용한 감정 표현 시스템 구성도

### 2.2 실험

다음의 표 1은 음성 인식시스템의 학습을 위해 사람의 감정에 따라 언급되는 단어들을 조사하여 정리한 것이다. 기본적인 감정 5가지(행복, 배고픔, 서운함, 화남, 놀람)에 대해 50명의 남녀(남: 30명, 여: 20명, 나이: 20~31) 대학원생에게 각 감정별 상황 속에서 하게 되는 말들을 하나씩 선정하게 하여 각 감정별 최다 우선순위 4개의 단어를 선정하였다. 이렇게 선정된 20개의 단어들은 그것들을 감정 데이터로 채택해도 될지 확인을 받아야 하기 때문에 녹음한 사람들 이외의 다른 10명에게 "다음의 단어가 어떤 감정을 포함하고 있는 것 같은가?"라는 질문을 하였고, 91.9%의 동의를 얻었다.

표 1. 감정별 언급되는 단어

| 감 정 | 단 어 |
|---|---|
| **a.행복할 때** | 아싸!! OK!! 야호!! |
| **b.배고플 때** | 배고파!! 밥먹자!! 뭐 먹을래? |
| **c.서운할 때** | 너무해~! 너무한다~! 그럴 수 있냐? |
| **d.화났을 때** | 젠장!! 제길!! 짜증나!! |
| **e.놀랐을 때** | 어떡해~! 어머~! 정말? |

그림 2는 LCD화면에 음성 인식 결과를 디스플레이한 결과이다. 그림의 a, b, c, d, e는 순서대로 행복, 배고픔, 서운함, 화남, 놀람에 대한 아바타(Avatar)들이다. 음성 인식 모듈을 이용한 감정 표현 시스템은 엔터테인먼트 로봇의 안면에 부착하는 것을 최종 목표로 한다.

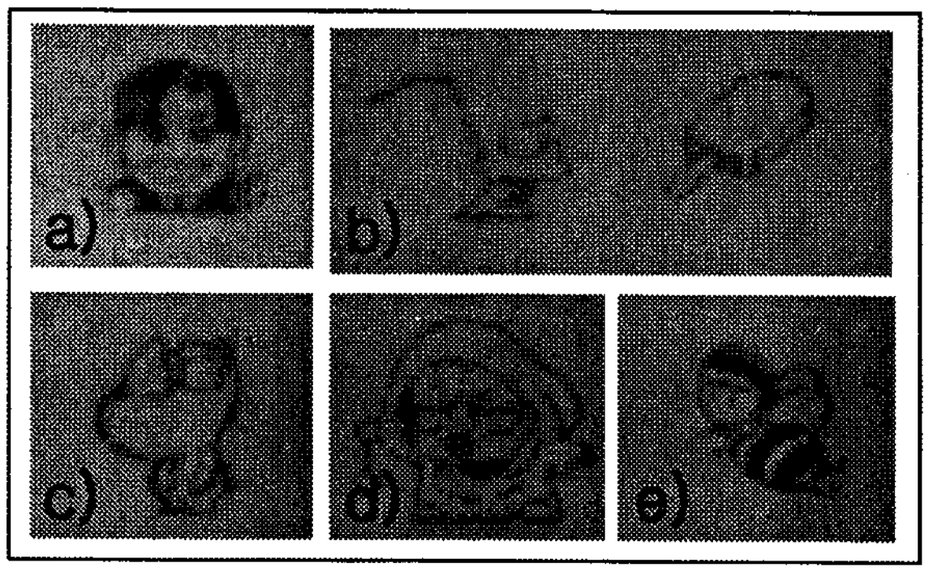


그림 2. LCD에 디스플레이된 음성인식 결과 아바타

## 3. ARM 플랫폼 기반의 음성 인식 및 감정 표현 시스템 개발

다음의 그림 3은 본 연구에서 두 번째 방법으로 제안한 ARM920T 플랫폼의 구성을 나타낸다. 시스템의 동작 과정은 ARM 프로세서에 연결된 마이크로폰을 통해 입력된 음성 신호로부터 소리의 크기, 섹션 개수, 피치의 평균, 피치의 최대점, IR(Increasing Rate), CR(Crossing Rate) 등 6가지의 특징을 검출하여 Bayesian Learning(BL)을 적용시켜 그 특징들의 패턴 분류를 하게 되고 그 결과는 동적 감정 공간의 입력값이 된다. 감정 표현 시스템은 동적 감정 공간의 입력값을 받아서 그 결과를 ARM 보드의 자체 LCD에 출력하게 된다. 본 연구에서는 삼성의 S3C 2440A (ARM920T 400Mhz 코어)를 사용하였다. S3C2440A는 최대 400MHz까지의 빠른 속도와 다양한 Peripheral을 가지고 있는 강력한 프로세서이며 파이프라인 구조에 의해 Mhz당 평균 1.1 MIPS(Million instructions per second) 능력을 가진다. 440 MIPS의 정수 연산능력을 가지며 64Mbyte의 내부 메모리와 2Gbyte의 외장 메모리 공간을 보유하고 있다.

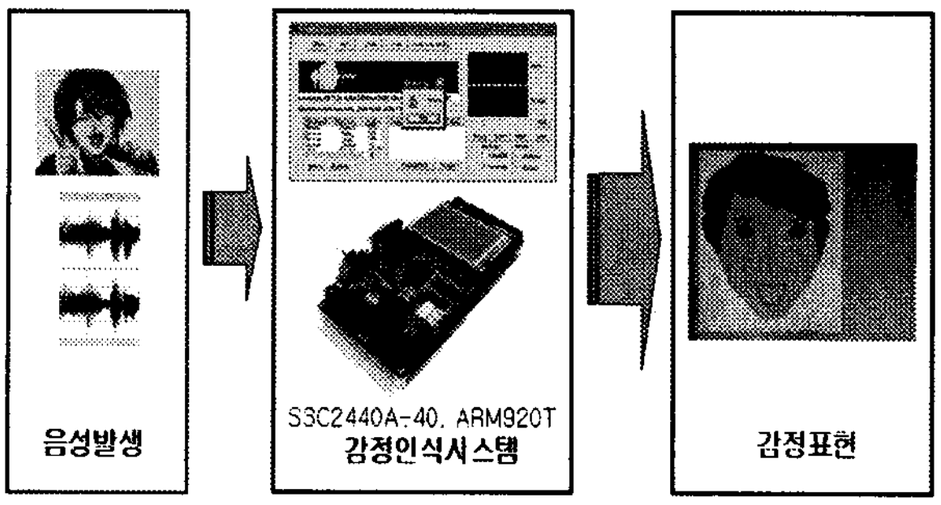


그림 3. ARM 플랫폼 음성 인식 및 감정표현 시스템 구성도

### 3.1 특징 추출

감정 인식기에 이용하는 특징들은 피치의 통계치, 소리의 크기, 섹션 개수 등이다. 피치의 추출 방법으로는 autocorrelation approach를 사용했으며 피치값은 0.1초마다 추출되어졌고, 그 값들의 평균을 pitch mean으로 정의 했다. 그리고 분산값 또한 동일한 데이터에서 얻었다. 소리의 크기는 magnitude estimation method에 의해서 구했다[5].

### 3.2 Bayesian Learning을 이용한 인식 실험

베이즈의 이론은 사전확률에 근거해 가설의 확률을 계산하는 방법을 제공하기 때문에 해당 문제에 대한 사전 확률을 구하기 위해서 400개의 샘플들을 사용하였다.

실험 결과 표 2와 그림 3의 결과로부터 알 수 있는 바와 같이 인식율이 사람마다, 그리고 감정별로 다른 것을 확인할 수 있었는데 이것은 사람마다 감정을 표현 하는 방식이 조금씩 다르기 때문이다[7].

표 2. BL을 이용한 감정 인식 결과(인식률)

| | Normal | Happy | Angry | Depress | Average |
|---|---|---|---|---|---|
| S1 | 57% | 40% | 80% | 70% | 62% |
| S2 | 90% | 73% | 80% | 94% | 84% |
| S3 | 70% | 51% | 89% | 91% | 75% |
| S4 | 67% | 56% | 91% | 85% | 75% |
| Average | 71% | 55% | 85% | 85% | 74% |

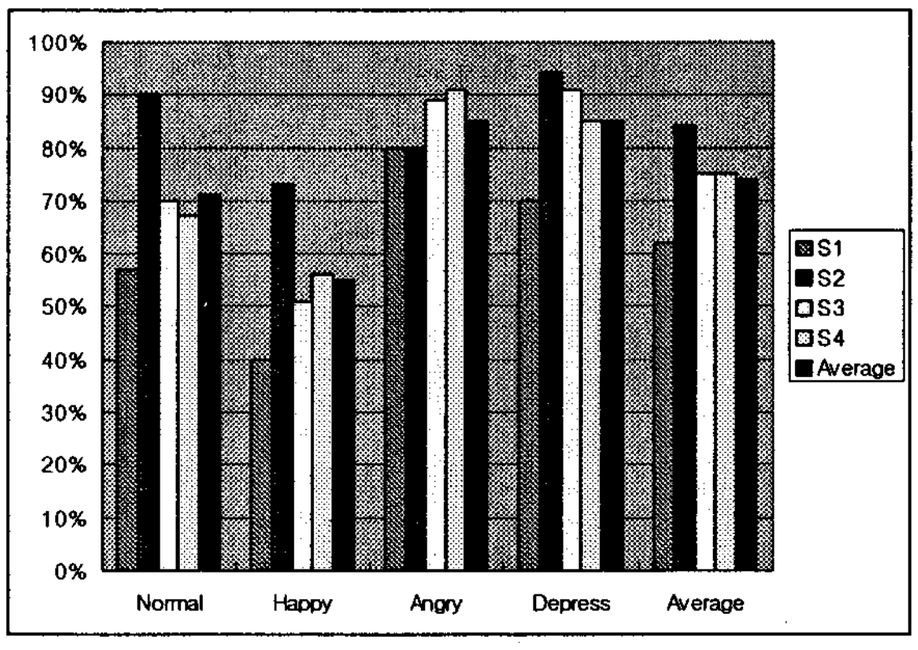


그림 4. BL을 이용한 감정 인식률 그래프

### 3.3 감정 표현 시스템

LCD로 출력되는 감정 표현 시스템은 동적으로 변화하는 2차원 감정 공간(emotion space)모델을 적용하였다. 이는 입력되는 감정의 가중치들의 조합과 학습에 의해 축적된 감정 빈도수를 이용하여 감정 공간의 좌표축을 변화시키고 감정 영역에 속하는 비율로서 얼굴 특징 파라미터들을 변화시켜 자연스러운 표정을 표현한다[8]. 다음의 그림 3은 ARM 플랫

폼 기반의 시스템에서 감정 표현 시스템인 ARM920T의 자체 LCD에 뿌려진 아바타이다.

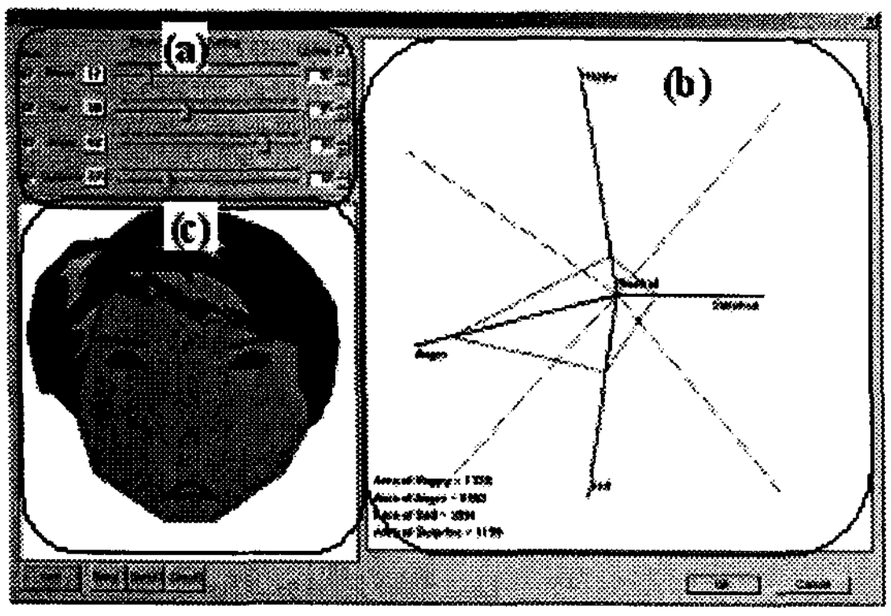


그림 5. ARM플랫폼 기반의 감정표현 시스템 결과 아바타

## 4. 결론 및 향후 연구 방향

본 연구에서는 음성인식 모듈을 사용하여 단어별 감정을 분류하여 AVR을 이용하여 감정 표현 시스템에 적용하는 것과 ARM 플랫폼 기반에서 음성 신호를 normal, happy, sad, surprise, anger 등 5가지의 감정 상태로 패턴 분류를 한 후 이것을 동적 감정 표현 시스템에서 표현할 수 있는 ARM 플랫폼 기반의 음성 인식 및 감정 표현 시스템 구현하였다. 첫 번째 방법으로 구현한 음성인식 모듈을 이용하여 감정을 인식하고 Display 장치를 통해 감정을 표현하는 시스템은 향후 엔터테인먼트 애완용 로봇의 안면에 부착되어 상용화되면 인간 친화적인 로봇 개발의 추세에 따라 인간과 로봇간의 상호작용을 통한 감성적 교류에 대한 연구에 조금이라도 이바지할 수 있을 것으로 생각한다. 현재 ARM 플랫폼 기반의 음성 인식 및 감정 표현 시스템은 기존의 Bayesian Learning 패턴 인식 알고리즘을 이용하였으나 PCA 알고리즘과 LBG 알고리즘을 이용한 연구를 진행중에 있으며, 인식률을 보다 향상 시킬 수 있는 알고리즘을 개발하기 위한 연구를 진행하고 있다.

## 참 고 문 헌

[1] Cynthia Breazeal, Brian Scassellati, "How to build robots that make ends and influence people, *IROS99*, pp.858-863, 1999.

[2] Hiroshi Kohayshi, et al, "Study on Face Robot for Active Human Interface - Mechanisms on Face Robot and Facial Expressions of 6 Basic Emotions -, *the Journal of the Robotics Society of Japan* Vol.12, No.1, pp.155-163, 1994.

[3] Hiroshi Kobayashi, Fumio Ham, "Real Time Dynamic Control of 6 Basic Facial Expressions on Face Robot, *the Journal of the Robotics Society of Japan* Vol.14, No.5, pp.677-685, 1996.

[4] B. Schuller, G. Rigoll, M. Lang, "Hidden Markov model-based speech emotion recognition", *Proc. of the IEEE ICASSP Conference*, pp. 1-4, April 2003.

[5] Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., and Taylor, J. G., "Emotion recognition in human-computer interaction", *IEEE Signal Processing magazine*, Vol. 18, No. 1, pp. 32-80, January 2001.

[6] Yi-Lin Lin, Gand wei, "Speech emotion recognition based on HMM and SVM, *Proc. of the Fourth International Conference*, 18-24 August 2005.

[7] Chang-Hyun Park., "The Pattern Recognition Methods for Emotion Recognition with Speech Signal", *Journal of Control, Automation and Systems Engineering*, Vol. 12, No. 3, March 2005.

[8] 심귀보, 변광섭, 박창현, "동적 감정 공간에 기반한 감정 표현 시스템", *한국퍼지 및 지능시스템학회 논문지*, 제15권, 제1호, pp. 18-23, 2005.