

인터넷 폰을 이용한 감성인식 시스템 구현

권병헌*, 서범석**

요약

본 논문에서는 감성 인식과 감성을 표현하는 캐릭터 애니메이션에 대한 내용을 소개한다. 본 논문에서 우리는 사용자의 감성을 표현하는 특성 파라미터를 찾는 방법과 패턴 매칭 알고리즘을 통해 감성을 추론하는 방법을 제안했다. 또한, 인터넷 폰에서 화자의 감성을 인식하여 캐릭터 애니메이션을 디스플레이 하는 플랫폼을 구현하여 감성 인식 시스템의 성능을 실험하였다.

Implementation of Emotion Recognition System using Internet Phone

Byeong-Heon, Kwon*, Burm-Suk Seo**

Abstract

In this paper, we introduces contents about the emotion recognition and character display expressing emotion. In this paper, we proposes on things like how to search characteristic parameters expressing user emotion, how to deduce emotion through pattern matching algorithm. Also, we implemented display platform recognizing the caller's emotion over internet phone. It display character animations expressing emotion as the result deduced by emotion recognition algorithm.

keyword : Internet Phone, Emotion Recognition, Contents

1. 서론

본 논문에서는 Internet Phone을 통해서 사용자의 음성을 인식하여 감성을 추론하고 그 결과를 이용해 감성을 표현하는 캐릭터를 디스플레이 하는 플랫폼 및 감성인식 모듈에 대한 내용을 소개한다.

본 논문에서는 감정 상태에 따라 분류된 한국어 음성 DB(Database)를 사용하여 사용자의 감정을 잘 표현하는 특징 파라미터를 찾고 패턴 비교 알고리즘을 통해 감성을 추론하는 방법을 제안한다.

또한, 추론된 결과에 따라 감정을 표현하는 다양한 캐릭터를 디스플레이 하여 Internet Phone

사용자가 상대방의 감정 상태를 인지하도록 하는 디스플레이 플랫폼을 구현하였다.

제안된 방법은 컴퓨터와 같은 기계가 인간에게 보다 친숙하게 다가갈 수 있도록 할 뿐만 아니라 음성 인식과 합성 기술이 발전함에 따라 인간의 음성으로부터 감정 상태를 파악하여 보다 고차원적이고 다양한 서비스를 제공함으로써 사용자의 만족을 극대화 시킬 수 있다.

2. 감성 인식 시스템

2.1 감성 표현 및 지각에 대한 한국어 문미역양의 영향

한국어의 특성 중 하나는 문법적으로 가장 중요한 요소가 다른 요소의 오른쪽에 오는 왼쪽 가지치기이다. 이러한 특성으로 문장의 마지막에 위치한 경계역양은 전체 문장의 의미와 결속될 가능성을 갖는다. 따라서 한국어에서 문미역양은 역양과 이의 의미 사이의 관계를 이해하는데 중요한 역할을 맡는다.

* 제일저자(First Author) : 권병헌

접수일자:2006년12월15일, 심사완료:2007년01월24일

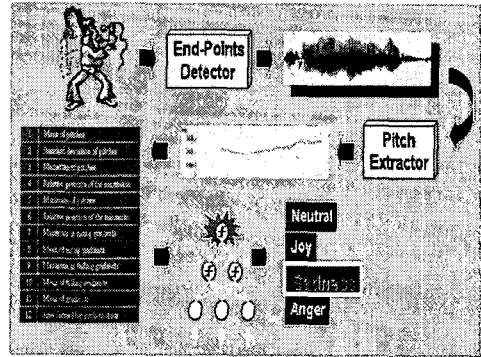
* 유한대학 정보통신학과

bhkwon@yuhan.ac.kr

** 유한대학 정보통신학과

지난 몇 년간 일단의 한국어 연구자가 화자의 감정과 태도에 문미역양이 미치는 영향을 조사하였다. 몇 가지 한국어 문미역양의 의미를 서술하였는데, 그 중에서 감정이나 태도와 연관된 문미역양의 패턴을 <표 1>과 같이 정리하였다[1].

여기서, 기호 L은 낮은 음조, H는 높은 음조를 나타내고 %는 문장의 끝을 의미한다. 위에서 살펴본 억양음운론 관련 연구결과에서 경계역양은 특정화자가 발성한 문장의 감정을 파악하는데 매우 중요한 역할을 한다는 것을 알 수 있다. 특히 한국어에서 억양에 의한 영향은 문미역양에 집중된다. 그러므로 한국어의 문미역양에 중요성을 부여한다면 보다 정확하게 화자의 감정을 인지할 수 있을 것이다.



(그림 1) 감정인식 시스템 처리 흐름도

<표 1> 문미역양 패턴

패턴	감정/태도
LHL%	설득, 주장, 확신, 귀찮음, 자극
LH%	귀찮음, 불쾌, 불신
HLH%	자신에 차 있어 청자의 동의를 구함
LHLH%	귀찮음, 자극, 불신
HLHL%	확신, 주장, 꺼림, 설득
LHLHL%	귀찮음 보다는 강한 감정

2.2 감성 인식 시스템 구현

본 논문에서 구현된 감성 인식 시스템은 감성 파라미터로 피치관련 특징을, 패턴 인식 방법으로 다층신경망을 사용한다. 이 시스템은 발화검출(End-Points Detection), 발화에서 피치곡선 추출(Pitch Extractor), 감정 특징 계산(12개 통계치), 감정 학습 및 인식(다층신경망)의 네 단계를 거쳐 평상, 기쁨, 슬픔, 화남의 네 가지 감정을 결정한다.

감정을 인식하기 위한 특징이 발화에 대해 계산되어야 하기 때문에 제일 먼저 입력된 소리 신호에서 문장을 검출해야 한다. 이를 위해 Rabiner와 Sambur의 방법[2]을 바탕으로 한 끝점 검출 알고리즘을 구현 시스템에 도입하였다.

피치곡선은 이전 단계에서 검출된 문장에서 추출된다. 구현 시스템의 피치추출 알고리즘은 David와 Niederiohn이 제안한 알고리즘[3]을 일부 수정하여 도입하였다.

<표 2> 구현시스템에 사용된 피치파생 특징

특징번호	설명
1	평균피치
2	피치 표준편차
3	최대피치
4	전체문장 내에서 최대피치의 상적위치
5	최소피치
6	전체문장 내에서 최소피치의 상대적 위치
7	최대상승기울기
8	평균상승기울기
9	최대하강기울기
10	평균하강기울기
11	평균기울기
12	밀변이 최소피치 위치인 피치곡선으로 형성된 면적

다층신경망은 <표 2>의 12가지 특징을 입력 패턴으로 하여 화자의 전형적인 감정을 학습하고 새로 입력되는 패턴을 학습된 감정으로 분류한다. 다층신경망은 하나의 입력계층, 0개 이상의 은닉계층, 하나의 출력계층을 갖는다. 입력계층은 패턴을 받아들이며, 은닉계층은 신경망 동작을 위한 학습능력을 결정하고, 출력계층은 대상감정의 인식점수를 산출한다. 각 계층은 하나 이상의 계산노드로 구성되며 한 계층 내의 모든 노드는 인접계층의 노드와 완전하게 연결된다. 다층신경망의 학습에 보편적으로 사용되는 오류역전과 알고리즘은 최대 기울기 감소 방법에 근간을 둔다[4]. 오류역전과 알고리즘은 다층신경망에게 기대된 동작을 실현하여 입력패턴을 대

응감정으로 분류한다. 이는 다층신경망의 현재 출력과 희망출력 간의 오차를 출력계층에서 입력 계층으로 역전파시켜서 이 오차가 최소가 되도록 다층신경망의 내부연결 가중치들을 조절하는 방법으로 실현된다.

2.3 감성 인식 시스템 모듈

1) 헤더 파일 및 함수 설명

```
#ifndef EMOTION_RECOGNIZER_INCLUDED
#define EMOTION_RECOGNIZER_INCLUDED

#include "PatternClip.h"

#define FALSE          0
#define TRUE           1
#define NO_OF_FEATURE 12
#define NO_OF_EMOTION 4

typedef struct ER ER;

typedef struct
/* 감정음성 데이터 학습을 위한 파라미터 */
{
    int nNoOfHidNod;
    /* MLP 은닉노드 개수 */
    float fLrnRate;
    /* EBP 학습속도 */
    float fObjErr;
    /* EBP 학습오류 목표치 */
} LrnPara;

/* 감정인식 기반 응용프로그램의 감정인식결과 처리를 위한 사용자정의 함수 */
typedef void (*ER_UserFunc)(void *pObj, float *fpEms);

ER *ER_Init();
/* ER 모듈 초기화 함수 */
void ER_Free(ER *opER);
/* ER 모듈 해제 함수 */
int ER_Save(ER *opER, char *szFile);
/* ER 모듈 파일저장 함수 */
ER *ER_Load(char *szFile);
/* ER 모듈 기록파일읽기 함수 */

/* 감정음성 데이터베이스 학습과 시험을 위한 함수 */
int ER_LearnEm(ER *opER, PC *opEm, LrnPara
```

```
oLrnPara, char *szOutFile);
int ER_TestEm(ER *opER, PC *opEm, char *szOutFile)

char **ER_GetEmStrs(ER *opER);
/* 현재 ER 모듈 내 감정문구 전달 함수 */
int ER_GetNumOfEm(ER *opER);
/* 현재 ER 모듈 내 감정개수 전달 함수 */

int ER_GetEm(ER *opER, char *szStr);
/* 사운드 소스에 대한 1회 감정인식 함수 */
void ER_GetEms(ER *opER, char*szStr, float *fpEmVals);
/* 사운드 소스에 대한 1회 감정인식 및 인식점수 전달 함수 */

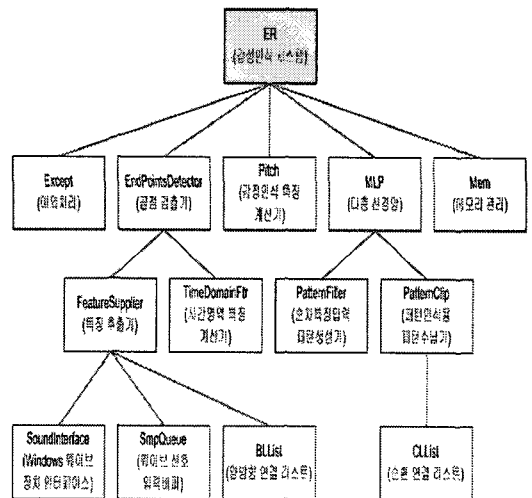
/* 마이크 입력에 대한 연속 감정인식 함수 */
int ER_GetContEm(ER *opER, ER_UserFunc pUserFunc, void *pObj);

void ER_StopContEm(ER *opER);
/* 연속 감정인식 중단 함수 */

#endif
```

2) 소프트웨어 모듈 구조

감성인식 시스템은 그림 2와 같은 C 모듈로 구성된다.



(그림 2) 소프트웨어 모듈 구조

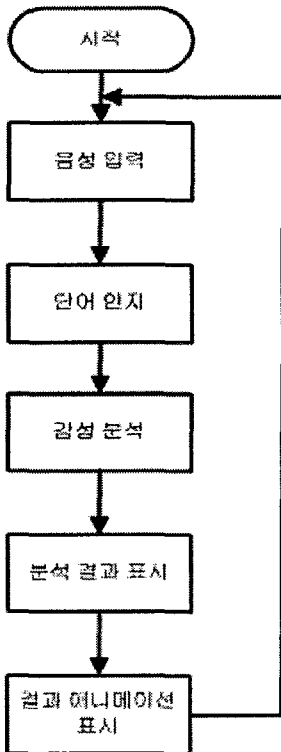
3. 감성 캐릭터 애니메이션 디스플레이

3.1 분석결과표시

m_Progress_N.SetPos() 함수를 이용하여 4개의 감정 표현 지수를 각각 그래프로 표시한다. 이 때 표시되는 값은 -1에서 +1 사이의 값을 가지며 그 해상도는 0.01 단위이다.

3.2 결과 애니메이션 표시

애니메이션 표시는 CVideo 클래스를 사용한다. 이미 동영상 파일이 표시되고 있는 상태이면 m_pVideo->Open() 함수를 사용하여 중지시키고 파일을 닫는다. 새로운 동영상 파일을 표시하기 위해 이전에 만들어진 클래스를 delete로 삭제한 후 다시 생성하고, m_pVideo->Open() 함수를 사용하여 동영상 파일을 연 후 m_pVideo->Play() 함수로 화면에 표시하게 된다.



(그림 3) 캐릭터 애니메이션 표시 처리흐름도

4. 실험

Internet Phone(네이버폰)을 이용하여 감성인식 모듈과 캐릭터 애니메이션 디스플레이 S/W 동작을 실험하였다. 사용자의 음성을 입력받아 추론된 감성인식 결과와 그 결과에 따라 디스플레이 되는 캐릭터 애니메이션이 일치하는지 평가하도록 한다.

4.1 Internet Phone 사용방법

네이버(www.naver.com)에 가입하여 네이버폰 설치파일을 다운로드한다. Direct X와 Window Media Player 버전을 확인하고 설치파일을 실행하여 네이버폰을 PC에 설치한다. 헤드셋(스피커+마이크)을 연결하여 통화 연결을 확인한다.

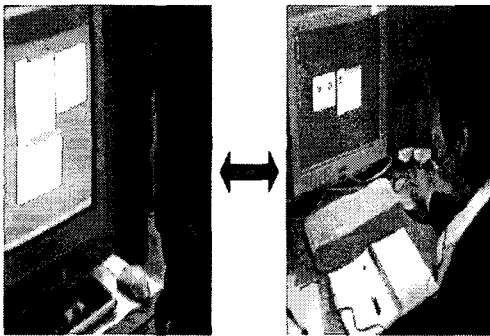


(그림 4) 네이버폰 사용자 인터페이스

4.2 실험 환경

PC 2대에 네이버폰을 설치하여 통화 연결을 확인한다. 네이버폰을 이용하여 음성 통화 시 상대방의 음성과 사용자의 음성을 구분하기 위해 각 PC에 사운드 카드를 2개씩 설치한다. 1번 사운드 카드는 네이버폰으로 음성 통화 시 사용자

의 헤드셋 마이크를 통해 음성을 입력받고 출력(상대방의 음성)은 2번 사운드 카드 입력으로 연결된다. 2번 사운드 카드에 입력되는 음성 신호는 감성인식 모듈이 수신하여 감정을 판별해 낸다. 2번 사운드 카드의 출력은 사용자의 헤드셋 스피커로 연결되어 음성 통화를 가능하게 한다. 인식된 감성 정보는 사용자 인터페이스를 통해 사용자에게 수치 정보와 함께 감성 캐릭터 애니메이션을 디스플레이 하게 된다.



(그림 5) 시스템 실험 환경

4.3 실험 결과

감성인식 모듈과 캐릭터 디스플레이 S/W로 구현된 감성 인식기(Emotion Recognizer)는 Animation Window, Button Input, Status Window, Graphic Window로 구성되어 있다.

1) Animation Window

Animation Window는 감성인식 결과에 대응하여 제작된 3D 캐릭터 애니메이션을 디스플레이 하는 창이다.

2) Button Input

감성 인식기 오른쪽 상단에 위치하며 인식기 동작 실행, 중지, 확인(종료) 명령을 입력 받는다.

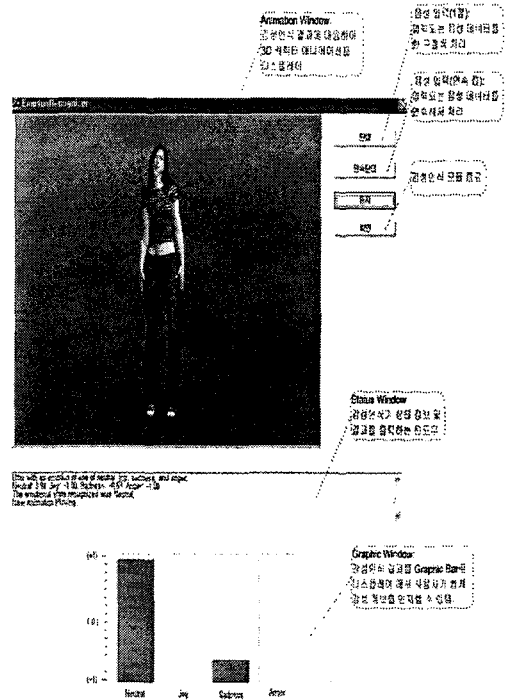
3) Status Window

Status Window에서는 감성 인식기 동작 상태를 확인할 수 있다. 입력된 음성 데이터를 분석하여 4가지 감정에 대한 인식 결과를 수치로 나타낸다.

4) Graphic Window

Graphic Window는 사용자가 감성인식 결과를 쉽게 인지할 수 있도록 수치 정보를 그래픽으로 보여준다. (그림 6)에서 중성(Neutral) 수치 0.94

와 슬픔(Sadness) 수치 -0.64가 Graphic Bar로 표현된 것을 확인할 수 있다. 즐거움(Joy)과 화남(Anger)은 수치가 -1이므로 Graphic Bar가 나타나지 않았다.



(그림 6) 감성 인식기

참고문헌

- [1]. Jun, S., K-ToBI Labelling Conventions, Ver. 3.1, <http://www.linguistics.ucla.edu/people/jun/ktobi/K-to-bi.html>, 2000.
- [2]. Rabiner, L. and Sambur, M. An Algorithm for Determining the Endpoints of Isolated Utterances, Bell System Technical Journal, Vol. 54, pp. 297~315, Feb 1975.
- [3]. Krubsack, D. A. and Niederjohn, R. J., An Autocorrelation Pitch Detector and Voicing Decision with Confidence Measures Developed for Noise-Corrupted Speech, IEEE Transactions on Signal Processing, Vol. 39, No. 2, pp. 319~329, Feb 1991.
- [4]. Bengio, Y., Neural Networks for Speech and Sequence Recognition, International Thomson Computer Press, 1995.



권병현

1987년 : 한국항공대학교 항공전자
공학과 공학사
1989년 : 한국항공대학교 대학원
항공전자공학과 공학석사
1995년 : 한국항공대학교 대학원
전자공학과 공학 박사

1995년 : (주)LG전자 선임연구원

현재 : 유한대학 정보통신학과 부교수

관심분야 : 영상신호처리, 통신, 3D displays.



서범석

1997년 : 한국항공대학교 항공전자
공학과 공학사
2001년 : 한국항공대학교 대학원
항공전자공학과 공학석사
2001년 - 현재 : 한국항공대학교
항공전자공학과 박사과정

관심분야 : 전자공학, LCD화질개선, 영상디스플레이(3D)등