**S/W멤버십 과제 기획서**

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

* 소 속 : 수원 소프트웨어 멤버십
* 작성자 : 최환종 (25-1)

안중환 (25-1)

정다비치 (25-1)

* 작성일 : 2015. 09. 21

**S/W멤버십 과제 기획서**

|  |  |  |  |
| --- | --- | --- | --- |
| **과 제 명** | STT, SVM을 이용한 청각장애인용 통화 SW개발 | | |
| **과제구분** | 창의과제(공모전대상) | | |
| **과제기간** | 2015. 10. 01 ~ 11. 30 (2개월) | | |
| **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 : 2015.10 ~ 2015.11 (2개월) | | |
| **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) 보완목표 6

[5) 시각화 6](#_Toc399930567)

[3. 기존 개발 내용 6](#_Toc399930570)

[1) 구성도 6](#_Toc399930567)

[2) 설계 7](#_Toc399930567)

(1) 안드로이드 [7](#_Toc399930567)

(2) Connection 서버 [8](#_Toc399930567)

(3) SVM서버 [8](#_Toc399930567)

[3) 주요 이슈 9](#_Toc399930567)

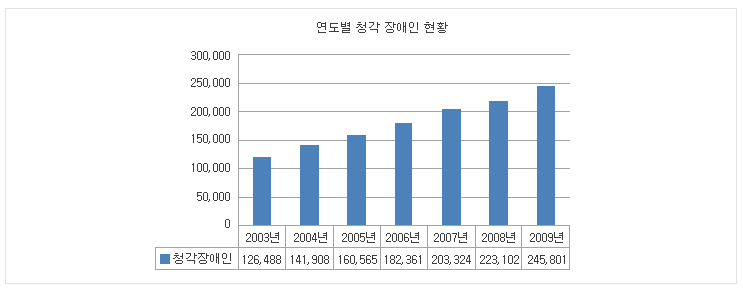
[4. 보완 내용 9](#_Toc399930571)

[5. 개발 일정 9](#_Toc399930571)

[6. 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.

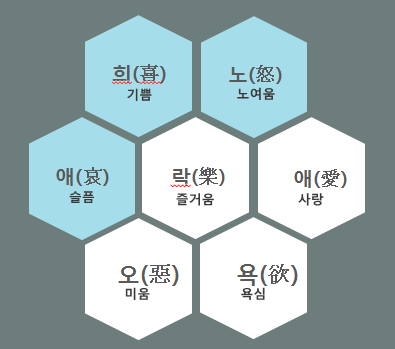
## 2. 개발 목표

## 1) 음성 TEXT 변환

1) Google STT 엔진을 사용한다.

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

## 2) 음성을 통한 감정분석

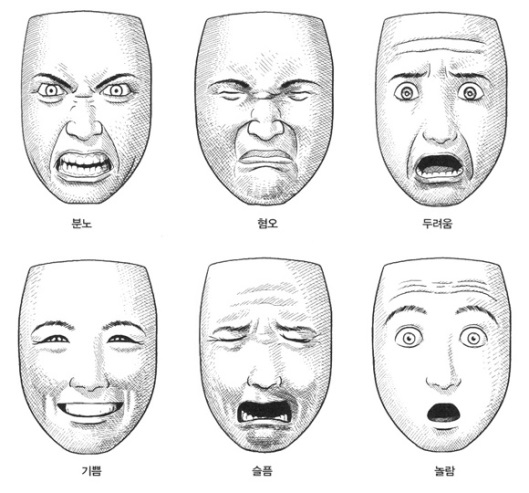


1) 기쁨(Happiness), 슬픔(Sadness), 화남(Anger), 보통(Neutrality)의 4가지 감정으로 분류한다.

[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.

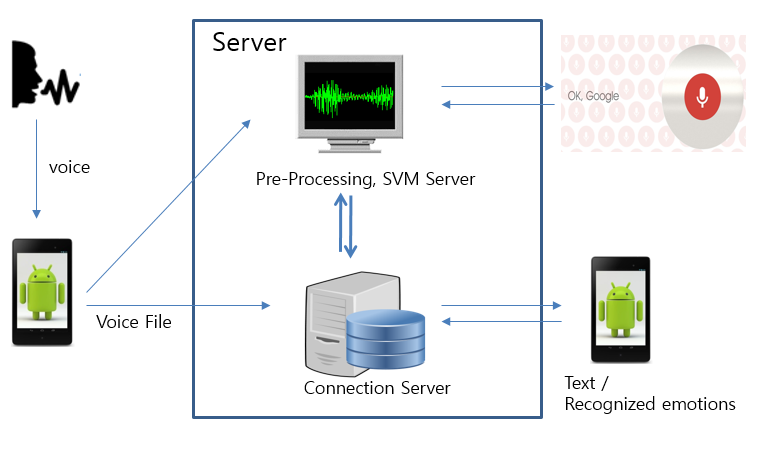
## 3) 시각화



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

# 3. 기존 개발 내용

## 1) 구성도



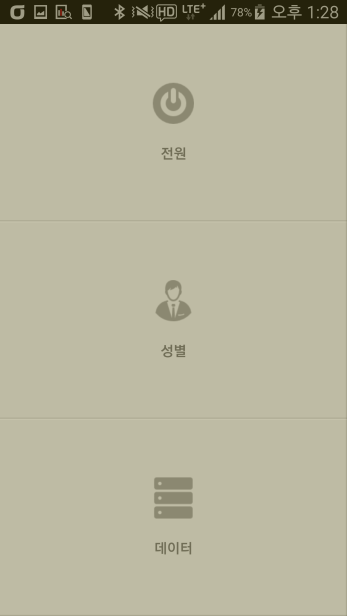
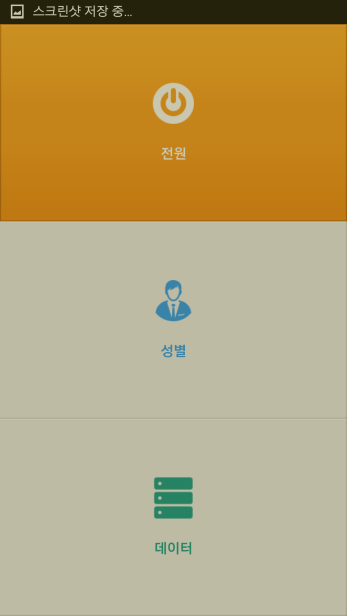
서버의 부하를 줄이기 위해 음성 전처리, SVM 서버와 모바일 클라이언트를 관리하는 서버를 따로 분산하여 구현하였다.

발신자의 음성은 전처리, SVM 서버로 전송하여 묵음 제거 및 잡음 제거 전처리 과정을 거친 뒤 Google stt(speech to text) 서버 전송과 동시에 음성 데이터를 PLP를 통해 피쳐를 추출하고 SVM(Support vector machine)을 통해 감정으로 분류한다. 그리고 이 결과 값(text, 감정)은 따로 따로 Connection 서버로 보내고 메시지 순서와 클라이언트를 구분하여 모바일로 전송한다.

모바일은 보통 Polling 방식을 많이 사용하지만 통화는 실시간을 요구하고, 3G / LTE / WIFI 로 IP가 항상 바뀌므로 통신 상태가 바뀌면 바로 새로운 Streaming을 교체해주는 방식으로 연결을 유지한다.

## 2) 설계

### (1) 안드로이드

1. 전원 / 성별 / 데이터 메뉴 설정창이다. 설정만 해줄 뿐 따로 앱을 통해서 통화를 하지 않아도 된다.

[전원] : ON을 하면 전화 통화 시 보조 UI가 나타난다.

[성별] : 자신의 성별을 설정해주는 기능이다. 성별을 나누는 것이 필요한 이유는 감정 분석에 사용되는 모델이 남녀로 구분 되어있기 때문이다.

[데이터] : 통화로그를 볼 수 있다.

1. 통화 앱이 아닌, Background에서 작동하는 방식으로 구현한다.

수신과 발신을 Background에서 감지하고 각각 사용자에게 알맞은 형태의 보조 UI로 제공한다.

1. 종합 감정 / 현재 감정 탭을 이용하여 실시간으로 감정을 시각적으로 표현 가능하다.

RaderChart, 애니메이션을 이용하여 종합적인 감정, 현재 감정을 실시간으로 표현 할 수 있으며, 채팅 창으로도 감정 확인이 가능하다.

1. 잡음 제거 녹음방식, signal을 줄여 처리속도 향상, 리 샘플링

### (2) Connection 서버

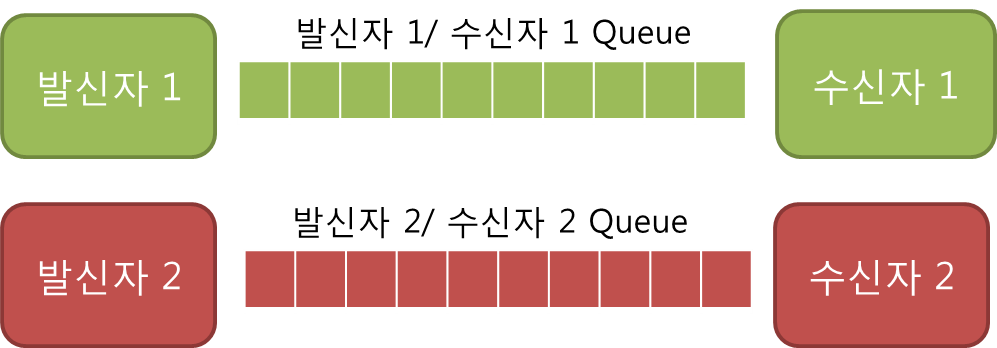
1. 각 사용자 별 설정에 따라 동작하도록 전화 발신, 수신 시 서버로 연결이 된다.
2. 발신자가 먼저 연결된 후 수신자가 전화를 받기를 기다린다. 전화를 끊거나 안받는 경우 접속을 종료한다.
3. 첫 접속 시 자신의 id, 상대방 id(id는 핸드폰번호), 보조 UI 동작 여부를 포함하여 패킷을 보내고, 모바일로 전송할 때는 어떤 클라이언트로 보내야 하는지에 대한 헤더 결과값을 포함하는 패킷을 보낸다.



* 첫 접속시



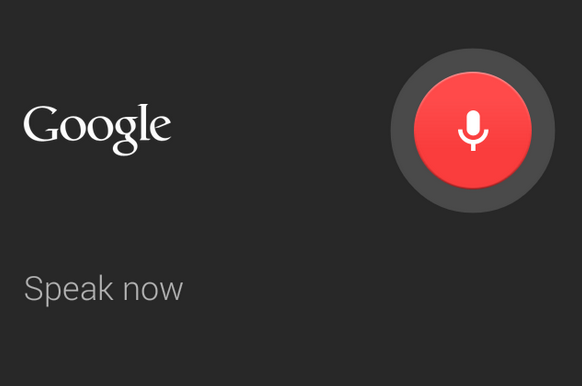
* 모바일로 전송시

1. 서버는 전화 연결당 하나의 큐를 가지고 있어서 해당 클라이언트에게 전송을 실패하면 패킷을 큐로 다시 복구하여 재 접속한 클라이언트에게도 보낼 수 있도록 설계하였다.  
     
   

### (3) SVM 서버

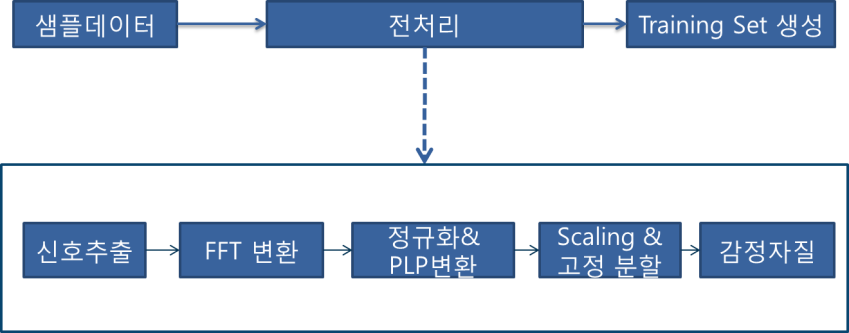
- svm 서버는 사용자로 부터 음성을 받고, STT 서버로 요청과 동시에 감정분석을 하고 그 결과를 Connection server로 전송한다.

### STT



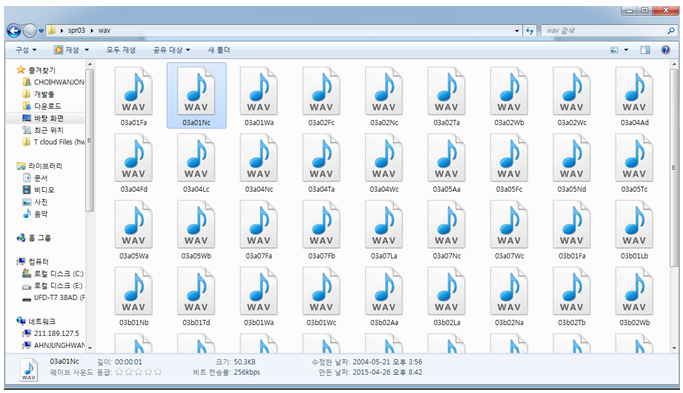
Google STT 엔진은 몇 가지 테스트를 통해 인식률이 뛰어난 것이 이미 확인 했다. 그래서 인식률 자체에는 크게 관여하지 않으며 이 STT를 실시간으로 음성을 받을 수 있도록 바꾸고, 변환된 text list 중 가장 적절한 것을 선택한다.

### SVM을 이용한 감정분류



전체적은 흐름은 위와 같고, 순서대로 살펴보면,

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



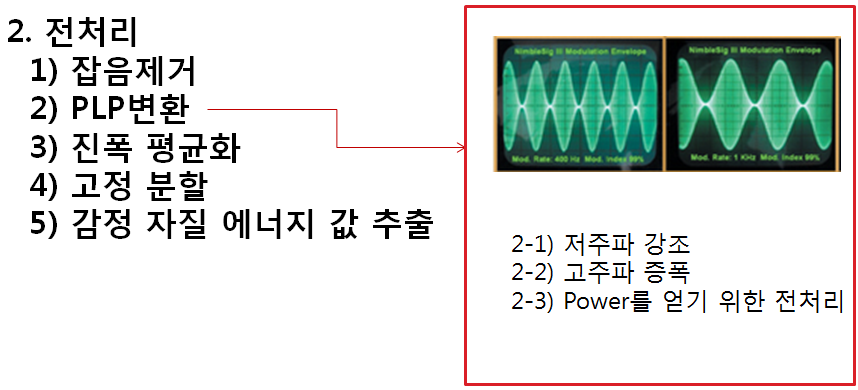
EMO에서 감정에 대한 샘플DB를 500개 정도 제공한다. 우리는 이 DB를 기준으로 모델을 선정하고, 부족하다면, 추가적으로 DB를 확보할 예정이다.

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.

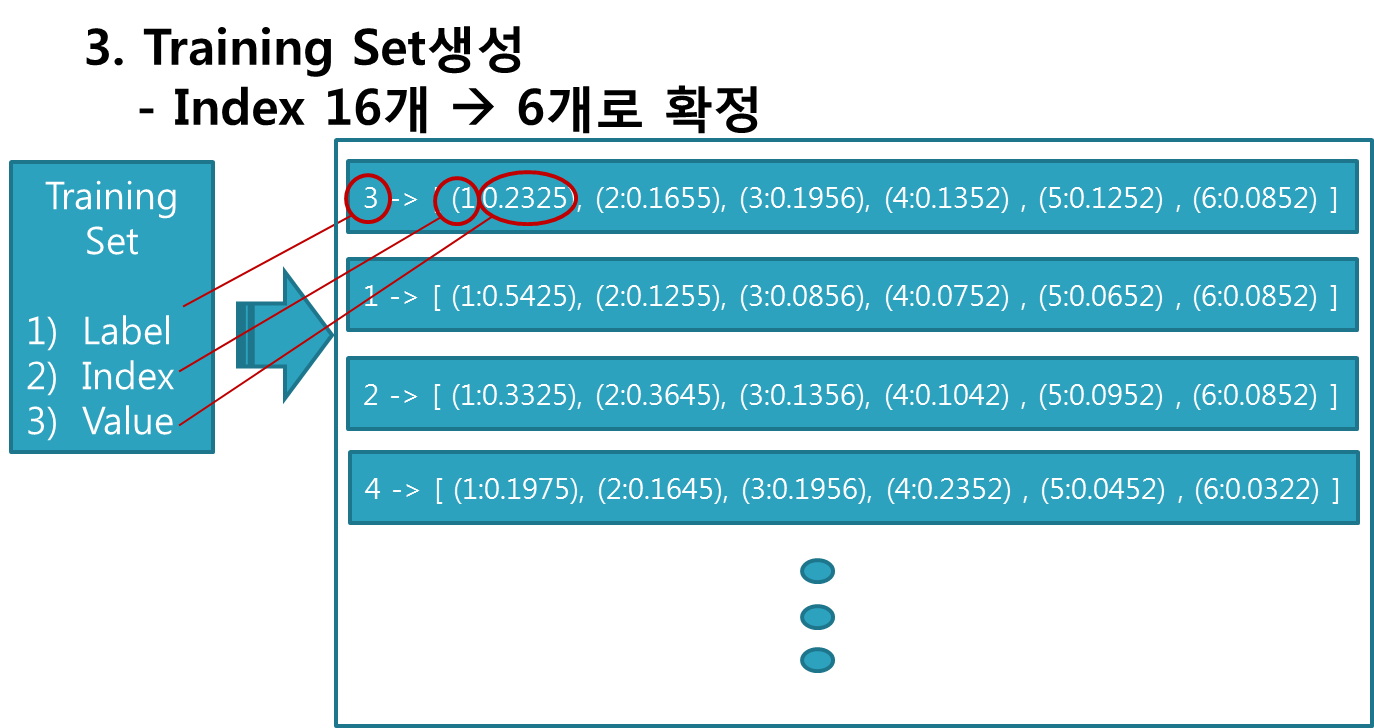
1. 음성 전처리



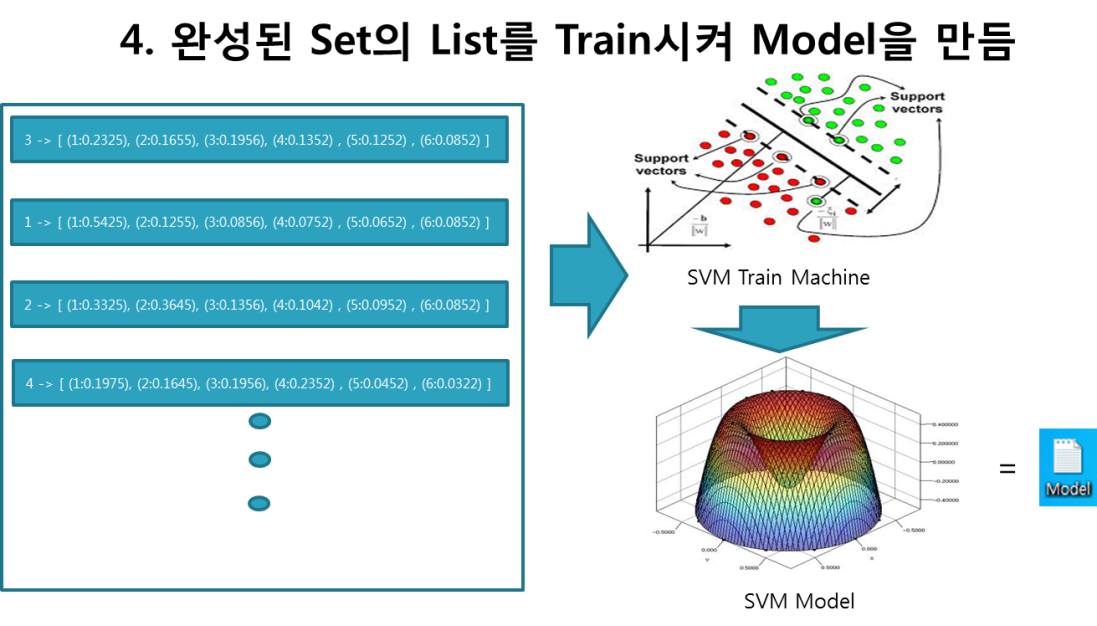
잡음제거→ 진폭 평균화→ 고정분할→ PLP분석→ 구간별 에너지 값 추출

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

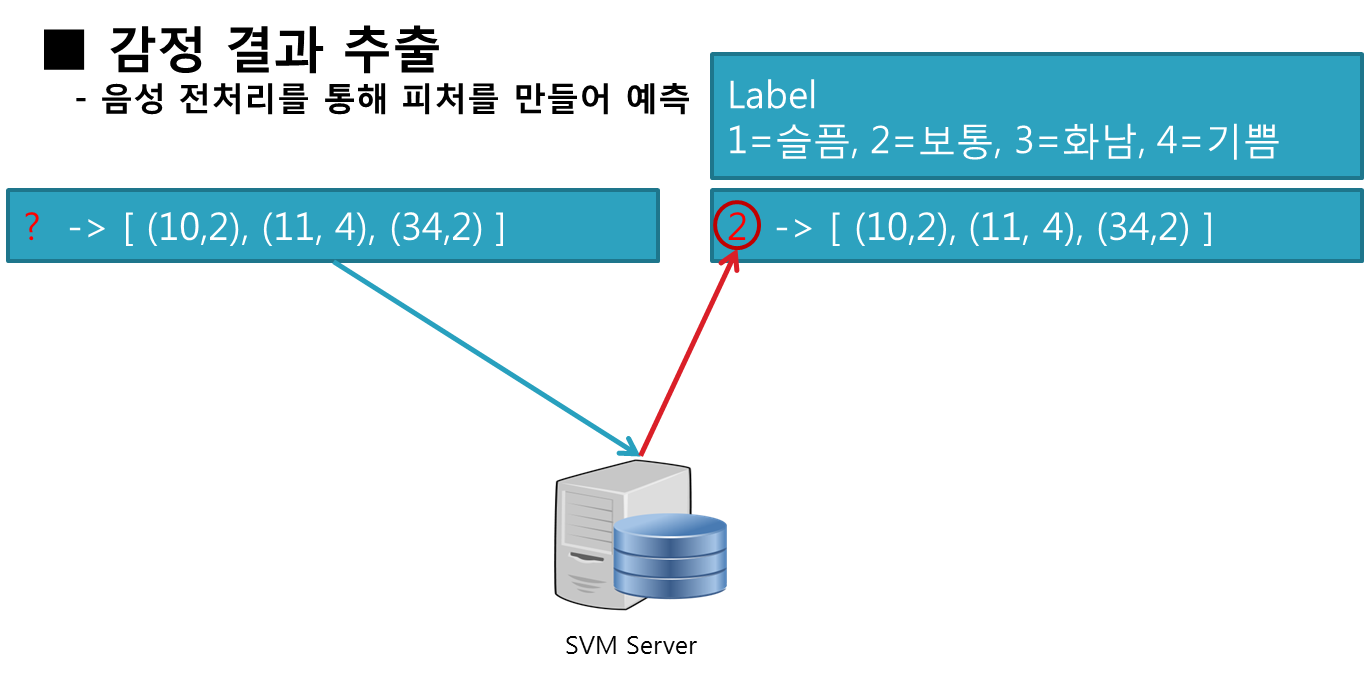
1. Training Set 생성 (Label 지도)  
   Training Set → [ Label { 감정자질 후보군, value} ]



1. 완성된 Set을 기계학습의 Training Set으로 사용  
   감정 분석을 위한 음성을 전 처리과정을 거쳐 Set을 만든 후 SVM에 Label요청을 한다.

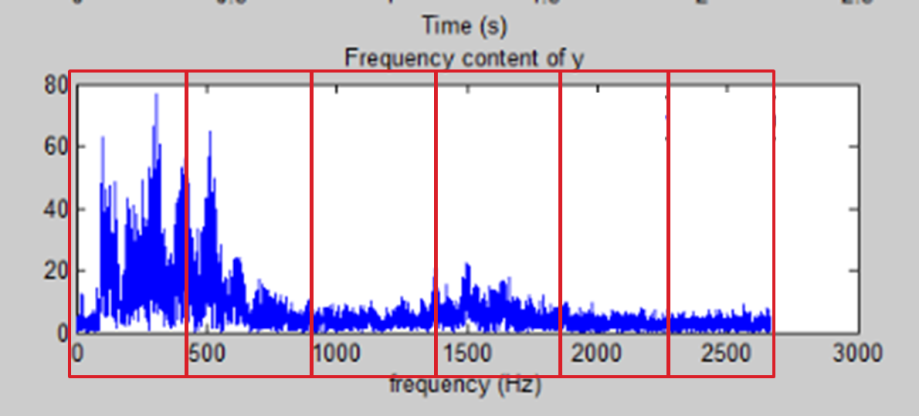


1. 감정 결과 추출

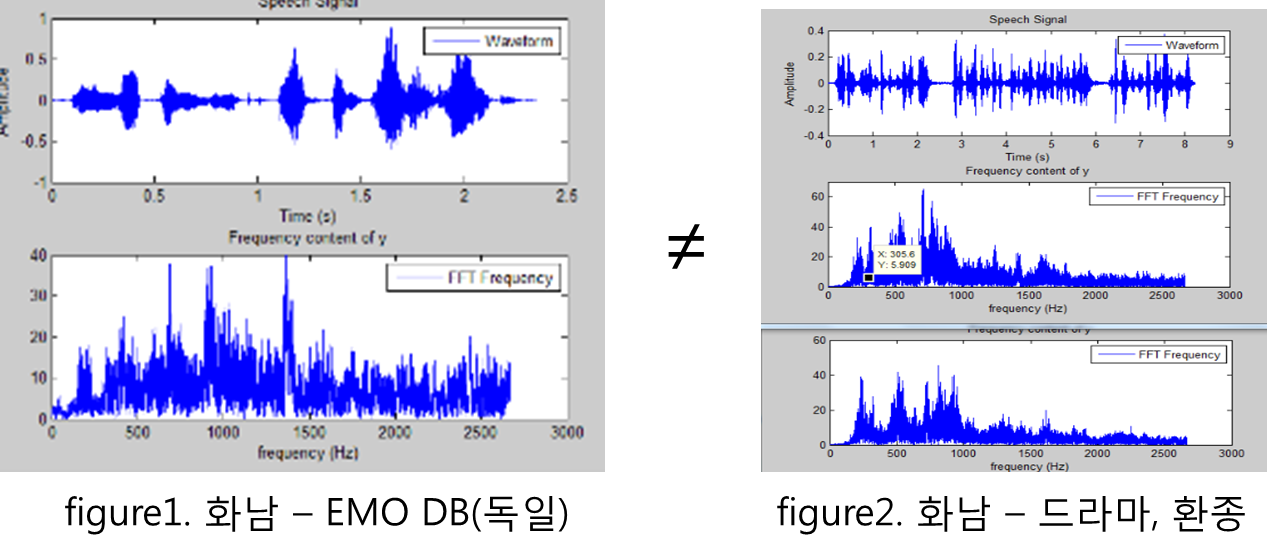


### 결과예측을 위한 실제 데이터

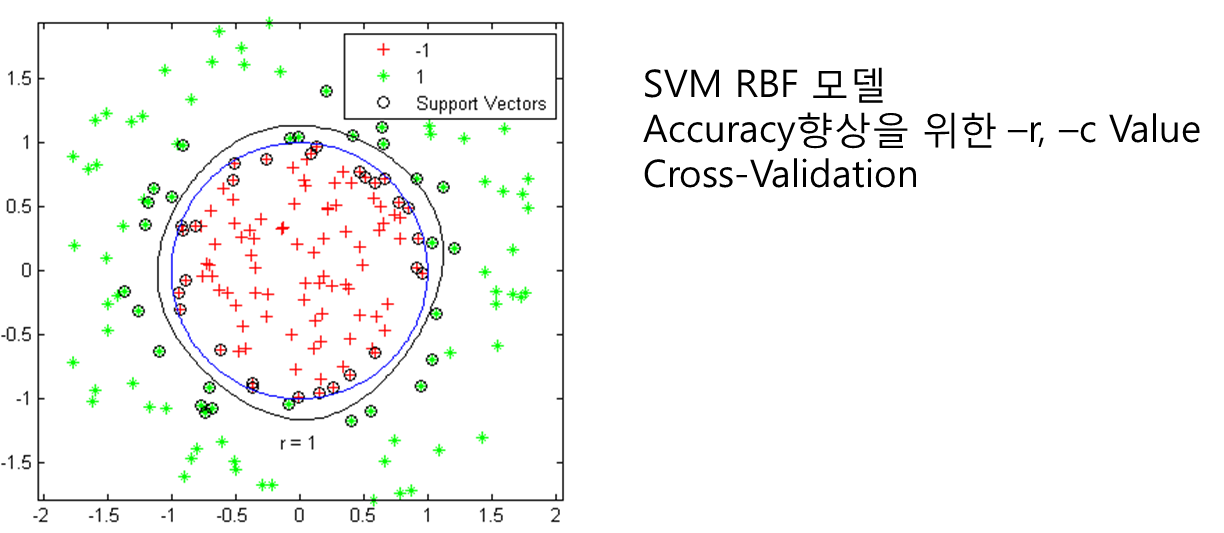
### Feature 개요



1. 정확한 Support Vector Line 정밀화 작업 - 특징 정확도



1. 정확한 Support Vector Line 정밀화 작업- SVM 라인 정확도



## 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. 보완목표

1) 감정정확도와 처리속도를 향상시킨다.

2) 화남, 기쁨의 유사패턴에 대한 구별알고리즘을 구현한다.

3) HTTP요청의 특성상 STT가 비실시간 처리속도를 보이므로 Native STT를 연동하여 음절단위실시간 처리가 되도록 보완한다.

4) 안드로이드 녹음방식을 개선하여 안정화되도록 구현한다.

# 5. 개발 일정

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **팀원** | **역할** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** |
| **최환종** | 감정정확도 및 속도 향상작업 |  |  |  |  |  |  |  |  |
| 화남/기쁨 감정분류 추가 알고리즘 |  |  |  |  |  |  |  |  |
| **안중환** | 안드로이드 녹음방식 안정화 |  |  |  |  |  |  |  |  |
| Connection 서버 최적화 |  |  |  |  |  |  |  |  |
| **정다비치** | STT Native 연결 및 속도향상 |  |  |  |  |  |  |  |  |
| UI 보완 |  |  |  |  |  |  |  |  |
| **공통** | 최종 시연 및 PPT준비 |  |  |  |  |  |  |  |  |
| 통합 및 디버깅 |  |  |  |  |  |  |  |  |
| 3차 훈격 심사 발표 |  |  |  |  |  |  |  |  |

# 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]