# HTML5기반 지능형 전자책 다국어 TTS 서비스에 관한 연구

류호빈*, 김하경*, 홍성용**
*한국과학기술원 전산학부
**한국과학기술원 산업 및 시스템공학과
e-mail:{ gosyhong}@kaist.ac.kr

# A Study on Intelligent ebook Multilingual TTS Service based on HTML5

Ho-Bin Ryu*, Ha-Kyung Kim*, Seong-Yong Hong**
*School of Computing, KAIST
**Dept of Industrial and System Engineering, KAIST

요 약

웹 기술의 발전은 인간 사회의 변화에 많은 영향을 주고 있다. 웹 발전 분야에서도 전자책 기술 발전은 확연하게 달라지고 있으며, 단순 텍스트 정보만을 제공하는 기술을 넘어 음성서비스를 위한 기술 분야에도 많은 연구가 활발하게 진행되고 있다. 따라서 본 논문에서는 웹 표준 기술인 HTML5기반 지능형 전자책 다국어 TTS 서비스를 위한 관련 연구 조사와 전자책 사용자의 편리성 및 접근성을 강화하기 위한 전자책 제작 및 서비스 방법에 대하여 연구하였다. 본 연구의 목적은 전자책 사용자들에게 텍스트를 포함한 전자책에서 텍스트 및 여러 객체들에 주제 혹은 설명을 TTS 서비스가 가능하도록 하고, 사용자의 선택적 영역과 언어에 따라 다국어 TTS가 자동 변환되어 서비스 할 수 있도록 연구 개발 하는 것이다.

## 1. 서론

최근 IT기술의 급격한 발전으로 소프트웨어 프로그램 원천기술에 활용도가 급격하게 증가하고 있다. 특히 TTS(Text-To-Speech), 음성인식, 인공지능, 빅 데이터, IoT 관련 분야가 많은 시스템과 산업분야에 접목이 되면서 그 활용도는 점점 증가하고 있는 것이다. 이러한 기술 분야는 다양한 활용 목적에 따라서 기술과 기술간에 협업이 발생하고, 기술에 조합을 통해 새로운 기술을 탄생시키기도 한다. TTS기술의 경우 과거와는 다르게 많은 응용 기술 분야에 활용이 되고 있는데, 로봇, 게임, 네비게이션, 무인 안내 시스템, 전자책 콘텐츠 서비스 등 응용 적용분야가 많이 있다. 따라서 본 연구에서는 웹을 기반으로 활성화되고 있는 다국어 TTS 서비스에 관한 연구를 진행하고, 현재 개발 진행하고 있는 HTML5기반 전자책 개발 연구[1]에 지능형 전자책 다국어 TTS 서비스가 적용될 수 있는 방안에 대하여 연구하고 프로토타입을 설계, 구현하는 것이다. 전자책에서 음성 서비스가 응용되어 시작된 것은 오래되었지만 단순한 텍스트의 음성을 출력하여 제공하는 것은 많은 호응을 얻지 못하였다. 그 이유는 전자책 개발의 초기에는 텍스트 콘텐츠(이미지, PDF 등)와 음성 파일(mp3, WAV 등)의 서비스가 별도로 제공되어 두 콘텐츠의 동기화 서비스를 하는 것이 쉽지 않았다. 또한 파일을 다운로드 하거나 별도의 플레이어를 설치해야하는 불편함이 있었던 것이다. 음성 파일은 대부분 인간의 음성을 직접 녹음하거나 전문 기계음 발생장치를 통해 음성합성을 하여 만들어 졌다. 그러나 최근에 음성출력 기술은 TTS 기술을 많이 사용하는데, 글자로 된 말을 소리로 바꾸어 주는  음성합성 기술로 정의할 수 있다[2]. 대표적으로 사용하는 TTS기술은 음편조합방식인데, 먼저 문장을 분석하고 분석된 결과에 맞는 음편들을 미리 만들어 저장해 놓은 대규모 음편 데이터베이스에서 조회하여 출력하게 된다. 따라서 다국어의 언어에 대한 음편 데이터베이스만 존재한다면 하나의 콘텐츠에서 여러 언어의 번역된 서비스가 가능한 것이다. 최근에 개발되는 전자책들은 대부분 TTS 기능을 포함하여 서비스하고 있는데, 하나의 콘텐츠에 한가지 언어만을 제공하는 것이 일반적이다. 또한 사용자가 원하는 언어적 선택이 불가능하고, 객체 인식 설명이 제공되지 않아, 실제 시각장애인이나 저 시력자에게 효용성이 부족한 단점이 있다.  따라서 본 연구의 목적은 HTML5기반으로 제작된 전자책을 사용하는 독자들에게 언어의 선택적 사용이 가능하고 전자책에 포함된 객체들에 대한 TTS서비스가 가능하도록 연구 개발 하는데 있다.

## 2. 관련 연구

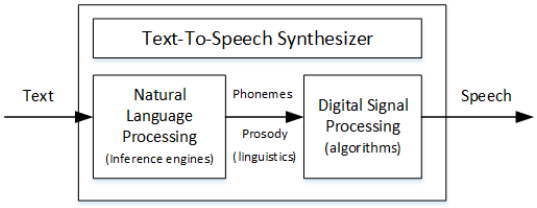
TTS기술의 연구는 크게 2가지 영역으로 나누어 학문 분

본 연구는 [중소기업 융·복합기술개발사업]의 지원을 받아 수행되었습니다.

야가 연구되고 있다. 첫 번째는 텍스트의 분석을 통해 인간의 음성구조와 매핑하여 좀 더 정확한 발음과 음성의 정확도를 높이는 연구 분야이다. 최근에 연구 동향은 인간의 음성을 인식하여 성별을 구분하거나 언어의 종류를 구분하여 TTS 서비스가 가능하도록 하는 연구가 많이 진행되고 있는데, 이러한 기술은 로봇 인간 상호작용이나 인공지능을 이용한 HCI분야에서 활발한 연구가 진행되고 있다. 또한 단순한 텍스트의 정보만을 인식하고 음성으로 변환하는 것이 아니라, 이미지 혹은 PDF와 같이 텍스트 형태의 파일을 광학문자인식(Optical Character Recognition, OCR) 기술을 통해 디지털로 인코딩하는 방법으로 이미지 포맷의 파일도 TTS서비스가 가능하도록 하고 있다. 광학문자인식 방법은 문자패턴의 표현 방법과 분류 방법에 따라 원형 정합(template matching) 방법, 통계적 방법, 분석적 방법 등으로 연구되어지고 있다[3]. 두 번째는 TTS API혹은 라이브러리를 이용하여 전자책 콘텐츠 서비스, 네비게이션 서비스와 같이 텍스트의 정보를 TTS로 활용하는 연구가 많이 진행되고 있다. 특히 2007년 장애인 차별금지법이 제정되고 국립 장애인도서관이 설립되면서 다양한 대체미디어들이 제작되고 있으나 실용성이 저조한 상태이다. 또한 스마트 기기의 급속한 발전으로 교육이나 행정분야에서는 전자책의 활용도가 높아져 TTS기술의 활용도는 점점 증가하고 있는 추세이다. 세계적으로 이러한 TTS기술 개발은 상당히 빠른 발전을 해 왔는데, 대표적으로 시각장애인을 위한 DAISY(Digital Accessible Information SYstem)는 저 시력인과 읽기에 장애가 있는 독자들을 위해 만들어진 XML기반의 전자책 제작 기술이다. 국내에서는 국내 음성도서 서비스를 위한 수식의 음성변환 기법연구도 있었으며, DAISY 포맷은 국제 기구인 DAISY Consortium에서 관리하고 있고, 약 50여개 국가에서 DAISY 표준에 참여하고 있다.

## 3. 다국어 TTS 서비스

음성합성(Speech Synthesis)은 인간의 언어(speech)를 인공적으로 만들어 내는 소프트웨어적 기술이다. 음성 합성의 정의는 기계적 장치 혹은 전자회로 또는 컴퓨터 시뮬레이션를 이용하여 자동으로 음성 파형을 생성하는 것이다. 그림 1은 인간의 텍스트를 어떻게 TTS로 변환하여 서비스하고 있는지에 대한 기능을 간략하게 설명하고 있다[4].




[그림1] TTS Synthesizer

음성합성은 여러 국가에서 자국의 언어에 대한 TTS 서비스의 사용을 위해 다양한 방법으로 연구 되고 있다. 국내에서도 한국어 TTS 기술 개발을 위한 연구는 다양한 방법으로 많이 시도되고, 연구되어지고 있었다. 사실 TTS의 기술을 연구하기 위해 언어학적 연구와 공학적 연구는 오래기간 지속되어 왔고, 최근 웹을 통한 TTS 서비스의 활용도는 점점 증가하고 있는 것이다. TTS API의 경우 전 세계적으로 이러닝(e-Learning)이 급속하게 확대되면서 많이 응용되어 개발되었는데 표1과 같은 TTS 소프트웨어 제공사들이 있다.

<표1. TTS 소프트웨어 제공사>

| TTS기술 제공사 명칭 | 한국어 지원 |
|---|---|
| ResponsiveVoice | YES |
| NaturalReader | NO |
| Zabaware Text To Speech Reader | NO |
| iSpeech | YES |
| Acapela Group Virtual Speaker | YES |
| TextSpeechPro | YES |
| AudioBookMaker | YES |
| Read The Words | NO |
| Voice Reader | YES |
| Ivona | NO |
| Neospeech | YES |

본 연구에서는 다국어 지원을 목적으로 개발되어진 TTS 소프트웨어의 조사와 분석을 통해 한국어 지원 여부, 음성합성의 인식정도 그리고 TTS API 사용의 유용성 등을 분석 하였다. 대부분에 TTS 소프트웨어 제공사들은 다국어 지원을 하고 있었으며, 한국어(korean language) 지원을 하지 않는 TTS도 일부 있었다. 또한 사용 방법으로 TTS 엔진 프로그램의 설치 혹은 플러그인(plug-in)을 통한 서비스를 하도록 하는 소프트웨어들이 많았으며, 온라인 API를 이용한 TTS 서비스 사용이 가능한 소프트웨어들도 있었다. 따라서 본 연구 목적으로 개발 하는 HTML5기반 지능형 전자책 다국어 TTS 서비스를 위해 가장 적합한 API로 ResponsiveVoice를 선택하여 연구 실험 및 프로토타입 구현에 사용하였다.

## 4. 지능형 전자책 다국어 TTS 서비스 설계

지능형 전자책 다국어 TTS 서비스 설계는 하나의 전자책 콘텐츠 작성으로, 여러 국가에서 사용하는 사용자 혹은 접속 국가를 자동으로 인식하여 해당 국가의 언어로 자동 TTS 서비스가 가능하도록 하는 것이다. 그림2는 본 연구에서 설계한 HTML5기반 지능형 전자책 다국어 TTS 서비스 구성도를 설명하고 있다.

[그림 2] 지능형 전자책 다국어 TTS 서비스 설계

전자책을 서비스하는 시스템은 N-스크린 디바이스에서 접속하는 사용자의 해당 국가 언어를 자동으로 확인하여 인식하고, 어떤 언어를 사용할 것인지 자동으로 선택하게 된다. 또는 사용자가 자신이 원하는 언어를 선택적으로 지정하여 TTS의 서비스를 사용할 수 있다. 또한 한국과 같이 영어를 많이 혼용하여 사용하는 국가의 경우, 한글의 콘텐츠를 영어로 읽어주는 TTS 서비스가 가능하고, 역으로 영어로 작성된 콘텐츠를 한국어로 읽어 주는 TTS 서비스가 가능하다. 예를 들어 전자책 콘텐츠에 텍스트는 한글로 보이지만, 영어 또는 다른 나라 언어로 읽어 주는 것이다. 이러한 전자책 TTS서비스 장점은 하나의 콘텐츠를 작성하여 다국어 TTS로 사용할 수 있을 뿐만 아니라, 해외에서의 전자책 활용 등에 큰 장점이 있을 것으로 기대한다.

## 5. 실험 및 프로토타입 구현

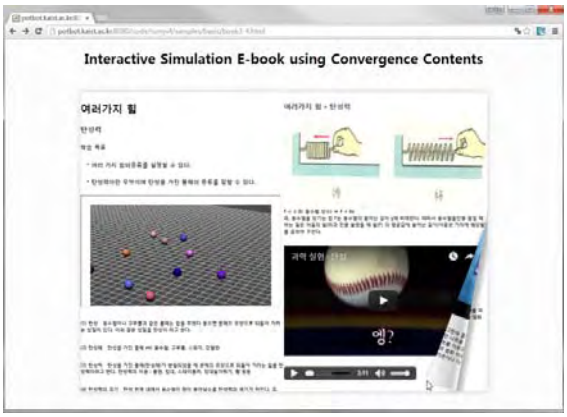
본 연구에 실험을 위하여 HTML5기반의 전자책을 구현하였으며, HTML5 문서의 구조는 EPUB3.0의 표준과 호환되도록 설계 제작하였다. 따라서 EPUB3.0 Reader가 없이도 일반 웹 브라우저에서 사용이 가능하다. 추가된 XML 시맨틱 태그 구조는 각 나라의 국가 명칭을 고려하여 한국어는 <KR>, 영어는 <EN>과 같이 설계하여 추가하였다. 그림3은 전자책에 포함되어 있는 한국어와 영어의 XML 시맨틱 태그 구조 일부를 나타내고 있다.

```
<div class='header'>
  <img src='./logo.gif' class='text_r' alt='<KR>카이스트
로고</KR><EN>KAIST logo</EN>' id='a' />
 <h1 class='text_r'> KAIST PORTAL </h1>
</div>

<div class='main'>
  <div class='nav'>
    <ul class='list'>
      <li class='text_r'><KR>이메일</KR>
                          <EN>E mail</EN></li>
      <li class='text_r'>Web ERP</li>
      <li class='text_r'><KR>도서관</KR>
                          <EN>Library</EN></li>
      <li class='text_r'><KR>학사시스템</KR>
       <EN>School affairs information system</EN></li>
      <li class='text_r'><KR>예약</KR>
                          <EN>Reservation</EN></li>
    </ul>
 </div>
 ... 생략
```

[그림3] HTML5기반 시맨틱 태그 구조

본 연구의 결과로 프로토타입 구현을 위해 ResponsiveVoice에서 제공하는 TTS API를 사용하였으며, HTML5, CSS3, 자바스크립트(javascripts), JQuery 언어로 전자책을 개발하였다. 구현 프로토타입의 중요 기능으로는 사용자의 텍스트 선택 행위에 따라서 TTS가 동작하도록 하였고, 한국어와 영어를 구분하여 TTS가 동작하도록 하였다. 그림4는 한국어와 영어가 동시에 전자책 콘텐츠로 작성되어 있는 사례를 보이고 있다. 사용자는 자신이 듣고 싶은 영역 텍스트에 터치를 하면 해당 국가의 언어로 들을 수 있다. 만약 한글로 보이는 텍스트라도 영어로 듣고 싶으면, 언어를 영어로 선택하여 번역된 음성으로 들을 수 있다. 또한 이미지나 동영상, WebGL객체 등에 대한 설명으로 해당 언어의 태그로 속성을 사용하면 TTS 서비스를 받을 수 있다.




[그림4] 지능형 전자책 다국어 TTS 서비스 구현 사례

## 6. 결론 및 제언

본 논문에서는 웹 표준 기술인 HTML5기반으로 제작된 전자책에 지능적으로 다국어 TTS 서비스를 가능하도록 하기 위한 관련 연구 조사와 설계 그리고 프로토 타입의 구현 내용을 설명하였다. 웹과 디바이스의 발전으로 전자책 사용자가 급증하고 있으며, 특히 시각장애인이나 저 시력의 문제로 책을 접하기 어려운 사용자들에게 사용자의 편리성 및 접근성을 강화하기 위한 지능형 전자책 다국어 TTS 서비스 방법에 대하여 연구하였다. 본 연구의 결과로 전자책 사용자들은 자신이 원하는 영역에 텍스트 혹은 객체의 콘텐츠를 선택할 수 있을 뿐만 아니라, 원하는 국가의 언어로 자동 변환하여 서비스 될 수 있도록 하였다. 이러한 연구의 결과는 다국어를 사용하는 국가나 하나의 콘텐츠를 여러 국가의 언어로 사용하는 곳에서 유용하게 사용될 수 있을 것으로 기대하며, 전자칠판을 많이 사용하고 있는 교육분야에서도 많이 활용할 수 있을 것으로 기대하고 있다. 향후연구로 전자책 콘텐츠에 작성되는 수학적 공식이나 표기 등에 대한 TTS 제공방식을 좀 더 구체적으로 연구하여 반영하는 것이 필요하다.

## 참고문헌

[1] 김태겸, 장현중, 홍성용 "3차원 상호작용기반 웹 표준 e-book 콘텐츠 저작도구 플랫폼 설계 및 구현" 한국정보처리학회 춘계학술발표, 2015, pp.985-988.

[2] Schroeter, Juergen. "Basic principles of speech synthesis." Springer Handbook of Speech Processing. Springer Berlin Heidelberg, 2008, pp.413-428.

[3] Mantas, J. "An overview of character recognition methodologies." Pattern recognition 19.6, 1986, pp.425-430.

[4] Dutoit, Thierry. "High-quality text-to-speech synthesis: An overview." Journal of Electrical and Electronics Engineering Australia 17, 1997, pp.25-36.