**S/W멤버십 과제 기획서**

**과제명 : STT, SVM을 이용한 청각장애인용 통화 SW개발**

* 소 속 : 수원 소프트웨어 멤버십
* 작성자 : 최환종, 안중환, 정다비치
* 작성일 : 2015. 05. 01

**S/W멤버십 과제 기획서**

|  |  |  |  |
| --- | --- | --- | --- |
| **과 제 명** | STT, SVM을 이용한 청각장애인용 통화 SW개발 | | |
| **과제구분** | 창의과제 | | |
| **과제기간** | 2015. 05. 01 ~ 07. 30 (3개월) | | |
| **지 역** | 수원 멤버십 | **참여인원** | 3명 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **회 원 명** | **학교** | **학과** | **학년** | **연락처** | **E-mail** |
| 안중환 | 경기대 | 컴퓨터과학과 | 4 | 010-8604-8894 | ajh8894@gmail.com |
| 최환종 | 명지대 | 컴퓨터공학과 | 4 | 010-2625-3577 | hwan7287@gmail.com |
| 정다비치 | 한국산업기술대 | 컴퓨터공학과 | 2 | 010-7440-9412 | ardabitchy02@naver.com |

|  |  |  |  |
| --- | --- | --- | --- |
| **구분** | **내용** | | |
| **개발 목적**  **및 동기** | 청각장애인 또는 듣기가 불편한 고령자들은 전화통화에 어려움을 겪는다. 이런 문제에 도움을 주기 위해 첫째로 상대방의 음성을 Text로 변환하여 전화중인 청각장애인에게 보여준다. 두 번째로 듣지 못하는 청각장애인은 텍스트만 보고는 상대방이 어떤 감정으로 말을 하였는지를 알기 어렵다. 우리는 상대방의 음성을 이용해 감정을 분석하여 청각장애인에게 감정까지 보여준다. 즉, 음성의 시각화를 목표로 한다. | | |
| **개발 환경**  **및 일정** | OS : Window 7  Tool : Eclipse, Matlab, Android SDK  Language : Java, MySQL  개발 일정 : 2015.05 ~ 2015.07 (3개월) | | |
| **창의성/**  **우수성** | 기존에는 SK 커뮤니케이션즈 / 한국정보문화진흥원에서 제공하는 메신저 음화상 서비스가 있다. 그러나, 이 프로젝트는 중간에 통신중계를 해주는 중계원이 꼭 있어야 하는 시스템 이다. 우리는 이런 중계원이 필요없는 시스템 일 뿐 만 아니라 중계원이 해주지 못하는 감정표현까지 간접적으로 해줄 수 있다. | | |
| **활용성/**  **사업성** | 기존의 청각장애인들이 이용하던 서비스의 대체 가능성을 제시한다.  구글에서 제공하는 STT는 학습에 따라 그 성능이 점점 향상되고 있다. 그러므로 STT 변환 기능은 점점 더 정확해 질 것이고 기존의 한정적이고, 비싼 영상 통화 요금제, 번거로운 중계 시스템을 이용하던 청각장애인들에게 큰 도움이 될 것이다. 더 나아가면 모바일 뿐만 아니라 상대방의 모든 음성에 대한 시각화가 가능해질 것 이다. | | |
| **지원부서** | S/W멤버십 | **기술지원 연구원** | 이재홍 |
| **전화번호** | 010-4333-7839 | **E-Mail** | ljh890322@gmail.com |

**목 차**

**- 내용 목차 -**

[1. 개발 목적 4](#_Toc399930566)

[1) 배경 4](#_Toc399930567)

[2) 목적 4](#_Toc399930567)

[3) 유사 프로젝트 5](#_Toc399930567)

[2. 개발 목표 5](#_Toc399930569)

[1) 음성 TEXT 변환 5](#_Toc399930567)

[2) 음성을 통한 감정 분석 6](#_Toc399930567)

3) 실시간 처리가 가능한 통신 6

[4) 시각화 6](#_Toc399930567)

[3. 개발 내용 6](#_Toc399930570)

[1) 구성도 6](#_Toc399930567)

[2) 설계 7](#_Toc399930567)

(1) 안드로이드 [7](#_Toc399930567)

(2) STT [7](#_Toc399930567)

(3) 전처리기 [8](#_Toc399930567)

(4) SVM서버 [8](#_Toc399930567)

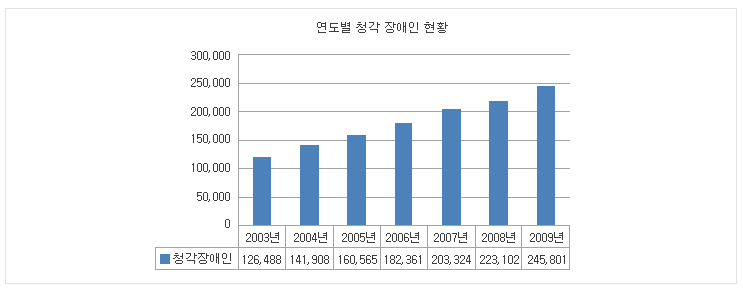
[3) 주요 이슈 9](#_Toc399930567)

[4. 개발 일정 9](#_Toc399930571)

[5. 참고 문헌 9](#_Toc399930573)

# 1. 개발 목적

## 1) 배경



국내 청각장애인은 해마다 증가하고 있는 추세이며, 2015년 현재 국내 청각장애인 현황은 약 28만명이다. 하지만, 각 이동통신사에서 제공하는 스마트 폰 요금제는 청각 장애인의 이동통신 소비 특성을 고려하지 않는 것으로 조사됐다. 한국소비자원 에서 스마트 폰 을 이용하는 청각장애인 중 장애인 편의에 맞춘 서비스를 이용하는 수는 8.3%에 불과했다.

|  |  |
| --- | --- |
|  | C:\Users\swssm\Desktop\KakaoTalk_20150421_171550275.jpg |

그 이유로 한국소비자원에서는“수화 전달을 위해서는 매우 선명한 영상이 제공되어야 하는데 현재 휴대폰 영상통화 품질은 수화를 전달하기에 충분하지 않으며, 상대와 휴대폰 기종이 다르면 영상통화가 실행되지 않는 등의 문제점이 있다” 를 들었다. 또한 현재 제공되는 청각장애인 서비스는 요구에 전혀 미치지 못했다고 했다.

## 2) 목적

청각장애인이 일반전화처럼 통화할 수 있는 SW를 개발한다.

첫 번째로 상대방의 음성을 Text로 변환하여 전화중인 청각장애인에게 보여준다.

두 번째로는 듣지 못하는 청각장애인이 텍스트와 UI만 보고도 상대방이 어떤 감정으로 말을 하였는지 파악할 수 있도록 상대방의 음성을 분석하여 감정까지 보여준다.

궁극적인 목적은 음성을 Text로 변환하고, 분석하여 효과적으로 시각화 하는데 있다.

## 3) 유사 프로젝트



통신중계서비스센터

기존의 통신중계서비스센터는 웹을 통해 제공되며, 수화가 가능한 제3자가 중계원으로 존재해야만 한다. 우리는 이런 중계원 역할을 서버가 대신해주고, 그 이상의 역할을 하게 하려고 한다.

   
청각장애인을 위한 통화 앱 – SupportEar

SupportEar은 통화 중에 채팅만 가능하게 해준 어플이다. 우리는 채팅은 청각장애인에게 말하는 수단으로만 제공하고, 상대방의 음성을 text로 실시간 변환한다.

## 2. 개발 목표

## 1) 음성 TEXT 변환

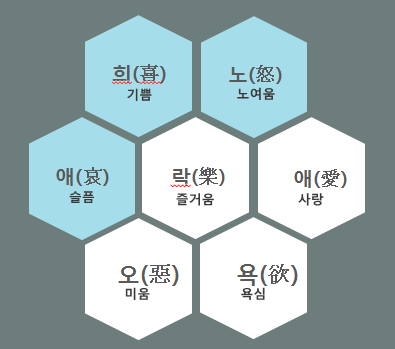
1) Google STT 엔진을 사용한다.

[1차 음성 분석 / 2차 방대한 학습데이터] 가 있기 때문에 인식률이 매우 좋다.

2) GCM 피드백 방식 사용

3G / LTE / WIFI 로 IP가 불안정하기 때문에 안정적인 통신인 GCM 피드백 방식을 사용한다.

## 2) 음성을 통한 감정분석

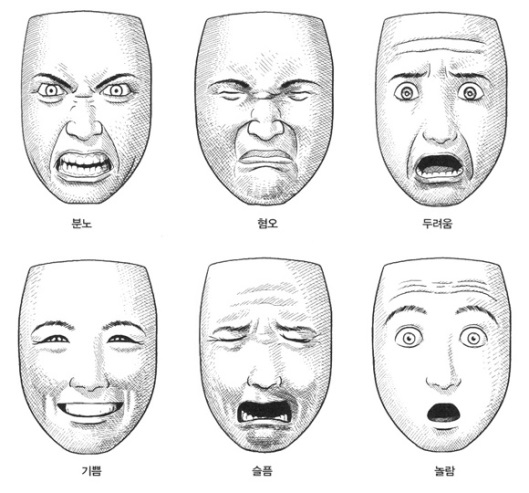


1) 기쁨(Happiness), 슬픔(Sadness), 화남(Anger), 보통(Neutrality)의 4가지 감정으로 분류한다.

[희 / 노 / 애 / 락 / 애 / 오 / 욕] 의 7가지 감정 중, 희 / 노 / 애의 감정과, 보통 감정의 4가지 Label로 분류한다.

2) 특징추출로는 에너지의 통계적 분포 (PLP: Perceptual linear prediction)를 사용하고, SVM으로 분류를 해주는 방법을 사용한다.

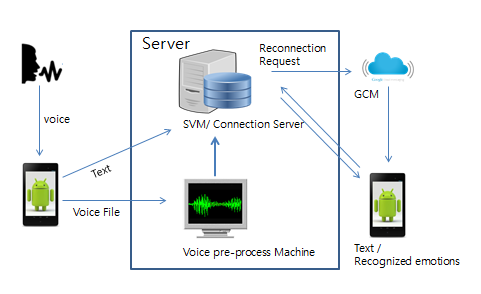
## 3) 시각화



텍스트로 보여주고, 감정으로 분류해줌과 동시에 애니메이션으로 표현 가능한 UI를 설계한다.

# 3. 개발 내용

## 1) 구성도



발신자의 음성을 안드로이드 내에서 STT를 이용하여 TEXT로 변환하고, 전송은 서버와 소켓 통신을 통해 이루어 진다. 하지만 모바일은 IP가 유동적으로 바뀌기 때문에 GCM으로 피드백을 받는 형식으로 통신 상태를 유지한다.

그리고 텍스트 변환, 전송과 동시에 음성 데이터를 서버의 MATLAB 전 처리기로 전송한다. 이 전치리 과정을 거친 음성이 SVM을 통해 감정으로 분류되며 이 결과와 TEXT 변환 결과를 결합하여 발신자에게 전송할 최종 데이터로 만들어진다. 마지막으로 수신자의 안드로이드폰에서 최종 결과를 효과적으로 UI에 표현해준다.

## 2) 설계

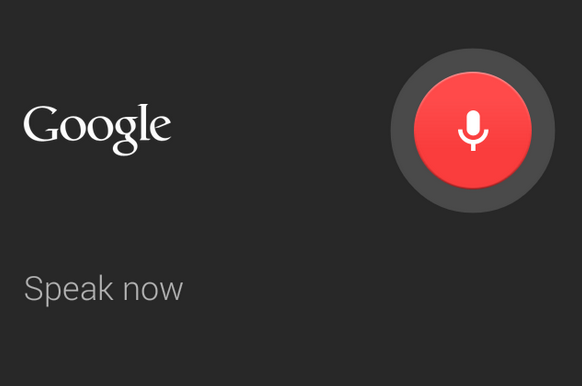
### (1) 안드로이드



안드로이드는 통화앱이 아닌 background에서 작동하는 방식으로 구현한다. 수신과 발신을 Broadcast를 통해 감지하고 각각 사용자에게 알맞은 형태의 보조 UI로 제공한다.

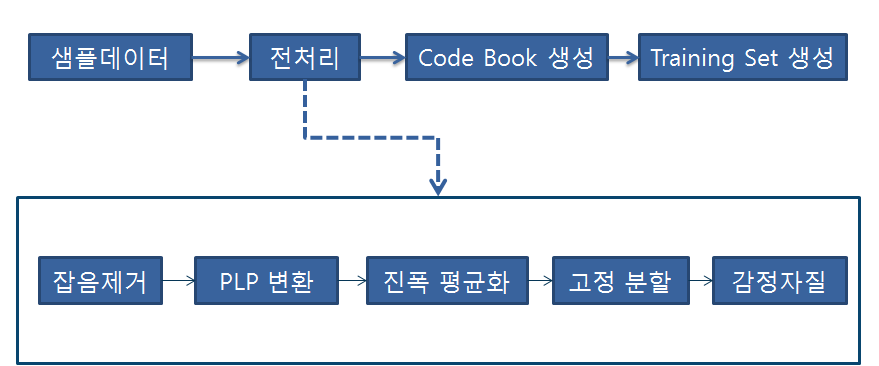
음성에 대한 처리는 녹음 기능을 통해 구현하기 위해 MediaRecorder의 AudioSource를 활용하려 한다. 하지만 음성 데이터의 전송은 실시간성을 떨어뜨릴 수가 있다. 그러므로 녹음된 음성 파일을 스트리밍화해서 보내는 방법과 RTP 프로토콜을 이용하여 실시간 음성 전송 방법 중 더 효과적이고 빠른 방법을 비교해서 사용할 것이다.

### (2) STT



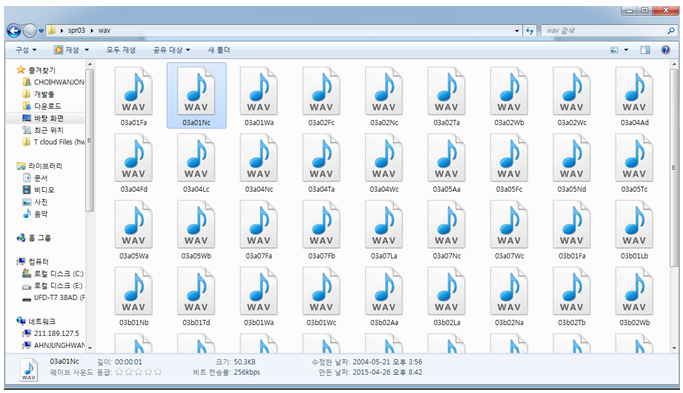
Google STT 엔진은 몇 가지 테스트를 통해 인식률이 뛰어난 것이 이미 확인 했다. 그래서 인식률 자체에는 크게 관여하지 않으며 이 STT를 실시간으로 음성을 받을 수 있도록 바꾸고, 변환된 text list 중 가장 적절한 것을 선택한다.

### (3) 전처리기



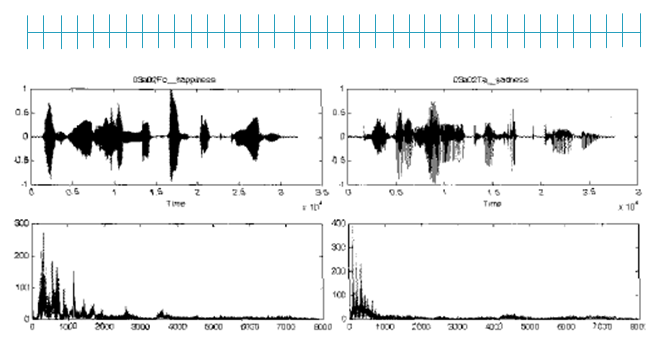
전처리기에서는 SVM 분류를 위한 음성에 대한 가공을 목표로 한다. 전체적은 흐름은 위와 같고, 순서대로 살펴보면,

1. 지도학습으로 EMO-DB의 Sample 음성 500개 학습 데이터를 이용하여 모델 선정



EMO에서 감정에 대한 샘플DB를 500개 정도 제공한다. 우리는 이 DB를 기준으로 모델을 선정하고, 부족하다면, 추가적으로 DB를 확보할 예정이다.

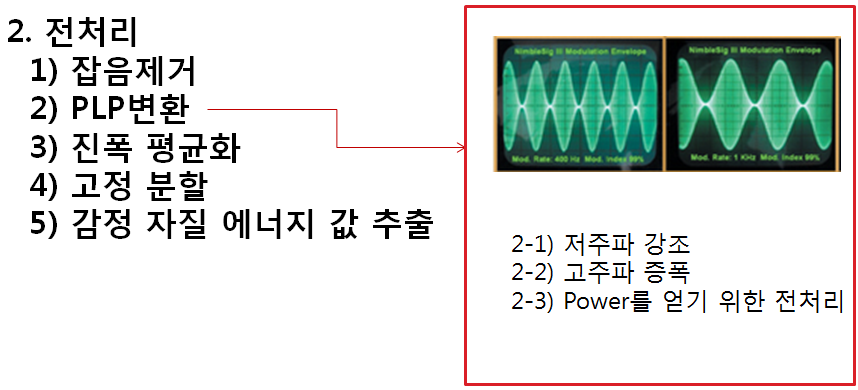
1. Matlab을 이용한 전처리



음성인식을 위하여 주로 사용되는 특징은 LPC cepstrum, PLP cepstrum, Mel frequency cepstral coefficient (MFCC), 필터뱅크 에너지 등이 있는데 우리는 PLP 특징 추출 방법을 사용할 것이다.

PLP는 에너지의 통계적 분포를 이용하여 특징을 추출해내는 방법이다.

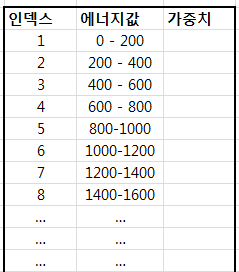
다음과 같은 순서로 진행 된다.



잡음제거→ 진폭 평균화→ 고정분할→ PLP분석→ 구간별 에너지 값 추출

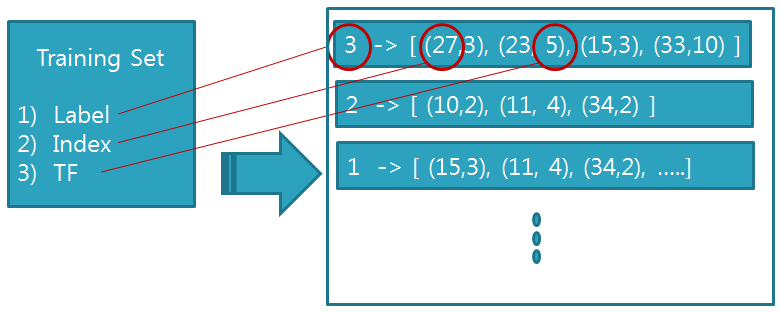
잡음 제거나 평균화를 하면 음성의 손실이 올 수도 있는데 여기서 가장 최적인 점을 찾아 내야 한다.

1. 에너지량 Code Book 생성 (Bag of word기법)



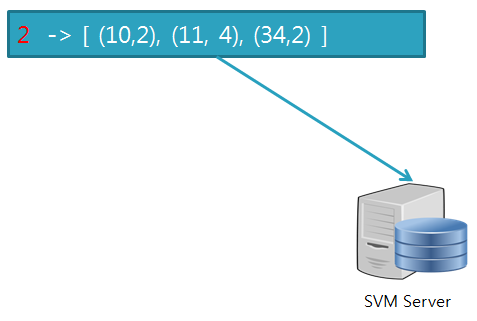
Bag of Words 방법은 먼저 음성에서 feature들을 뽑은 후, 이들 feature들을 대표할 수 있는 값(code)들로 구성되는 코드북(codebook)을 생성한다. 보통 코드북은 다수의 음성들로부터 추출한 feature들 전체에 대해 분류를 통해 획득한 대표 feature들로 구성된다.

1. Code Book을 이용하여 Training Set 생성 (Label 지도)



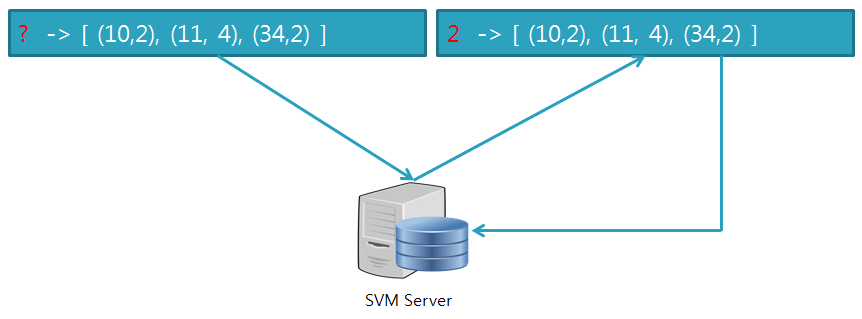
Training Set → [ ① Label { ②감정자질 후보군 ③감정자질 TF } ]

1. 완성된 Set을 기계학습의 Training Set으로 사용



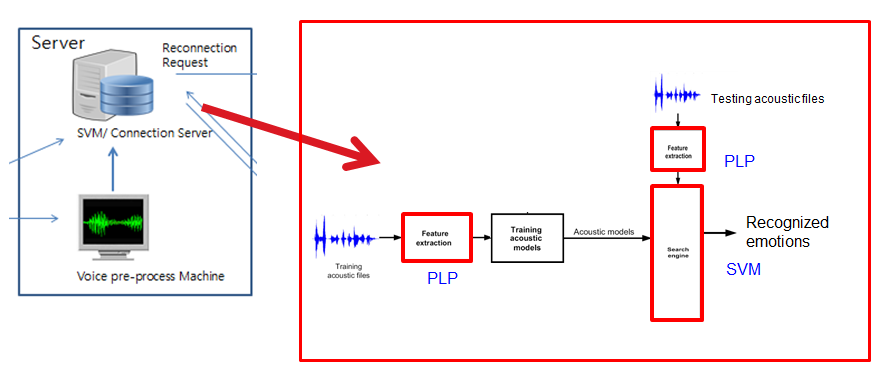
감정 분석을 위한 음성을 전 처리과정을 거쳐 Set을 만든 후 SVM에 Label요청을 한다.

1. 감정 결과 추출



추가적인 학습을 위해 Label이 포함된 Training Set을 다시 학습시킨다.

### (4) SVM 서버



서버에서 안드로이드로부터 TEXT를 받고, 전처리기를 통해 받은 음성데이터를 SVM을 통해 나온 감정을 최종 결과로 만들어낸다.

또 안드로이드 기기와의 연결을 유지하고, 받은 데이터가 손실되지 않도록 버퍼 형태로 유지 관리하게 개발해야 한다.

SVM 라이브러리는 다음 링크에서 가장 알맞은 것을 선택하여 사용할 것이다.

<https://docs.google.com/spreadsheet/ccc?key=0Aunb9cCVAP6NdDVBMzY1TjdPcmx4ei1EeUZNNGtKUHc&hl=en#gid=0>

## 3) 주요 이슈

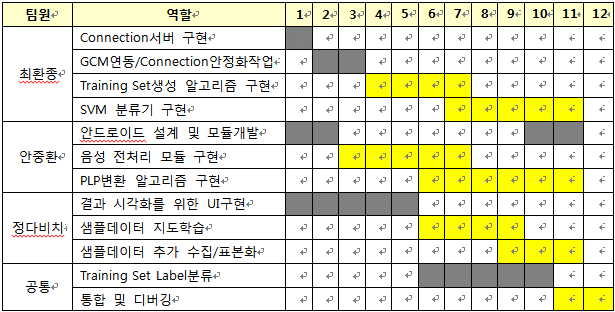
1. 저주파에서의 남녀 감정차이 분별 (보통, 슬픔)

PLP 특징 추출 방법은 통계적 에너지 분포에 의존하기 때문에 남녀의 차이가 확실하다. 하지만 만약 문제가 될 경우 남녀를 구분 해야 하는 API를 사용하거나 SVM을 남녀로 분리하여 설계해야 할 수도 있다.

2. 감정분석결과를 전송하는 실시간성

실제로 구현해보지 않는 한 실시간성은 장담할 수가 없다. 하지만 음성에 대한 TEXT 전송은 어느 정도 실시간성을 보장할 수 있으므로 음성 분석과 어느 정도 시간차를 두고 싱크를 맞춰준다면 해결할 수 있다고 본다.

# 4. 개발 일정



# 5. 참고 문헌

## 1) 청각장애인에게 기존 서비스가 무용지물인 이유 [뉴스]

[**http://beminor.com/news/view.html?section=1&category=3&no=6778**](http://beminor.com/news/view.html?section=1&category=3&no=6778)

## 2) Google STT API 관련 소스 [블로그]

[**http://blog.acronym.co.kr/317**](http://blog.acronym.co.kr/317)

## 3) 음성신호를 이용한 감정인식 [논문]

## 4) 감성기반 서비스를 위한 통화 음성 감정인식 기법 [논문]

## 5) 음성인식 기술을 이용한 일본드라마 감성 분석[영남대학교 연구 자료]