한국음향학회지 제40권 제5호 pp. 523～529 (2021)
The Journal of the Acoustical Society of Korea Vol.40, No.5 (2021)
https://doi.org/10.7776/ASK.2021.40.5.523

pISSN : 1225-4428
eISSN : 2287-3775

다음색 감정 음성합성 응용을 위한 감정 SSML 처리기

An emotional speech synthesis markup language processor for multi-speaker and emotional text-to-speech applications

유세희,[1] 조희,[2] 이주현,[3] 홍기형[4†]

(Se-Hui Ryu,[1] Hee Cho,[2] Ju-Hyun Lee,[3] and Ki-Hyung Hong[4†])

[1]성신여자대학교 대학원 미래융합기술공학과, [2]한국전력공사 전력연구원, [3]CJ 올리브네트웍스 AI-DT 연구소,
[4]성신여자대학교 서비스디자인공학과

(Received July 31, 2021; accepted September 2, 2021)

초 록: 본 논문에서는 감정 마크업을 포함하는 Speech Synthesis Markup Language(SSML) 처리기를 설계하고 개발하였다. 다양한 음색과 감정 표현이 가능한 음성합성 기술이 개발되고 있으며 다양한 음색 및 감정 음성합성의 응용 확대를 위하여 표준화된 음성 인터페이스 마크업 언어인 SSML을 감정 표현이 가능하도록 확장한 감정 SSML (Emotional SSML)을 설계하였다. 감정 SSML 처리기는 그래픽 사용자 인터페이스로 손쉽게 음색 및 감정을 원하는 텍스트 부분에 표시할 수 있는 다음색 감정 텍스트 편집기, 편집 결과를 감정 SSML 문서로 생성하는 감정 SSML 문서 생성기, 생성된 감정 SSML 문서를 파싱하는 감정 SSML 파서, 감정 SSML 파서의 결과인 다음색 감정 합성 시퀀스를 기반으로 합성기와 연동하여 음성 스트림의 합성 을 제어하는 시퀀서로 구성된다. 본 논문에서 개발한 다음색 감정합성을 위한 감정 SSML 처리기는 프로그래밍 언어 및 플랫폼 독립적인 개방형 표준인 SSML을 기반으로 하여 다양한 음성합성 엔진에 쉽게 연동할 수 있는 구조를 가지며 다양한 음색과 감정 음성합성이 필요한 다양한 응용 개발에 활용될 것으로 기대한다.

핵심용어: 음성합성, 마크업언어, 감정, 다음색 합성

ABSTRACT: In this paper, we designed and developed an Emotional Speech Synthesis Markup Language (SSML) processor. Multi-speaker emotional speech synthesis technology that can express multiple voice colors and emotional expressions have been developed, and we designed Emotional SSML by extending SSML for multiple voice colors and emotional expressions. The Emotional SSML processor has a graphic user interface and consists of following four components. First, a multi-speaker emotional text editor that can easily mark specific voice colors and emotions on desired positions. Second, an Emotional SSML document generator that creates an Emotional SSML document automatically from the result of the multi-speaker emotional text editor. Third, an Emotional SSML parser that parses the Emotional SSML document. Last, a sequencer to control a multi-speaker and emotional Text-to-Speech (TTS) engine based on the result of the Emotional SSML parser. Based on SSML which is a programming language and platform independent open standard, the Emotional SSML processor can easily integrate with various speech synthesis engines and facilitates the development of multi-speaker emotional text-to-speech applications.

Keywords: Text-to-speech, Markup language, Emotion, Multiple voice colors

PACS numbers: 43.72.Ja, 43.70.Ep

†**Corresponding author:** Ki-Hyung Hong (khhong@sungshin.ac.kr)
Department of Service Design Engineering, Sungshin Womens University, 34 da-gil 2, Bomun-ro, Seongbuk-gu 02844, Republic of Korea
(Tel: 82-2-920-7525, Fax: 82-2-920-7157)

Copyright ⓒ2021 The Acoustical Society of Korea. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

I. 서 론

음성 합성 및 인식 기술의 발전에 따라 음성인터페이스를 기반으로 하는 응용 서비스가 확대되고 있다. 기존의 음성 자동응답시스템은 초기에는 음성 녹음을 기반으로 하였고, 낭독체 음성합성으로 충분하였다. 최근에는 로봇 사용자 인터페이스나 동화 읽어 주기 등에서 다양한 개성과 감정 표현이 가능한 다음색 및 감정 음성 합성 기술이 요구되고 있다.

기존의 하드코딩으로 개발하였던 음성 사용자 인터페이스는 VoiceXML[1]과 같은 개방형 웹 기반 표준의 등장과 VoiceXML 처리기 개발[2]을 통하여 음성 인터페이스 시나리오의 변경 비용이 크게 줄고, 유지관리 편이성이 높아짐에 따라 다양한 분야에서 음성 기술의 활용성을 크게 확대하였다.[3]

Speech Synthesis Markup Language(SSML)[4]은 W3C에서 음성 브라우저 및 다양한 응용에서 음성합성을 지원하기 위하여 개발한 개방형 Extensible Markup Language(XML) 기반 마크업 언어이다. 음성합성음 생성에 필요한 음색, 발음열, 피치, 운율, 합성음 빠르기 등의 다양한 특징값을 표준 마크업으로 표현하여, SSML 문서를 작성하고 SSML 파서가 이를 분석하여 음성합성엔진에 전달하여 음성합성음 생성의 제어를 수행할 수 있다. 대표적으로는 구글 클라우드 Text-to-Speech(TTS)의 SSML 서비스[5]가 있다.

VoiceXML, SSML과 같은 XML 기반 개방형 표준의 사용은 음성합성엔진과 음성 응용 서비스 개발이 서로 독립적으로 가능하게 한다. 음성 응용 서비스는 결국 SSML 문서 또는 VoiceML 문서로 개발되며 SSML 처리기 또는 VoiceXML 브라우저가 문서를 해석하여 음성합성엔진과 상호작용하여 최종 서비스가 이루어지도록 한다. 음성응용서비스를 다시 개발하지 않아도 음성합성엔진의 업그레이드가 가능하고, 보다 성능 좋은 합성엔진으로 교체가 가능하도록 한다.

최근에는 로봇 사용자 인터페이스, 개인 맞춤형 음성 서비스 등 단일 음색 낭독체 음성합성에서 더 나아가 다음색 및 감정 음성합성 기술에 대한 요구가 증대하고 있으며, 다음색 및 감정 음성합성엔진의 연구 개발이 이루어지고 있다.[6,7]

본 논문에서는 SSML에 감정 표현 마크업을 추가하여 확장한 감정 SSML(Emotional SSML)을 설계하고 감정 SSML 처리기를 설계 구현하였다. 감정 SSML 처리기는 그래픽 사용자 인터페이스를 가진 다음색 감정 텍스트 편집기, 감정 SSML 문서 생성기, 감정 SSML 파서, 그리고 음성합성기(TTS Engine)와 연동하는 시퀀서로 구성하였다. 개방형 표준인 SSML 문서, 그리고 음성합성기와의 연동에서 또 다른 개방형 데이터 교환 표준인 Javascript Object Notation(JSON)[8]을 채용함으로써, 음성합성기의 개발 및 운영 플랫폼에 독립적으로 응용개발을 가능하게 한다.

II. 감정 SSML

먼저, 다음색(voice color or multi-speaker)은 이미 SSML에 포함되어 있으며, 다음의 SSML 엘리먼트 <*voice*>로 명시할 수 있다. Fig. 1 <voice> 엘리먼트의 name 특성 값에 음성합성엔진이 지원하는 음색(위 예에서는 "유미")을 지정한다.

감정 표현은 SSML만으로는 할 수 없고, 감정 표현에 관한 웹 표준인 EmotionML[9]의 엘리먼트를 함께 사용해야 한다. EmotionML은 감정 표현이 필요한 다양한 영역에서 사용할 수 있도록 마련된 웹 표준으로 감정 음성을 포함하여 표정, 동작 등의 감정 표현에도 활용할 수 있다.

EmotionML에서는 먼저 표현할 수 있는 감정의 종류와 각 감정의 명칭을 정의하는 감정표현 정의 XML 문서를 작성하여야 한다. Fig. 2는 '화남', '보통', '기쁨', '슬픔'의 4가지 감정으로 사용할 감정명칭을 각각 'anger', 'neutral', 'happiness', 'sadness'로 정의한 예제 문서(emotiontts_categories.xml)이다. 여기서 정의한 감정명칭의 집합을 감정명칭공간(TTS-specific emoition name space)이라고 한다.

본 논문에서는 SSML 1.1[4]의 명칭공간과 EmotionML 1.0[9]의 명칭공간, 그리고 연동할 다음색 감정합성기

```
<voice name="유미" gender="female" variant="2" age="20"
    required="gender variant" ordering="name gender variant age"
    onvoicefailure="processorchoice">
```

Fig. 1. (Color available online) A 〈voice〉 element including a TTS-specific voice name.

```
<emotionml version="1.0"
    xmlns="http://www.w3.org/2009/10/emotionml">
    <!-- CATEGORIES -->
    <vocabulary type="category" id="emotiontts">
        <item name="anger"/>
        <item name="neutral"/>
        <item name="happiness"/>
        <item name="sadness"/>
    </vocabulary>
</emotionml>
```

Fig. 2. (Color available online) An XML document to define a name-space for TTS-specific emotions (emotiontts_categories.xml).

```
1 <?xml version="1.0"?>
2 <speak xml:lang="ko" version="1.1"
3     xmlns="http://www.w3.org/2001/10/synthesis"
4     xmlns:emo="http://www.w3.org/2009/10/emotionml">
5     <p>
6         <emo:emotion category-set="emotiontts_categories.xml">
7             <emo:category name="neutral" value="0.4"/>
8         </emo:emotion>
9         안녕하세요.
10        <s>
11            지금은 로봇 음성 합성기와
12            <w xml:lang="en-US">EmotionSSML parser</w>
13            의 연동을 위하여 회의 중입니다.
14        </s>
15        오늘은
16        <say-as interpret-as="date" format="ymd">2019/06/19</say-as>
17        수요일 입니다.
:         ... [중략] ...
41        <voice name="유미" gender="female" variant="2" age="20"
42            required="gender variant" ordering="name gender variant age"
43            onvoicefailure="processorchoice">
:             ... [중략] ...
48            <emo:emotion category-set="emotiontts_categories.xml">
49                <emo:category name="anger" value="0.4"/>
50            </emo:emotion>
51            회의 시간에 30분이나 지각을 하다니, 제정신이야!
:             ... [중략] ...
54        </voice>
55        <s>
56            <prosody pitch="high" volume="+6dB">다음부터 늦지마.</prosody>
57        </s>
58        <s>
59            <prosody rate="default" pitch="low">그럼 회의를 시작하지.</prosody>
60        </s>
61    </p>
:    ... [중략] ...
68 </speak>
```

Fig. 3. (Color available online) A sample emotional SSML document.

에서 제공하는 감정(TTS-specific emotions)들을 XML 명칭으로 정의한 Fig. 2와 같은 합성기 지원 감정명칭 공간을 사용하는 문서를 감정 SSML 문서라고 한다.

감정 SSML 문서는 문장(또는 단어나 구)에 대하여 합성을 원하는 음색과 감정을 명시한 XML 문서이다. Fig. 3은 감정 SSML 문서의 예이다.

감정 SSML 문서는 Fig. 3의 3번째, 4번째 줄과 같이 <speak> 엘리먼트에 SSML 명칭공간(xmlns="http://www.w3.org/2001/10/synthesis")과 EmotionML 명칭공간(xmlns:emo="http://www.w3.org/2009/10/emotionml")이 명시되어 있어야 하며, 기본 명칭공간은 SSML이다. EmotionML에서 정의한 명칭(엘리먼트 이름, 특성값의 이름 등)은 접두어 'emo:'으로 시작하도록 정의하였다.

합성기 지원 감정명칭공간은 Fig. 3의 6번째 줄의 <emo:emotion> 엘리먼트의 'category-set' 특성의 값으로 지정한다. Fig. 3의 문서에서는 감정명칭공간을 Fig. 2에서 보인 "emotiontts_categories.xml"을 사용하였다. 합성기 지원 감정명칭공간에서 정의한 감정명칭 중에서 원하는 감정명칭을 사용하여 다음에 나타나는 문장의 감정을 명시할 수 있다. Fig. 3의 7번째 줄 <emo:category> 엘리먼트는 '보통' 감정(neutral), 정도 값은 0.4로 '안녕하세요' 문장 (9번째 줄)을 음성합성하라는 뜻이다.

사용 가능한 음색의 종류와 감정은 연동하고자 하는 다음색 감정 음성합성기에 따라 달라진다. 감정 SSML에서 음색의 표시는 SSML의 <value> 엘리먼트의 name 특성의 값으로 연동하고자 하는 음성합성기에서 지원하는 이름을 사용한다. 또한, 음성합성기에서 지원하는 감정의 종류와 표현은 Fig. 2의 예에서 보인 것과 같이 감정명칭공간으로 정의하여 사용한다.

감정 합성이 가능한 SSML 처리기에는 아마존의 SSML[10]과 마이크로소프트의 SSML[11] 등이 있다. 아마존과 마이크로소프트의 SSML 처리기는 각 기업에서 개발한 합성기에서 지원하는 감정표현만을 지원한다. 예를 들어 아마존의 경우, 'excited'와 'disappointed'의 2가지 표현만을 처리한다. 이에 비하여, 본 논문의 감정 SSML은 EmotionML 웹 표준에 기반하여 Fig. 2와 같이 사용자가 감정표현을 정의하여 사용할 수 있다.

III. 감정 SSML 처리기 설계 및 구현

본 논문에서 개발한 감정 SSML 처리기는 음성합성기와는 독립적으로 설계 구현하였다.

Fig. 4는 감정 SSML 처리기의 블록도이다. 감정 SSML 처리기는 크게 다음색 감정문서텍스트편집기, 감정 SSML 문서 생성기, 감정 SSML 파서, 그리고 시퀀서의 4개 모듈로 구성하였다.

3.1 다음색 감정 텍스트 편집기

마크업 언어에 대한 전문 지식이 없어도 쉽게 합성하고자 하는 텍스트에 원하는 음색과 감정을 원하는 부분에 명기하기 위하여, 그래픽 사용자 인터

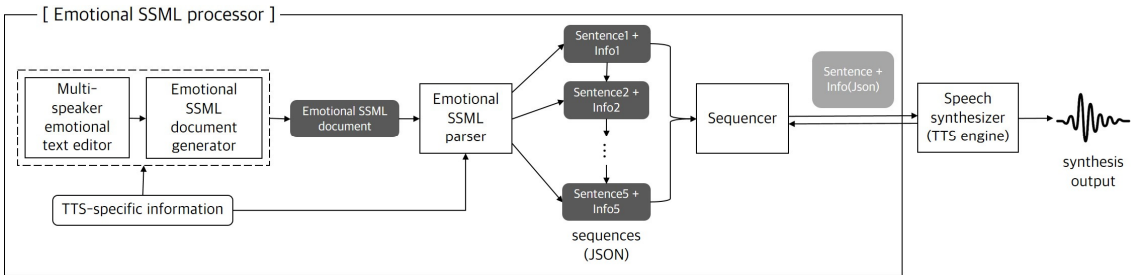

Fig. 4. Emotional SSML processor.

Fig. 5. Multi-speaker emotional text editor.

페이스를 가진 편집기를 Fig. 5와 같이 설계 구현하였다.

편집기의 Fig. 5의 (1)은 합성할 텍스트를 입력하는 부분으로 미리 준비한 텍스트 파일명을 입력하여 읽어 들이거나, 직접 입력할 수 있다. Fig. 5의 (2)는 텍스트를 문장단위로 자동 분리하여 문장 목록으로 보여준다. 문장 목록에서 음색이나 감정을 표시할 문장을 선택한다.

선택한 문장에 대한 음색은 Fig. 5의 (3)에 보인 바와 같이 음색 이름, 성별, 그리고 연령을 선택하여 지정한다. 선택 문장에 대한 감정은 Fig. 5의 (4)에서와 같이 감정명칭, 정도를 선택 또는 입력한다. 음색 이름과 감정명칭은 연동하고자 하는 음성합성엔진에 따라 다르므로, 음성합성엔진 특성 정보를 제공받아서 음색 선택 목록, 감정 명칭 목록을 제시할 수 있도록 구현하였다. 음성합성엔진 특성 정보는 해당 음성 합성 엔진이 지원하는 음색 이름 목록과 감정명칭공간을 정의한 XML 파일(Fig. 2 참고)을 의미한다.

3.2 감정 SSML 문서 생성기

다음색 감정 텍스트 편집이 완료되면, 사용자가 텍스트에 명기한 음색과 감정을 토대로 감정 SSML

```
<voice name="유미" gender="female" variant="2" age="20"
    required="gender variant" ordering="name gender variant age"
    onvoicefailure="processorchoice">
    ... [중략] ...
    <emo:emotion category-set="emotiontts_categories.xml">
        <emo:category name="anger" value="0.4"/>
    </emo:emotion>
    회의 시간에 30분이나 지각을 하다니, 제정신이야!
    ... [중략] ...
</voice>
```

Fig. 6. (Color available online) An automatic generation result of a sentence with a specific speaker and emotion.

문서를 자동 생성한다.

예를 들어 “회의 시간에 30분이나 지각을 하다니, 제정신이야!”의 문장에 대하여 음색은 ‘유미’, 감정은 ‘화남’, 감정의 정도는 ‘0.4’로 설정한 경우, 감정 SSML 문서 자동 생성기가 만든 결과는 Fig. 6과 같다.

<voice> 엘리먼트의 특성 name 값이 “유미”로 지정되어 있으며, <emo:emotion>에서 사용할 감정명칭 공간을 “emotiontts_categories.xml”(Fig. 2 참고)로 지정하였다. 그리고 <emo:category> 엘리먼트의 특정 name의 값이 ‘anger’, value의 값이 ‘0.4’로 지정된 것을 알 수 있다. 텍스트 전부를 자동변환한 결과는 Fig. 3과 같다.

3.3 감정 SSML 문서 파서

감정 SSML 문서는 감정 SSML 문서 파서(Emotional SSML parser)에 의하여 파싱되고, 그 결과는 JSON 객체의 시퀀스로 음성합성기와 상호작용을 담당하는 시퀀서에 전달한다.

감정 SSML 문서 파서는 Fig. 7과 같이 감정 SSML 파싱과 시퀀스 생성의 2단계로 구성하였다.

먼저 감정 SSML 파싱은 입력된 감정 SSML 문서를 파싱하여 Fig. 7의 (2)와 같은 Document Object Model (DOM)[12] 트리(Parsing DOM tree)를 구성한다. DOM

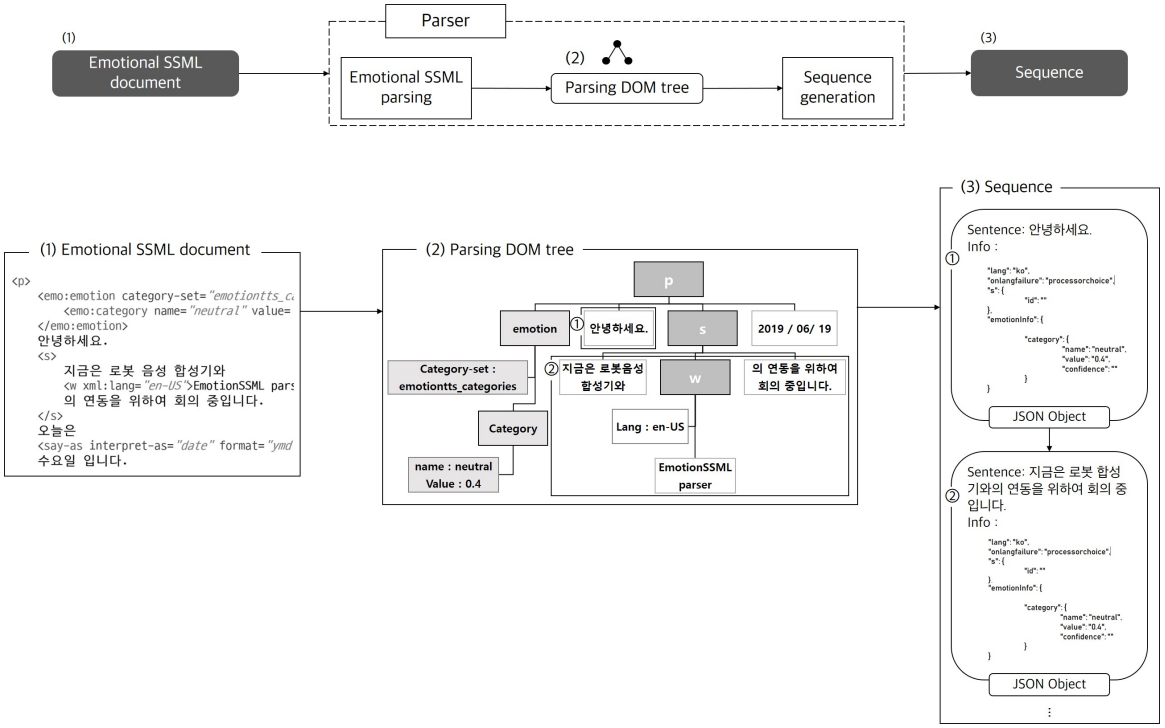

Fig. 7. Emotional SSML document parser.

트리는 XML 문서의 처리를 위한 표준 모델이다. 파싱 결과는 시퀀스 생성 단계에서 연동하고자 하는 음성합성기의 특성 정보를 바탕으로 음성합성기에서 처리할 수 있는 합성 단위로 분해하고 각 합성단위를 감정 SSML 문서에 기술된 순서에 따라 순서화한 시퀀스를 생성한다.

음성합성기 특성 정보는 음성합성기가 합성할 수 있는 언어, 합성오류 처리 방법, 음색의 종류 및 명칭, 감정 표현 개수와 음성합성기에서의 감정 이름 등을 JSON 객체로 기술한 것이다. 감정과 음색에 대한 정보는 편집기에서 사용자의 선택 목록을 제시하기 위해서도 사용한다.

Fig. 7 (2)의 DOM 트리를 문장단위 합성기와 연동하기 위하여 문장 단위로 JSON 객체를 생성하고 이를 감정 SSML 문서에 기술된 순서에 따라 순서화한 시퀀스는 Fig. 7 (3)과 같다. Fig. 7 (2)의 DOM 트리에서 ①과 ②문장을 각각 Fig. 7 (3)의 ①과 ②의 JSON 객체 시퀀스로 변환한 것이다. 각 JSON 객체는 합성할 문장, 언어, 오류 처리 방법, 합성 음색, 감정 표현 정보를 담고 있다.

3.4 감정 SSML 시퀀서

감정 SSML 파서가 생성한 시퀀스는 연동하고자 하는 음성합성기와 직접 상호작용하는 시퀀서에서 시퀀스의 각 JSON 객체를 순서대로 음성합성기에 전달하여 합성기가 음성신호를 생성하도록 한다. 시퀀서의 주요 역할은 다음과 같다.

- 합성하고자 하는 텍스트의 길이가 긴 경우, 합성기가 음성 신호를 생성하기까지 소요되는 시간이 길어질 수 있다. 연동하고자 하는 합성기가 실시간으로 음성합성할 수 있는 텍스트 길이를 단위로 합성 대상 전체 텍스트를 나누어 시퀀스의 JSON 객체로 구성할 수 있도록 한다. 시퀀서가 시퀀스의 JSON 객체를 파라미터 순서대로 음성합성기를 호출함으로써 합성기의 성능에 맞추어 합성 지연 시간을 최소화하는 합성을 가능하게 한다.
- 음성합성기에서 발생하는 오류는 합성기 자체에서 처리하는 경우와 오류를 합성기를 호출한 모듈로 처리를 위탁하는 경우가 있다. 후자의 경우에는 전체 응용 프로그램이 음성합성기의 오류로 인하

Table 1. Implementation environment.

Computer (CPU)	Intel Xeon 4116 x86
OS	Ubuntu 18.04.2 LTS
Language	Python 3.7
Management	Git

여 비정상 종료하는 것을 방지하여야 한다. 시퀀서는 합성기로부터 수탁한 오류를 처리하여 전체 응용 프로그램의 비정상 종료를 방지한다.

3.5 구현환경

본 논문에서 구현한 감정 SSML 처리기의 구현 환경은 Table 1과 같다. Ubuntu 운영체제에서 개발되었으며, 처리기 및 그래픽 사용자 인터페이스 구현에 Python 3.7 버전을 사용하였다. 감정 SSML 처리기를 사용하기 위해서는 최소 Python 3.5 이상이 설치된 환경이 요구된다. 처리기 구현에 사용된 데이터들과 코드, 이들의 사후 관리를 위해 Git을 통해 형상관리를 한다.

감정 SSML 처리기는 Reference [6]에서 개발한 다음색 감정 음성합성엔진과 연동하여 시험하였다. 본 논문에서 설계 구현한 감정 SSML 처리기와 시험에서 사용한 다음색 감정 음성합성기는 GitHub[13]에 공개하였다.

IV. 결 론

본 논문에서는 로봇 대화 인터페이스, 동화 구연 등의 다음색 감정 음성합성 응용에서 사용할 수 있는 감정 SSML 처리기를 설계 구현하였다. SSML[4]과 EmotionML[9]의 명칭 공간을 이용하여 다음색 및 감정 마크업이 가능한 감정 SSML을 설계하였다. 감정 SSML 처리기는 크게 다음색 감정문서텍스트편집기, 감정 SSML 문서 생성기, 감정 SSML 파서, 그리고 시퀀서의 4개 모듈로 구현하였다.

감정 SSML 문서 편집기는 합성 대상 텍스트에 사용자가 원하는 감정 및 음색을 그래픽 사용자 인터페이스를 통하여 표시할 수 있고, 사용자가 표시한 감정과 음색에 따라 감정 SSML 문서 생성기에서 감정 SSML 문서를 자동 생성한다.

연동하고자 하는 다음색 감정 합성기에서의 합성음 생성까지의 지연 시간을 참고하여, 감정 SSML 파서는 감정 SSML 문서를 합성기의 성능에 적합한 합성단위로 분해하고 각 합성 단위를 JSON 객체로 변환한다. 입력된 전체 감정 SSML 문서에 기술된 순서에 따라 합성 단위 JSON 객체를 순서화한 시퀀스를 시퀀서에 전달한다. 시퀀서는 감정과 음색 정보가 명기된 각 JSON 객체를 파라미터로 연동하고자 하는 다음색 감정음성 합성기를 호출하여 음성합성을 제어한다.

본 논문에서 설계 구현한 감정 SSML 처리기는 감정 SSML 파서와 시퀀스 모듈만 독립적으로 활용할 수 있다. 감정 SSML 파서는 웹 표준인 XML 기반의 감정 SSML을 입력으로 하며, 시퀀서는 JSON이라는 프로그래밍 언어에 종속되지 않는 개방형 표준을 사용하여 음성합성기와 연동할 수 있다. 로봇 대화 인터페이스나 동화 구연 등의 응용은 감정 SSML 문서를 작성하는 것으로 개발할 수 있다. JSON 객체를 받아들이는 음성합성기는 그 개발 언어와 플랫폼에 무관하게 본 논문에서 개발한 감정 SSML 처리기와 연동할 수 있다.

프로그래밍 언어 독립적이며 플랫폼 독립적인 표준을 사용한 감정 SSML 처리기는 응용 프로그램 개발 언어와 사용할 음성합성기에 대한 제약이 없으며, 다음색 감정 음성 합성 응용의 확산에 기여할 것으로 기대한다.

감사의 글

이 논문은 산업통상자원부의 산업기술혁신사업으로부터 지원을 받아 수행된 연구입니다(지원번호: 10080667, 음원 다양화를 통하여 로봇의 감정 및 개성을 표현할 수 있는 대화음성합성 원천기술 개발).

References

1. *World Wide Web Consortium (W3C) Recommendation Voice Extensible Markup Language (VoiceXML) Version 2.0*, https://www.w3.org/TR/voicexml20, (Last Viewed July 2, 2021).
2. K. R. Kim and K. H. Hong, “An extension of the VoiceXML platform for push-based voice applications”

(in Korean), J. Acoust. Soc. Kr. **21**, 27-36 (2002).
3. J. Rouillard, "Web services and speech-based applications around VoiceXML," J. Networks. **2**, 27-35 (2007).
4. *World Wide Web Consortium (W3C) Recommendation Speech Synthesis Markup Language (SSML) Version 1.1*, https://www.w3.org/TR/speech-synthesis11/, (Last Viewed July 2, 2021).
5. *Google Cloud Text-to-Speech Synthesis Markup Language (SSML)*, https://cloud.google.com/text-to-speech/docs/ssml, (Last Viewed July 2, 2021).
6. H. J. Choi, S. J. Park, J. U. Park, and M. S. Hahn, "Emotional speech synthesis for multi-speaker emotional dataset using wavenet vocoder," Proc. IEEE ICCE. 1-2 (2019).
7. S. S. Oh, S. Y. Um, I. S. Jang, C. H. Ahn, and H. G. Kang, "Determination of representative emotional style of speech based on k-means algorithm" (in Korean), J. Acoust. Soc. Kr. **38**, 614-620 (2019).
8. *The JSON data interchange syntax (ECMA-404) 2nd Ed*, https://www.ecma-international.org/publications-and-standards/standards/ecma-404/, (Last Viewed July 4, 2021).
9. *World Wide Web Consortium (W3C) Recommendation Emotion Markup Language (EmotionML) 1.0*, https://www.w3.org/TR/emotionml/, (Last Viewed July 2, 2021).
10. *Alexa Developer Documentation SSML Reference*, https://developer.amazon.com/en-US/docs/alexa/custom-skills/speech-synthesis-markup-language-ssml-reference.html#amazon-emotion, (Last Viewed August 30, 2021).
11. *Microsoft Azure Speech Service*, https://docs.microsoft.com/ko-kr/azure/cognitive-services/speech-service/speech-synthesis-markup?tabs=csharp, (Last Viewed August 30, 2021).
12. *World Wide Web Consortium (W3C) Recommendation Document Object Model (DOM) 4.1*, https://www.w3.org/TR/dom41/, (Last Viewed July 4, 2021).
13. *EmotionTTS*, https://github.com/emotiontts/emotiontts_open_db, (Last Viewed July 4, 2021).

저자 약력

▸ 유 세 희 (Se-Hui Ryu)

2021년 2월: 성신여자대학교 정보시스템 공학과 학사
2021년 3월 ~ 현재: 성신여자대학교 대학원 미래융합기술공학과 석사과정

▸ 조 희 (Hee Cho)

2019년 2월 : 성신여자대학교 정보시스템 학과학사
2021년 2월 : 성신여자대학교 컴퓨터학과 석사
2021년 3월 ~ 현재 : 한국전력공사 전력연구원 일반연구원

▸ 이 주 현 (Ju-Hyun Lee)

2019년 2월 : 성신여자대학교 정보시스템 학과학사
2021년 2월 : 성신여자대학교 컴퓨터학과 석사
2021년 3월 ~ 현재 : CJ 올리브네트웍스 연구원

▸ 홍 기 형 (Ki-Hyung Hong)

1985년 2월 : 서울대학교 전자계산기공학과 학사
1987년 2월 : KAIST 전산학과 석사
1994년 2월 : KAIST 전산학과 박사
1994년 3월 ~ 1998년 2월 : 한국전자통신연구원 선임연구원
1998년 3월 ~ 현재 : 성신여자대학교 서비스디자인공학과 교수