

---

# HEEAS: 감정표현 애니메이션 알고리즘과 구현에 관한 연구

## HEEAS: On the Implementation and an Animation Algorithm of an Emotional Expression

---

민용식\*, 김상길\*\*  
호서대학교 컴퓨터공학부\*, 한영신학대학교\*\*

Yong-Sik Min(ysmin@office.hoseo.ac.kr)\*, Sang-Kil Kim(ksknet2000@hytu.ac.kr)\*\*

---

### 요약

본 논문은 음성이 인간에게 전달되어 나타나는 여러 가지 감정 표현 중에서 단지 4가지 감정 즉 두려움, 싫증, 놀람 그리고 중성에 대한 감정 표현이 얼굴과 몸동작에 동시에 나타내는 애니메이션 시스템인 HEEAS(Human Emotional Expression Animation System)를 구현하는데 그 주된 목적이 있다. 이를 위해서 본 논문에서는 감정 표현이 풍부한 한국인 20대 청년을 모델로 설정하였다. 또한 입력되어진 음성 신호를 통해서 추출된 감정표현에 대한 데이터를 얼굴코드와 몸동작코드를 부여하고 이를 데이터베이스화 하여 실제 애니메이션 구현을 하기위한 처리의 시간을 최소화하였다. 즉, 입력되어진 음성 신호를 이용해서 원하는 결과인 얼굴, 몸동작에 대한 자료를 이진 검색을 이용해서 데이터베이스에서 찾으므로 검색 시간을 최소화하였다. 실제 감정 표현에 대한 문제들을 실험을 통해서 얻은 결과가 99.9%의 정확도임을 알 수가 있었다.

■ 중심어 : | 애니메이션 | 음성감성인식 | 멀티미디어 HCI 응용 |

### Abstract

The purpose of this paper is constructed a HEEAAS(Human Emotional Expression Animaion System), which is an animation system to show both the face and the body motion from the inputted voice about just 4 types of emotions such as fear, dislike, surprise and normal. To implement our paper, we chose the korean young man in his twenties who was to show appropriate emotions the most correctly. Also, we have focused on reducing the processing time about making the real animation in making both face and body codes of emotions from the inputted voice signal. That is, we can reduce the search time to use the binary search technique from the face and body motion databases, Throughout the experiment, we have a 99.9% accuracy of the real emotional expression in the cartoon animation.

■ Keyword : | Animation | Speech Emotion Recognition | Multimedia HCI Application |

---

\* 본 연구는 2005년도 호서대학교 학술연구비 지원에 의해 수행되었습니다.

접수번호 : #051114-002

심사완료일 : 2006년 01월 10일

접수일자 : 2005년 11월 14일

교신저자 : 민용식, e-mail : ysmin@office.hoseo.ac.kr

## I. 서론

면접이나 상담을 할 때 내담자의 음성언어뿐만 아니라 비 음성언어에 귀 기울이는 것이 중요하다. 이유는 비음성언어가 “누설적인” 성격을 띠기 때문이다[10]. 내담자가 어떤 말을 하는 도중에 뺨이나 이마 혹은 목 부분에 홍조를 띠면 이는 그가 수치감 또는 당황을 경험하고 있다는 것을 “누설”하는 셈인데 이 때 이런 비음성적인 단서들을 민감하게 관찰하고 알아차리는 능력은 상담자에게 매우 중요하다. 비음성언어의 경청은 전문가나 비전문가에게 매우 중요한 일이다. 그러나 전문가는 변화를 직접적으로 유도하지만, 비전문가는 변화를 간접적으로 유도하기 때문에 비전문가는 내담자의 마음과 생각을 분명히 아는 것이 더욱 필요하다.

Albert Mehrabian은 전달되는 메시지의 전체적인 영향중에서 약 7%가 말(단지 이야기)에 의한 것이고 38%는 목소리(음조 및 억양, 그 밖의 다른 소리를 포함하여)에 의한 것이며, 55%가 비음성적인 것이라는 것을 알아내었다. 또한 Birdwhistel도 얼굴을 맞대고 하는 대화중에서 음성적 구성 요소는 35% 미만이며 의사 전달의 65% 이상이 비음성적으로 이루어진다고 밝힌 바 있다[10].

또한 음성은 휴먼 인터페이스 중에서 가장 많은 정보를 포함하고 사용이 편리하다는 장점 때문에 여러 분야에서 활용되고 있다[10]. 하지만 음성은 주변의 소음이 일정수준 이상이 되면 전달되기 어려운 특성을 가지고 있다. 따라서 음성만 전달하는 것 보다 음성 정보를 포함한 모습이나 억양, 표정 그리고 웃 등을 통해서도 자신의 성격을 드러내고 자신이 전달하고자 하는 내용을 더욱 효과적으로 전달한다.

입력된 음성으로 인해 표출되는 인간의 감정 카테고리가 무수하게 많다. 이러한 점을 고려하여 본 논문에서는 단지 4개의 감정 카테고리 즉 중성, 두려움, 싫증 및 놀람의 감정이 얼굴과 몸동작으로 나타나는 애니메이션 시스템을 구현하려고 한다. 이렇게 하기 위해서

본 논문에서는 첫째 감정표현(음성, 얼굴, 몸동작)들의 훈련자료를 만들어 각각의 감정표현 종류에 따라서 다르게 코드를 부여하고 감정별 데이터베이스로 구축한다. 둘째 입력된 음성 시그널이 4가지의 감정(중성, 두려움, 싫증, 놀람)의 어느 부류에 속하는지 알기 위해 여러 차원의 판별장치를 설치했다.

셋째 판별장치를 통한 감정별로 구분된 음성 신호를 가지고 이진 검색 기법을 사용하여 데이터베이스에서 찾는다. 마지막으로 데이터베이스에서 찾아진 자료들이 구분되어(얼굴, 얼굴+몸동작) 화면으로 나타나는 얼굴과 몸동작 만화 애니메이션 시스템을 구현한다. 이 구현을 통해서 결과는 99.9%의 정확도를 얻을 수 있었음을 알 수 있었다. 이렇게 만들어진 애니메이션 시스템은 날로 발전하는 산업사회에 비음성적인 메시지에 대한 지식이 부족한 사람도 비음성적 메시지를 인지하는데 많은 도움이 될 수 있을 것이다.

## II. 감정표현 데이터베이스

### 1. 훈련자료 수집

본 논문에서 사용되어질 4가지의 감정(중성, 두려움, 싫증, 놀람)에 해당하는 자료를 수집하기 위해 영화필름을 대상으로 데이터를 수집했다. 수집한 필름에서 4가지의 감정(중성, 두려움, 싫증, 놀람)에 해당하는 부분만을 발췌했다. 발췌 시 그 길이를 10초가 넘지 않도록 했다. 그 이유는 영화필름을 선택할 때 두려움, 싫증, 놀람의 장면이 많은 영화를 선택했기 때문이다. 그렇게 모여진 필름은 평균길이가 약 10초인 것이 모두 3,000개이다. 이같이 수집된 데이터를 이용해서 이 음성의 특징 점을 발췌하고 분석하는데 이용하고자 한다.

### 2. 훈련자료 코드 부여

감정의 강도에 따라 %로 분류되어진 4가지의 감정(중성, 두려움, 놀람, 싫증) 자료를 데이터베이스에 저장하기 위해 각각에 코드를 부여했다. 코드 부여는 다음의 방법에 의한 블록코드 분류법을 사용했으며 그 체계는 4단계로 나누어진다[표 1]. 나누어진 구성을 보면 1

단계 데이터 종류로 문자 1자리 크기, 2단계 감정의 종류로 문자 1자리 크기, 3단계 감정의 강도로 숫자 1자리 크기, 4단계 3자리크기의 일련번호가 된다. 단계별 내용을 보면 1단계는 음성자료에 “S”, 영상자료에 “M”, 얼굴자료에 “F”, 몸동작 자료에 “B”를 넣어 구성한다. 2단계는 감정종류의 분류로 중성에 “N”, 두려움에 “F”, 놀람에 “S”, 싫증에 “D”를 넣어 구성한다. 3단계는 감정의 강도로 0%에 “1”, 25%에 “2”, 50%에 “3”, 75%에 “4”, 100%에 “5”를 부여했다. 4단계는 각각의 감정별 자료 데이터 건수를 표시하는 것으로 3자리의 일련번호를 부여했다. 예를 들어 음성 자료에서 두려움의 25%는 SF2001~SF2090까지가 된다.

표 1. 훈련자료 코드 체계

| 데이터 종류 | 감정의 종류 | 감정의 강도 | 일련번호 |
|--------|--------|--------|------|
| 1      | 1      | 1      | 3    |

(1,1,1,3은 자릿수 크기)  
 DB 종류 : S(음성), M(영상), F(얼굴), B(몸동작)  
 감정의 종류 : 중성(N), 두려움(F), 싫증(D), 놀람(S)  
 감정의 강도 : 0%(1), 25%(2), 50%(3), 75%(4), 100%(5)  
 일련번호 : 데이터 건수

### 3. 감정표현 데이터베이스 구조도

음성자료로 준비한 필름에서 음성과 영상을 분리하여 저장했다. 이 영상자료는 나중에 결과를 확인하기 위해 사용되어진다. 또한 얼굴자료와 몸동작 자료는 음성에 따른 감정이 얼굴과 몸동작으로 표현되는 모습을 일러스트레이터를 사용해 그림으로 그리고 코드를 부여하여 저장한다.

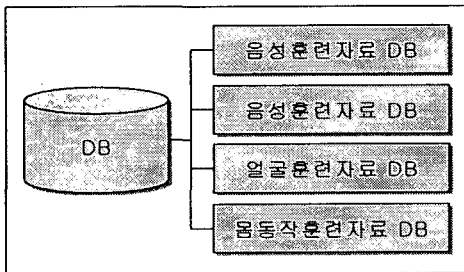


그림 1. 데이터베이스(DB:DateBase) 구조도

### III. 음성자료 데이터베이스

입력된 음성 중에서 4가지 감정(중성, 두려움, 싫증, 놀람)만을 선별하여 음성자료 데이터베이스를 구축한다. 이때 음성에 관련된 어조의 특징 점은 [5][11]에서 사용한 것처럼 음향의 두드러진 점을 발췌하여 감정에 음성을 삽입시키도록 하고자 한다. 각각의 어조에 대해 다음과 같은 요소를 가지고 16 차원의 벡터를 이용하였다.

- 리듬과 관련된 통계: 말하는 속도, 소리가 나는 지역 사이의 평균 길이, 최대 횡수/(최소 + 최대)횡수, 위경사의 횡수/경사의 횡수
- 평온한 음높이 시그널에 대한 통계: 최소, 최대, 중앙, 표준 편차
- 평온한 음 높이의 미분계수에 대한 통계: 최소, 최대, 중앙, 표준 편차
- 개별 소리 부분에 대한 통계: 최소 평균, 최대 평균
- 개별 경사에 대한 통계: 평균 양수 편차, 평균 음수 편차

#### 1. 어조의 분류

얼굴과 몸동작 애니메이션을 작동시키기 위해 연속 음성을 사용했다. 또한 감정이 짧은 기간에서 안정적인 수 없기 때문에 여러 차원의 판별장치를 설치하고 음성을 적당한 부류로 분류했다. 각각의 구로 분절된 구에 대해서는 감정의 비율을 나타내는 인지 연산을 적용시켰다. 분절연산은 “감정의 강도는 주어진 구내에서는 불변하고, 구가 침묵 또는 분류기가 구의 감정 분류를 정할 수 없을 때는 중성이다”는 가정 하에 이루어진 것이다. 연속적인 음성 감정인지의 검증의 판별장치는 [5][11]에 제시된 방법에 의해서 이루어졌다.

이러한 과정을 통해 얻어진 어조의 분류를 4가지(중성, 두려움, 싫증, 놀람)로 구분했다[표 2]. 우리는 감정의 강도에 따라 나타나는 것을 구분하기 위해 4가지(중성, 두려움, 싫증, 놀람)의 감정을 0%, 25%, 50%, 75%, 100%로 상세하게 구분했다.

표 2. 각 감정에 대한 샘플을 다룬 횟수

| 중성  | 두려움 | 놀람  | 싫증  |
|-----|-----|-----|-----|
| 311 | 384 | 205 | 197 |

## 2. 분류된 어조의 채택 작업

구분되어진 감정들의 진위를 가리기 위해 어조를 잘 구별할 수 있는 5명의 애니메이션 전문가를 선정했다. 5명의 전문가가 어조를 구별하는 정확성을 높이기 위해 전문가들을 서로 격리시키고 작업하였다. 서로 격리되어진 장소에서 감정별 어조를 듣게 했다. 어조를 듣고 그들에게 느끼는 감정을 분류하게 했다. 최종으로 분류되어진 감정들의 정확도를 높이기 위해 선정한 5명의 감정이 일치된 것만 채택했다[표 3].

표 3. 감정을 상세하게 분류한 횟수

| 구분  | 0%  | 25% | 50% | 75% | 100% |
|-----|-----|-----|-----|-----|------|
| 중성  | 311 | 0   | 0   | 0   | 0    |
| 두려움 | 0   | 90  | 96  | 96  | 102  |
| 놀람  | 0   | 50  | 50  | 52  | 53   |
| 싫증  | 0   | 45  | 50  | 62  | 40   |

## 3. 채택된 자료에 코드부여

음성 자료는 앞에 “S”를[표 4] 영상 자료는 앞에 “M”[표 5]을 부여했다. 감정종류의 분류는 중성에 “N”, 두려움에 “F”, 놀람에 “S”, 싫증에 “D”를 부여했다. 감정의 강도의 분류는 0%에 “1”, 25%에 “2”, 50%에 “3”, 75%에 “4”, 100%에 5를 부여했다. 각각의 감정별 %에는 3자리의 일련번호를 부여했다. 예를 들어 두려움의 25%는 SF200 .....까지가 된다.

훈련자료로 음성 자료가 필요하며 영상자료는 증명자료로 사용하기 위해 서로를 분리 저장했다. 구별을 위해 음성 자료에는 앞에 “S”를 영상자료의 앞에는 “M”을 부여하여 나누고 데이터베이스로 구축하기 위해 저장했다. 코드가 부여된 음성 자료를 음성 데이터베이스로 구축했다.

표 4. 감정별, 감정의 강도별 음성자료 코드표

| 구분  | 0%              | 25%             | 50%             | 75%             | 100%            |
|-----|-----------------|-----------------|-----------------|-----------------|-----------------|
| 중성  | SN1001<br>..... | 0               | 0               | 0               | 0               |
| 두려움 | 0               | SF2001<br>..... | SF3001<br>..... | SF4001<br>..... | SF5001<br>..... |
| 놀람  | 0               | SS2001<br>..... | SS3001<br>..... | SS4001<br>..... | SS5001<br>..... |
| 싫증  | 0               | SD2001<br>..... | SD3001<br>..... | SD4001<br>..... | SD5001<br>..... |

표 5. 감정별, 감정의 강도별 영상자료 코드표

| 구분  | 0%              | 25%             | 50%             | 75%             | 100%            |
|-----|-----------------|-----------------|-----------------|-----------------|-----------------|
| 중성  | MN1001<br>..... | 0               | 0               | 0               | 0               |
| 두려움 | 0               | MF2001<br>..... | MF3001<br>..... | MF4001<br>..... | MF5001<br>..... |
| 놀람  | 0               | MS2001<br>..... | MS3001<br>..... | MS4001<br>..... | MS5001<br>..... |
| 싫증  | 0               | MD2001<br>..... | MD3001<br>..... | MD4001<br>..... | MD5001<br>..... |

## IV. 얼굴과 몸동작 데이터베이스

본 논문에서 사용될 감정별(중성, 두려움, 싫증, 놀람)로 얼굴과 몸동작 훈련자료를 수집하여 얼굴과 몸동작 데이터베이스를 구축한다.

### 1. 모델 선정

만화 얼굴, 몸동작 애니메이션의 실행을 위해선 입력된 음성을 얼굴과 몸동작으로 내기위한 훈련 데이터가 필요하고, 훈련 데이터를 만들기 위해서는 실제로 입력된 감정을 나타내 줄 모델이 필요하다. 그런데 감정의 표출은 사람의 문화에 따라 차이가 있다는 점을 고려하여 본 논문에서는 20대의 한국청년을 모델로 선정했다.

### 2. 훈련자료 사진작업

분류된 음성 자료를 모델에게 들려주고 모델에게 느낀 감정을 표현하도록 했다. 모델의 감정 표출로 나타난 얼굴과 몸동작의 모습을 750만화소의 디지털

카메라로 촬영했다. 사진을 4가지(중성, 두려움, 싫증, 놀람)의 감정인 0%, 25%, 50%, 75%, 100%로 상세하게 구분(몸동작과 얼굴을 각각 %별로)하여 촬영했다. 촬영 방법은 음성자료를 20가지의 종류별로 반복하여 모델에게 들려주고 촬영했다. 촬영 시 짧은 시간에 같은 음을 반복하여 듣게 되면 자연스러운 표정이 나오지 않을 것을 고려하여 하루에 3차례(오전, 오후, 저녁)촬영했다. 사진은 약 60일 동안 약 3000장을 촬영했다. 이렇게 구분하여 촬영한 감정들의 사진을 선별하기위해 선정한 5명의 애니메이션 전문가들에게 감정을 0%, 25%, 50%, 75%, 100%로 상세하게 구분하여 선별하게 했다. 이때 사용된 자료의 전체는 1,097장을 이용하였다.

### 3. 얼굴과 몸동작 훈련자료 만들기

입력되는 음성의 감정이 얼굴과 몸동작으로 나타난 만화 얼굴 몸동작 애니메이션의 실행을 위해 얼굴과 몸동작자료를 만든다. 얼굴과 몸동작 그림은 사진작업을 통해 감정별, 감정의 강도별(20가지)로 선별된 얼굴과 몸동작 사진을 일러스트레이터 전문가에게 주어 그리게 했다. 그림을 그릴 때 몸동작은 팔의 동작을 중점을 두어 그리고, 얼굴 표정사진은 입, 눈, 코를 중점을 두어 그리게 했다. 그려진 그림을 분류하여 정리한 몸동작을 그린 것[그림 2], 얼굴 표정을 그린 것[그림 3]이며, 몸동작과 얼굴 표정을 합하여 그린 것[그림 4]이다.

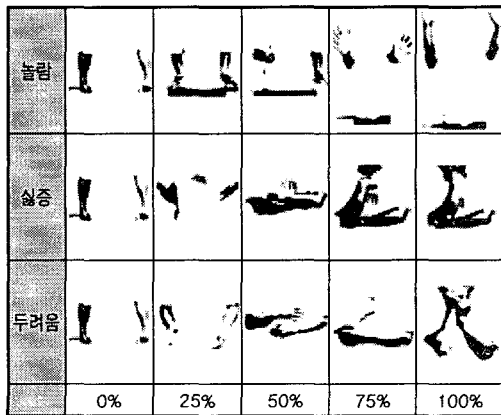


그림 2. 몸동작(중성, 두려움, 싫증, 놀람)

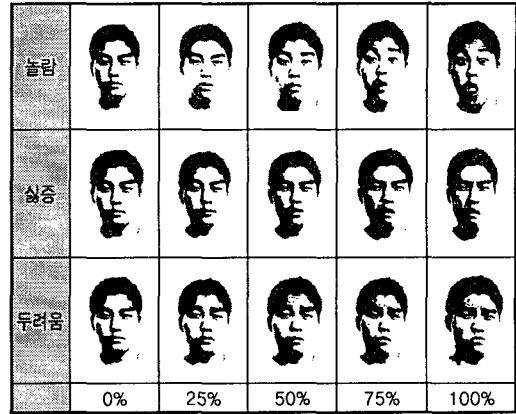


그림 3. 얼굴(중성, 두려움, 싫증, 놀람)

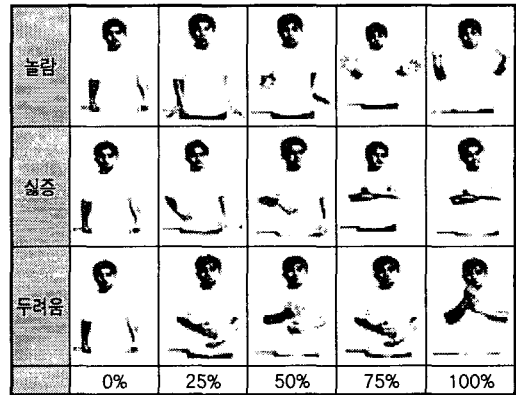


그림 4. 몸동작과 얼굴(중성, 두려움, 싫증, 놀람)

### 4. 채택된 얼굴과 몸동작에 코드 부여

4가지의 감정(중성, 두려움, 놀람, 싫증)을 감정의 강도에 따라 %별로 분류되어진 자료를 데이터베이스에 저장하기위해 각각에 코드를 부여했다. 코드 구성의 1단계에 얼굴 자료는 "F"를[표 6] 몸동작 자료는 "B"[표 7]을 부여했다. 코드 구성의 2단계는 음성 자료와 같게(중성(N), 두려움(F), 두려움(F), 놀람(S), 싫증(D)) 부여하였다. 코드 구성의 3단계도 음성 자료와 같게(0%(1), 25%(2), 50%(3), 75%(4), 100%(5))를 부여했다. 코드 구성의 마지막 단계인 4번째는 음성 자료와 마찬가지로 3자리의 일련번호를 주어 해당 자료의 수량을 알 수 있게 했다. 예를 들어 얼굴 자료에서 두려움의 25%는 FF2001.....까지가 된다. 코드가 부여된 자료를 얼굴 데이터베이스로 구축했다.

표 6. 얼굴자료 코드표(감정별, 감정의 강도별)

| 구분  | 0%               | 25%               | 50%              | 75%              | 100%             |
|-----|------------------|-------------------|------------------|------------------|------------------|
| 중 성 | FN1001...<br>... | 0                 | 0                | 0                | 0                |
| 두려움 | 0                | FF2001...<br>...  | FF3001...<br>... | FF4001...<br>... | FF5001...<br>... |
| 놀림  | 0                | FS2001...<br>...  | FS3001...<br>... | FS4001...<br>... | FS5001...<br>... |
| 싫증  | 0                | FD2001...<br>...5 | FD3001...<br>... | FD4001...<br>... | FD5001...<br>... |

표 7. 몸동작자료 코드표(감정별, 감정의 강도별)

| 구분  | 0%               | 25%              | 50%              | 75%              | 100%             |
|-----|------------------|------------------|------------------|------------------|------------------|
| 중 성 | BN1001...<br>... | 0                | 0                | 0                | 0                |
| 두려움 | 0                | BF2001...<br>... | BF3001...<br>... | BF4001...<br>... | BF5001...<br>... |
| 놀림  | 0                | BS2001...<br>... | BS3001...<br>... | BS4001...<br>... | BS5001...<br>... |
| 싫증  | 0                | BD2001...<br>... | BD3001...<br>... | BD4001...<br>... | BD5001...<br>... |

V. 얼굴과 몸동작 애니메이션 알고리즘

인간의 감정표현을 위해서 본 논문의 HEEAS의 전체적인 알고리즘은 다음과 같다. 필요한 초기화 작업으로는 영화에서 추출을 할 수 있는 각 감정별 데이터들을 데이터베이스를 통해서 구축을 하면 된다. 또한 얼굴의 표정과 몸동작에 대한 표정이 각 상황별로 데이터베이스화 되어져 있다고 하자.

그리고 음성을 인지한 후 각 인지 비율에 따라 얼굴과 몸동작에 대한 데이터베이스에서 이진 검색을 통해서 조사하여 필요한 부분을 출력하는 방법을 이용한 것이다.

```

procedure HEEAS
// 감정을 동반한 음성을 이용한 얼굴, 얼굴+몸동작 애니메
이션 생성 방법 //
{
1. // 음성 인식//
Recognize_continuous_speech( );
2. //감정이 수반된 얼굴과 얼굴+몸동작 음성동기화 구성//
Synchronize_with_emotion_and_speech( );
3. // 원하는 결과를 원하는 위치에 표현 //
Output
}
    
```

다음의 알고리즘은 연속된 음성을 적당한 분류로 분류하기 위한 여러 차원의 판별장치를 설치하여 사용을 하였다. 여러 판별장치는 음성이 짧은 기간에서 안정적인 수 없기 때문에 문제가 되는 것을 방지하기 위함이다. 판별장치를 통해서 얻어진 분류된 음성에 분절 연산을 통해 음성을 구로 분절 시켰다. 분절 연산은 “감정의 강도는 주어진 구내에서는 불변이고 구가 침묵 또는 분류기가 구의 감정 분류를 정할 수 없을 때는 중성이다”라는 가정 하에서 이루어진 것이다. 분절 연산을 수행하기위해서 본 논문의 시스템에서는 다른 특징 사이의 상호 관계를 고려하지 않고 빨리 다룰 수 있는 SVM[5][11]을 사용했다.

```

procedure Recognize_continuous_speech
( // 연속 음성 감정을 인식하기 위한 순서//
1. // 판별장치 설치//
Classifying proper ranges depending on each
person's speech
2. // 연속 음성 구로 분절//
Divide the sentence(or phrase) from
continuous speech because a shor period
will provide a stable position
3. // 감정 인지 비율 연산 적용//
Calculate and represent the percentage of
each sentence(or phrase)
)
    
```

이같이 인지된 음성은 음성 데이터베이스를 통해서 원하는 자료로 검색이 되고 이를 바탕으로 해서 얼굴, 입, 몸동작에 대한 데이터베이스로 부터 데이터를 추출을 하여서 원하는 결과를 얻을 수가 있게 된다. 이때 사용되는 검색방법은 이진 검색을 이용하여서 처리를 하였다. 그리고 이진 검색을 통해서 검색되는 시간을 줄여주기 위해서 각각의 데이터를 [표 3][표 4][표 5][표 6]에 제시된 형태로 미리 본 논문의 자체의 블록 코드화를 만들었다. 이같이 수행되는 알고리즘은 다음과 같다.

```

procedure Synchronize_with_emotion_and_speech( )
(// 감정과 음성, 입, 얼굴 몸동작을 합성한 부분//
1. //코드화 시스템//
Block_Coded_System()
2. //각 데이터 베이스에서 원하는 결과 검출//
Searching exact data from each DataBase
(voice, face, body, lip etc.) using Binary
Search
3.// 입, 얼굴, 몸동작과 각 감정별로 일치//
(1) Lip_Sync_with_emotion()
(2) Face_and_Lip_Sync_with_emotion()
(3)Body_and_Face_and_Lip_Sync_with_
emotion ( )
)
    
```

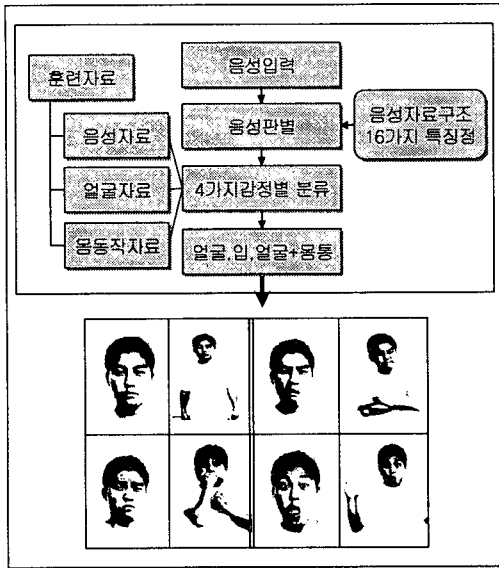


그림 5. HEEAS 시스템 구성도

## VI. 실험 결과 및 평가

만화 얼굴과 몸동작 애니메이션을 작동시키기 위해 Pentium 4 cpu 2.80GHz 컴퓨터와 여러 차원의 판별장치를 설치하여 사용하였다. 이를 위한 본 논문의 시스템 구성은 [그림 5]와 같이 제시하고 있다.

얼굴과 몸동작 만화 애니메이션을 실행하기 위해 음성을 입력하면 입력된 음성은 판별장치인 SVM을 통해 구로 분절되어 4가지 감정을 식별한다. 식별된 감정은 음성 데이터베이스와 연결된다. 여기서 감정에 해당하는 음성자료 코드 값을 취한다. 즉, 음성자료 코드의 첫 번째 자리를 "B"로 변환하여 몸동작 데이터베이스에서 해당하는 자료를 찾고, 다시 "F"로 변환하여 얼굴 데이터베이스에서 해당하는 자료를 찾는 방법이다.

이때 취해진 코드 값을 가지고 이진 검색 기법으로 입, 얼굴과 몸동작 데이터베이스에서 입력된 음성에 대한 감정이 표현된 얼굴과 몸동작 그림을 찾는다. 찾어진 감정이 나타난 몸동작 그림을 Morphing 기법[6]으

로 화면에 애니메이션 한다. [그림 6]은 몸동작그림을, [그림 7]은 얼굴그림을 화면에 애니메이션 한 것이다.



그림 6. 몸동작(두려움100%)



그림 7. 얼굴(두려움100%)

실제 실험을 위해, 얼굴과 몸동작 애니메이션을 작동시키기 위해 영화 "터미네이터"를 선택했다. 터미네이터는 두려움, 싫증, 놀람의 장면이 비교적 많으며 감정의 구분이 선명하여 식별하기가 용이하다. 영화 터미네이터에서 나온 감정 전체를 표현한 것이 [표 8]과 같고 여기서 추출한 각 감정별 즉, 중성([표 9]참조), 놀람([표 10]참조), 싫증([표 11]참조) 두려움([표 12]참조)의 실험결과를 볼 수가 있다.

본 논문에서 목적인 것은 화자의 음성을 듣고 말로 표현하지 않고 비음성적표현으로 입술과 얼굴, 몸동작에 표현하는 애니메이션 알고리즘 구현이었으며, 효과적인 애니메이션 알고리즘으로 평가할 수 있었다.

실험의 정확도를 확인하기 위해 훈련자료를 만들 때 함께 해주신 애니메이션 전문가 5분에게 실험 결과를 평가해달라고 부탁했다. 애니메이션 전문가들이 평가한 실험 결과를 보면 평온한 중성과 놀람의 음성을 듣고 얼굴과 몸동작으로 표현되는 것은 100%로 나타났으며, 두려움과 싫증은 99.8%로 나타났다.

본 논문에서 구현하고자한 만화 얼굴 몸동작 애니메이션은 평균 99.9%의 정확도([표 8]과 [표 13] 참조)를 보였다. 여기서 0.2%씩 오차를 보인 두려움과 싫증의 원인을 분석해 보았더니 음성인식부분과 그 느낌을 표현하는 과정에서 구별하기 힘든 요소가 작용한 것으로 보인다.

표 7. 기존 방법과의 비교

| 구분                | 기본개념                                   | 문제점   | 개선점  |
|-------------------|--|---|--|
| Hong Chen의 방법 [9] | -얼굴 스케치만을 생성                           | -다양한 얼굴 특징점 추출 부족<br>-입력된 윤곽선만을 변형해서 생성<br>-웹상에서 사용불가능                |  |
| Li 외 방법 [10]      | -감성과 더불어서 음성을 지닌 얼굴 애니메이션              | -감성 분류를 단지 4개로 제한(노여움, 행복, 슬픔 과 중성)<br>-감정 인식 정확도 63.5%, 연속 음성인지도 90% | -웹상에서 사용가능<br>-감정과 음성을 얼굴에 표현                    |
| 민 응식의 방법 [11]     | -감성과 더불어서 음성을 지닌 얼굴 애니메이션              | -단순히 얼굴만이 아닌 몸동작도 같이 구현을 해야만 함  | -감정의 범위를 3개 더 확대<br>-감정 인식 정확도 70%, 연속 음성인지도 95% |
| 본 논문 제시 방법        | -감성과 더불어서 음성을 지닌 얼굴, 몸동작, 얼굴+몸동작 애니메이션 | -감성의 분류를 4개로 제한(중성, 두려움, 싫증, 놀람)을 두었다                                 | -모든 감성의 인지가 가능할 수 있는 애니메이션                       |

[표 7]에서 볼 수 있듯이 다른 논문들이 주로 얼굴만을 다루고 있는 반면에 본 논문은 음성을 통해서 받아들인 구나 문장을 분석하여서 이 단어의 감정을 얼굴, 입과 몸동작에 이르기까지 표현을 했다는데 큰 의미를 부여하고 있다.

또한 이때 각각의 감정별로 검색되어지는 시간을 빠르게 하기 위해서 각 감정별 각 부위별로 블록 코드화 함과 동시에 검색기법 가운데 빠른 이진 검색을 통해서 이루어졌다.

표 8. 실험결과





| 감성의 종류 | 실험 내용      | 실험 결과 | 감정 추출 에러율 | 얼굴 애니메이션  | 몸+얼굴 애니메이션 결과 |
|--------|------------|-------|-----------|---|---------------|
| 중성     | 영화 터미네이터에서 | 100%  | 0%        |  |               |
| 두려움    | 영화 터미네이터에서 | 99.9% | 0%        |  |               |
| 싫증     | 영화 터미네이터에서 | 99.8% | 0.2%      |  |               |
| 놀람     | 영화 터미네이터에서 | 100%  | 0%        |  |               |

표 9. 중성

| 입력된 음성 | 이리와 놀자   |  |   |   |
|--------|--|--|---|---|
| 음성 파형  |  |  |   |   |
| 촬영한 얼굴 |   |  |  |  |
| 훈련 자료  |   |  |  |  |
|        | 입술   | 얼굴   | 몸동작   | 입술+얼굴+몸동작   |

표 10. 놀람

| 입력된 음성 | 비행기다 엎드려!!  |  |   |   |
|--------|---|--|---|---|
| 음성 파형  |   |  |   |   |
| 촬영한 얼굴 |   |   |   |   |
| 훈련 자료  |  |  |  |  |
|        | 입술  | 얼굴   | 몸동작   | 입술+얼굴+몸동작   |

표 11. 싫증

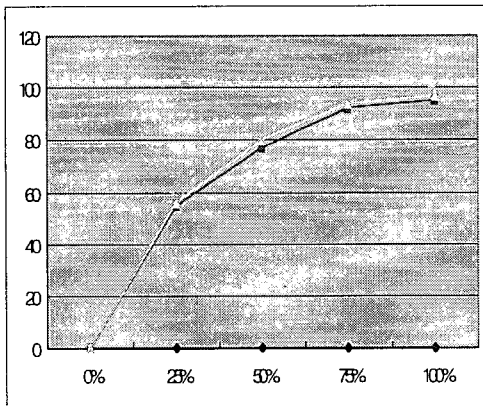
| 입력된 음성 | 여기서 기다려  |  |   |   |
|--------|--|--|---|---|
| 음성 파형  |  |  |   |   |
| 촬영한 얼굴 |   |  |  |  |
| 훈련 자료  |   |  |  |  |
|        | 입술   | 얼굴   | 몸동작   | 입술+얼굴+몸동작   |



표 12. 두려움

|        |           |    |     |           |
|--------|-----------|----|-----|-----------|
| 입력된 음성 | 바이러스야 조심해 |    |     |           |
| 음성 파형  |           |    |     |           |
| 활영한 얼굴 |           |    |     |           |
| 출력 자료  |           |    |     |           |
|        | 입술        | 얼굴 | 몸동작 | 입술+얼굴+몸동작 |

표 13. 감정별 인원별 분포도(정확도)



## VII. 결론

본 논문은 귀에 들려진 음성의 감정이 감정의 종류에 따라 다르게 얼굴과 몸동작으로 표현되는 얼굴 몸동작 애니메이션을 구현했다. 애니메이션을 하기위해 음성을 입력되면 SVM을 통해 구로 분절되어 감정을 4가지(중성, 두려움, 싫증, 놀람)로 구분하여 식별한다. 식별된 감정은 음성 데이터베이스에 연결된다. 음성 데이터베이스에서 감정에 해당하는 음성자료 코드 값을 취한다. 취해진 코드 값을 가지고 이진검색 기법으로 얼굴과 몸동작 데이터베이스에서 입력된 음성에 대한 감정

이 표현된 얼굴과 몸동작 그림을 찾는다. 찾아진 그림을 Morphing 기법으로 화면에 애니메이션 한다.

이러한 얼굴 몸동작 만화 애니메이션은 현재 음성을 듣고 그 느끼는 감정이 얼굴과 몸동작으로 표현이 자연스럽게 이루어 질수 있도록 합과 동시에 감정의 인식을 표현하고자 하는 경우에 매우 유용하게 사용될 것이다. 따라서 본 논문은 실험 결과가 99.9%로서 믿음직하고 우수한 성능을 가지고 있음을 실험결과를 통해 입증되었다.

향후 계획은 음성을 인식하고 분석하는 부분과 감정을 표현하는 부분에서 좀 더 세밀하게 인식되는 편한 얼굴 몸동작 부분과 인간의 전체적인 모습 애니메이션 구현을 위해 더 많은 연구를 계속할 것이다.

## 참고 문헌

- [1] J. Cassell, C. Pelachaud, N. I. Badler, M. Steedman, B. Achorn, T. Beckett, B. Douville, S. Prevost, and M. stone, "Animated conversation: rule-based generation of facial display, gesture and spoken intonation for multiple conversational agents," In Proc. ACM Siggraph94, pp.413-420, 1994.
- [2] E. Chang, J. L. Zhou, S. Di, C. Huang, and K. F. Lee. "Large vocabulary mandarin speech recognition with different approaches in modeling tones," In Proc. ICSLP, pp.983-986, 2000.
- [3] F. Dellaert, T. Polzin, and A. Waibel, "Recognizing emotion in speech," In Proc. ICSLP, pp.855-860, 1996.
- [4] ISO/IEC JTC1/SC29/WG11/MPEG97 N1820, "SNHC Verification Model 5.0," July, 1997.
- [5] Y. Li, F. Yu, Y. Q. Xu, E. Chang, and H. Y. Shum, "Speech-Driven Cartoon Animation with Emotions," IEEE Computer Animation, pp.365-371, June, 2002.

[6] Y. Li and H. Y. Shum, "Animating cartoon face from video," In 6th International Conference on Control, Automation, Robotics and vision, pp.840-855, 2000.

[7] R. Parent, *Computer Animation: Algorithms and Techniques*, Morgan Kaufmann Publishers, 2002.

[8] W. H. Sumby and I. Pollack, "Visual contribution to speech intelligibility in noise," *Journal of the Acoustical Society of America* Vol.26, pp.212-215, 1954.

[9] Q. Summerfield, A. MacLeod, M. McGrath, and M. Brooke, "Lips, teeth, and the benefits of lipreading," *Handbook of Research on Face Processing*. Elsevier Science Publishers, pp.223-233, 1989.

[10] <http://www.wdfamily.or.kr>

[11] 민용식, 김상길, "FAES: 감성 표현 기법을 이용한 얼굴 애니메이션 구현", *한국콘텐츠학회논문지*, 제5권, 제2호, pp.147-155, 2005.

김 상 길(Sang-Kil Kim)

정회원



- 1990년 2월 : 경희대학교 대학원 졸업(이학석사)
- 2006년 2월 : 호서대학교 대학원 컴퓨터공학과 졸업(공학박사)
- 1998년 3월~현재 : 한영신학대학교 교수

<관심분야> : 컴퓨터그래픽스, 컴퓨터애니메이션

### 저 자 소개

민 용 식(Yong-Sik Min)

정회원



- 1981년 2월 : 광운대학교 전자계산학과 졸업(이학사)
- 1983년 2월 : 광운대학교 대학원 전자계산학과 졸업(이학석사)
- 1991년 2월 : 광운대학교 대학원 전자계산학과 졸업(이학박사)
- 1987년 3월~현재 : 호서대학교 뉴미디어학과 교수

<관심분야> : 컴퓨터알고리즘, 컴퓨터애니메이션, 컴퓨터그래픽스