

청각장애인을 위한 감성자막 편집기 개발

*김현순 *오주현

*한국방송공사 미디어기술연구소

*{soon71, jhoh}@kbs.co.kr

Development of an emotional subtitle editor
for the deaf and hearing impaired people

*Hyunsoon Kim, *Juhyun Oh

KBS Media Technology Research Institute

요약

방송의 디지털화에 따른 비장애인 대비 소외 계층의 정보 접근성의 부족은 소외 계층에 대한 정보 격차를 심화시킬 수 있다. 이에 캐릭터 수화 방송, 자막 방송 등 장애인을 위한 방송 서비스의 양적, 질적 개선에 관한 연구가 진행되고 있다.

자막 방송 서비스의 경우, 관련 법령에 따라 서비스를 실시하고 있으며 지상파 UHD 방송의 경우에도 본 방송을 시작한 이래 폐쇄 자막 서비스 시스템을 구축하여 서비스를 제공하고 있다. 이러한 기존 자막 서비스는 텍스트 형태의 단조로운 내용 전달 방식이어서 다양한 스타일로 풍부하게 내용을 전달하는 것에 대한 요구가 있다.

이에 본 논문에서는 지상파 UHD 방송을 대상으로 개선된 형태의 자막 서비스인 감성자막 서비스를 소개하고 이를 위한 감성 자막 편집기 기술 개발에 대하여 다룬다. 감성자막 서비스는 화자의 감정 정보를 자막 메타데이터에 추가적으로 제공하여, 감정에 따라 다양한 이모티콘이나 다른 종류의 폰트 스타일로 자막 서비스가 가능하게 하는 서비스이다. 감성자막 편집기는 이러한 감성 자막 메타데이터를 추가, 편집하고 감성자막 파일로 생성하기 위한 시스템으로, 지상파 UHD 송출 시스템 및 폐쇄 자막 표준을 고려하여 개발하였다.

1. 서론

장애인 방송 서비스에 대한 비율을 매년 확대하고자 하고 있으며, 이를 위하여 방송통신위원회에서는 2011년 ‘장애인방송 편성 및 제공 등 장애인 방송접근권 보장에 관한 고시’[1]를 제정하고 지속적인 개정을 통하여 장애인방송 대상사업자, 편성비율 및 제공기준과 방법 등에 필요한 사항을 규정하고 있다. 이에 따라 국내 지상파 UHD 방송의 경우 본 방송을 실시한 이래 표준에 맞게 폐쇄 자막 서비스 시스템을 구축하여 서비스를 제공하고 있다[2].

기존 자막 서비스는 단순한 텍스트 형태의 내용을 전달하는 방식이다. 이러한 방식에서 벗어나 다양한 스타일로 풍부하게 내용을 전달할 수 있도록 개선하고자 하는 연구가 진행되어 왔다[4]. 이러한 개선된 서비스의 일환으로 화자의 감정 등 감성 정보를 지상파 UHD 기반에서 제공하기 위한 표준화도 최근 진행하고 있다[6].

본 논문에서는 지상파 UHD 방송을 대상으로 이러한 보다 개선된 형태의 자막 서비스인 ‘지상파 UHD 기반 감성자막 서비스’를 제공하기 위한 편집기 개발 결과를 다룬다.

먼저 감성자막 서비스 및 이를 제공하기 위한 전체 시스템에 대한 개요를 기술한다. 이어 이를 제공하기 위한 편집기 개발 상세 내용 및 서비스 실시 예에 대하여 다룬 후, 향후 개선 연구 항목을 도출한다.

2. 서비스 및 시스템 개요

감성자막 서비스는 기존의 단조로운 자막 서비스에서 벗어나 화자의 감정 등 감성 정보를 자막 메타데이터에 추가하여 제공하는 서비스이다. 화자의 감정에 따라 자막의 폰트 종류, 크기, 색상 등을 다양화하여 풍부하게 정보를 전달하거나 감성 정보를 기술하는 텍스트나 이모티콘으로 제공할 수 있다. 이 외에도 발화자 구분을 위한 화자 이름 표시, 화자의 위치에 따른 자막의 위치 조정, 의성어 및 의태어 표현, 전화벨 소리 및 노래 가사 등의 장면 이해에 도움이 되는 부가적인 장면 설명 정보를 추가적으로 제공할 수 있다.

감성자막 서비스는 지상파 UHD 자막 서비스에 추가적인 정보를 제공하는 것이므로, 지상파 UHD 자막 표준인 IMSC(TTML Profiles for Internet Media Subtitles and Captions)[8]에 기반한 자막 메타데이터에 감성 정보(화자의 감정, 화자 위치, 의성어 및 의태어 정보 등)를 추가하여 전송하도록 설계되어야 한다.

감성자막 서비스도 자막의 추가적인 서비스이므로 사용자의 요구사항을 수집/분석하고, 이를 토대로 자막 시스템 개발 및 자막 입력이 일관적으로 수행되도록 하는 최소한의 가이드라인이 필요하다. 이에 자막 서비스를 일관성 있고 가독성 있게 제공하기 위한 가이드라인을 제시하고 있는 BBC, Netflix 등의 자막 가이드라인 사례를 분석하였다. 이러한

사례 분석을 기반으로 규격화된 방식으로 감성자막을 신규 삽입하기 위한 기본 기준을 설정하여 감성자막 편집기 개발의 시스템 요구 사항으로 설정하여 실험하였다.

아래는 자막 예시와 가이드라인 분석 결과 예시이다.

```
<?xml version="1.0" encoding="UTF-8" ?>
<tt xmlns:lang="kr" ttp:timeBase="clock" ttp:frameRate="1000" ttp:clockMode="utc"
xmlns="http://www.w3.org/ns/ttml" xmlns:ttp="http://www.w3.org/ns/ttml#styling"
xmlns:tp="http://www.w3.org/ns/ttml#parameter">
<head>
<metadata xmlns:ttml="http://www.w3.org/ns/ttml#metadata">
<ttml:title>TTML-Kor</ttml:title>
</metadata>
<ttp:profile use="http://www.w3.org/ns/ttml/profile/imsc1/text"/>
<styling>
<style xml:id="s1" ttp:fontSize="50px" ttp:backgroundColor="black" ttp:color="white"
ttp:lineHeight="33px" ttp:textAlign="center"/>
<style xml:id="s2" ttp:fontSize="50px" ttp:backgroundColor="black" ttp:color="yellow"
ttp:lineHeight="33px" ttp:textAlign="center"/>
<style xml:id="s3" ttp:fontSize="50px" ttp:backgroundColor="black" ttp:color="green"
ttp:lineHeight="33px" ttp:textAlign="center"/>
</styling>
<layout>
<region xml:id="bottom" ttp:origin="20% 30%" ttp:extent="50% 50%" ttp:lineHeight="33px"
ttp:displayAlign="before"/>
</layout>
</head>
<body>
<div>
<p xml:id="CC075" region="bottom" style="s1" begin="419742:57:57:000" end="419742:57:59:000">
여긴 어떻게?<br/></p>
<p xml:id="CC076" region="bottom" style="s1" begin="419742:57:57:001" end="419742:57:59:000">
GPS!<br/></p>
<p xml:id="CC077" region="bottom" style="s1" begin="419742:57:57:002" end="419742:57:59:000">
어디 있는 찾아낼 수 있단나가요.</p>
</div>
</body>
</tt>
```

그림 1. IMSC 표준 기반 지상파 UHD 자막 예

<가이드라인 분석 결과>

① 화자 구분

- 발화자가 화면에 보이지 않는 경우 이름 표기
- 발화자가 다른 사람의 말을 인용할 경우, 상황에 따라 작은따옴표('), 큰따옴표(")를 기입
- 화자는 색을 이용하여 특정할 수 있음

② 감정 표현

- 말하는 사람이 화면에 보이지 않아 표정으로 감정을 알 수 없을 경우 감정 상태를 자막으로 표현
- 글꼴, 글자 크기, 글자 색상, 이모티콘 등으로 구분 가능하도록 구현

③ 장면 설명용 의성어, 의태어 표현은 풍부하게 자세히 기입

④ 라인 수는 (특별한 경우가 아니면) 3라인을 넘지 않을 것, 2라인 권장

⑤ 자막 딜레이

- A/V와 자막과의 시간차를 50msec 이하로 유지하는 것 권장

앞서 언급한 바와 같이 감성자막은 화자의 감정 정보 데이터를 추가적으로 제공한다. 본 논문에서는 TTA 표준인 '맞춤형 방송 서비스' 표준에서 정의하는 8가지(Happiness, Anger, Excitement, Sadness, Fear, Surprise, Disgust, Neutral) 감정을 우선 대상으로 선정하여 감성자막 편집기에서 감정 정보를 입력하고 표현하도록 하였다.

감성자막 편집기는 지상파 UHD 폐쇄 자막 서비스에 기반하여 감성자막 서비스를 제공하기 위한 시스템이다. UHD 폐쇄 자막 표준인 IMSC 표준을 지원하며 기존의 자막 텍스트를 포함하여 감정을 표현하는 등의 추가 정보를 편집하는 기능을 수행한다. 기존의 지상파 UHD 방송 시스템의 경우 폐쇄 자막 인코더에서 IMSC 표준을 따르는 XML 형태의 자막 데이터를 송출하며, 감성자막 편집기의 경우 기존 폐쇄 자막

인코더로 추가적인 감정 정보를 포함하는 감성자막 데이터를 송출하는 인터페이스를 제공한다.

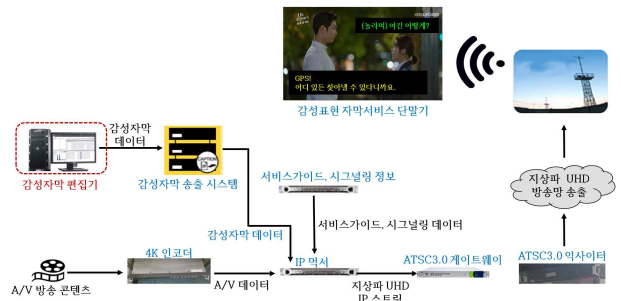


그림 2. UHD 지상파 방송 기반 감정표현 자막 서비스 개념도

3. 감성자막 편집기 개발

감성자막 편집기는 영상과 자막 파일을 입력받아 화자의 감정 등 감정 정보를 입력 및 편집하는 편집기 모듈과, 서비스 전에 동영상과 함께 동기화하여 재생해 보는 재생기 모듈을 제공한다.

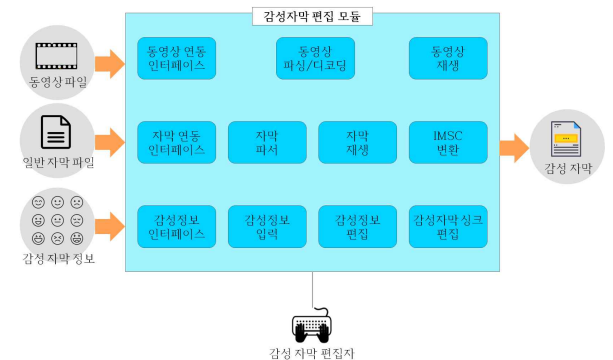


그림 3. 감성자막 편집 모듈 기능도

감성자막 편집기 모듈은 지상파 UHD 자막 표준인 IMSC 표준을 지원하며, 기존의 자막 텍스트를 포함하여 감정 표현 정보를 편집하는 기능을 수행한다. 감성자막 편집 모듈에서는 동영상, 자막 파일을 입력받고 시스템 사용자인 감성자막 편집자로부터 감정 정보, 화자 정보, 감성스타일, 자막 위치, 자막 표현 시간 정보 등을 입력 받아 감성자막 파일을 생성, 출력, 저장하는 기능을 제공한다. 자막 정보, 동영상 파일, 오디오 파일을 파싱하고 시스템 사용자가 직관적으로 편집할 수 있도록 UI(User Interface)를 제공한다. 감성자막 정보(감정 정보, 시간, 스타일, 화자 정보) 표시, 자막 표시 시간대의 영상 표시, 오디오 파형 표시, 표시 시간에 따른 자막 정보 표시 기능을 제공한다.

자막 편집의 기본은 자막 텍스트와 자막 표시 시간의 편집이다. 여기에 감정 정보를 포함하여 감성자막을 편집한다. 자막 기본 정보인 자막 텍스트와 표시 시간에 대한 편집이 가능하고, 감성자막 정보인 자막 위치, 화자, 감정 표현 이미지에 대한 설정이 가능하다. 감성자막 편집기 내에서 재생 영상 및 음성 파형에 맞게 자막 동기화를 할 수 있다. 자막을 편집할 때마다 자주 사용하는 스타일 등에 대해 미리 등록하고 이를 통해 조금 더 손쉽게 설정할 수 있다. 감성자막 편집기 화면 구성은 자막 정보 영역, 비디오 영역, 오디오 파형 영역으로 구분된다.

감성자막에서 사용하는 감성 정보는 감성 스타일, 화자, 자막 표시 위치, 이모티콘 등 감성 표현 이미지이다. 각각의 감성 정보에 대해 별도의 감성 정보 등록이 가능하다. 감정에 따른 폰트 스타일 설정으로 화자 및 상황에 따른 스타일을 설정할 수 있다.

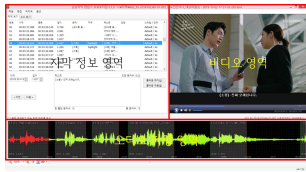


그림 4. 편집기 메인 UI

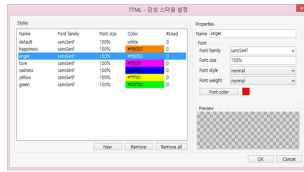


그림 5. 자막 스타일 설정 기능

```

<styling>
  <style xml:id="default" ts:font="com.microsoft.sans" ts:fontWeight="normal" ts:fontSize="80%" ts:fontFamily="sansSerif" ts:color="white" />
  <style xml:id="화자구분(색상)" ts:font="com.microsoft.sans" ts:fontWeight="normal" ts:fontSize="80%" ts:fontFamily="sansSerif" ts:color="#FF0000" ts:fontSize="120%" />
  <style xml:id="감정표현" ts:font="com.microsoft.sans" ts:fontWeight="normal" ts:fontSize="80%" ts:fontFamily="sansSerif" ts:color="#FF0000" ts:fontSize="120%" />
  <style xml:id="화자구분(색상)" ts:font="com.microsoft.sans" ts:fontWeight="normal" ts:fontSize="80%" ts:fontFamily="sansSerif" ts:color="#000000" ts:fontSize="120%" />
</styling>
<layout>
  <region ts:extent="80% 40%" ts:origin="10% 10%" ts:displayAlign="center" ts:textAlign="center" ts:fontWeight="bold" ts:fontSize="120%" />
  <region ts:extent="80% 40%" ts:origin="10% 10%" ts:displayAlign="center" ts:textAlign="center" ts:fontWeight="bold" ts:fontSize="120%" />
</layout>
</head>
<body style="default">
  <div xml:lang="ko">
    <p begin="20.103" xml:id="p0" end="20.85">[화자구분(색상)] 심만! />
    <p begin="53.085" xml:id="p12" end="55.97" style="color: green">[화자구분(색상)] 아</p>
    <p begin="45.165" xml:id="p95" end="45.667" style="color: red">[감정표현] [화자구분(색상)] 아</p>
    <p begin="47.5.103" xml:id="p101" end="47.667" style="color: red">[감정표현] [화자구분(색상)] 아</p>
    <p begin="47.7.222" xml:id="p102" end="47.779" style="color: red">[감정표현] [화자구분(색상)] 아</p>
    <p begin="482.498" xml:id="p104" end="483.238" style="color: red">[감정표현] [화자구분(색상)] 아</p>
    <p begin="210.413" xml:id="p43" end="212.04" style="color: red">[화자구분(색상)] 아</p>
    <p begin="222.188" xml:id="p48" end="225.078" style="color: red">[화자구분(색상)] 아</p>
    <p begin="236.81" xml:id="p53" end="238.857" style="color: red">[화자구분(색상)] 아</p>
    <p begin="287.166" xml:id="p58" end="293.682" style="color: red">[화자구분(색상)] 아</p>
  </div>

```

그림 6. 감성자막 파일 예

아래 그림은 시간에 따라 자막 폰트의 크기를 다르게 설정하여 감정을 표현한 경우와, 화자를 구분하기 위해 화자의 이름을 표시하고 화자의 위치에 맞추어 자막의 위치를 다르게 배치하는 경우의 예이다.

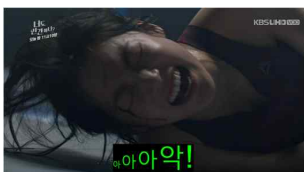


그림 7. 감정 표현 예



그림 8. 화자 구분 예

4. 결론

본 논문에서는 단순한 텍스트 위주의 폐쇄 자막 서비스를 개선하기 위해, 화자의 감정 등에 따라 풍부하게 자막으로 표현할 수 있는 지상파 UHD 기반 감성자막 서비스를 소개하고, 이를 제공하기 위한 전체 서비스 시스템 구성에 대하여 기술하였다. 특히 감성자막을 제공하기 위한 핵심 시스템인 감성자막 편집기 기술에 대한 세부 개발 결과를 제시하였다.

본 논문에서 제시한 개발 결과로 감성자막을 편집, 생성하는 기본 기능은 구축되었으나, 이를 기존 UHD 송출 시스템으로 송출하기 위한 인터페이스 개발에 대한 연구를 지속적으로 수행할 예정이다. 또한 화자의 감정 정보를 자동으로 추출할 수 있는 AI 시스템에 대한 연구가 필요

하며, 이를 통하여 일부 자동화가 실시되면 서비스를 위한 시간과 노력을 감축할 수 있을 것이다.

Acknowledgement

본 연구 논문은 과학기술정보통신부 및 정보통신기획평가원의 출연금으로 수행하고 있는 한국전자통신연구원 시청각 장애인의 방송시청을 지원하는 감성표현 서비스 개발(2019-0-00447)의 연구결과입니다.

참고문헌

- [1] “장애인방송 편성 및 제공 등 장애인방송접근권 보장에 관한 고시,” 방송통신위원회고시, 제2011-53호.
- [2] Yunhyoung Kim, “Implementation of Closed Captioning System for Terrestrial UHD based on ATSC 3.0,” SMPTE 2017 Annual Technical Conference & Exhibition.
- [3] TTA, “지상파 UHDTV 방송 송수신 정합,” TTA.KO-07.0127/R4, 2019.
- [4] 양승준, 안충현, “청각장애인을 위한 동적인 감성 자막에 관한 연구,” 한국통신학회 추계종합학술발표회, pp. 85-86, 2014.
- [5] 장인선, 안충현, 차지훈, “장애인방송 서비스 및 기술동향,” 전자통신동향분석, 2015, pp. 21-30.
- [6] TTA, “맞춤형 방송 서비스,” TTA.KO-07.0093/R2, 2018.
- [7] TTA, “시청각 장애 보조 방송 서비스,” TTA.KO-07.0093/R2, 2018.
- [8] “TTML Profiles for Internet Media Subtitles and Captions 1.0 (IMSC1),” W3C, <http://www.w3.org/TR/ttml-ims1.0.1/>
- [9] BBC, “Subtitle Guidelines,” <https://bbc.github.io/subtitle-guidelines/>
- [10] Netflix, “Timed Text Style Guide,” <https://partnerhelp.netflixstudios.com/hc/en-us/articles/215758617-Timed-Text-Style-Guide-General-Requirements>