

장애인방송 기술개발 현황

Status of Development of Broadcasting Technology for the Disabled

안충현 (C.H. Ahn, hyun@etri.re.kr)

실감AV연구그룹 책임연구원

ABSTRACT

In the digital era, broadcasting plays a very large role as a means of communication, as it no longer merely provides information or entertainment media. Particularly in the case of the visually and hearing-impaired, broadcasting is the primary means of acquiring information, so its role as a public service needs to be expanded. The development of digital technology enables the quantitative expansion of traditional methods of disability broadcasting, i.e., closed caption, sign language, and descriptive video service, with an acceptable level of quality. In this study, we review the current trends of, and concerns related to, broadcasting services for disabled people, as well as technological trends, and proposals for increasing visibility and accessibility to broadcasting for hearing-impaired people.

KEYWORDS Barrier-free, Closed Caption Service, Descriptive Video Service

1. 서론

방송은 단순히 정보제공과 오락 매체의 기능을 뛰어넘어 커뮤니케이션 수단으로서 역할을 수행하는 매체로 점차 변화하고 있으며, 디지털 시대에 있어 커뮤니케이션 수단으로서의 중요성은 점차 커지고 있다. 그렇지만 시장 주도적인 논리가 우선시되면서 방송의 공공서비스 역할이 위축되어 커뮤니케이션 장의 역할에 한계를 보이고 있다. 이러한 공적 역할의 축소는 신체적인 장애로 인하여 매체 이용

에 제한을 받는 시·청각장애인의 고립과 소외를 더욱 심화시키는 매체로 작용할 수도 있다[1].

방송통신융합이 진행되면서 미디어 생태계는 생활패턴, 취향 등에 발맞추어 끊임없이 진화 중이다. 더는 방송사가 제공하는 편성 시간표에 의존하는 이른바 본방사수에서 벗어나 주말이나 쉬는 날 몰아서 선호하는 콘텐츠를 보는 새로운 소비행태가 등장하고 있는데, 이를 빈지 뷰잉(Vinge Viewing: 좋아하는 방송을 몰아 보는 새로운 시청 형태)이라 한다. 이러한 소비 행태는 스마트폰과 인터넷

* DOI: 10.22648/ETRI.2019.J.340301

* 본 연구는 미래창조과학부 및 정보통신기술연구진흥센터의 정보통신·방송 연구개발사업의 일환으로 수행하였음[B0192-16-1001, 시·청각장애인 방송접근권 향상을 위한 디지털자막·음성해설 서비스 기술 개발].

기술의 발달로 확대되고 있으며, 서비스 플랫폼도 Video on Demand(VOD), Over the Top(OTT) 등 다양화되고 있다. 그렇지만 대부분의 콘텐츠가 시·청각장애인이 소비할 수 있는 적절한 수단(화면해설, 자막, 수화)으로 제공되지 않기 때문에 매체의 다양화에도 불구하고 시·청각장애인이 접근하기에는 많은 어려움이 있다. 일례로, 2017년 유료 방송 플랫폼별 VOD 보유 내역을 분석한 결과, 유료방송 사업자가 장애인용 VOD를 보유하고 있는 비율이 0.0011% 미만으로 장애인의 방송접근권에 대한 외면이 심각한 수준인 것으로 파악되고 있다[2]. 인터넷 기반의 온라인 동영상 서비스 확대로 방송과 온라인 동영상 간의 경계가 점차 희미해지고 있어[3], 현재 방송 위주의 장애인방송 편성 기준을 VOD나 OTT 서비스에도 확장하여 적용함으로써 공익성을 해치지 않도록 최소한의 장치가 필요하다. 참고로 미국 Federal Communications Commission(FCC)은 2012년 8월 온라인 동영상 및 인터넷 TV 서비스로 제공되는 동영상 콘텐츠에 대해서도 청각장애인을 위한 자막 삽입을 의무화한 바 있다[3]. 지난 2011년 ‘장애인방송 편성 및 제공 등 장애인 방송접근권 보장에 관한 고시’[4]의 제정으로 국내 장애인방송 서비스가 공식적으로 시작되어 비교적 양호하게 제도가 정착되고 있다. 현재 장애인방송 고시에 의한 장애인방송 의무편성 비율에 따르면 지상파의 경우 전체 프로그램 중 자막 방송 100%, 화면해설방송 10%, 수화방송 5%만 편성하면 되고, 종편이나 보도채널은 화면해설방송 8%, 수화방송 4%만 편성하면 된다. 장애인방송 의무화를 통해 시·청각장애인의 방송접근권이 크게 개선되었고, 그 결과 소외계층의 정보 접근과 소통, 문화 향유에 있어 상당한 기여를 하였다. 이를 더 확대하기 위해서는 의무편성 비율에서 수화와 화면해설방송 비중의 상향조정이 필요하다. 자막

방송에 비해 수화와 화면해설방송은 의무편성 비율 자체가 미미한 수준이며, 그나마 의무편성 비율만 겨우 넘기는 경우가 많아 의무편성 비율의 상향 조정에 대해서도 정책적인 검토와 이해당사자 간의 합의가 필요하다. 장애인방송 시청 행태에 대한 최근 조사[5]에 의하면, 시각장애인의 채널별 화면해설방송 만족도는 5점 척도를 기준으로 할 때 대부분 3점에 못 미치는 수준으로 나타났으며, 편성 프로그램 수가 적다는 것과 프로그램 다양성이 부족하다는 것이 주요 불만 사항으로 파악되고 있다. 청각장애인은 자막 및 수화통역방송 편성 프로그램 수가 부족한 것을 방송 이용에 있어 가장 큰 어려움으로 꼽았으며, 편성 시간대, 프로그램 수, 자막이나 수화의 빠른 속도 등을 또 다른 이유로 들고 있음을 알 수 있다. 양적 확대와 더불어 일부 장애인방송에서 나타나는 사례(자막방송에서 오타나 띄어쓰기 무시, 정보 누락 등의 현상과 수준이 미달하는 수화 통역사를 활용, 화면과 불일치하는 화면해설방송 등)를 방지하기 위해서 개별 프로그램과 채널 내에서의 편성에 대한 질적 평가 기준을 마련하여 시·청각장애인의 실질적인 방송접근권을 보장할 방안을 고민해야 할 시기를 맞고 있다. 2012년 이후 영국의 규제기관인 OfCom은 라이브 자막방송의 질적 서비스를 평가하는 방법을 제시한 바 있다. 여기에서는 라이브 방송의 자막서비스에 대한 속도, 정확도, 지연시간, 오류의 종류에 대한 분석을 다루고 있다[6].

한편, 딥러닝 기술의 발전으로 시·청각장애인의 일상생활을 돕는 많은 서비스가 등장하고 있다. 방송산업에서는 ‘제작-유통-시청자 대응’과 같은 영역에서 Artificial Intelligence(AI)를 활용하는 시도를 하고 있으며, 콘텐츠 제작을 돕는 AI부터 개인 맞춤형 방송 콘텐츠를 추천하는 큐레이션 서비스에 이르기까지 AI 기반의 다양한 서비스가 진화를 거

듭하며 방송산업의 새로운 가능성을 열어가고 있다[7].

대표적으로는 음성인식 기술을 이용하여 방송 자막을 생성하는 서비스를 들 수 있으며, 입 모양을 보고 텍스트를 생성하는 립리딩(Lip reading) [8], 음성정보를 입 모양의 CG와 동기화하는 기술 [9]이 발표되고 있다. 이들 기술은 딥러닝을 이용한 음성합성 기술과 결합하여 훨씬 적은 비용으로 빠른 콘텐츠를 제작하거나 많은 음성서비스에 활용될 가능성이 높다. 또한, 동영상에 기반한 감정 인식 결과는 자막, 수화, 음성합성에 적절한 표현을 부여하는 정보로 사용되어 시·청각장애인이 내용을 입체적으로 파악하는 데 도움이 될 것으로 판단된다.

한편, 초고화질 방송인 Ultra High Definition(UHD) 방송서비스는 High-Definition Television(DTV) 이후의 차세대 주요 방송서비스로 대두되고 있으나, 기존 장애인방송 표준을 그대로 적용할 수 없기 때문에 방송 미디어 환경변화에 따른 장애인 방송 개선방안에 대한 보완이 필요하다 [10].

여기에서는 자막방송, 수화방송, 화면해설방송으로 대표되는 장애인방송 현황, 장애인방송기술 개발현황에 대해 서술한다.

II. 장애인방송 현황

방송통신위원회에서는 ‘장애인방송 편성 및 제공 등 장애인 방송접근권 보장에 관한 고시’ 제10조 제3항 제5호에 따라 매년 장애인방송 고시 의무사업자 지정을 위한 ‘장애인방송물 평균 제작비 산정기준’을 공고하고 있다. 또한, 정부에서는 시·청각장애인의 방송접근권을 보장, 장애인 등 방송소외계층의 방송을 통한 정보·지식 획득 기회를 제

공, 장애인과 비장애인 간의 방송접근 격차를 해소, 비장애인과 동등하게 방송을 시청할 수 있도록 매칭 펀드의 형태로 방송사업자의 자막·수화·화면해설방송 제작 지원 사업을 시행하고 있다. 이와 함께 매년 전년도 장애인방송 편성 평가를 실시하여 목표를 달성하지 못한 방송사업자에 대해서는 행정지도를 실시하고, 차년도 장애인방송 제작 지원 시 의무 편성 미달률을 반영하여 정부 제작지원금을 차등 지원하는 정책을 펴고 있다.

III. 기술개발 현황

1. 자막방송 서비스

자막은 방송 프로그램이나 영화 속에서 등장인물의 대화를 글로 표현한 것을 말하며, 보통 외국어를 우리말로 번역하거나 청각장애인이 방송내용을 이해할 수 있도록 하기 위해 존재한다. 불과 몇 년 전까지만 해도 방송 프로그램의 내용을 단순하게 문자의 형태로 전달하는 것이 주 기능이었지만, 이제 자막은 보조자의 역할뿐만 아니라 출연자, 시청자, 제작진의 입장을 모두 대변하는 ‘제3의 역할’을 해내고 있어 단순한 보조적인 역할을 벗어나 주요 내용표현방식을 대체하고 있다[11]. 등장인물의 대사 외에도 어떠한 상황이 발생하면 제작진들은 출연자가 그 상황에서 느꼈을 법한 감정을 토대로 자막 처리를 해 출연자의 감정을 극대화한다. 일반적으로 소리 효과와 함께 느낌표나 물음표, 또는 ‘당황’, ‘두려움’ 등의 감정 표현으로 효과를 주거나, 시청자들이 공감할 만한 내용에 대해 글자의 크기나 색, 형태를 감정에 따라 조절함으로써 동적인 감정을 느끼면서 프로그램을 즐길 수 있다. 이러한 자막은 개방형 자막(Open caption)의 형태로 제공되고 있어, 청각장애 유무와 관계없이 누구나 볼 수 있으며, 과도하지 않은 경우 방송내용을 쉽

게 이해할 수 있도록 도와준다.

기존 HDTV에서 CEA-708 규격을 사용하는 폐쇄 자막(Closed caption)에서는 텍스트만 가능하기 때문에 개방형 자막과 같은 다양한 효과를 제공할 수 없으며, 청각장애인이 자막방송에 대해 갖는 불만사항(자막 위치, 크기 등)을 해결할 수 없다. 텍스트에 다양한 효과를 주어서 청각장애인의 방송 시청을 지원하고자 하는 노력이 시도되어 왔으나[12-14], 연구 단계를 벗어나지 못하였다. 기존 CEA-708 규격에 따라 자막을 영상의 부가정보로써 함께 인코딩하여 전송하는 방식과 달리 마크업 방식의 자막 방식은 마크업 언어인 Timed Text Markup Language(TTML) 등을 기반으로 하여 텍스트의 형태로 제작하여 전달하는 방식이며, 확장 가능한 규격이다. 최근 ATSC 3.0의 등장은 자막을 비디오 패킷에 삽입하여 전달하는 기존 방식 외에 IP망을 전송로로 사용할 수 있는 규격이 추가되었으며, 현재 수도권에서 실시되고 있는 4K 본방송에서는 이 전송로를 이용하여 자막을 제공하고 있다. ATSC 3.0에서 자막규격으로 채용하고 있는 IMSC1을 활용하여 청각장애인의 자막 가독성 및 내용이해도를 향상시키기 위해 텍스트 이외의 감정적인 표현을 가능하게 하는 이미지나 다른 애니메이션 등을 통해 보다 자막의 표현을 풍부하게 하는 서비스에 대해 검토할 필요가 있다.

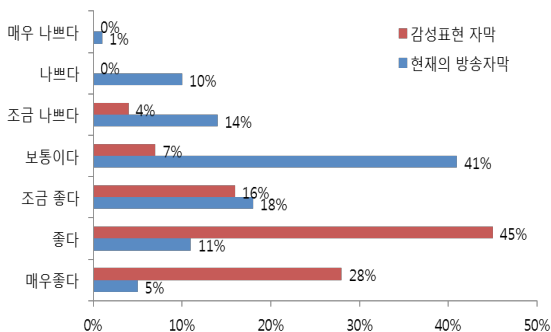


그림 1 CEA-708 자막과 개량형 자막의 비교

그림 1은 한국전자통신연구원에서 한국농아인 협회와 함께 2014년 수도권 9개 복지관의 청각장애인 100명을 대상으로 기존방송자막과 감정적인 표현을 추가한 개량형 자막에 대한 평가 결과를 보여 준다. ‘좋다/매우 좋다’의 비율은 기존 자막(16%)에 비해 개량형 자막(73%)으로 전체적으로 기존자막보다 감정적인 표현이 추가된 개량형 자막에 대한 선호를 알 수 있다.

그림 2는 평가에 사용된 개량형 자막의 일례를 보여준다.

개량형 자막은 속기에 의해 실시간으로 제공되는 폐쇄자막과는 달리 자막에 감정적인 표현을 추가하기 위해서는 제작에 드는 시간이 있기 때문에, 사전제작 또는 본 방송 후 기존 자막을 활용한 추가 제작을 통해 재방송이나 VOD와 같은 인터넷 기반의 동영상 서비스를 통해 청각장애인에게 제공될 필요가 있다. 이러한 서비스를 위해서는 화면 내 등장인물에 대한 이해 및 식별, 감정에 따른 자막(폰트, 크기, 색), 동적인 자막의 표현, 이모티콘을 통한 상황의 전달 등 청각장애인을 대상으로 한 의견수렴을 바탕으로 한 자막 표현에 대한 추가적인 개발이 요구된다. 이러한 개량형 자막은 비디오 화면과는 별도의 전송로를 통해 전달되며, 텍스트 파일로 제작되기 때문에 시청자의 기호에 맞추어 표현을 달리 할 수도 있다. 해외에 수출하는 경우 자막 표현은 변경하지 않고 해당 국가의 언어로



그림 2 청각장애인 평가에 사용된 개량형 자막

만 번역해서 제공 가능하기 때문에 영상의 재편집이나 자막을 다시 제작하는 데 소요되는 비용도 절감할 수 있을 것으로 기대된다.

2. 수화방송 서비스

수화방송은 청각장애인이나 언어장애인을 위해서 방송 화면과 함께 수화 통역을 통해 방송 내용 또는 대화를 소개해 주는 서비스이다. 일반적으로 수화 화면은 주화면의 오른쪽 하단에 전체 화면의 1/16에 불과한 크기로 청색이나 녹색 배경을 갖는 원 안에 수화 통역사를 표출하고 있다. 수화방송은 비디오 화면에 수화 화면이 함께 인코딩되어 송출되기 때문에 비장애인의 경우에는 수화 화면이 주화면을 가리는 것에 대한 불만이 있다. 또한, 청각 장애인도 여러 이유에 의해 수화방송에 대한 불만을 갖고 있다. 수화방송에 대한 청각장애인의 만족도는 매우 낮은 편으로 수화방송의 이해도를 조사한 결과 응답자의 79%는 절반 이하밖에 이해하지 못한다고 하며, 82%가 만족하지 못한다는 의견을 보였는데, 주된 요인은 화면의 크기, 수화의 속도, 통역사의 표현방식과 잘못된 표현 등이 지적되었다[15]. 최근 방송통신위원회에서는 현재 제공되고 있는 수화방송의 불편함을 개선하여 방송 영상과 수화방송 영상을 각각 방송망과 인터넷망으로 송출하되, TV에서 이를 합성하여 화면에 동시에 표시함으로써 수화 화면의 위치나 크기를 조절할 수 있는 ‘스마트 수화방송 서비스’를 계획하고 있다[16].

현재 수화방송의 편성 기준은 5%로 뉴스나 시사 프로그램에 집중되어 제공되고 있다. 수화방송의 문제점에 있어 화면의 크기 외에 수화 방식의 차이, 수화 통역사의 수화 동작 오류, 수화 속도 등이 제시되고 있다. 또한 한정된 수화 통역사

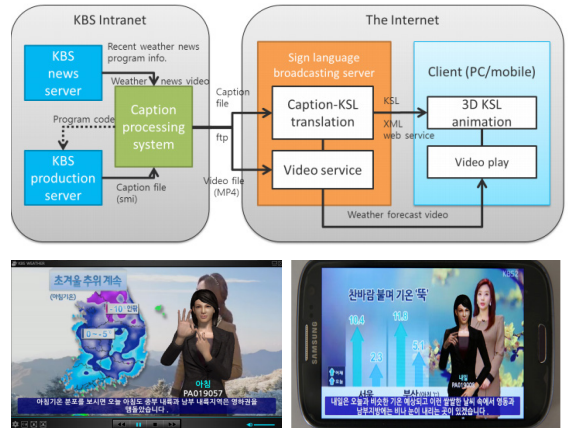


그림 3 KBS의 일기예보 수화서비스 시스템 및 서비스 화면

출처 Juhyun Oh et al., "An Avatar-Based Weather Forecast Sign Language System for the Hearing-Impaired," 10th IFIP International Conference on Artificial Intelligence Applications and Innovations (AIAI), 2014.

에 의존해서는 양적 확대도 어려운 형편이다. 이를 극복하기 위해서는 정확한 정보의 전달이 요구되는 분야가 아닌 경우에는 발전하고 있는 그래픽기술을 이용하여 수화방송을 확대할 수 있는 방안을 고려할 필요가 있다. KBS에서는 2014년 방송 자막을 기반으로 그래픽 캐릭터를 이용한 수화방송시스템을 개발하여 일기예보에 적용하여 실시한 바 있다(그림 3 참조)[17]. 전체적인 번역률은 65% 정도로[18], 최상배의 연구[19]에서 조사된 뉴스와 시사보도 분야의 9개 방송 프로그램에서 방영된 수화 통역의 통역률(평균 80%)에 비하면 낮은 수준을 보인다. 기계번역에 의해 자동생성하는 수화방송 서비스의 번역률은 최소한 수화 통역사의 번역률 정도는 되어야 할 것으로 생각한다.

또한, 그래픽 기반의 수화 표현에 대한 청각장애인의 만족도를 향상시키기 위해서는 단순한 손(몸) 동작 외에 내용의 이해를 도울 수 있는 표정 전달, 알아보기 쉽고 분명한 동작, 방송내용의 정확한 전달이 요구된다[20]. 따라서 새롭게 시도될 그래픽

기반의 수화 생성 시스템은 손동작 외에도 표정, 입술의 움직임을 통하여 내용 전달과 함께 감정 전달도 함께할 수 있도록 데이터가 구성되어야 할 것이다. 이를 위해서는 한국어에 대응하는 수화 말뭉치에 대해 동작, 표정, 입술의 움직임 등을 디지털화하여 데이터베이스를 구축할 필요가 있다. 아울러 수화 말뭉치의 디지털화는 시간과 비용이 소요되기 때문에 꾸준히 진행되어야 한다.

따라서, 향후 디지털화를 수행하는 주체와 관계 없이 수화 동작의 통일성을 유지하고 양적 확대 및 많은 분야에의 적용을 확장해 나가기 위해서는 수화 동작의 디지털화에 대한 표준화가 필요하다. 특정 분야에 대한 한정되는 서비스가 아닌 대국민 공공서비스, 실생활 전반적인 분야로의 확대, 딥러닝 기반의 자동 수화 인식 등의 장애와 관계없는 소통 플랫폼(Barrier-Free Communication Platform)을 위해서는 국가적 차원에서 한국어 수화 말뭉치에 대한 디지털 데이터를 꾸준히 확장시켜 나가야 한다.

3. 화면해설방송

화면해설방송은 대사, 인물의 움직임, 소리 등 화면에 나타나는 정보를 바탕으로 시각장애인이 해당 영상을 이해할 수 있도록 화면해설 작가가 대본을 쓴 뒤 성우가 녹음하고 이를 오디오 엔지니어가 메인 오디오에 믹싱하여 보조 오디오로 제공하는 서비스를 말한다. 화면해설방송은 화면해설 작가의 작가적 능력과 시각장애에 대한 이해 정도가 중요하며 단순히 자막을 읽는 식으로는 제작되기 어렵다. 장인선의 연구[20]는 화면해설방송 콘텐츠를 제작하는 데 시간적, 경제적 요소를 줄이고, 콘텐츠의 양적 확대를 위한 방법의 하나로 소프트웨어를 이용한 방안을 제시하고 있다(그림 4

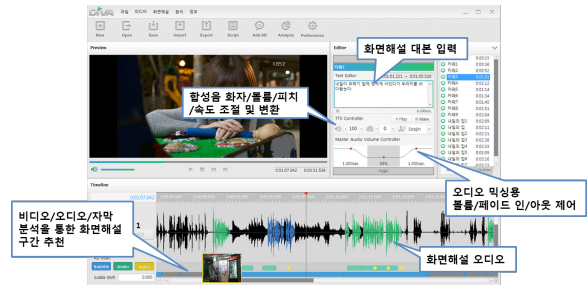


그림 4 화면해설 콘텐츠 저작도구 실행 화면

참조).

한편, 지상파를 비롯한 많은 매체에서 해외 영화를 시청할 수 있으나 대부분 시각장애인이 감상할 수 없는 자막을 제공하고 있는 경우가 많다. 외국 작품을 보는 데 있어서 자막은 본 작품을 느낄 수 있는 장점이 있지만, 외국어에 익숙하지 않은 이들에게는 몰입도가 떨어지는 단점이 있다. 한편, 자막을 대체할 수 있는 더빙은 본 작품과 다른 구석이 있지만, 시청 집중도는 더 높일 수 있다. 시각장애인의 입장에서 외화 더빙은 단순한 편의 영화를 감상하는 데 그치는 것이 아니라 시각장애인의 문화 향유권을 담보하는 중요한 수단으로, 외국의 우수한 외화들을 도움 없이 감상할 수 있는 거의 유일한 수단이다.

이를 극복하는 방법으로는 대부분의 프로그램에서 자막을 제공하고 있어, Text to Speech(TTS) 기술을 이용하여 자막을 음성 합성을 통해 오디오로 제공하는 이른바 음성 자막(Spoken Subtitle)을 고려해 볼 수 있다. 다만, 사용할 수 있는 음성이 제한되어 있고, 대부분이 낭독체이기 때문에 감정을 주고받는 대화체 위주의 방송에 적용하기 위해서는 감정표현이나 인물의 특성을 어느 정도 반영하는 진화된 음성합성기술이 필요하다. 최근 구글과 같은 글로벌 기업을 필두로 대용량 음성데이터에 바탕을 둔 딥러닝 기반 음성합성기술이 비약적으로 발전하고 있다.

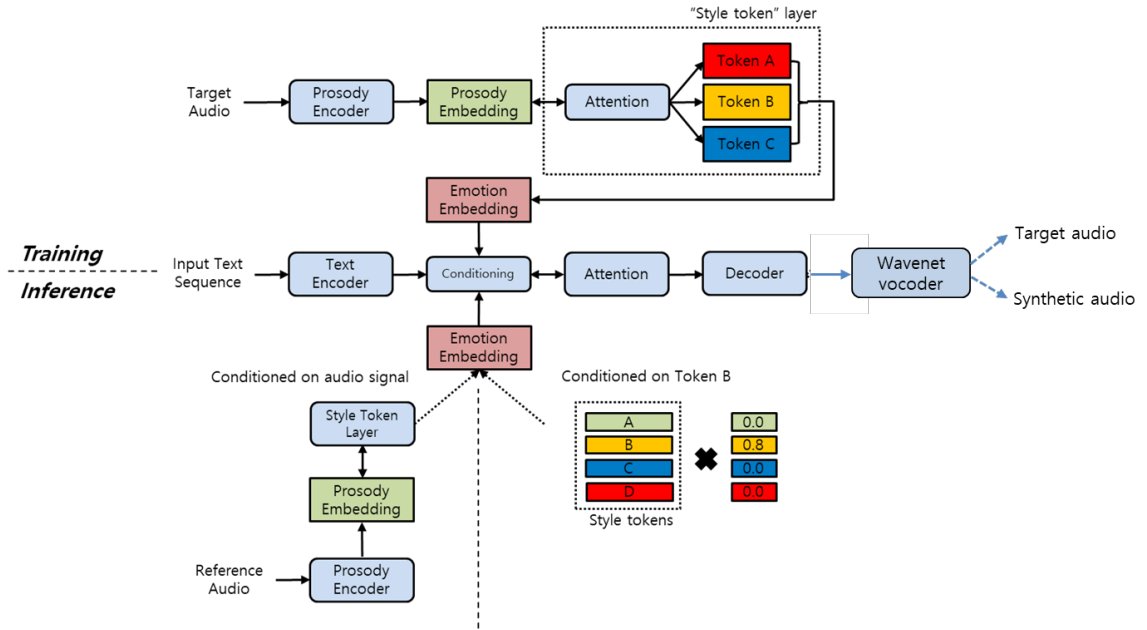


그림 5 한국전자통신연구원의 감정음성합성 시스템 구조

구글의 Tacotron은 Sequence-to-Sequence 모델에 기반을 두고 있으며 가장 자연스러운 음성을 표현한다. 네이버에서는 음성 데이터양을 1/10로 줄여 적은 음성 데이터만으로 음성 합성이 가능하고 웨이브넷 보코더를 결합시켜 음질을 개선한 하이브리드 음성합성엔진을 발표하고 있다[21]. 이러한 음성합성기술의 발전은 목소리에 감정을 추가하여 표현할 수 있는 중단 간 감정음성합성시스템에 대한 연구로 이어지고 있다[22,23].

한국전자통신연구원에서는 Tacotron-2와 Global Style Token을 이용하여 감정표현을 가능하게 감정 음성 합성기술에 대해 개발을 진행하고 있으며, 향후 딥러닝을 이용한 감정음성 합성기술과 콘텐츠 저작 도구의 연결로 전문 화면해설작가가 작성한 대본이나 방송 자막을 기반으로 마치 성우가 녹음한 것과 같은 품질을 갖는 화면해설방송이나 음성 자막을 제작, 서비스할 수 있는 날이 머지않은 것으로 판단된다(그림 5 참조).

4. AI 활용 콘텐츠 제작

현재의 장애인방송에 있어 편성 목표만 달성하고자 하는 수동적인 자세로는 시·청각장애인의 만족도를 높이기 어렵다. 콘텐츠의 제작에 있어서도 감성적인 측면을 고려하는 것이 질적 만족도를 향상시킬 수 있을 것이다.

또한, 최근 디지털콘텐츠의 폭발적 증가는 효과적인 자료 검색을 위해 특정 자료의 내용 및 성격 등에 대한 메타데이터 구축이 필수적이다. 기존에는 직접 영상을 살펴보면서 메타데이터를 입력하였지만, 최근에는 딥러닝을 이용하여 영상/오디오/자막을 분석하고, 등장인물의 표정 등을 식별하여 자동으로 메타데이터를 구축하는 서비스도 등장하고 있다. 검색을 위한 메타데이터의 구축과 더불어 콘텐츠에 감성적인 효과를 반영함으로써 질적 만족도를 높이기 위해서는 감성 또는 감정과 관련된 정보를 자동으로 추출하고(표 1 참조), 이를 서비스 대상 콘텐츠에 적절하게 표현해 줌으로

표 1 미디어에서 추출할 수 있는 감정 관련 정보

미디어	대상	추출 가능한 정보	
영상	표정	감정(화남/기쁨/슬픔/중립 등)	
	자막	화면에서의 자막 표현 위치(검출된 얼굴 위치에 상대적 위치), 감정상태에 따른 자막 표현(폰트, 크기, 색 등)	
	화자	성별(남/녀), 연령대, 배우명(실명 또는 배역명)	
오디오	음성	음성	감정(화남/기쁨/슬픔/중립 등)
		자막	음성인식을 통한 자막 텍스트 생성, 자막 시작/종료시간, 자막/오디오 동기화, sentiment
		화자	성별(남/녀), 연령대, 배우명(실명 또는 배역명)
	비음성	이벤트	노크, 전화벨, 천둥소리 등

써 앞서 언급한 개량형 자막, 수화, 음성 자막에서 등장인물의 감정과 내용을 효과적으로 시·청각장애인에게 전달할 수 있다.

영상에서 특정 인물을 찾거나 표정 분석을 통해 감정을 인식하는 서비스로 마이크로소프트사의 ‘Cognitive Service’가 알려져 있다. 오디오 분야에서는 복수의 등장인물이 등장하는 오디오에서 화자 구분, 음성 영역 추출을 위해 딥러닝을 이용한 ‘Speaker Diarization’에 대한 연구가 이루어지고 있다.

오디오로부터 추출되는 정보는 음성인식에 의해 생성되는 자막과 함께 내용에 따라 폰트의 크기나 색 등을 다르게 가져가며, 영상에서는 등장인물의 구분과 위치, 표정을 인식하고 인물을 가리지 않는 부분에 자막을 표현할 수 있도록 자막 위치를 정의할 수 있다. 이는 영상과 오디오로부터 등장인물들의 대화, 캐릭터 특징, 감정과 관련된 정보를 추출하고 이를 시·청각장애인을 대상으로 하는 콘텐츠 생성에 활용함으로써 상황에 대한 이해를 쉽게 전달할 수 있다.

5. 장애인 방송 모니터링 시스템 기술

가. 모니터링 시스템

장애인방송 고시 이후 방통위에서는 매년 전년의 장애인방송 편성 실적에 대한 평가를 시행하고 있다. 이러한 평가를 상시로 시행하기 위해서는 모니터링 시스템의 구성이 요구된다. 이를 위해 한국전자통신연구원에서는 방송 신호 수신 Set Top Box(STB)를 개량하여 장애인방송 수신 여부를 확인하고, 장애인방송인 경우 해당 서비스에 적합한 분석을 수행하고, 통계적인 데이터를 산출하는 방송 모니터링 시스템을 개발하였다(그림 6 참조). 이 시스템은 단순히 방송 신호부터 장애인방송에 대한 정보를 체크하여 방송 여부를 판단하는 것이 아니라, 장애인방송에 대해서는 실제 콘텐츠 분석까지 실시하는 것이 특징이다[24].

현재의 장애인방송 고시에 따르면 편성 비율만 규정하고 있을 뿐 내용의 정확도와 같은 질적인 부분에 대해서는 내용이 없기 때문에 방송 화면과 맞지 않는 자막이 송출되는 경우도 발생한다. 따라서, 실제 방송된 콘텐츠에 대한 분석을 통해 양질의 서비스를 보장할 수 있는 방법을 모색할 필요가 있다.

나. 화면해설방송 분석

일반적으로 화면해설 오디오는 그림 7과 같이

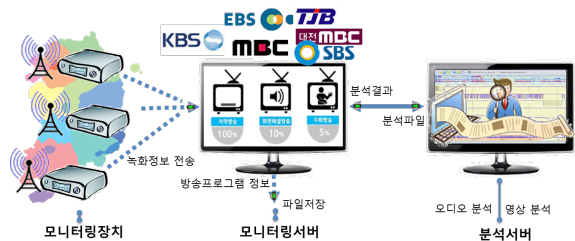


그림 6 장애인방송 모니터링 시스템 구성

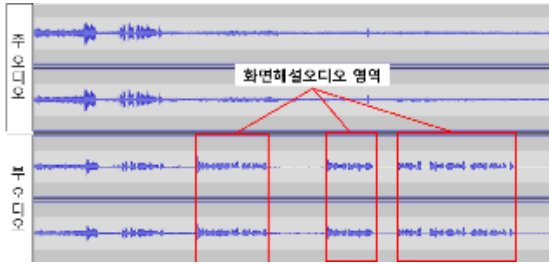


그림 7 부 오디오 내 화면해설 오디오 영역

주 오디오의 음성이 없는 부분에 삽입된다[25].

화면해설 오디오에 대한 분석은 주 오디오와 부 오디오 간의 두 오디오의 차이를 비교하여 실제 화면 해설에 대한 오디오의 분량이 어느 정도 인지를 파악하는 것이다. 두 오디오의 동기가 정확한 경우에는 오디오 프레임마다 특징 벡터를 추출하여, 비교하고 전체 오디오에 대해 화면 해설에 대한 오디오로 판정된 부분의 합계를 비율로써 산출한다.

$$\text{화면해설오디오 분량}(\%) = \frac{\sum_{i=1}^n (\text{추출한 화면해설 오디오 영역시간})}{\text{전체 프로그램 시간}} \times 100$$

보통 다큐멘터리와 같이 외국어가 포함된 프로그램에 있어서 외국어는 주 오디오에 위치하고, 우리말로 통역된 오디오는 부 오디오에 수록된다. 따라서 단순히 오디오의 특징 차이를 비교하는 방법에서는 전형적인 화면해설 오디오와 통역 오디오를 구분할 수 있는 방법이 없으므로, 다른 방법을 고안하는 것이 필요하다.

그림 8은 102개의 지상파 프로그램에 대해 직접 청취에 의한 프로그램 내의 화면해설분량과 소프트웨어를 통해 검출한 화면해설구간을 분석한 화면해설분량을 비교한 결과이다.

이와 같은 방식에 따라 2017년 10월 24일부터 2018년 2월 26일까지 KBS1, KBS2, EBS1, MBC, SBS에서 방송된 총 957개(750.34시간)의 화면해설 방송 콘텐츠를 분석한 결과는 그림 9와 같다. 각

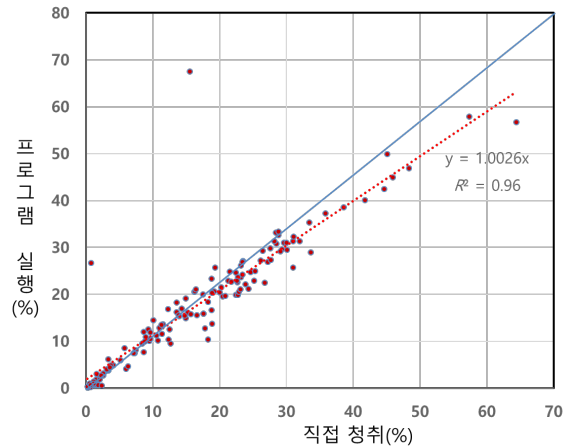


그림 8 직접 청취에 의한 화면해설 오디오 분량과 프로그램에 의해 검출된 화면해설 오디오 분량 비교

방송 프로그램 장르에서 화면해설 오디오가 차지하는 평균 비율은 예능(11.36%), 뉴스(1.88%), 교양(8.25%)을 제외하고는 15~20% 정도로 파악되었다.

SBS의 경우에는 예능에, MBC의 경우 뉴스에 화면해설방송을 편성하고 방송하고 있는 것으로 분석되었다. 다만 뉴스에서의 화면해설 오디오의 비율은 약 2% 정도로 화면해설 방송이 갖는 실질적인 효과에 대해서는 의문을 가질 수밖에 없다.

다. 수화방송 분석

자막이나 화면해설방송의 경우에는 방송 시그널

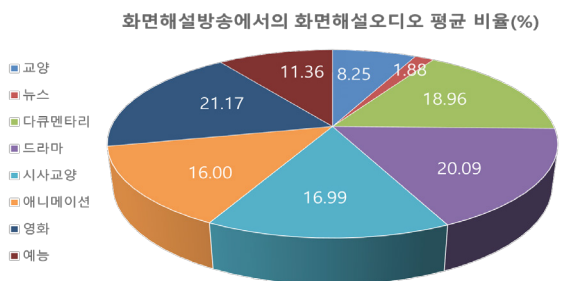


그림 9 방송프로그램 장르별 화면해설오디오 평균 비율

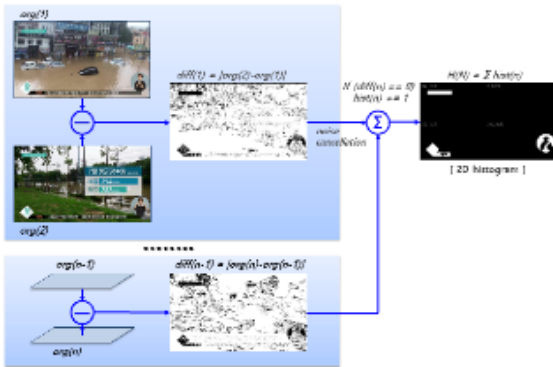


그림 10 수화방송 분석 알고리즘 개념도

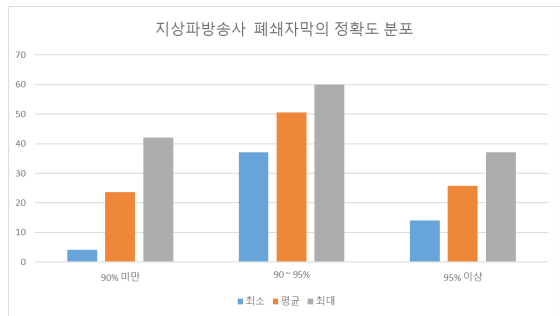


그림 11 폐쇄 자막의 정확도 분포도

링 정보에 서비스를 구분하는 디스크립터를 통해 최소한 현재 방송되는 프로그램이 장애인방송에 해당되는지 확인할 수 있다. 그렇지만 수화방송의 경우에는 서비스 유무를 표시하는 정보가 방송 시그널링 정보에 포함되고 있지 않기 때문에 방송화면에서 수화영역을 인식하여 판단할 수밖에 없다.

수화방송의 수화 영역은 일반적으로 방송 화면의 우측 하단에 위치하고 있고, 방송사별 또는 방송프로그램별로 위치가 일정하지 않은 경우가 많다. 또한, 최근에는 수화 화면이 주 화면을 가리는 불만을 해소하기 위해 배경 없이 수화 통역사만 분리하여 화면에 제시하는 경우도 많아지고 있다. 최지훈의 연구[24,26]는 방송프로그램별 방송영상 캡처 이미지 셋을 이용하여 수화 영역을 자동으로 검출하여 수화방송 유무 판단 오류를 줄일 수 있는 방법을 제안하고 있으며, 약 93% 정도의 정확도로 보인다(그림 10 참조).

라. 자막방송 분석

현재 자막방송은 지상파의 경우 100% 이루어지고 있으며, 속기에 의한 자막 시간 지연은 2~4초, 정확도는 98% 이상인 것으로 알려지고 있으나[27] 공식적으로 보고된 것은 없다.

자막의 표기 방법은 화자의 원 발화를 그대로 읊

겨 적는 것을 원칙으로 하고 중간에 단절되는 부분이 없도록 하며, 방송 프로그램 매 회당 오·탈자 및 띄어쓰기 등의 입력 오류는 2% 이내인 것이 바람직한 것으로 장애인방송 프로그램 제공 가이드라인을 제시하고 있다[28]. 2017년 10월 23일부터 방송된 8,371개의 폐쇄 자막을 기초로 분석한 결과는 그림 11과 같다. 전체 자막의 76% 정도가 90% 이상의 정확도를 갖는 것으로 파악되었으며, 이 중 26% 정도가 95% 이상의 정확도를 갖는 것으로 분석되었다.

방송 자막은 몇 개의 단어로 이루어진 단문이 많아 단어 사전에 기초한 자막의 오·탈자 정도만 파악할 수 있다. 따라서 단어 자체는 맞더라도 문맥에 맞지 않는 경우에 대해서는 분석되지 않아 한국어 자연어처리기술의 적용도 요구된다. 한편, 자막 지연 시간에 대해서는 오디오의 시간과 자막의 Presentation Time Stamp(PTS)를 이용하여 분석 가능할 것이다. 이를 위해서는 해당 자막이 나오는 오디오의 시작 부분을 찾는 오디오와 폐쇄 자막을 동기화하는 기술도 필요하다. 최근 지상파에서 드라마나 예능프로그램을 12부로 나누는 쪼개기 편성이 많아지면서 실제 오디오와 폐쇄 자막이 갖는 PTS의 간격이 더욱 커져 오디오에 기반한 자막 동기화에 애로사항으로 작용할 수 있다.

IV. 결론

현재 50만 명을 넘는 시·청각장애인의 편리한 TV 시청을 지원하여 보편적 권리인 방송접근권을 향상시키기 위해서는 꾸준한 정책적 지원과 지속적인 기술개발이 필요하다. 또한 장애인 방송 프로그램의 양적 확대와 함께 좋은 품질의 장애인방송 서비스를 위해서는 사전 제작을 통해 시·청각장애인의 눈높이와 요구수준에 맞춘 콘텐츠 제작이 요구된다. 이를 위해서는 콘텐츠 제작에의 딥러닝 기술의 적용과 콘텐츠 분석을 통해 전체적인 품질 향상에 대한 방안을 모색할 필요가 있다. 또한 방송 시청 환경의 변화로 적극적으로 반영하여 전문적으로 장애인방송을 제공하는 VOD 서비스 도입도 정보 접근권 강화의 수단으로 검토할 필요가 있다.

용어해설

장애인방송 장애인의 방송접근권 보장을 위하여 방송사업자가 제공하는 자막방송·수화방송·화면해설방송 등 시청편의 서비스

자막방송 방송 음성 및 음향을 TV 화면상에 텍스트의 형태로 보여주는 서비스

수화방송 청각장애인의 방송시청을 돕기 위해 방송 내용을 수화를 제공해 주는 서비스

화면해설방송 시각장애인의 시청을 돕기 위해 현재 방송되고 있는 장면과 관련된 등장인물의 행동, 표정, 위치, 의상 등의 내용을 음성으로 제공해 주는 서비스

약어 정리

AI	Artificial Intelligence
FCC	Federal Communications Commission
OTT	Over The Top
PP	Program Provider
PTS	Presentation Time Stamp
SO	System Operator
STB	Set Top Box

TTML	Timed Text Markup Language
UHD	Ultra High Definition
VOD	Video On Demand
OTT	Over The Top

참고문헌

- [1] 송종길, 이용주, "디지털시대 장애인방송 활용론," 보고서번호: 2003-08, 2003.12, pp. 1-269.
- [2] 채수용, "유료방송장애인용VOD 보유비율 0.001%에 불과," 디지털데일리, 2017.9.22.
- [3] KCA, "온라인 동영상(OTT)중심의 신 미디어 트렌드를 바라보는 주요국 규제기관의 대응현황," 동향과 전망: 방송·통신·전파, 통권 제59호, 2013, p. 23.
- [4] 방송통신위원회, "장애인방송 편성 및 제공 등 장애인 방송접근권 보장에 관한 고시," 제2015-4호, 2015.5.28.
- [5] 주정민, "해외 선진국과 국내의 장애인방송 운영제도 비교 연구," 방통융합정책연구 KCC-2014-29, 방송통신위원회, 2014, p. 439.
- [6] Ofcom, "Measuring Live Subtitling Quality," 2015, https://www.ofcom.org.uk/_data/assets/pdf_file/0011/41114/qos_4th_report.pdf
- [7] 한국방송통신전파진흥원, "심층리포트: 방송산업의 인공지능(AI) 활용 사례 및 전망," KCA Media Issue & Trend, 2017.10.
- [8] T. Afouras, J. S. Chung, A. Zisserman, "Deep Audio-visual Speech Recognition," *IEEE Trans. Pattern Analysis Mach. Intell.*, Early Access, 2018, Doi: 10.1109/TPAMI.2018.2889052.
- [9] S. Suwajanakorn, S. M. Seitz, I. Kemelmacher-Shlizerman, "Synthesizing Obama: Learning Lip Sync from Audio," *ACM Trans. Graphics*, vol. 36, no. 4, 2017, Article no. 95.
- [10] 박동영 외, "환경변화에 따른 장애인방송의 역할 및 개선방안 연구," 방통융합정책연구 KCC-2017-37, 방송통신위원회, 2017, p. 160.
- [11] 정윤희, 손원준, "TV 예능 프로그램의 자막 유형 연구," 한국디자인트렌드학회, 34권, 2012, pp. 267-276.
- [12] R. Rashid, "Representing Emotions with Animated Text," M.S. thesis, Univ. of Toronto, 2005, p. 114.
- [13] J. Ohene-Djan, R. Shipsey, "E-Subtitles: Emotional Subtitles as a Technology to assist the Deaf and Hearing-Impaired when Learning from Television and Film," in *Proc. Int. Conf. Adv. Learn. Technol.*, Kerkrade, Netherlands, July 5-7, 2006, pp. 464-466.
- [14] J. Ohene-Djan, J. Wright, K. Combie-Smith, "Emotional Subtitles: A System and Potential Applications for Deaf and Hearing Impaired People," in *Conf. Workshop Assitive Technol. People Vision Hearing Impairments (CVHI)*, Granada, Spain, Aug. 28-31, 2007.

- [15] 김태욱, “한국TV 수화통역방송에 대한 농인의 인지도 조사연구,” 석사학위논문, 나사렛대학교, 2017.
- [16] 채수웅, “방통위, 스마트 수어방송 상용화 추진계획,” 디지털데일리, 2018.2.21.
- [17] J. Oh, S. Jeon, M. Kim, H. Kwon, I Kim, “An Avatar-Based Weather Forecast Sign Language System for the Hearing-Impaired,” in *IFIP Int. Conf. Artif. Intell. Applicat. Innovations (AIAI)*, Rhodes, Greece, Sept. 19-21, 2014, pp. 519-527.
- [18] 오주현 외, “실시간 그래픽 수화방송 시스템 개발,” 한국방송공학회 하계학술대회, 2015.6.
- [19] 최상배, “TV수화언어 방송통역 분석,” 특수아동교육연구, 제13권 제3호, 2011, pp. 77-99.
- [20] 장인선, 임우택, 안충현, “화면해설방송 콘텐츠 저작 기술,” 한국방송공학회 추계학술대회, 2014, p. 175-176.
- [21] 유창선, “네이버, 8월 음성합성 기술 상용화 ... 구글보다 빨라,” 전자신문, 2018.7.4.
- [22] Y. Lee, S.Y. Lee, A. Rabiee, “Emotional End-to-End Neural Speech Synthesizer,” in *Conf. Neural Inform. Process. Syst. (NIPS 2017)*, Long Beach, CA, USA, Dec. 4-9, 2017.
- [23] S. Kim, H. Choi, “Emotional Voice Conversion Using Generative Adversarial Networks,” 2017. http://mac.kaist.ac.kr/~juhan/gct634/2018/finals/emotional_voice_conversion_using_generative_adversarial_networks_report.pdf
- [24] 최지훈, “수화방송 모니터링 기술,” 한국방송-미디어 공학회 추계학술대회, 2017.
- [25] 안충현, 장인선, “화면해설방송 콘텐츠 분석기술 연구,” 한국방송-미디어공학회 하계학술대회, 2017
- [26] 최지훈 외, “수화방송 모니터링을 위한 수화영역 검출 방법 연구,” 2018년 한국방송-미디어공학회 추계학술대회. 2018.
- [27] 홍경희, “청각장애인의 나침반 ‘자막방송 속기사,’” 시사뉴스, 2005.7.28. <http://www.sisa-news.com/news/article.html?no=2861>
- [28] 시청자미디어재단, “장애인방송 프로그램 제공 가이드라인,” 2017. <https://kcmf.or.kr/cms/lib/view2.php?idx=10409>