

# 한국어 음성 합성을 이용한 이메일 매니저

조규상<sup>o</sup> 이영훈 이병렬 서대영  
한국산업기술대학교 컴퓨터공학과  
{glamit032<sup>o</sup>, yhoon2da, Kmode, seody}@kpu.ac.kr

## Using of The Korean Language Voice Synthesis For E-Mail Manager System

Gyu-Sang Jo<sup>o</sup> Young-Hoon Lee Byeong-Ryeol Lee Dae-Young Seo  
Dept. of Computer Engineering, Korea Polytechnic University

### 요 약

IT 관련 산업의 발전에 의한 저변의 확대로 장애우들의 IT 사용 수요가 늘고 있다. 본 논문에서는 IT분야에서 가장 기초적으로 활용되는 E-Mail을 시각 장애우가 활용 하는 데에 불편함이 없도록 하는 이메일 매니저 개발에 관련된 기법에 대해 논하고자 한다. TTS(Text-To Speech : 문자 텍스트를 음성으로 전환하여 들려줌)와 음성키보드(키보드 입력 시 입력한 문자를 음성으로 알려줌) 기능으로 시각 장애우가 이메일을 사용함에 있어 불편함을 느끼지 않도록 하였으며 본 시스템의 TTS 알고리즘은 국어 표준발음법을 참고로 하여 자바로 구현 하였다.

주제어 : TTS, 시각 장애인, 음성 합성

## 1. 서 론

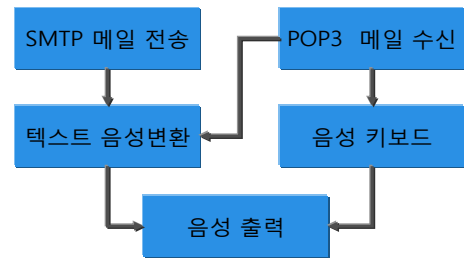
최근 IT 관련 산업의 발전이 급격하게 이루어지면서 이전에는 혜택을 보지 못했던 층에게도 저변이 확대되고 있다. 이러한 연유로 장애우들의 IT 활용 수요도 증가하게 되었다. 점차 나아지고 있는 추세이지만 현재 우리의 소프트웨어 시장은 이러한 장애우들의 IT 활용 수요 증가에 비해 개발된 소프트웨어의 수가 적고, 이미 개발된 소프트웨어의 경우에도 이를 알리기 위한 홍보가 부족하여 정작 장애우들이 혜택을 누리기가 어려운 실정이다. 일례로, 대표적인 포털사이트중 하나인 다음에서 제공하는 한메일 익스프레스 서비스 중 시각 장애우를 위한 기능을 들 수 있다. 베타버전임에도 불구하고 비교적 높은 완성도로 시각 장애우가 이용하기에 편리함을 지니고 있고, 실제 장애우의 이용 동영상과 체험 캠페인을 통한 홍보를 펼치는 등의 노력을 했지만, IT를 활용할 수 있는 계층을 타겟으로 한 홍보였기에, 정작 IT 소외 계층에 해당하는 시각 장애우 들은 직접적인 홍보를 받지 못하였고, 아직까진 많은 사람들이 알지 못하고 있는 실정이다. 이와 같이 완성도 높은 프로그램들이 존재하더라도 적절한 홍보의 부족으로 IT 소외계층에 해당하는 시각 장애우가 활용할 수 있는 종류의 소프트웨어를 제대로 접하지 못하고 있는 상황이다. 이에 착안하여 IT 서비스 중 많은 이들이 사용하며 기본이 된다고 볼 수 있는 이메일 서비스를 장애우가 활용함에 있어 불편함이 없도록 하는 이메일 매니저 개발에 착수하게 되었다.

장애우를 위한 이메일 매니저를 개발함에 있어 특히 염두에 둔 점은, 다른 장애와 달리 시각장애의 경우가 이메일 사용이 가장 많은 제한을 주기에 TTS(Text-To

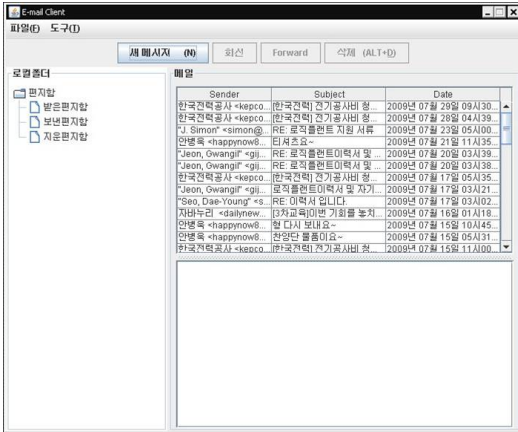
Speech)와 음성키보드 기능의 지원으로 이메일 작성 제한요소를 최소화하여 일반인이 사용하는 환경에 최대한 근접시키는 것이다. 단축키를 통하여 이메일 매니저 사용 시 편의를 증대시켰으며, 메일 작성 시 도움 마법사 기능을 부가하여 메일 작성의 어려움을 줄이고자 하였다. 또한, 일반적으로 많이 사용되는 웹 기반의 이메일 서비스 제공 사이트나 이메일 관리 프로그램으로 대표적인 Outlook Express 등의 환경과 유사한 인터페이스로 접근성의 장벽을 낮추고 편의성을 높이는 것을 목표로 하였다.

위에서 언급한 바와 같이 본 시스템은 시각장애우의 이메일 매니저 사용에 초점을 맞추고 개발하였으며, 그 까닭은 우리가 사용하는 보편적인 이메일 서비스 환경에 비추어 볼 때 시각장애 및 읽기장애를 지닌 사람들이 사용가능한 환경을 갖추게 된다면 대다수의 사람들이 이용함에 있어 어려움이 없으리라 여겨졌기 때문이다.

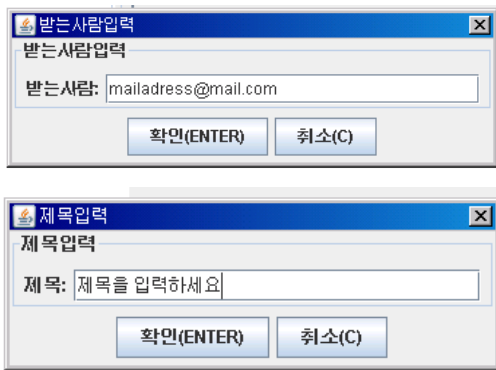
본 시스템의 프로세스와 메인 화면은 아래 그림과 같다.



<그림 1>



<그림 2>



<그림 3>

메일 작성 도움 마법사의 경우, 메일을 받을 사람의 메일 주소를 입력하고 메일의 제목을 입력 한 후 메일 내용을 입력하는 순서로 진행된다. 모든 과정에 음성 서비스를 제공하여 시각 장애우의 메일 작성 시 편의를 최대한 고려하였다.

## 2. TTS (Text-To Speech)

TTS는 도움말 파일이나 웹페이지와 같은 컴퓨터 문서의 내용을 사람이 읽어주는 소리로 만들어주는 음성합성 시스템의 한 종류이다. TTS는 시각 장애가 있는 사람들을 위해 컴퓨터 화면에 나타난 정보를 대신 읽어주는 것도 가능하다.

알려진 TTS의 합성방법으로 Concatenative(연결) 합성, Unit Selection(단위 선택) 합성, Diphone(소리 강도의 인접 쌍) 합성, Domain-Specific(도메인 특정) 합성, Formant(모음의 음정 등 주파수 세기 분포) 합성, Articulatory(조음) 합성, HMM-based(흙-기반) 합성, Sinewave(사인파) 합성 등이 존재한다. 이중 본 시스템

에서는 음절 단위로 분리된 데이터베이스를 합성하여 음성을 출력 해주는 Unit Selection을 기반으로 구성되었다. 일부 매뉴얼이나 도움말 음성의 경우에만 미리 작성된 문장을 출력하는 Domain-Specific 방식을 적용하였다.

본 시스템은 사용자로부터 텍스트를 입력 받아 분석하여 발음 소리로 변환하여 음성을 내보내는 방식이다. 음소 변환은 국어사전의 표준어 규정(문교부 고시 제88-2호 : 1988. 1. 19.) 제2부 표준 발음법을 참고하였다.

텍스트 분석은 모두 유니코드로 이루어지며 유니코드로 인해 텍스트 분석이 끝나면 바로 텍스트 분리에 들어간다.

분리는 초성, 중성, 종성 3가지의 성으로 분리하며 이것들을 가지고 표준 발음법에 의거 바로 전 중성과 현 초성을 가지고 발음을 결정하여 기존의 성을 음소로 변환하게 된다. 여기서 합성어 처리나 한자어 처리는 애매하므로 데이터베이스화 하여 처리하거나 파일처리를 이용하여 구현하여야 한다.

<분리 공식>

$$\begin{aligned}
 c &= ch - 0xAC00; \\
 a &= c / (21 * 28); \\
 c &= c \% (21 * 28); \\
 b &= c / 28; \\
 c &= c \% 28;
 \end{aligned}$$

초성[a], 중성[b], 종성[c].

이렇게 분리하여 음소로 변환된 성을 가지고 다시 합성하여 소리를 내보내게 된다. 이것에 대한 공식도 다음과 같다.

<합성 공식>

$$\text{int unicode} = 44032 + a * 21 * 28 + b * 28 + c; // \text{문자를 결합 및 변환}$$

$$\begin{aligned}
 44032 &= \text{가} \text{의 유니코드} \quad | \quad a = \text{초성번호} \quad | \quad b = \text{중성번호} \\
 & \quad | \quad c = \text{종성번호} \quad 21 = \text{자음의 갯수} \quad 28 = \text{모음의 개수}.[2]
 \end{aligned}$$

사용자들이 숫자나 영문을 입력할 것을 대비하여 숫자는 한글로 먼저 변환 후 위의 프로세스에 따랐다. '단위(單位)'란, 길이, 무게, 수효, 시간 따위의 수량을 수치로 나타낼 때 기초가 되는 일정한 기준으로 이 단위를 표현하는 단위명사가 있다. 단위명사 앞에는 서수가

오기 때문에 본 시스템에서는 단위명사 앞의 숫자는 서수로 읽도록 하였다.

<단위명사의 예>

개 : (숫자 다음에 써서) 하나씩 떨어져 있는 물건들의 수를 나타내는 단위.

가락 : 가느스름하게 토막진 엇가락과 같은 물건의 날개를 세는 단위.

꾸러미 : 달걀 10개를 꾸려서 싼 것. 꾸려서 싼 것을 세는 단위.

끼 : 끼니를 쉴 때 쓰는 말.

뿔 : 앞, 쇠붙이로 만든 돈, 가마니같이 납작한 물건을 날날의 뜻으로 세는 말.

모금 : 액체나 기체를 한 번 머금은 분량.[11]

본 논문에서 제시한 소프트웨어에 쓰인 발음 소리 변환은 국어 대사전의 표준 발음법을 참고하여 구현하였다. 국어 대사전에 명시된 주요 표준 발음법을 참고하여 구현한 알고리즘은 다음과 같으며 표준 발음법에 의거 일반 텍스트가 변환된 것도 예시에 명시하였다.

<주요 표준 발음법 예시>

- 제4장 받침의 발음

제8항 받침소리로는 ‘ㄱ, ㄴ, ㄷ, ㄹ, ㅁ, ㅂ, ㅇ’의 7개 자음만 발음한다.

- 제6장 된소리되기

제26항 한자어에서 ‘ㄹ’받침 뒤에 연결되는 ‘ㄷ, ㅅ, ㅈ’은 된소리로 발음한다.

예) 갈등[갈똥] 발동[발똥] 절도[절뜨]

<본 논문에서 미 구현된 사항 추가>

제27항 관형사형 ‘-(으)ㄹ’ 뒤에 연결되는 ‘ㄱ, ㄷ, ㅂ, ㅅ, ㅈ’은 된소리로 발음한다.

예) 할 것을[할꺼슬] 갈 데가[갈떼가] 할 바를[할빠를]

- 제7장 소리의 첨가

제29항 합성어 및 파생어에서, 앞 단어나 접두사의 끝이 자음이고 뒤 단어나 접미사의 첫 음절이 ‘이, 야, 여, 요, 유’인 경우에는, ‘ㄴ’소리를 첨가하여 [니, 냐, 녀, 뇨, 뉴]로 발음한다.

예)숨-이불[숨 : 니불] 흙-이불[훈니불] 막-일[망닐][3]

<알고리즘(ㅍ → ㅂ)>

```
if(JongSung[c] == 'ㅂ' || JongSung[c] == 'ㅍ' ||
JongSung[c] == 'ㅃ')
    if(nextch != 'ㅇ' || nextch == ' ')
        //JongSung[c] = 'ㅂ';
```

c = 17;

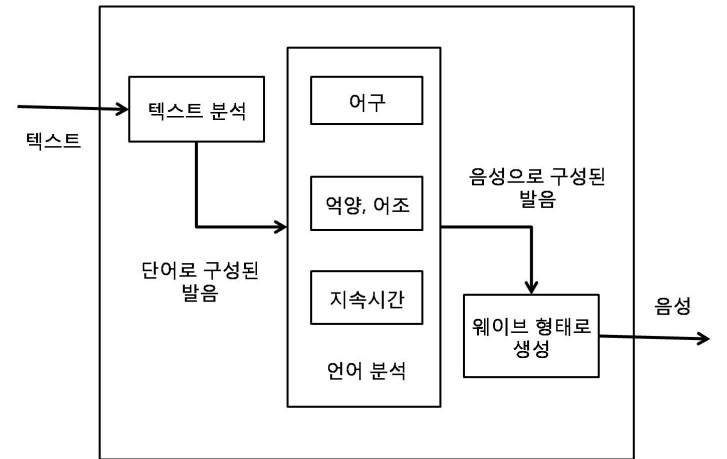
<일반 텍스트>

지난 8월 한글을 공식문자로 받아들인 첫 민족이 나오면서 한글의 우수성이 다시 한 번 주목을 받게 된 사건이 있었다.

<시스템을 통해 변환된 발음 소리>

지난 파열 한그를 공식문자로 바다드린 천 민조기 나오면서 한그레 우수성이 다시 한 번 주목을 받게 된 사건이 일었다.

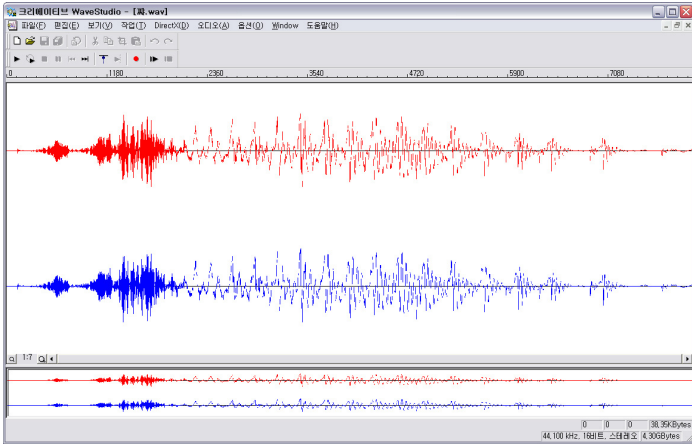
TTS시스템에 대한 프로세스는 여러 가지 있으나 본 시스템은 아래의 프로세스를 기반으로 구현 하였다.



<그림 3>[4]

영문은 실제 미국 원문 발음으로 하지 않았다. 다만 A를 ‘에이’, B를 ‘비이’로 표현 했을 뿐이다. 영문은 따로 발음 변환을 하지 않았으며, A~Z까지 일일이 하나씩 녹음하여 발음을 내보내는 형식으로 구현 하였다. 기타 괄호나 특수 문자 등 도 영문과 마찬가지로 하나씩 녹음하여 그 소리를 가지고 내보내는 형식으로 구현하였다.

음소를 녹음 할 때에는 각 0.4초의 간격으로 녹음하였으며, 음높이도 조절 하였다. 음소 중 0.4초를 넘어가는 것은 없었으며 모두 0.4초가 안되었지만 0.4초에서 발음 웨이브를 제외한 부분에 묵음 처리를 하였다. 영문이나 특수기호는 한 음절 당 0.4 초를 재생하는 형식으로 하였다. 음절의 음파 형태는 아래 그림과 같다.



<그림 4>

본 시스템에서 메일 다운로드 시 간혹 HTML이 포함된 메일을 받을 때도 있으나 대부분이 스팸메일이거나 그림 형식이 많으므로 메일 내용 중 HTML은 모두 제거하여 텍스트 원문만을 나타내었다.

### 3. 음성키보드 & 녹음 (Voice Keyboard & Recording)

음성키보드는 사용자가 본 시스템에 문자 입력 시 이를 알려주기 위하여 도입된 것이며 문자를 받아들여 음소로 변환하는 것과 음성을 들려주는 것을 동시에 하기 위하여 스레드를 이용 본 시스템의 트래픽을 최소화 하였다.[5]

소리 출력 방식은 사용자가 공백을 입력할 때마다 공백 전에 입력된 문자가 출력된다. 다만, 맨 마지막 글자도 공백을 입력해야 나오지만 본 시스템에서는 Trim을 이용하여 양 끝에 공백을 제거하여 메일을 보내게 된다.[6]

본 시스템에서는 녹음도 도입 되었는데 사용자가 텍스트를 입력하기 어려운 환경일 경우 녹음 프로세스를 이용하여 사용자의 음성을 WAV로 녹음한 뒤 그 파일을 첨부하여 메일을 보내는 방식으로 구현하였다. 녹음메일이 도착 했을 시에는 해당 메일을 열어 재생버튼을 통해 재생하며, WAV파일을 제외한 모든 파일은 다운로드 되지 않도록 하였다.

### 4. 개발환경

본 논문에서 제시한 소프트웨어는 Win32-Eclipse를 이용하여 프레임을 구현하였으며, 시스템의 구현은 JDK6.0을 이용하여 구현 하였다.

그래픽은 순수 JAVA에 Swing만을 이용하였으며 Sound API를 이용하여 녹음 및 소리 출력부분을 구현하였다. 기타 메일부분은 Mail API를 이용하여 구현하였다.[7]

음성 WAV 데이터베이스 구성은 HCI Lab에서 제공하는 TTS 음성을 활용했으며, WAV 파일의 편집에는 쉐어웨어 프로그램인 Gold Wave와 Creative Wave Studio를 사용하였다.[8]

## 5. 결론 및 고찰

지금까지 시각 장애우를 위한 이메일 매니저에 관련된 내용과 기법들에 대해 기술하였다. 본 시스템은 시각 장애우의 사용 환경에 중점을 두고 개발되었지만, 일반적인 이메일 매니저의 기능을 모두 포함하고 있어 음성서비스를 제공받고 싶은 모든 이들에게 유효하다고 볼 수 있다.

또한, 이메일 매니저를 처음 접할 확률이 높은 다수를 대상으로 하였기에 일반적인 PC환경이라면 특별히 설치가 복잡하지 않고, 시스템의 구성에 있어 직관적이고 심플한 인터페이스로 접근성을 높이는데 주력했다.

TTS의 구현이 완벽하지 못하고, TTS의 완성도를 높이기 위해서는 매우 많은 시간의 노력이 요구되며 음성 WAV 데이터베이스의 구성과 분량, 적용방식에 따라 TTS의 성능이 좌우되는 경향은 아쉬운 점이다.

인터넷을 통한 정보와 지식이 중요해지는 사회적인 상황에서 그동안 혜택을 누리지 못했던 장애우들이 본 이메일 매니저를 통하여 비장애인과 소통을 이루는 데에 일조 할 수 있기를 바란다.

### [참고문헌]

- [1] 박상언, 코퍼스 기반 한국어 음성합성 시스템의 합성음 자연성 향상, 전남대학교 학위논문, 2001
- [2] 한국어 분리 합성 공식  
<http://blog.naver.com/jwmoon74>
- [3] 이기문 감수, 동아 새국어사전, 동아출판, 1997
- [4] Wikipedia  
[http://en.wikipedia.org/wiki/Speech\\_synthesis](http://en.wikipedia.org/wiki/Speech_synthesis)
- [5] 김윤명, 뇌를 자극하는 Java 프로그래밍, 한빛미디어, 2008
- [6] 김승현, 열혈강의 Java Programming, 프리텍, 2007
- [7] 한글 API Site  
<http://javapia.co.kr/api/jdk-1.5/index.html>

[8] 음성인식 전문업체 HCI Lab

<http://www.hcilab.co.kr/>

[9] 자바누리

<http://javanuri.devpia.com/>

[10] Herbert Schildt, James Holesmes 공저/ 문봉재

역, Art of Java, 정보문화사, 2004

[11] 단위명사

<http://chohamuseum.net/230>