

기술보고서

FBMF-TR-017

제정일: 2024. 11. 29.

2024 방송 미디어 서비스 기술동향

2024 Trends in Broadcast Media
Service and Technology



표준초안 검토 위원회 방송IP하이브리드분과위원회

표준안 심의 위원회 운영위원회

	성명	소 속	직위	위원회 및 직위
기술보고서(과제) 제안	김정덕	KBS	수석	방송IP하이브리드분과 / 분과위원장
기술보고서 초안 에디터	이동관	MBC	부장	
	한철	CBS	부장	위원
	최윤진	KBS	책임	위원
	이경렬	SBS	차장	위원
	이학주	SBS	차장	간사
사무국 담당	함상진	미방포럼	사무총장	-

본 문서에 대한 저작권은 미래방송미디어표준포럼에 있으며, 미래방송미디어표준포럼과 사전 협의 없이 이 문서의 전체 또는 일부를 상업적 목적으로 복제 또는 배포해서는 안 됩니다.

발행인 : 미래방송미디어표준포럼 의장

발행처 : 미래방송미디어표준포럼

06130, 서울특별시 강남구 테헤란로 7길 22 신관 1108호

Tel : 02-568-3556, Fax : 02-568-3557

발행일 : 2024.12

서 문

1 기술보고서의 목적

본 기술보고서의 목적은 빠르게 진화하는 방송 미디어 환경의 맥락에서 2024년 방송 미디어 분야에서 주목할 만한 서비스 및 기술 동향을 소개하는 것이다. 이 보고서는 방송 미디어의 미래를 형성하고 있는 몇 가지 주요 기술과 서비스에 대한 포괄적인 개요를 제공하는 것을 목표로 한다.

여기에는 방송미디어 분야의 AI 글로벌 동향, 방송 음성·자막 처리를 위한 AI 기술현황, 방송 제작을 위한 뮤직 오토태깅, 라디오 콘텐츠 소비 다양화와 온라인 스트리밍 통계, FAST 서비스 광고 삽입 기술, 그리고 다양한 네트워크망을 활용한 고정밀 위치정보 서비스 기술이 포함된다. 본 보고서에서 기술한 주요 영역을 상세히 설명함으로써 방송미디어 분야에서 향후 표준화 기술을 개발하고, 서비스와 기술에 대한 미래 방향을 제시할 수 있는 통찰력을 제공함을 목표로 한다.

2 주요 내용 요약

본 기술보고서는 6가지 주요 영역에 초점을 맞춰 방송 미디어 산업 내 AI 발전을 포함하여 전반적인 방송 미디어분야의 최신 서비스와 기술을 포함하여 포괄적인 동향을 제공하고 있다. 우선 NAB2024 및 IBC2024와 같은 주요 산업 행사에서 선보인 AI 기술의 글로벌 동향을 조사하여 AI가 방송 미디어에 미치는 혁신적인 영향을 살펴 본다. 또한 본 기술보고서에서는 음성 및 자막 처리 분야의 AI 애플리케이션을 살펴보고 보다 효율적인 제작 및 접근성을 지원하며, 딥러닝 기반 뮤직 오토태깅을 통해 미디어 제작을 위한 콘텐츠 분류를 향상시키고자 하는 내용을 다루고 있다. 그리고 온라인 스트리밍 통계에 중점을 두고 디지털 시대의 정확한 시청자 지표에 대한 필요성을 반영하여 라디오 콘텐츠 소비의 변화를 검토한다. TV 분야에서 콘텐츠 배포를 위한 새로운 수익모델을 제공하는 무료 FAST TV 서비스를 위한 광고삽입 기술의 발전도 다루고 있다. 끝으로, GNSS, ATSC 3.0 등 다양한 네트워크 기술을 활용한 고정밀 위치 서비스와 기술에 대해서 분석하고 있다.

3 인용 기술보고서와의 비교

해당사항 없음

Preface

1 Purpose

The purpose of this technical report is to introduce significant trends in services and technology within the broadcast media field as of 2024, set against the backdrop of the rapidly evolving media landscape. This report aims to provide a comprehensive overview of key technologies and services that are driving the future of broadcast media. The focus areas include global AI trends in the broadcasting sector, the current status of AI applications for broadcast voice and subtitle processing, music auto-tagging for production purposes, diversification in radio content consumption and online streaming analytics, advertisement insertion technology for FAST services, and high-precision location information services leveraging various network technologies. Through an in-depth explanation of these core areas, this report seeks to contribute to the development of future standardization technologies in broadcast media and provide insights that can guide the future direction of services and technology in the industry.

2 Summary

This technical report focuses on six key areas and provides a comprehensive overview of trends, including the latest services and technologies across the broadcast media sector, with a particular focus on AI developments within the industry. First, it examines the transformative impact of AI on broadcast media by analyzing global trends in AI technology showcased at major industry events such as NAB 2024 and IBC 2024. Additionally, this report explores AI applications in voice and subtitle processing, which support more efficient production and accessibility, and addresses advancements in deep learning-based music auto-tagging aimed at improving content classification for media production. It then examines shifts in radio content consumption, with a focus on online streaming statistics and the growing need for accurate audience metrics in the digital age. The report also covers developments in advertising insertion technology for the free ad-supported streaming television (FAST) service, offering a new revenue model for content distribution in the TV sector. Lastly, it analyzes high-precision location services and technologies using various network technologies, such as GNSS and ATSC 3.0.

3 Relationship to Reference Standards

None.

목 차

1	적용 범위	1
2	인용 표준	1
3	용어 정의	1
4	약어	x
5	방송 미디어 AI 글로벌 동향	x
6	방송 음성·자막 처리를 위한 AI 기술현황	x
7	방송 제작을 위한 뮤직 오토태깅	x
8	라디오 콘텐츠 소비 다양화와 온라인 스트리밍 통계	x
9	FAST 서비스 광고 삽입 기술	x
10	다양한 네트워크망을 활용한 고정밀 위치정보 서비스	x
부록	I -1 시험인증 관련 사항	xx
	I -2 참고 문헌	xx
	I -3 영문기술보고서 해설서	xx
	I -4 기술보고서의 이력	xx

2024 방송 미디어 서비스 기술 동향

(2024 Trends in Broadcast Media Service Technology)

1 적용 범위

본 기술보고서는 AI가 제작 및 접근성에 미치는 영향에 초점을 맞춘 방송미디어 산업의 중요한 영역을 포함하고 있다. 운영의 효율성과 접근성을 높이는 AI 기반 음성 및 자막 처리 기술과 최적화된 콘텐츠 분류를 위한 딥러닝 기반의 뮤직 오토태깅 기술을 범주로 다루고 있다. 또한 본 보고서는 온라인 스트리밍 지표에 초점을 맞춰 라디오 콘텐츠 소비의 변화를 다루고, 새로운 수익 모델로서 FAST 서비스 광고삽입 기술을 탐구하고 있다. 그리고 향후 방송 애플리케이션에 영향을 미칠 수 있는 GNSS 및 ATSC 3.0 기술이 지원하는 고정밀 위치 서비스 기술을 포함함으로써 방송미디어 분야 전문가들과 이해관계자들에게 최신의 방송미디어 서비스와 기술에 대한 유용한 정보를 제공하도록 작성되었다.

2 인용 표준

해당 사항 없음

3 용어 정의

3.1 커넥티드카(Connected Car)

통신망에 연결된 자동차. 통신 수단으로 무선랜(Wi-Fi), 엘티이(LTE) 이동통신 등이 사용된다. 커넥티드카는 다른 차량이나 교통 및 통신 인프라, 보행자 단말 등과 실시간으로 통신하며 운전자의 편의와 교통 안전을 돕고 인터넷의 다양한 서비스를 제공한다. 커넥티드카 자체가 통신 기기가 된다는 의미에서 ‘거대한 사물 인터넷(IoT) 기기’라고도 한다.

[출처] TTA정보통신용어사전

3.2 안드로이드 오토(Android Auto)

Android Auto는 자동차 내부에 있는 디스플레이 장치에 최적화된 UI로 Android의 실행 환경을 재구성하여 미러링하는 플랫폼이다. 다시 말해서 안드로이드 기기의 UI를 그대로 보여주는 것이 아니라 Android Auto 앱을 통해서 운전자에게 최적화된 UI로 안드로이드 앱이 실행되도록 화면을 재구성하는 역할을 한다.

[출처] 나무위키

3.3 안드로이드 오토모티브(Android Automotive)

Android Automotive는 사전 설치된 IVI(In-Vehicle Infotainment) 시스템 Android 애플리케이션과 선택적 2 차 및 타사 Android 애플리케이션을 실행하는 기본 Android 플랫폼이다. Android Automotive는 Android이며, Android와 동일한 코드를 베이스로 한다. Android를 인포테인먼트를 위한 모든 기능을 갖춘 플랫폼으로 구축하는 과정에서 자동차 관련 요구 사항, 기능 및 기술에 대한 지원을 추가한다.

[출처] <https://chuuu-devcamp.tistory.com/71>

3.4 미들웨어 (Middleware)

응용프로그램과 그 프로그램이 운영되는 환경 간에 원만한 통신이 이루어질 수 있게 하는 소프트웨어

[출처] 네이버 지식백과

3.5 자연어 처리 (Natural Language Processing, NLP)

컴퓨터를 이용해 사람의 자연어를 분석하고 처리하는 기술.

[출처] TTA정보통신용어사전

3.6 모달(Modal)

사람이 기계와 상호 작용하며 입출력에 사용하는 정보의 유형.

모달(modal)은 언어학, 음악 등 다양한 분야에서 ‘형태’나 ‘유형’의 의미로 사용한다. 예를 들어 음악 분야에서 모달은 특정 음계를 중심으로 음악을 구성하는 방식을 말한다. 언어학에서는 화자의 주관적인 태도와 판단을 표현하며 미묘한 뉘앙스를 전달하는 문법적 요소를 의미한다.

[출처] TTA정보통신용어사전

3.7 은닉 마르코프 모델 (Hidden Markov Model, HMM)

은닉 마르코프 모델(HMM)에서는 상태(state) 정보가 숨겨져 있고 출력(output)된 정보만 관찰된다. 따라서 출력된 정보만을 가지고 숨겨진 상태(state) 정보를 추정한다. 은닉 마르코프 모델(HMM)은 음성 인식, 자연어 처리, 몸짓 인식(gesture recognition) 등과 같이 대량으로 출력된 데이터를 통계적으로 패턴 분석하여 입력된 정보를 추론하는 데에 응용된다.

[출처] TTA정보통신용어사전

3.8 인공신경망(Artificial Neural Network)

사람 또는 동물 두뇌의 신경망에 착안하여 구현된 컴퓨팅 시스템의 총칭.

[출처] TTA정보통신용어사전

3.9 순환 신경망(Recurrent Neural Network, RNN)

시계열 데이터(time-series data)와 같이 시간의 흐름에 따라 변화하는 데이터를 학습하기 위한 심층 기계 학습(Deep learning) 모델.

[출처] TTA정보통신용어사전

3.10 규칙 기반 기계 번역(Rule-Based Machine Translation, RBMT) / 통계적 기계 번역(Statistical Machine Translation, SMT) / 신경망 기계 번역(Neural Machine Translation, NMT)

전통적으로 기계 번역에는 단어 사전과 번역 규칙을 이용하는 규칙 기반 기계 번역(RBMT: Rule-Based Machine Translation)과 원문과 해당 번역문에 대한 대량의 말뭉치(corpus) 데이터를 활용하는 통계적 기계 번역(SMT: Statistical Machine Translation)이 주로 사용되어 왔다. 2013년에 원문과 해당 번역문 사이의 관계를 인공 신경망으로 학습하는 방법이 개발되었으며, 기존의 통계적 기계 번역(SMT)보다 우수한 성능을 보였다.

[출처] TTA정보통신용어사전

3.11 온라인 동영상 서비스(Over-The-Top, OTT)

인터넷을 통해 영화나 드라마, TV 방송 등의 동영상을 제공하는 서비스.

[출처] TTA정보통신용어사전

3.12 양자화(Quantization)

입력 신호를 유한한 개수의 값으로 근사화 하는 것.

[출처] TTA정보통신용어사전

3.13 푸리에 변환(Fourier Transform)

신호를 주파수 성분으로 분해하는 조작.

[출처] TTA정보통신용어사전

3.14 광고 차단(Ad Block)

PC나 공유기에서 hosts 파일을 변경하는 방법, 브라우저의 플러그인 사용 등의 방법으로 광고 서버와 통신을 차단하여 시청 디바이스에서 광고 재생을 차단하는 기법

[출처] 나무위키

4 약어



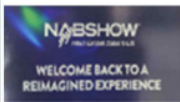


OTT	Over The Top
AOD	Audio On Demand
EPG	Electronic Program Guide
ANN	Artificial Neural Network
ASR	Automatic Speech Recognition
HMM	Hidden Markov Model
LSTM	Long Short-Term Memory
NLG	Natural Language Generation
NLP	Natural Language Processing
NLU	Natural Language Understanding
NMT	Neural Machine Translation
RBMT	Rule-Based Machine Translation
RNN	Recurrent Neural Network
SMT	Statistical Machine Translation
HLS	HTTP Live Streaming

5 방송미디어 AI 글로벌 동향

본 장에서는 2023년에 방송미디어 업계에 부상하였던 AI가 올해 미국과 유럽의 글로벌 양대 방송미디어 전시회인 NAB2024와 IBC2024에서 그 중심에 올라서게 된 내용을 살펴보기로 한다.

5.1 NAB 2024

2024년 NAB에서는 작년 “AI의 부상”이라는 주제에 이어서 올해는 “AI 중심”이라고 할 정도로 전시장 및 컨퍼런스에서 최상위 키워드로 자리잡았다. AI와 함께 여전히 클라우드, XR, IP 제작 “NextGen TV”로 일컫는 차세대방송 분야가 국내 방송미디어 시장에 관심을 가질만한 주제였다.

2018 NAB	2019 NAB	Covid19로 2020/ 2021 행사 없음	2022 NAB	2023 NAB	2024 NAB
M.E.T 융합	이야기의 시작		재창조된 경험	시대를 위한 축하	새로운 관객을 잡아라
Media, Entertainment and Technology	Every Story Starts Here		Welcome Back to A Reimagined Experience	A Celebration for the Ages * NAB 100주년	Capture New Audience
					
UHD 고해상도 HDR 선명영상 ATSC 3.0 차세대방송 IP 제작 클라우드 미디어 서비스	8K 고해상도 HDR 선명영상 ATSC 3.0 차세대방송 5G IP 제작 클라우드 미디어 서비스		클라우드 IP 제작 라이브 제작 AI/원격제작 XR 실감영상 8K 고해상도 ATSC 3.0 차세대방송	AI/원격제작 클라우드 라이브 제작 스트리밍 XR 실감영상 NextGen TV 차세대방송	AI 클라우드 스트리밍 XR 실감영상 IP 제작 NextGen TV 차세대방송 미디어 서비스

(그림 5-1) NAB2024의 중심에 선 “AI”

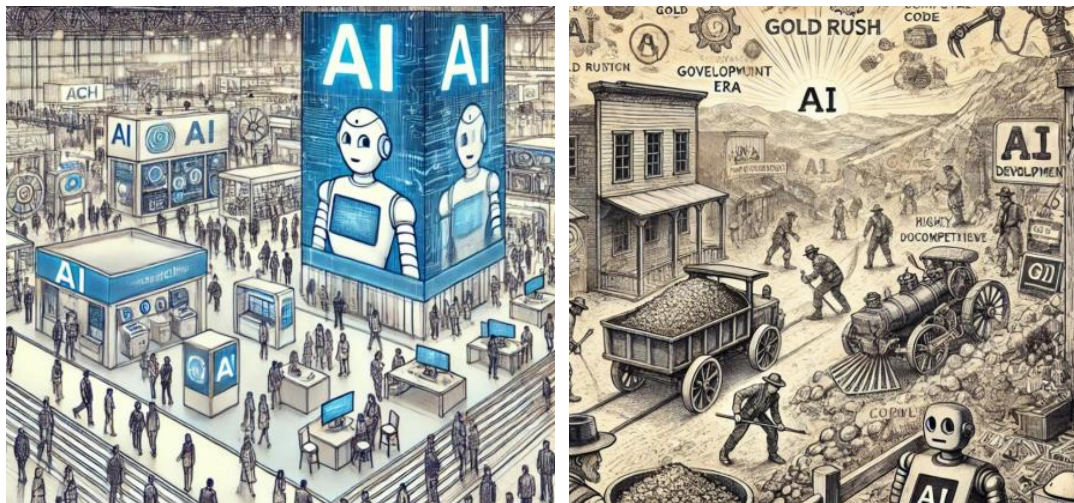
NAB2024 전시회 및 컨퍼런스의 중앙무대에 오르면서 강조된 미디어 산업 내의 AI의 영향, 잠재력과 도전 과제에 대한 주요 내용은 다음과 같이 요약할 수 있다.



(그림 5-2) NAB2024 컨퍼런스

- 1) AI는 NAB 2024 토론을 지배했으며, 기업들이 미디어 및 엔터테인먼트 산업에서 AI의 혁신적인 잠재력을 탐구하면서 행사 전반에 걸쳐 초점이 되었다. 거의 모든 조직은 AI를 중추적인 기술 발전으로 인식하면서 AI를 통합하거나 통합할 계획을 갖고 있었다.
- 2) AI는 상당한 잠재력과 과제를 동시에 제시한다. 파괴적이고 혁신적인 솔루션을 제공하여 미디어 워크플로우에 혁신을 가져올 것을 약속하지만 효과적인 구현에 필요한 비용과 리소스는 상당히 크다. 업계는 AI의 기능을 최대한 활용하기 위해 이러한 요구에 대비해야 할 것으로 보고 있다.
- 3) AI 기술이 미디어 프로세스에 더욱 뿌리내리면서 “신뢰성” 이슈가 중요한 문제로 대두되었다. AI로 생성된 콘텐츠를 신뢰할 수 있도록 보장하는 것이 중요하며, 콘텐츠 관리 및 전달에 있어 AI의 효율성은 원활한 통합을 위해 높은 기준을 충족해야 할 것으로 진단하고 있다.
- 4) AI는 “미디어 워크플로우”를 재편하고 복잡한 프로세스를 단순화하며 생산 및 유통을 최적화하고 있다. 그러나 AI를 기존 운영에 통합하는 것은 어려운 문제라는 것으로 인식하고 있다. 기업은 새로운 요구 사항을 해결하고 워크플로 효율성을 개선하기 위해 적응하고 지속적으로 혁신할 것을 주문하고 있다.
- 5) AI 개발의 현재 상태는 빠르고 경쟁적인 발전과 규제 부족이 특징인 “서부 금광 개척 시대”와 비교가 되고 있다. 이러한 어려움에도 불구하고 AI의 미래는 미디어 혁신의 초석이 될 것이라는 기대와 함께 낙관론이 존재하고 있다.
- 6) AI는 새로운 창의적인 접근 방식과 개인화된 콘텐츠를 지원함으로써 스토리텔링과 교육경험을 향상시키고 있다. 고급 AI 도구를 통해 교육자와 스토리텔러는 더욱 매력적이고 역동적인 학습 환경을 조성하여 창의성과 맞춤형 육성에 있어 AI의 가치를 입증할 수 있을 것으로 보고 있다.
- 7) 광고분야에서 AI는 기업이 수익을 늘리고 성과를 향상시키는 데 도움을 준다. AI는 광고 타겟팅 및 배치를 최적화함으로써 광고 재고가치를 극대화하고 기업이 보다 효과적이고 개인화된 광고 전략을 통해 더 나은 재정적 성과를 달성할 수 있도록 지원할 수 있다.
- 8) AI는 또한 AI 관련 기술을 갖춘 전문가에 대한 수요가 증가함에 따라 미디어 채용 및 온보딩을 재편하고 있다. 또한 AI는 맞춤형의 혁신적인 콘텐츠 제작을 위한 새로운 기회를 제공하고 업계 내에서 창의적인 가능성을 더욱 확대함으로써 제작자에게 힘을 실어줄 것으로 보고 있다.

- 9) AI가 보안조치에서 더 큰 역할을 하기 때문에 미디어 콘텐츠를 보호하는 것이 중요한 요소이다. AI는 무단 액세스를 방지하고 불법 복제 위험을 완화함으로써 지적 재산의 무결성을 유지하고 콘텐츠 보호에 관해서 업계에서 점점 커지고 있는 우려 중 하나를 해결하는 데 도움을 줄 것으로 보고 있다.
- 10) ATSC 3.0 및 NextGenTV와 같은 방송표준의 발전은 미디어 환경 현대화에서 AI의 역할을 강조할 수 있다. 기술이 발전함에 따라 방송에 AI를 통합하면 시청 경험이 향상될 수 있다고 보고 있다. 다만, 시장 및 규제 문제는 여전히 이슈로 남아 있을 것으로 본다.



(그림 5-3) 전시장 중앙무대의 AI와 서부 금광시대에 비유한 AI(ChatGPT)

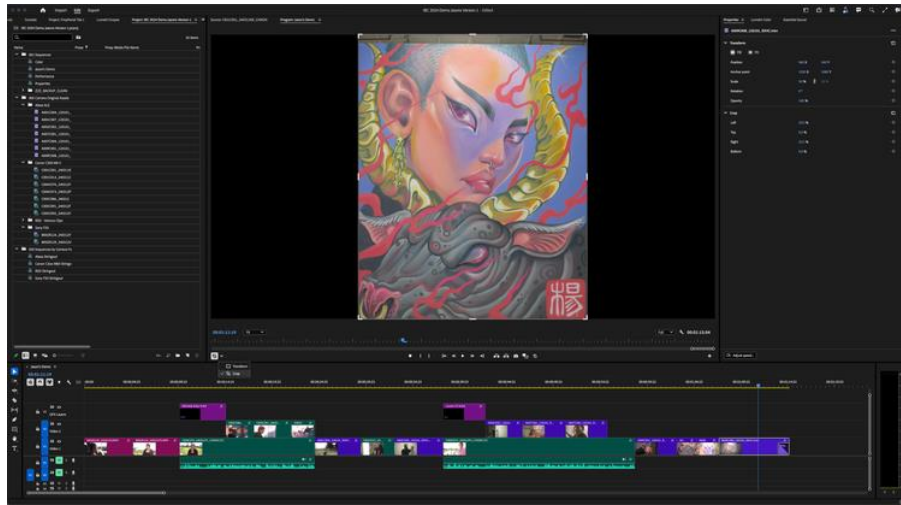
이와 같은 전방위적 AI의 적용에 대한 관심과 함께 좀 더 세부적으로 살펴보더라도 NAB2024에서는 AI 기술이 제작, 편집, 송출, 데이터분석 등 방송제작 현업으로 확대되었고 특히 생성형 AI를 통한 영상제작, 콘텐츠 요약, 자동 자막 생성 및 번역 등 AI의 방송 미디어 결합분야가 점점 확대되고 있었다. 또한 올해의 경우에는 방송제작 분야 중에서 라이브 스포츠 중계에도 AI를 적용한 사례가 있는데 농구경기를 라이브 중계하면서 “슛”을 AI가 자동으로 인식하여 스코어에 자막을 넣는 사례를 볼 수 있었다.

5.2 IBC 2024

2024년 9월, 유럽의 IBC2024에서도 NAB2024와 같이 EBU와 IBC가 “AI Tech Zone” 부스 존을 만들어 운영할 정도로 AI가 중심 키워드로 전시장과 컨퍼런스 전방에 녹아 들어 있었다. IBC2024의 AI 동향에 대해서는 주요 분야에 대하여 사례를 중심으로 정리하였다.

- 1) AI 기반 콘텐츠 제작: Adobe는 기능을 향상하고 제작 프로세스를 간소화하는 것을 목표로 비디오 생태계에 새로운 AI 기반 기능을 도입했다. 주요 도구에는 자

동화된 편집, 장면 감지, 색상 교정이 포함되어 있어 편집자가 반복 작업에서 시간을 절약할 수 있다. 이번 업데이트를 통해 제작자는 더욱 효율적으로 작업 흐름을 가속화하고 창의적인 제어 기능을 선보였다.



(그림 5-4) Adobe Video Ecosystem

[출처] <https://store.epicgames.com/ko/blog/path-of-exile-2-the-long-road-to-exceeding-a-decade-of-expectations>

- 2) 실시간 분석 및 시청자 인사이트: Google Cloud는 시청자 선호도와 행동에 대한 실시간 통계를 제공하는 AI 기반 분석 도구를 도입하여 방송사가 시청자 동향과 인구통계를 기반으로 콘텐츠 전략을 조정할 수 있다.
- 3) 가상 제작 및 AI로 강화된 워크플로우: Epic Games는 복잡한 작업을 자동화하고 가상환경을 실시간으로 생성하는 AI 도구를 갖춘 가상 제작용 Unreal Engine을 선보였다. NVIDIA는 또한 공동 제작 워크플로를 위한 Omniverse 플랫폼을 시연하여 가상 제작 프로세스를 간소화하였다.
- 4) 콘텐츠 보안 및 불법 복제 방지: Irdeto는 라이브 스트리밍 및 주문형 비디오 콘텐츠를 보호하기 위해 AI기반 불법복제 방지 솔루션을 도입했으며, Viaccess-Orca는 무단 콘텐츠 배포를 실시간으로 모니터링하고 방지하는 AI기반 도구를 선보였다.
- 5) 향상된 콘텐츠 검색 가능성: Comcast는 향상된 메타데이터 생성 및 더 나은 권장 사항을 포함하여 사용자가 콘텐츠를 더 쉽게 찾을 수 있도록 하는 Xfinity 플랫폼의 AI 개선 사항을 선보였다.
- 6) 청중참여 및 상호작용: Futuri는 방송사에 매력적인 라이브 방송을 위해 소셜 미디어 데이터를 기반으로 인기 있는 주제를 제공하는 TopicPulse AI 시스템을 선보였다.

7) 접근성을 위한 AI: Microsoft는 실시간 캡션 및 다국어 번역을 포함하여 Azure Media Services 내에서 새로운 접근성 기능을 발표하여 더 많은 청중이 콘텐츠를 더욱 포용하게 했다.

8) AI 최적화를 통한 지속 가능성: Akamai는 에너지 소비를 줄이고 미디어 전송 효율성을 높이는 데 도움이 되는 CDN(콘텐츠 전송 네트워크)에 대한 AI 기반 최적화를 공유했다. Sky Group은 AI를 사용하여 방송 운영 전반에 걸쳐 전력 사용을 모니터링하고 최적화하여 지속 가능성 목표를 지원하고자 했다.

이 밖에 Sony를 비롯하여 Canon, AstroDesign 등 AI 기반의 자동 추적 기능 등의 AI 솔루션을 개발하거나 선보이는 등 방송 콘텐츠의 제작, 편집, 데이터 가공 및 전송에 이르기까지 방송미디어의 각 분야에서 AI의 활용이 점점 많아졌다. 아직 AI가 초기단계라는 것이 전문가들의 이야기지만 최근의 전시와 컨퍼런스의 발표 내용으로 볼 때 점점 더 혁신이 가속화될 것으로 전망된다.

6 방송 음성·자막 처리를 위한 AI 기술현황

6.1 개요

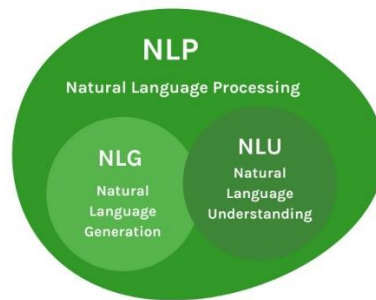
방송 제작 환경이 디지털 전환, 실시간 스트리밍 증가, 글로벌 서비스 확대 등으로 빠르게 변화하면서 AI 기술 활용의 중요성이 커지고 있다. 특히 방송 서비스 환경, 시청자 요구사항, 제작 환경 측면에서의 변화가 AI 기술 도입을 가속화하고 있다. AI 기반의 음성 인식, 음성 합성, 자막 처리 기술은 방송 콘텐츠의 접근성을 크게 향상시키고, 제작 효율성을 높이며, 새로운 형태의 방송 경험을 제공할 수 있다는 점에서 AI 기술 적용과 발전이 기대되는 분야이다. 그 배경에는 방송 서비스 환경 변화가 있다. 특히 한국 방송 시장에서는 OTT(Over-The-Top)를 통한 스트리밍 서비스가 일반화되고 있고 K-콘텐츠의 글로벌화로 인한 다국어 서비스 수요가 급증하고 있다. 더불어 이러한 OTT 플랫폼의 성장은 콘텐츠 제작과 유통 방식의 변화를 이끌고 있다. 이와 함께 시청자들의 요구사항도 날로 다양화되고 있다. 장애인방송 의무 편성 비율 확대로 고품질의 자막과 수어 서비스 수요가 증가하고 있으며, 고령화 사회에 대응하기 위한 음성·자막 서비스의 품질 향상이 필수적인 상황이다. 여기에 다문화 가정을 위한 다국어 방송 서비스의 필요성도 높아지고 있다. 이러한 서비스 다변화와 시청자 요구의 증가는 제작 환경의 혁신을 필요로 하고 있다. 방송 콘텐츠의 제작 주기가 단축되고 신속한 처리가 필수적이며, 인건비 상승으로 인한 제작 비용 효율화가 절실하고, 방대한 양의 아카이브 자료를 효율적으로 관리하고 활용하기 위한 기술적 지원도 요구되고 있다. 이처럼 다양한 환경 변화에 대응하기 위해 방송사들은 AI 기술을 도입하여 음성·자막 처리의 자동화, 효율화를 하나 둘씩 추진하고 있다. 본 보고서는 현재 방송 현장에서 활용되고 있는 AI 기술 중 음성 인식, 합성, 자막 분야의 현황을 분석하고, 실제 적용 사례와 한계점, 발전 방향을 살펴보고자 한다.

6.2 AI 음성 인식 및 합성, 자막 기술 현황

6.2.1 자연어 처리(NLP) 기술 개요

자연어 처리는 컴퓨터가 인간의 언어를 이해할 수 있도록 돕는 컴퓨터 과학, 인공지능 연구의 분야 중 하나이다. 자연어 처리 연구는 자연어 이해(NLU)와 자연어 생성(NLG)으로 구분할 수 있다.

자연어 이해 기술(NLU)은 다량의 데이터를 학습하여 텍스트에서 의미 있는 정보를 추출하고 해당 상황에서의 텍스트 내용이나 감정 등을 파악하는 것을 목적으로 한다. 예를 들면 자동 음성 인식(ASR)은 사람의 음성을 인식하여 텍스트로 변환할 수 있는 기술이다. 반면 자연어 생성 기술(NLG)은 사람의 언어인 자연어를 출력, 생성하는 프로세스로 이미지 캡션, 챗봇 등으로 구현된다.



(그림 6-1) NLP, LNU, NLG 관계 개념도

[출처] AX SEMANTICS <https://www.ax-semantic.com/en/blog/natural-language-generation-explained>

또 자연어 처리 기술들을 task 기준으로 분류했을 때 음성을 텍스트로 변환하는 Speech-to-text(STT)기술, 텍스트를 음성으로 변환하는 Text-to-speech(TTS) 기술, 그리고 텍스트를 이미지로 변환하는 기술(Text-to-image)은 Text-to-Data분야로 분류될 수 있다. 이 분야는 텍스트와 더불어 음성 등 다양한 모달을 활용하는 추세로 최근 학계에서 많이 논의되고 있는 세부 분야이다.



(그림 6-1) 자연어 처리 기술 분류

[출처] Medium <https://medium.com/nlplanet/two-minutes-nlp-33-important-nlp-tasks-explained-31e2caad2b1b>

6.2.2 관련 AI 기술 발전

AI 기반 음성 인식, 자막 생성, 자막 번역, 음성 합성 기술 발전 과정은 AI의 발전과 맞물려 있으며, 특히 딥러닝과 대용량 데이터 학습, 트랜스포머(Transformer) 기반 아키텍처의 도입이 큰 기여를 한 것으로 평가된다.

음성 인식(ASR) 기술은 사용자의 음성을 텍스트로 변환하는 기술로, 다양한 언어와 억양을 이해하고 정확하게 분석할 수 있도록 발전해왔다. 초기의 음성 인식 기술은 특정 화자에 맞춰진 제한된 시스템이었으며, 주로 은닉 마르코프 모델(HMM)과 같은 통계적 모델을 사용하여 음성 신호를 패턴 인식하는 방식이었다. 그러나 2010년대 이후 딥러닝

이 도입되면서 음성 인식의 정확도와 효율성은 비약적으로 향상되었다. 인공 신경망(ANN)과 순환 신경망(RNN), Long Short-Term Memory(LSTM) 모델이 도입되며 음성 데이터의 복잡한 패턴을 더 정교하게 학습할 수 있게 되었고, 최근에는 트랜스포머(Transformer) 계열 모델을 기반으로 한 End-to-End 방식의 음성 인식 기술이 널리 쓰이며, 다양한 방언과 억양을 보다 정확하게 인식하고, 특정 사용자 맞춤형 학습이 가능해졌다.

자막 생성 기술도 음성 인식과 밀접한 관계가 있다. 초기 자막 생성 시스템이 인력에 의존하여 자막을 제공하는 방식이었다면, ASR 기술의 발전으로 실시간 자막 생성이 가능해졌고 문장을 분할하고 구두점을 자동으로 추가하는 등 이제는 자막 생성 시스템이 자연스러운 문장 구성을 자동으로 처리할 수 있게 되었다. 자막 번역 기술은 다국어 콘텐츠 수요가 늘면서 큰 주목을 받고 있다. 초기에는 규칙 기반 기계 번역(RBMT)이 주로 사용되었으나, 이는 자연스러움과 정확성에서 많은 한계가 있었다. 이후 통계적 기계 번역(SMT)이 대량의 번역 데이터를 학습하여 언어 패턴을 분석하는 방식으로 등장했지만 문맥 이해가 부족했다. 최근에는 신경망 기계 번역(NMT)이 도입되면서 자막 번역의 정확성과 자연스러움이 크게 향상되었다. 특히 트랜스포머(Transformer) 모델 기반의 신경망 기계 번역(NMT)은 문맥을 이해하고 자연스러운 문장 구조를 만들어내 자막 번역에서도 매우 중요한 역할을 하고 있다.

음성 합성 기술은 텍스트를 음성으로 변환하여 청각 정보를 제공하는 기술로, 초기에는 문자나 단어 단위로 녹음된 음성을 조합하는 방식이었지만, 음성의 자연스러움에 한계가 있었다. 딥러닝 도입 이후, WaveNet과 같은 모델이 개발되면서 실제 사람의 목소리와 유사한 음성을 생성할 수 있게 되었고, 다양한 감정과 억양을 표현하는 합성 음성도 가능해졌다. 최근에는 특정 화자의 목소리를 학습하여 특정 인물의 목소리로 음성을 합성하거나, 적은 데이터를 이용하여 voice cloning 하는 음성 합성도 가능해졌다.

다음 장부터는 음성 인식과 자막 기술, 음성 합성 관련 기술 동향을 차례대로 좀더 세부적으로 살펴보고자 한다.

6.3 음성 인식 및 자막 기술 동향

미디어 분야에서 AI 기반 음성 인식 및 자막 기술의 발전은 콘텐츠 제작, 전달, 소비 방식에 큰 변화를 가져오고 있다. 특히, 방송 및 OTT 서비스가 활성화되면서 이러한 기술은 전 세계 시청자에게 다양한 언어로 신속하게 자막을 제공하거나, 청각장애인을 포함한 다양한 시청자가 콘텐츠에 쉽게 접근할 수 있도록 돕고 있다.

6.3.1 음성 인식(Automatic Speech Recognition, ASR)

실시간 방송에서는 음성을 텍스트로 변환하여 자막으로 제공하는 기술이 적용될 수 있는데, 최근 급속히 성장한 딥러닝 기반의 음성 인식 기술이 이러한 실시간 자막 생성에 핵심 역할을 할 수 있다. 음성 인식 기술 발전으로 방대한 방송 데이터와 대규모 음성 데이터를 학습하여 빠르고 정확한 인식이 가능해졌기 때문이다. 특히, 한국어처럼 발음과 억양이 다양한 언어에서도 높은 정확도를 달성할 수 있는 최신 음성 인식 기술이 제안되

어 적용되고 있다.

6.3.2 자막 생성 기술

음성 인식을 통해 얻어진 텍스트를 자막으로 처리하는 기술도 미디어 분야에서 활용성이 높은 기술이다. 자막은 청각장애인을 위한 접근성 강화만이 아니라, 언어가 다른 시청자들에게도 핵심적인 요소로 작용한다. 최근 OTT 서비스로 방송 프로그램을 시청하는 활동이 증가하면서 화면에 자막을 켜고 프로그램을 시청하는 인구도 크게 증가했다. 원래 장애인의 시청권을 보장하기 위해 제공되던 폐쇄자막(closed caption)이 비장애인에게까지 유용한 기능으로 인식된 데에는 넷플릭스 등 OTT 시청시에 배속 시청이나 대중교통 내에서 소리 없이 시청하는 활동이 증가한 요인도 있어 보인다. 이에 따라 최근에는 SBS ‘법쩐’, ‘모범택시 시즌2’ 재방송과, MBC ‘수사반장 1958’ 본방송에서 자막이 들어간 영상을 제작하여 송출할 만큼 프로그램 제작과 송출에서 자막이 차지하는 영향이 커지고 있는 것으로 보인다.

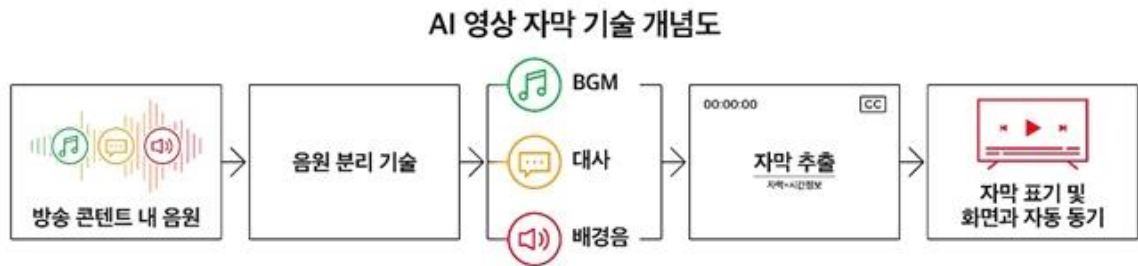
자막 생성 기술은 ASR 결과에 따라 문맥을 이해하고, 시청자에게 더 자연스럽게 명확하게 의미를 전달할 수 있도록 최적화되고 있다. 실시간 방송에서 이러한 자막 생성 시스템을 활용하면 기존 대본이 없는 상황에서도 자동으로 자막을 제공할 수 있으며, 이런 기능은 뉴스, 스포츠 중계, 예능 프로그램 등 실시간 정보가 중요한 방송 콘텐츠에서 매우 유용할 수 있다.

6.3.3 다국어 자막 번역 기술

OTT 이용이 글로벌화 되면서 방송 콘텐츠의 자막을 번역하고 이를 포함하여 유통시키는데 AI 기술을 활용하고자 하는 시도도 다양하게 시도되어 왔다. OTT 서비스와 같은 글로벌 미디어 플랫폼에서는 하나의 콘텐츠를 다양한 언어로 번역하여 제공하는 것이 중요한 과제일 것이다. 자막 번역 기술은 딥러닝을 활용한 신경망 기계 번역(NMT)을 통해 발전해왔으며, 특히 트랜스포머(Transformer) 모델이 등장하면서 번역의 자연스러움과 정확성이 크게 향상되었다. 미디어 자막 번역은 단순히 문장 구조를 변환하는 것에 그치지 않고, 문화적 차이를 반영하여 시청자에게 이해하기 쉬운 번역을 제공하는 것이 중요한데, 이를 위해 최근 AI 모델은 대량의 번역 데이터로 학습하여 문맥과 뉘앙스를 인식할 수 있도록 기술을 발전시키고 있다.

6.3.4 AI 기반 음성인식 및 자막 기술 방송 활용

JTBC스튜디오와 SKT는 딥러닝 기술을 활용하여 영상에 대한 한국어 자막을 자동으로 생성하도록 기술 개발을 진행하였다. SKT는 방송에 특화된 AI 음성인식 기술과 잡음제거 등의 기술을 개발하여 2021년 골프대회 영상을 통해 생방송에 적용한 바 있다.



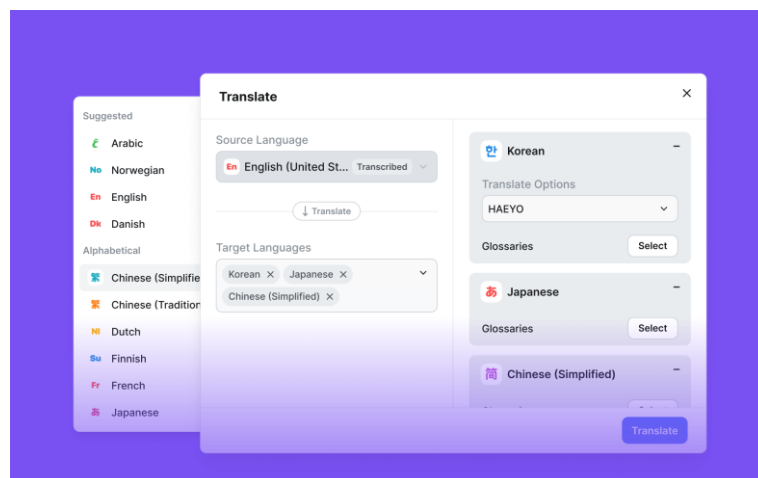
(그림 6-3) SKT AI 영상 자막 기술 개념도

[출처] SKT <https://news.sktelecom.com/171867>

더불어 SKT는 SK브로드밴드(SKB)와 함께 AI 기술을 활용한 한글 자막 생성 시스템을 개발에 참여하였고, SKB는 2023년 IPTV인 Btv와 모바일 Btv에서 VOD시청시 ‘AI 자막’ 기능 통하여 한글 자막 서비스를 이용할 수 있도록 서비스를 제공하였다. 보도에 따르면 AI 기술 적용에 따라 한글 자막 제공 시간이 3~4일에서 12시간 이내로 단축되었다.

KBS는 방영되는 콘텐츠의 음성을 속기사가 듣고 타이핑해서 폐쇄형 자막으로 제작하는 프로세스를 거쳐 방송을 송출한다. 이 때 폐쇄형 자막으로 만들어진 자막은 사내에 저장되며, 프로그램 영상을 홈페이지 등 인터넷에 유통할 때 해당 자막 정보가 함께 제공된다. 그런데 속기사가 타이핑할 때 발생하는 시점(싱크)이 영상과 맞지 않는 문제점이 존재하였고 이를 해결하기 위해 AI 기술을 활용한 ‘자막 동기화 시스템’을 개발하여 KBS 뉴스 홈페이지에 적용하였다. 그 결과 2024년 7월 기준 약 7만개 이상의 뉴스시가 자막이 오류 없이 제공되고 있다.

OTT 시청이 글로벌화 하면서 자막의 번역에도 AI 기술을 활용한 효율화가 진행되고 있다. 자막 시장에서 2022년 기준 전 세계 15% 점유율을 차지하고 미국 자막/더빙 1위 업체를 인수한 아이유노SDI미디어 그룹은 AI 기술을 활용하여 자막 제작의 완성도를 높이고 있다. 아이유노SDI는 자회사(XL8)에서 개발한 AI 번역을 위한 엔진을 활용하여 초벌 번역하는데 해당 엔진은 주요 OTT 콘텐츠의 음성과 자막을 기반으로 학습을 진행하여 회화나 구어체 번역에 특화 되어있다. 그 결과 일반적인 목적의 AI 번역에 비해 비디오 영상에서 번역 성능을 높였다.



(그림 6-2) XL8 사의 AI 번역기

[출처] XL8 <https://www.xl8.ai/products/mediacat-translate>

6.3.5 AI 기반 음성인식 및 자막 처리의 기술적 한계 및 과제

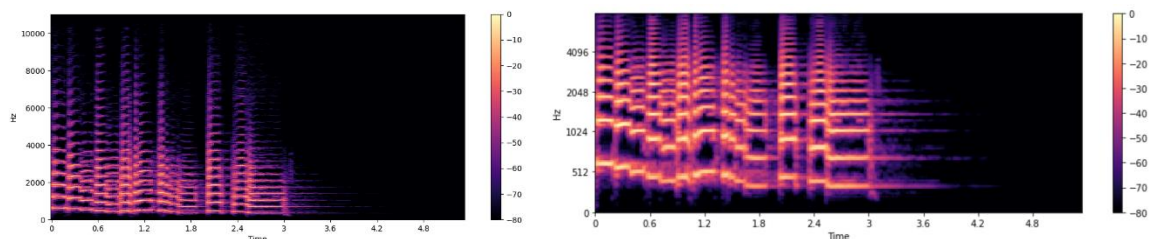
현재 AI 기술이 접목된 다양한 서비스들이 시도되고 있으나 복잡한 음향 환경에서의 인식을 저하, 감정 표현이 포함된 발화의 처리 한계, 동시다발적 대화 상황에서의 인식 정확도 저하 등은 여전히 해결해야 할 과제로 남아있다. 또한 방언이나 전문 용어의 경우 충분한 학습 데이터 확보의 어려움이 존재한다. 그 중 주요한 내용은 실시간 처리 속도와 정확도의 균형, 다양한 환경에서의 안정적인 성능 확보, 전문 분야별 특화 모델 개발 등이 주요 해결 과제이다. 특히 방송 환경에 특화된 학습 데이터의 지속적인 확보와 모델 최적화가 필요하다.

향후 해당 분야의 AI 기술 발전 방향으로, 자기 지도 학습과 전이 학습을 통한 적은 데이터로도 높은 성능을 발휘할 수 있는 모델 개발, 멀티모달 학습을 통한 영상-음성 통합 처리 기술 발전, 에지 컴퓨팅을 활용한 실시간 처리 성능 향상 등을 기대할 수 있다. 또한 AI 기술이 좀 더 고도화되면 감정 인식, 맥락 이해 등 더욱 섬세한 음성 처리가 가능할 것으로 전망된다.

6.4 음성 합성 기술 동향

6.4.1 AI 음성 합성 기반 기술

음성 신호 처리 기반 기술은 음성 신호의 특징을 추출하고 분석하는 기술로, 음성의 주파수, 진폭, 피치 등의 특성을 디지털 신호로 변환하고 처리한다. 음성 신호는 공기의 파동으로 아날로그 신호이므로 이를 컴퓨터에서 인식할 수 있는 데이터로 전환하기 위해서는 샘플링(sampling)과 양자화(quantization)을 통해 디지털화 하는 과정이 필요하다. 이렇게 디지털화 된 음성 신호는 푸리에 변환(Fourier Transform)을 활용하면 주파수 성분으로 분리할 수 있으며 역 변환 과정을 통해 원본 신호로 복원할 수 있게 된다. 최근 딥러닝 모델에서는 음성 특징을 포함하는 Acoustic feature를 이용하여 학습을 진행하게 되는데, 이 때 시간에 따라 변하는 신호의 주파수 스펙트럼을 시각적으로 표현한 스펙트로그램(spectrogram), 그 중 사람이 인지할 수 있는 정도로 여러 주파수를 묶어 데이터를 압축한 멜스펙트로그램(Mel-spectrogram) 형식으로 데이터를 가공하여 학습에 이용한다.



(그림 6-5) 딥러닝에 활용되는 음성 데이터인 Spectrogram<좌>과 Mel-spectrogram<우> 예제

[출처] https://huggingface.co/learn/audio-course/chapter1/audio_data

최근 음성 신호 처리 기반 기술로 트랜스포머(Transformer) 구조를 기반으로 한 음성 인식 모델이 주류를 이루고 있다. 이에따라 긴 시퀀스의 음성 데이터도 효과적으로 처리할 수 있게 AI 모델이 발전되고 있으며, 문맥 정보를 고려한 정확한 인식이 가능해졌다.

6.4.2 AI 음성 합성 기술 방송 현장 활용

KBS는 20204년 한국 최초로 생성형 AI ‘제니크’가 진행하는 ‘스테이션X’ 프로그램을 선보였다. 이는 네이버클라우드 기술을 적용한 것으로, 라디오 제작자, 진행자가 관여하던 선곡, 원고 작성, 진행, 보도자료를 망라한 다양한 업무를 AI DJ가 맡아 수행하는 것으로 기획되었다. 해당 프로그램은 매일 새벽 1시부터 1시간 방송이 송출되고 있다.

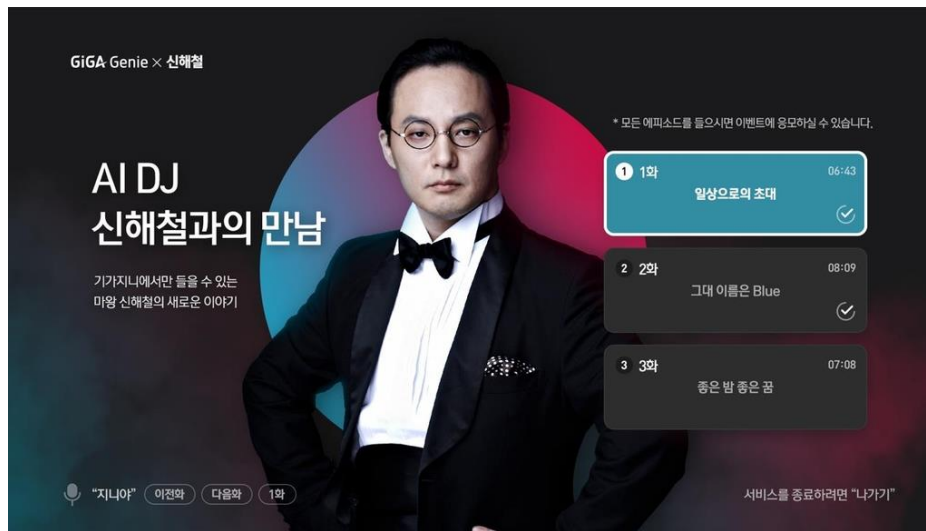


(그림 6-3) KBS 라디오 프로그램 ‘스테이션X’ 포스터

[출처] MyloveKBS

https://mylovekbs.kbs.co.kr/index.html?source=mylovekbs&sname=mylovekbs&type=blog&contents_id=70000000402009

KT는 2021년 AI 음성 합성 기술을 이용하여 고인이 된 가수 신해철의 음성을 복원하고 라디오방송 3편을 제작하였다. 학습에는 11년간 고인이 진행한 ‘신해철의 고스트스테이션’ 방송이 활용되었으며 KT가 개발한 개인화 음성합성 기술(P-TTS, Personalized Text to Speech)이 적용되었다. 한편 2024년에는 넥스트유나이티드가 고인의 방송, 강연, 공연 등의 데이터를 활용하여 새롭게 학습한 AI 모델을 'AI 신(新)해철' 이름으로 공개하였다.



(그림 6-7) KT에서 공개한 ‘AI DJ 신해철과의 만남’ 소개 포스터
[출처] <https://www.yna.co.kr/view/AKR20210912013400017>

SBS는 2023년 배우 최화정 목소리를 학습하여 ‘AI화정’을 만들고 8월 ‘AI 뮤직 페스티벌’을 통해 선보였다. 해당 방송에서는 SBS 미디어기술연구소에서 개발한 ‘AI Voice’를 활용하여 최화정 라디오 방송 데이터를 학습하였다. 해당 서비스는 제작자가 원하는 텍스트를 입력하면 AI로 학습된 ‘AI 화정’ 모델이 최화정 DJ의 목소리, 속도, 톤을 유지하여 발화하는 방식으로 수행되었다.



(그림 6-8) SBS AI 뮤직 페스티벌 소개 포스터 및 AI로 구현한 ‘AI 화정’
[출처] SBS

https://news.sbs.co.kr/news/endPage.do?news_id=N1007290829&plink=COPYPASTE&cooper=SBSNEWSEND
https://ent.sbs.co.kr/news/article.do?article_id=E10010274427

6.4.3 AI 음성 합성 기술적 한계 및 과제

AI 기술로 합성된 음성을 활용한 방송이 등장하고 있지만 여러 한계점이 여전히 존재한다. 한국 최초 AI DJ가 진행하는 KBS 라디오 프로그램 ‘스테이션X’에 대해서도 진행자의 목소리가 어색하다는 평가가 나오고 있으며, 이에 대해 담당 제작자 또한 “청취자들의 거부감이 있을 것이라고 예상해 일부러 심야 시간에 프로그램을 배치했다”고 인터뷰에서 밝힐 정도로 AI 음성 합성 기술이 아직 완벽한 상황은 아닌 것으로 보인다.

또한 아직 AI 합성 기술 자체가 장시간 발화에서의 음질 일관성 유지, 미세한 감정 표현의 한계, 실시간 처리 시의 품질 저하 등의 부분에서 부족함이 존재하며, 화자 특성의 완벽한 재현에도 한계가 존재하는 것이 사실이다. 또 자연스러운 운율이나 감정 등을 제어하기 위해서 이용자의 입력을 받으려 하는 경우 제어의 복잡성이 높아지기도 하므로 AI 기술 적용의 문턱을 낮추는 기술 개발이 더 가속화되어야 할 것으로 보인다.

향후 음성 합성 기술은 더 빠르게 발전이 가속화될 것으로 기대되는데, 특히 멀티모달 학습을 통한 상황 인지형 음성 합성, 초개인화 된 음성 복제 기술, 실시간 감정 변조 기술 등의 발전이 예상된다. 또한 적은 데이터로도 고품질의 음성을 합성할 수 있는 기술의 발전이 앞으로 기대된다.

6.5 결론

최근 방송 환경에서 AI 기술은 방송 제작 효율성과 콘텐츠의 접근성을 높이기 위해 다양하게 시도되고 도입되고 있다. 특히 음성 인식, 음성 합성, 자막 생성에 이용되는 AI 기술은 실시간 자막, 다국어 자막 서비스, 가상 DJ를 효율적으로 지원할 수 있는 툴이다. 해당 분야의 향후 AI 기술은 방송 현장에서 실시간 처리 성능의 최적화, 감정 표현 및 문맥 이해의 고도화, 다국어 자막 번역의 자연스러움 향상을 목표로 발전할 것으로 기대된다. 특히 AI와 멀티모달 학습 등의 발전으로 방송 콘텐츠의 품질을 높이고, 다양한 시청 환경에 적합한 맞춤형 서비스 제공을 가능하게 할 것으로 기대된다. 따라서 방송 업계는 앞으로도 방송 콘텐츠의 접근성, 다양성, 시청자 맞춤화를 실현하기 위해 이러한 기술적 시도를 지속하며 발전을 이어갈 필요가 있다.

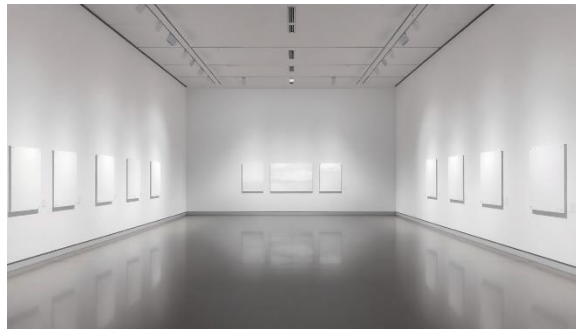
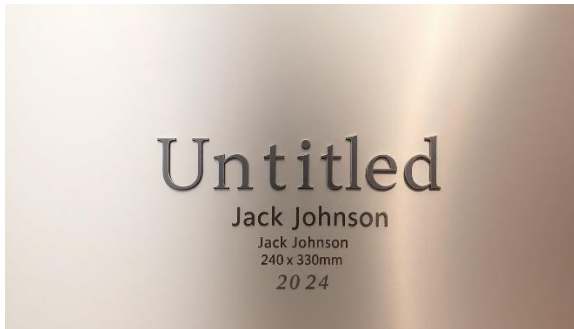
7 방송 제작을 위한 뮤직 오토태깅

7.1 뮤직 오토태깅이란

7.1.1 꼬리표, 메타데이터(Metadata)

캡션이 없는 미술관에 갔다고 상상해 보자. 작품은 사방에 널려 있는데 제목도 작가도 재료도 작품에 담긴 의도도 알 수가 없다. 관람객들이 작품을 감상하는 것보다 캡션을 읽는 데 더 많은 시간을 들인다는 지적도 있고 캡션이 작품에 대한 선입견을 형성한다는 비판도 있지만, 캡션이 없는 미술관은 웬만한 전문가나 애호가도 혼란스럽게 느낄 것이다.

캡션에 담긴 정보는 작품이 만든 세계로 들어가는 관문 역할을 한다. 심지어 캡션은 작품에 담긴 의미를 강화하거나 비트는 주요 정보를 제공함으로써 작품의 일부가 되기도 한다.



(그림 7-1) 미술작품 캡션의 예(생성형 인공지능 Flux로 생성) (좌)

(그림 7-2) 메타데이터 없는 미술관의 모습(생성형 인공지능 Flux로 생성) (우)

캡션에 담긴 정보들은 일종의 ‘메타데이터’라 할 수 있다. ‘메타데이터’는 ‘데이터에 대한 데이터’로, 말 그대로 데이터에 대한 정보를 제공하거나, 데이터로의 접근을 돕는 역할을 한다. 미술 작품을 정보의 집합체라고 본다면, 관람객의 감상행위란 예술적 체험을 하기에 알맞은 정보(의미)의 탐색이라고 할 수 있다. 그리고 그 탐색에 있어서 제목이 무엇인지, 누가 언제 만들었는지, 무엇을 재료로 만들었는지 등의 메타데이터는 이정표와 같은 역할을 해낼 수 있다.

7.1.2 메타데이터(Metadata)의 힘

메타데이터는 주로 청각 정보로 이루어진 음악 데이터에서 더 큰 힘을 발휘한다. 사람은 한눈에 넓은 장소를 볼 수 있기 때문에, ‘캡션 없는 미술관’에서도 여러 작품을 동시에 볼 수 있고, 제목은 모를지 언정 원하는 작품을 한눈에 찾아낼 수 있다. 책에서 특정 구절을 찾는 일도 마찬가지다.

책장을 앞뒤로 넘기면서, 여러 장을 한꺼번에 넘기며, 한눈에 여러 줄을 읽으면서 특정 구절을 찾는 일은 그리 어렵지 않을 것이다. 반면 시계열 청각 정보인 음악 데이터에 대

해서는 그렇게 하기 어렵다. 재생/되감기/빨리감기는 책장을 넘기는 속도나 미술관 한쪽 벽면을 한눈에 보는 속도에 비견할 수 없고, 여러 소리를 동시에 들으며 특정 악구를 찾는 것은 불가능에 가까울 것이다.

그런데 미술 작품에 캡션을 달듯 음악을 이루는 순간순간마다 꼬리표를 달 수 있다면 어떨까? 그리고 그 꼬리표에 담긴 정보를 바탕으로 원하는 악구를 찾는다면 어떨까? 이번에는 꼬리표를 책장 넘기듯 한눈에 읽고, 꼬리표에서 확인한 시간 정보를 바탕으로 꼭 필요한 재생, 되감기, 빨리감기만 할 수 있게 될 것이다. 꼬리표의 내용이 자세하고 정교하다면 여러 곡이나 악구를 동시에 찾는 일도 가능할 것이다.



(그림 7-3) 음악에 꼬리표(태그)가 달린 모습

7.1.3 뮤직 오토태깅 = 딥러닝(Deep Learning)이 달아주는 꼬리표

메타데이터의 힘이 이토록 강력한데도 모두가 음악에 시간대별로 태그를 달지 않았던 것은, 그 작업이 효과에 비해 매우 비효율적인 작업이기 때문이었을 것이다. 사람이 음악에 꼬리표를 달기 위해서는 위에서 언급한 바로 그 작업, ‘음악에서 특정 악구 찾기’ 같은 비효율적이고 지루한 반복작업을 견뎌내야만 한다. 그 작업을 위해 들어갈 시간과 노력의 크기는, 음악에 시간대별로 태그를 단다는 상상을 재산성 없는 광산처럼 여기게 했을 것이다.

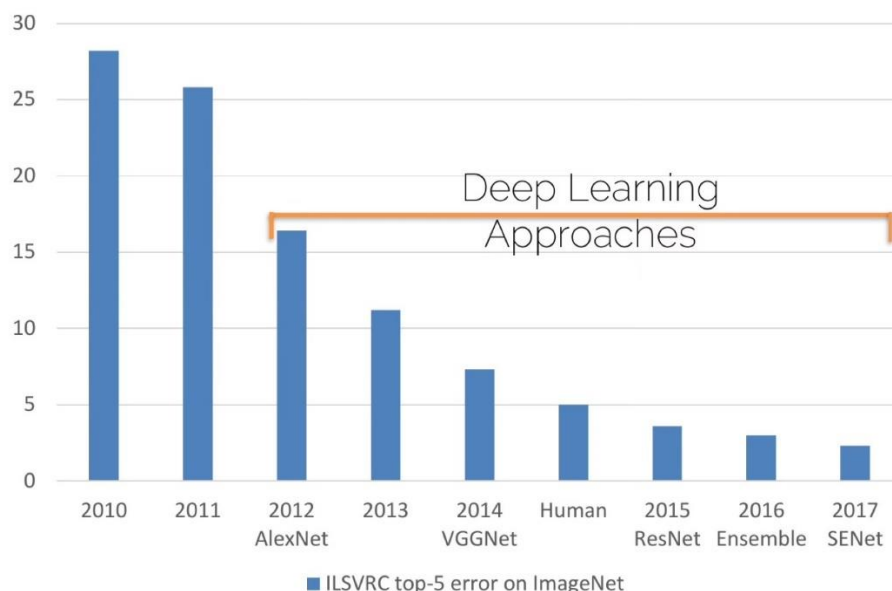
그런데 정보기술, 특히 딥러닝의 비약적 발전은 음악에 시간대별로 꼬리표를 붙이는 일도 쉬운 일로 만들었다. ‘뮤직 오토태깅’은 정보과학으로써 음악을 연구하는 ‘음악정보학’의 한 연구 분야로, 오디오에 자동으로 문자 태그를 다는 기술을 의미하는데, 이미 좋은 성과를 낸 연구가 많이 발표되어 현재는 ROC-AUC 기준 90% 이상의 성능을 낸 논문도 드물지 않게 찾아볼 수 있다.

‘뮤직 오토태깅’을 해볼만한 과제로 만든 기술은 역시 딥러닝이었다. 잠깐 본론에서 벗어나 딥러닝이 무엇이며 얼마나 발전했고, 딥러닝의 발전이 뮤직 오토태깅 등 음악정보학의 다른 연구 분야와 어떤 관계에 있는지 설명하고자 한다.

7.2 딥러닝(Deep Learning) 발전사와 음악정보학

올해 노벨물리학상 수상자는 물리학자가 아니었다. 평생 AI를 연구한 제프리 힌튼 교수가 노벨물리학상 수상자로 선정됐다. 화학상도 단백질 생성 AI '알파폴드'를 개발한 구글 딥마인드 CEO 등에게 돌아갔다. 분야를 막론하고 학계의 중심축은 딥러닝을 활용한 연구로 이동하는 중이다. 물리학이나 화학을 정석으로 연구한 학자들이 수상하지 못한 데 대해 물리학자들과 화학자들 사이에서 비판의 목소리도 나왔지만, 딥러닝 분야의 연구 성과가 학계의 다른 분야로 퍼져나가는 것은 10여 년 전부터 예견된 일이나 마찬가지였다.

가장 먼저 딥러닝이 '대세'가 된 분야는 컴퓨터 비전이었다. 컴퓨터 비전은 기계의 '시각'을 연구하는 컴퓨터 과학의 한 분야다. 인간이나 동물의 시각을 기계를 통해 재현하고자 하는 시도는 오래전부터 있었지만, 발전 속도는 더디었다. 컴퓨터 비전의 가장 대표적이며 오랜 연구 과제였던 '이미지 분류' 문제에서마저 발전 속도는 느렸다. 개 사진을 보고 '개' 카테고리로, 고양이 사진을 보고 '고양이' 카테고리로 분류하는 것은 인간에게는 너무나도 당연하고 쉬운 작업이지만, 기계에게 가르치는 것은 매우 힘들었다. 2009년 무려 천만 장의 이미지를 각각 1000개 카테고리로 분류한 데이터셋 'ImageNet'이 등장했고, 2010년부터 기계를 이용해 이 이미지들을 올바르게 분류하는 대회가 열렸다. 연구자들은 당시 기준 최첨단 기술들을 동원해 이미지 분류 기계를 만들었고, 이 대회에서 성과를 거뒀다. 당시 ImageNet만큼 규모가 크고 정교한 이미지 분류 데이터셋은 없었다. 그러므로 이 대회의 승자가 당시 세계에서 가장 뛰어난 이미지 분류 기계였을 것이다. 2010년 첫 대회의 우승자는 28.2%의 오분류율을, 2011년 우승자는 25.7%의 오분류율을 기록했다.



(그림 7-4) ImageNet 챌린지 역대 우승자별 오류율을 나타낸 그래프

[출처] Andreas Geiger, Tuebingen Univ.에서 발췌

오류율을 수 퍼센트나 낮추는 것은 지금이나 그때나 대단한 연구 성과였다. "이제 이미

지 분류 문제는 더 이상 발전할 데가 없다"는 의견이 나올 정도로 ImageNet에서의 25%는 엄청난 연구 성과였다. 그런데 이듬해인 2012년, 그 전까지 이미지 분류 문제에서 쓰이지 않던 아키텍처가 해성처럼 등장해 우승을 차지한 건 딥러닝이었다. 우승자인 AlexNet은 오분류율 16.4%를 기록했다. 이듬해에도, 그 다음에도, 그 다음 해에도 딥러닝 방식이 1위를 차지했다. 급기야 2015년에는 오분류율이 5%를 밑돌았다. 사람이 직접 분류해도 5.1%는 틀린다고 하니, 인간을 넘어선 셈이다. 그 이후로 ImageNet 분류 문제에는 아무도 매달리지 않게 됐다. 여기서 정확도 100%를 향해 연구하는 것은 더 이상 의미가 없었기 때문이다.

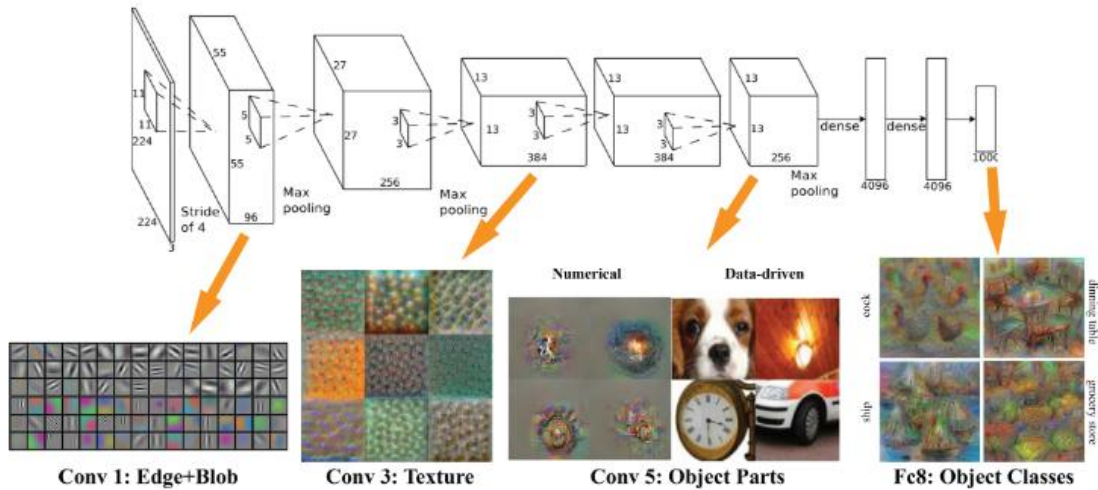
대신, 딥러닝의 연구 성과는 점차 다른 분야로 확산해 갔다. 2016년에는 구글 딥마인드의 '알파고'가 바둑에서 이세돌 9단을 이겼고, 이 연구는 결국 올해 노벨 화학상을 수상한 '알파폴드'의 바탕이 됐다. 불과 2년 전 시장에 등장한 'ChatGPT'는 매주 2억 5천만 명이 이용하는 서비스가 됐다. 생성형 인공지능 붐은 시장을 휩쓸고 있다. 올해 1월부터 8월까지 전세계 AI 앱 다운로드 수는 22억건에 달한다.

음악 분야에서도 딥러닝 기술이 빠르게 자리잡으며 성과를 내고 있다. 컴퓨터과학, 정보과학을 이용해 '시각'을 연구하는 게 컴퓨터 비전이라면, '음악'을 연구하는 것은 '음악 정보학'(Music Information Retrieval)이다. 음악정보학을 주제로 하는 세계적인 학회 ISMIR(International Society of Music Information Retrieval, 국제 음악정보학회)는 10년 전까지만 해도 컴퓨터과학 분야의 연구자들만이 아니라 음악학, 사회학 등 다양한 배경을 가진 연구자들이 모이는 학회였다. 그런데 10년 전부터 서서히 딥러닝 연구자들이 자리를 대신하기 시작하더니, 지금은 거의 모든 연구가 딥러닝을 통해 이뤄지고 있다.

7.2.1 ‘딥러닝’과 특징(Feature)

왜 ‘딥’러닝이라는 이름이 붙었을까? 최초의 CNN 아키텍처인 LeNet보다 3개 층 더 깊이 쌓아 만든 인공신경망이었기 때문이다. 일반적으로 은닉층을 깊게 쌓을수록 인공신경망의 성능은 좋아진다. 그러나 사실 현재 ‘딥러닝’으로 불리는 흐름의 핵심은 깊게 쌓아 만든 신경망이라는 점보다는 ‘기계가 특징(feature)을 알아서 추출해 학습’한다는 데 있다. 그리고 2012년 ImageNet 챌린지에서 우승하며 딥러닝의 시대를 연 AlexNet 이전에도 이러한 흐름은 있었고, ‘표현 학습(Representation Learning)’ 또는 ‘특징 학습(Feature Learning)’이라고 불렸다. ‘딥러닝’이라는 이름은 실패한 연구 분야로 악명 높았던 ‘Representation Learning’ 대신 연구에 대한 승인을 받기 위한 ‘마케팅용’ 이름으로서 탄생했다는 뒷얘기가 있다.

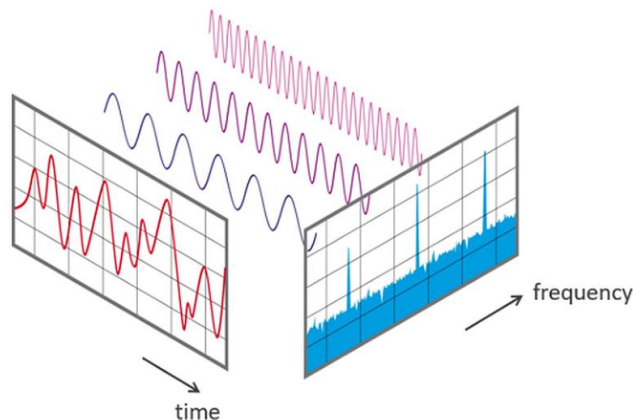
아래 그림은 ‘특징 학습’의 과정에서 기계가 추출하는 특징(feature) 또는 표현(representation)을 시각화한 것이다. CNN(Convolution Neural Network) 모델의 입력층과 가까운 층에서는 경계면(edge)이나 덩어리(blob) 등 원시적인 패턴을 특징으로서 추출해 내고, 중간쯤 되는 층에서는 이러한 패턴이 모여 형성한 질감(texture)을, 출력층과 가까운 층에 이르러서는 특정 개체의 부분을 인지하는 식이다.



(그림 7-5) CNN – Convolution and Pooling

그렇다면 ‘특징 학습’ 모델은 음악에서 어떤 특징을 추출하게 될까? 음악을 음고 (pitch), 음의 길이, 음의 세기 등의 정보를 담은 기호의 조합으로 본다면, 그러한 기호들의 배열과 조합, 패턴은 모델이 학습할 대상이 된다. 이 경우 음악은 NLP(자연어 처리) 연구에서의 언어처럼 다뤄지게 된다. 한편 STFT(Short-Time Fourier Transform) 등의 알고리즘을 통해 오디오 데이터를 스펙트로그램으로 변환한 뒤 스펙트로그램 속의 시각적 패턴을 학습하는 방법도 있다. 이 경우 음악은 CV(컴퓨터비전) 연구에서처럼 이미지로서 다뤄지게 된다. 이 글에서는 후자인 스펙트로그램(Spectrogram) 방식을 소개하고자 한다.

7.2.2 스펙트로그램(Spectrogram)

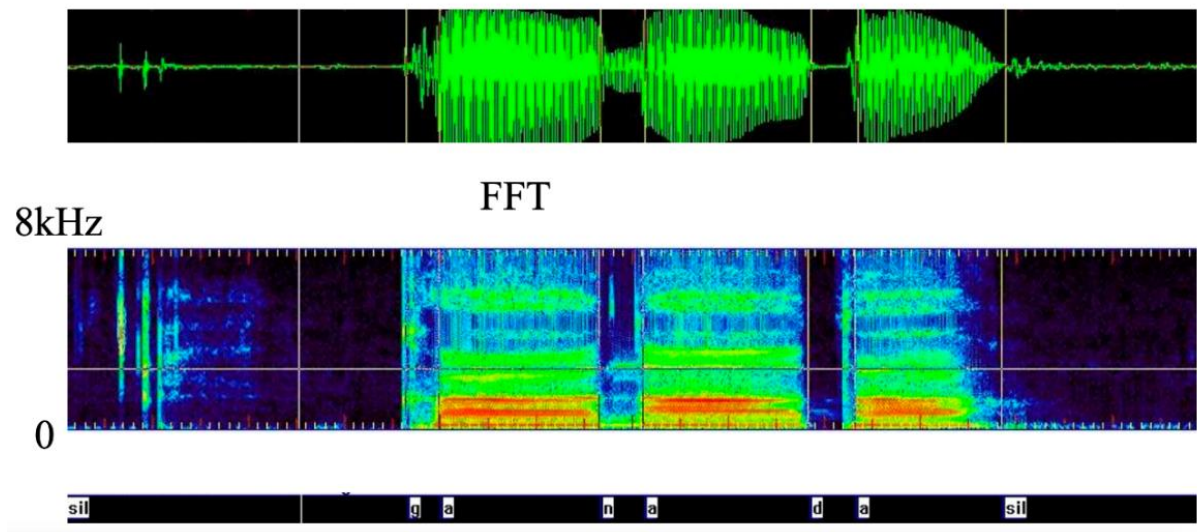


$$\text{STFT}\{x[n]\}(m, \omega) \equiv X(m, \omega) = \sum_{n=-\infty}^{\infty} x[n]w[n-m]e^{-i\omega n}$$

(그림 7-6) 일반적인 푸리에 변환

변환의 결과에는 매 수십 밀리 초(ms)마다 오디오 신호를 구성하는 주파수 성분들이

모두 담겨 있다. 시간, 주파수, 진폭의 3차원 데이터가 되는 셈이다. 이를 그래프로 나타낸 것이 스펙트로그램(Spectrogram)이다. 일반적으로 가로축은 시간, 세로축은 주파수 정보를 나타내는 데 쓰고, 그래프를 평면상에 나타내기 위해 진폭은 색상으로 표시한다.

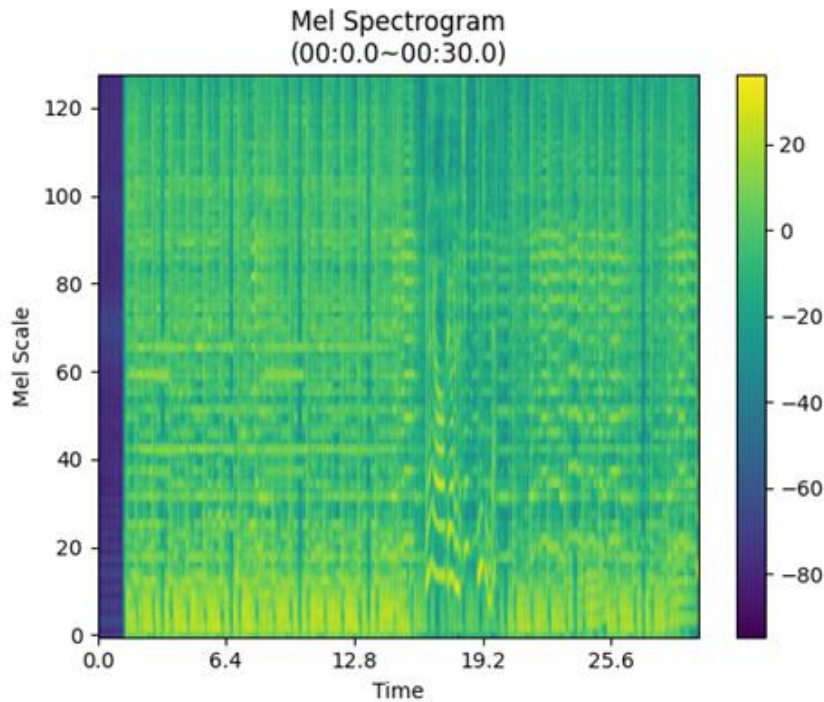


(그림 7-7) 아날로그 음성의 변환 과정

7.2.3 스펙트로그램(Spectrogram) 속 특징(Feature)들

스펙트로그램(Spectrogram)을 통해 음고(Pitch), 음가(Note Value), 음색, 소리의 크기 등 음악에서 중요한 소리의 요소들을 모두 확인할 수 있다. 음고(Pitch)는 스펙트로그램(Spectrogram)의 세로 막대 한 줄의 주파수 성분과 배음의 패턴을 통해 알아낼 수 있고, 음색은 주파수 성분의 분포로 알아낼 수 있고, 소리의 크기는 색상으로, 음가는 이러한 패턴들의 가로 길이를 재서 알아낼 수 있다.

특징 학습 모델이 스펙트로그램(Spectrogram)에서 추출해낼 특징은 일반적으로 사람도 육안으로 확인할 수 있는 정도의 패턴일 것이다. 예를 들어서, 오디오에서 음성과 음악을 분류하는 모델이 학습할 패턴은 우리 눈에 어떻게 보일까? 아래는 <김용신의 그애와 여는 아침>의 시작 부분의 오디오 30초를 Mel Spectrogram으로 변환하여 나타낸 결과이다. 2초경 오프닝 음악이 시작해서, 16초부터 오프닝 멘트, “안녕하세요, 그대와 여는 아침, 김용신입니다”가 발화되고, 20초쯤부터 다시 오프닝 음악이 나온다. 음성과 음악 16초부터 20초 사이에는 스펙트로그램(Spectrogram)에서 세로 패턴이, 나머지 시간대에서는 가로 패턴이 두드러진다는 것을 확인할 수 있다. 음성-음악 분류 모델은 이러한 패턴을 학습하게 될 것이다.



(그림 7-8) Spectrogram의 응용(세로축-Mel Filter Bank를 통과한 값)

7.3 오토태깅 모델의 구현

딥러닝, 음악정보학, 뮤직 오토태깅 등에 대해 알아보았으니 이제 뮤직 오토태깅 모델을 직접 만들어봄으로써 비용, 실용성 등을 따져볼 차례다. 모델 아키텍처로는 Auto-Tagging 연구에서 아직도 널리 쓰이고 있는 CNN을, 학습용 데이터셋으로는 MagnaTagATune을 사용했다.

7.3.1 모델 구현을 위한 데이터셋

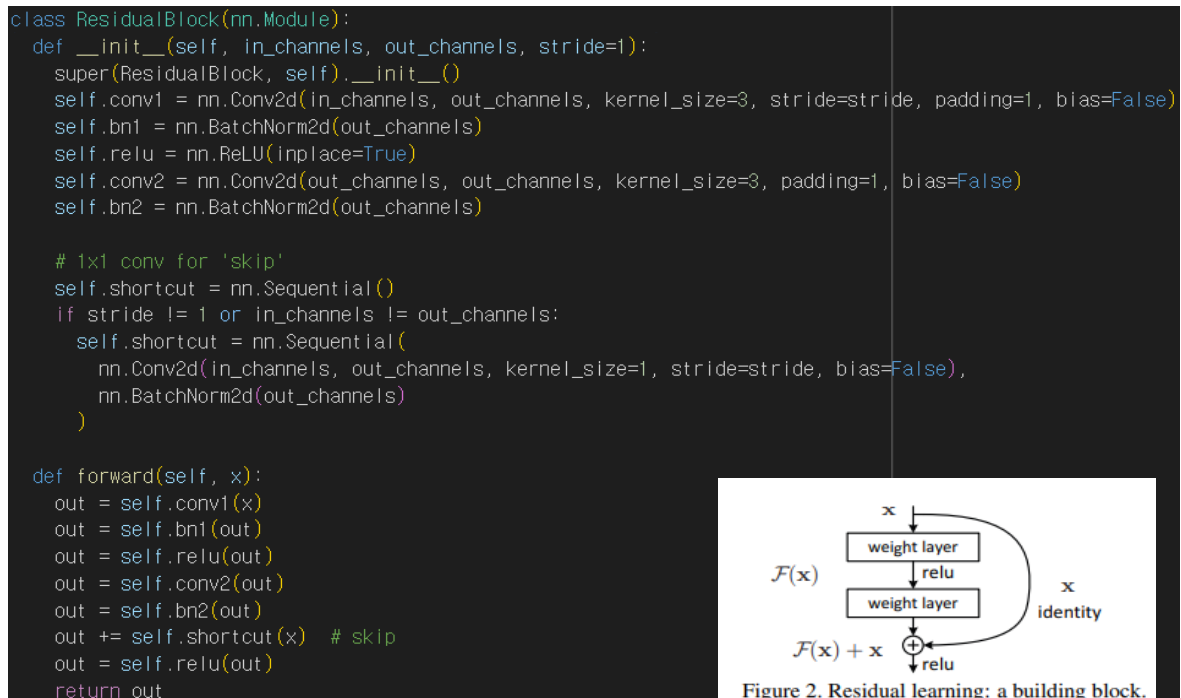
MagnaTagATune 데이터셋에는 25,863개의 음악 클립이 들어있다. 각 클립은 5223개의 노래, 445개의 앨범, 230명의 아티스트 중 하나에 속하는 29초 길이의 발췌본이다. 이 클립들은 클래식, 뉴에이지, 일렉트로니카, 록, 팝, 월드, 재즈, 블루스, 메탈, 펑크 등 다양한 장르를 포함하고 있다. 각 오디오 클립에는 188개의 태그에 대한 이진 주석 벡터가 제공된다. 이 주석들은 사람이 두 명에서 온라인 게임 TagATune을 하면서 얻은 것이다. 이 게임에서 두 플레이어는 동일하거나 다른 오디오 클립을 제시 받고, 각자의 클립에 대해 태그를 만들어야 한다. 그 후 플레이어들은 서로의 태그를 보고 동일한 오디오 클립이 제시되었는지 결정하게 된다. 두 명 이상의 플레이어가 동의한 경우에만 태그가 할당된다. 주석에는 '가수 있음', '가수 없음', '바이올린', '드럼', '클래식', '재즈' 같은 태그가 포함된다. 일반적으로 가장 인기 있는 50개의 태그가 평가에 사용되며, 이는 각 태그에 충분한 학습 데이터를 확보하기 위해서다. 데이터셋은 16개 파트로 나뉘며, 연구자들은 보통 파트 1-12를 학습에, 파트 13을 검증에, 파트 14-16을 테스트에 사용한다.

7.3.2 오디오 데이터 전처리

다만 구축 환경의 컴퓨터 성능 문제로 인해, MagnaTagATune 데이터셋에서 클립을 8000개, 태그는 빈도순으로 상위 50개만을 발췌해 학습, 검증, 테스트를 진행했다. 오디오 클립과 태그 데이터 쌍 중 5000개는 학습에, 1000개는 검증에, 나머지 2000개는 테스트에 사용했다. 학습과 검증 등의 과정을 고속으로 수행하도록 데이터셋 내의 모든 오디오 파일을 sampling rate 16kHz로 리샘플링해 메모리에 로드하였는데, 이렇게 데이터를 전처리해 메모리에 로드하는 데 2분 20초 정도가 소요됐다. 이 덕분에 학습에 소요되는 시간은 90% 가까이 경감했으나, 학습 및 검증 과정에서 프로세스가 점유한 메모리는 32GB에 육박했다.

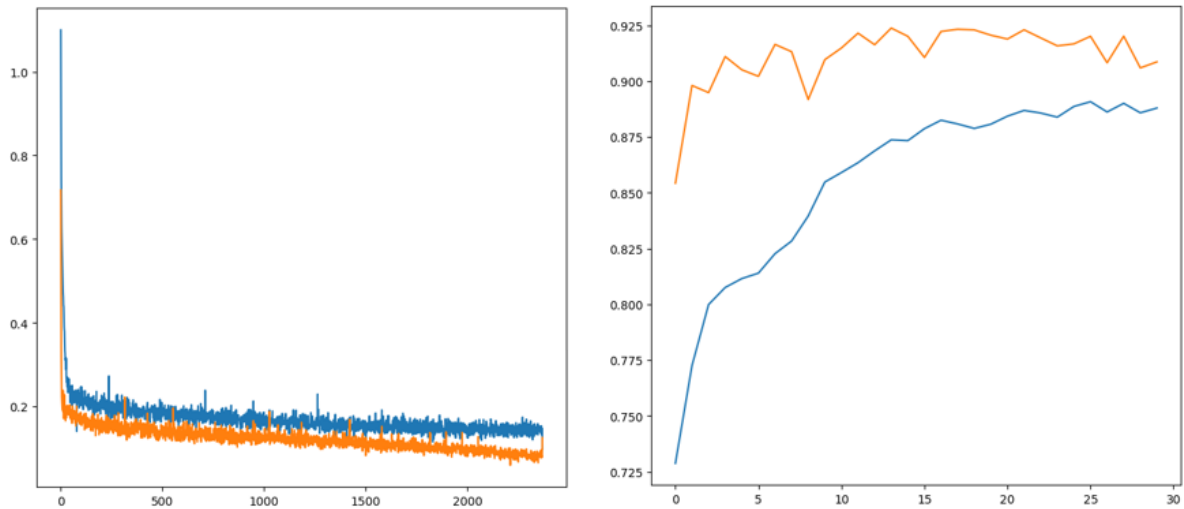
7.3.3 모델 학습

이렇게 정제한 데이터를 CNN 아키텍처의 가장 발전된 형태 중 하나인 ResNet를 통해 학습하도록 했다. Residual Network, 즉 ResNet은 레이어 입력을 기준으로 잔차(residual) 함수를 학습한다. 즉, 기존의 레이어들이 직접적으로 매핑 $F(x)$ 를 학습하도록 하는 대신, 잔차 네트워크는 레이어들이 잔차 매핑 $F(x) + x$ 를 학습하게 만든다. 이렇게 함으로써, 각 레이어들은 매핑 $F(x)$ 자체보다는 입력 x 와 매핑 $F(x)$ 의 차이를 학습하게 된다. 이 모델은 ResNet 계열의 모델 중 가장 연산량이 적은 ResNet-18 아키텍처를 따랐다.



(그림 7-9) Residual Block의 개념도(He 2015)와 Residual Block을 구현한 Python 코드

7.3.4 성능



(그림 7-10) CNN과 ResNet-18의 성능비교

좌측의 그래프는 (Conv-Pool-ReLU)x3 구조(청)와 ResNet-18구조(주황)에서의 Training Loss(BCE)의 비교를 보여주며 우측은 Validation Accuracy를 비교하고 있다. ResNet ah모델이 최대 92.56%의 정확도를 보이고 있다.

테스트셋에서의 ROC_AUC는 91.10%를 기록했다.

7.3.5 한계점

테스트 성능이 90%를 넘는데도 출력된 태그의 질은 상용화 가능한 수준에 미치지 못했다. MagnaTagATune 데이터셋 자체에 한계가 있기 때문인 것으로 판단된다. 예를 들자면, Oboe Quartet 곡인 Mozart KV. 370의 일부를 입력했더니 출력된 태그는 다음과 같았다.

['no piano', 'classical', 'string', 'violin', 'slow', 'no singer', 'cello', 'flute']

먼저, 'no piano', 'classical', 'string', 'violin', 'slow', 'no singer', 'cello'와 같이 7개 태그는 모델이 제대로 예측해낸 태그다. 하지만 이 곡에 flute는 등장하지 않는다. 원본 데이터셋에서 이 오디오 클립에 매겨진 태그를 확인해 보니 'classical', 'string', 'violin', 'slow', 'flute'가 있었다.

classical	string	violin	slow	flute
1	1	1	1	1

(그림 7-11) 추출된 실험 결과

'flute' 태그는 데이터셋 제작 과정에 참여한 사람들의 착각에 의해 붙었을 것이다.

MTAT 데이터셋의 제작에 교차 검증 과정이 있었는데도 걸러내지 못한 오류다. 오디오-태그 데이터셋의 질은 이처럼 레이블을 붙이는 사람의 감각에 종속될 수밖에 없다. 그리고 딥러닝 모델의 성능은 데이터셋의 질에 좌우될 수밖에 없다.

다른 한편으로는, 일반적으로 오디오 데이터를 다루는 데 고성능 컴퓨터가 필요하다는 점도 한계라고 할 수 있다. 이번 실험에서도 하드웨어 성능의 한계로 원본 데이터셋에서 오디오 클립은 8000개만, 태그는 빈도순으로 상위 50종류만 추출해내 사용했다. 이렇게 추출한 50개의 태그 중 ‘oboe’가 없었기 때문에 KV. 370 오디오 입력 후의 예측 결과에서 ‘oboe’ 태그를 기대할 수는 없었다. 또 이번 실험을 위해 사용한 컴퓨팅 자원들은 Google Colab의 Cloud상의 컴퓨팅 자원이어서, 이번에 구현한 오토태깅 모델을 영속적으로 사용할 수 없다는 점도 한계다.

7.4 결론

뮤직 오토태깅 기술은 딥러닝을 활용한 음악 정보 처리의 새로운 가능성을 제시하며, 음악 데이터를 효과적으로 분류하고 분석하는 데 중요한 기술이다. 앞으로 어떤 부문에서 이 기술이 활용될 수 있을지, 소요 비용과 성능 사이에서 최적의 선택지를 어떻게 찾을 수 있을지 고민하고 적용할 차례다.

ChatGPT, Claude 등 LLM 기반 생성형 모델들은 매우 유용하고 대중성 있는 기술이지만, 환각 등 고질적인 문제 때문에 모든 작업에서 만능키처럼 쓰일 수는 없다. 예를 들어 라디오 방송에서 음성만을 추출하거나, 청취자 메시지에서 신청곡만을 추출해야 한다면, 이 작업을 LLM에게 맡길 수는 없다. 환각 문제 때문에 결과를 신뢰하기 어려울 것이기 때문이다. 그 결과가 우리가 분류하라고 입력한 데이터에서 비롯된 것인지, 환각에서 비롯된 것인지 의심해야만 한다면, LLM을 쓰는 것보다는 사람이 직접 분류하는 것이 더 효율적일 수도 있다. 반면, 오히려 소형의 맞춤형 분류 모델은 이러한 작업에 있어서 LLM이나 인간 분류자의 대안이 될 수 있다. 적어도 분류 모델은 ‘데이터를 생성하는 법’을 알지 못하기 때문에, LLM의 환각과 같은 문제를 일으키지 않는다.

온디바이스(on-device) AI의 등장도 우리에게는 희소식이다. 온디바이스 AI는 서버 대신 스마트폰 등 기기 자체의 컴퓨팅 자원을 사용해 모델을 구축하고 이용한다는 취지인데, 온디바이스 AI의 등장은 이제 AI 관련 학계와 업계가 모델의 성능뿐만 아니라 경량화를 위해서도 애쓰고 있다는 증거가 된다. 실제로 Apple은 2023년 10월, ‘ReLU Strikes Back’이라는 도발적인 제목의 논문에서 활성화 함수 ‘ReLU’가 메모리를 큰 폭으로 절약하면서도 성능면에서는 뒤떨어지지 않기 때문에, 그 변형인 ELU, GELU, SiLU 등보다 더 유리하다는 주장을 펼쳤다. 삼성전자가 온디바이스 AI를 탑재한 스마트폰을 출시한 상황에서, 이 논문의 발표는 Apple 또한 AI 모델의 경량화에 애를 쓰고 있다는 의미로 해석된다. 이러한 업계의 동향은 경량화된 고성능 아키텍처의 발전으로 소규모/저예산 환경에서도 도전할 수 있는 영역이 넓어질 것을 시사한다.

향후 연구는 더 정교하고 품질 높은 데이터셋을 활용하고, 경량화된 모델 아키텍처를 도입해 실용성을 높이는 방향으로 나아가야 한다. 이를 통해 비용과 성능의 균형을 이루며, 딥러닝 기반 뮤직 오토태깅 기술이 방송 제작과 음악 정보 분석에 필수적인 도구로 자리 잡을 수 있을 것으로 생각한다.

8 라디오 콘텐츠 소비 다양화와 온라인 스트리밍 통계

8.1 라디오 콘텐츠 소비의 다양화

8.1.1 스마트폰 청취 앱

국내 주요 지상파 라디오 방송사들은 라디오 청취자의 청취 행태 변화에 맞춰 이미 오래전부터 온라인 라디오 스트리밍 청취 앱을 개발해 서비스를 제공해 왔다. 라디오 온라인 청취의 다양한 형태 중 큰 비중을 차지하고 있으며 SBS의 고릴라, KBS 콩, MBC 미니, CBS의 레인보우가 대표적인 서비스다. 방송사 자체적으로 어플리케이션을 개발하여 서비스하고 있으며, 대부분 음성 대체광고 및 디스플레이 배너 광고 등 각 사별로 별도의 수익 모델을 가지고 있다.

라디오 방송사들이 정확한 수치를 공개하고 있지 않지만 스마트폰 어플리케이션 시장 분석 회사의 자료에 따르면 각 사 어플리케이션의 월간 활성 사용자(MAU: Monthly Active Users)가 적게는 50만 이상에서 100만 명 이상에 이르는 것으로 조사된 바 있다. 한 명의 사용자가 여러 방송사의 앱을 중복으로 사용하는 경우를 고려하면 라디오 온라인 청취 앱으로 라디오를 청취하는 사용자는 전체적으로 약 250만명 정도로 추정된다.

한편, 지역이나 방송사 규모에 따라서 자체적으로 라디오 청취 어플리케이션을 갖출 수 없는 방송사들도 있어 모든 라디오 청취자에게 보편적인 서비스를 제공해야 한다는 측면에서 정책적으로 라디오 통합 어플리케이션 개발에 대한 논의가 오랫동안 있어 왔으나, 방송사 간 이해 관계 및 추진 주체와 예산 등의 여러 현실적인 문제로 인해 현재는 논의가 중단된 상황이다.

8.1.2 PC플레이어 및 라디오 홈페이지

온라인 라디오 스트리밍 어플리케이션 이전에는 PC나 노트북에서 주로 온라인 라디오를 청취했다. 별도의 클라이언트 프로그램을 설치하는 형태로 SBS의 고릴라, KBS 콩, MBC 미니, CBS의 레인보우 모두 처음에는 PC플레이어 형태로 서비스를 시작했다. 예전에 비해 사용량이 많이 줄긴 했지만 현재도 적지 않은 사용자가 PC플레이어를 통해 온라인 라디오를 청취하고 있다.

또한 라디오 방송사 마다 별도의 홈페이지를 운영하며 편성 정보를 제공하고 청취자 게시판도 관리하고 있다. 방송사마다 다른 경우가 있지만 일반적으로 홈페이지에서 자사의 온에어 채널을 청취할 수 있도록 서비스하는 경우가 많다.

8.1.3 스마트 디바이스

2019년 경부터 통신사를 중심으로 국내에 AI스피커가 보급되기 시작하면서 AI스피커를 통해 온라인으로 라디오를 소비하는 형태가 확산되기 시작했다. SK텔레콤의 ‘누구’ 서비스를 비롯하여 KT의 ‘지니’, LU유플러스 ‘글로벌’ 와 같은 통신사를 시작으로 카카오의 ‘헤이 카카오’, 삼성의 ‘빅스비’까지 다양한 회사와 제휴를 통해 라디오 온라인 스트리밍

서비스를 제공하고 있다. 최근에는 스피커 형태 외에도 셋톱박스, 스마트TV 등의 다양한 형태의 스마트 디바이스로 서비스를 확장해 나가고 있다. 이런 형태의 서비스 제휴는 온라인 서비스 제공에 따른 CDN 비용만 고려한다면 별도의 애플리케이션 개발이 필요 없어 부담이 적고 청취에 대한 소비량에 대한 측정도 비교적 용이하여 활발하게 확장되고 있다.

8.2 온라인 라디오 청취의 확대

8.2.1 유튜브

최근 2년동안 라디오를 유튜브로 소비하는 사용자가 크게 증가했다. 처음에는 방송사 채널별로 하이라이트나 재미있는 부분을 편집하여 유튜브 클립으로 올리는 형태가 대부분이었으나, 이후 개별 방송 프로그램별로 방송 편성시간과 동일한 시간에 유튜브 라이브를 동시에 진행하는 경우도 많아졌다. 이러한 현상은 특정 방송사의 한 시사 프로그램이 이슈가 되면서 청취자가 대거 유튜브로 이동한 이후 많은 시사 프로그램들이 유튜브 라이브를 동시에 진행하면서 가속화된 측면이 있지만 지금은 시사에 관심이 많은 중년층 뿐 아니라 유튜브에 친숙한 젊은 세대를 중심으로 사용자가 증가하고 있는 추세이다.

채널	구독자 현황	설정
SBS 에라오	147 만	음악 예능 큐레이션
KBS 쿨FM	142 만	KBS 예능 라디오
MBC 라디오 시사	134 만	MBC 시사
KBS 1라디오	130 만	KBS 시사
CBS 뉴스쇼	128 만	CBS 라디오 시사
MBC 모험즈	71.9 만	MBC 큐레이션
SBS 시교라	45.1 만	SBS 시사교양 채널

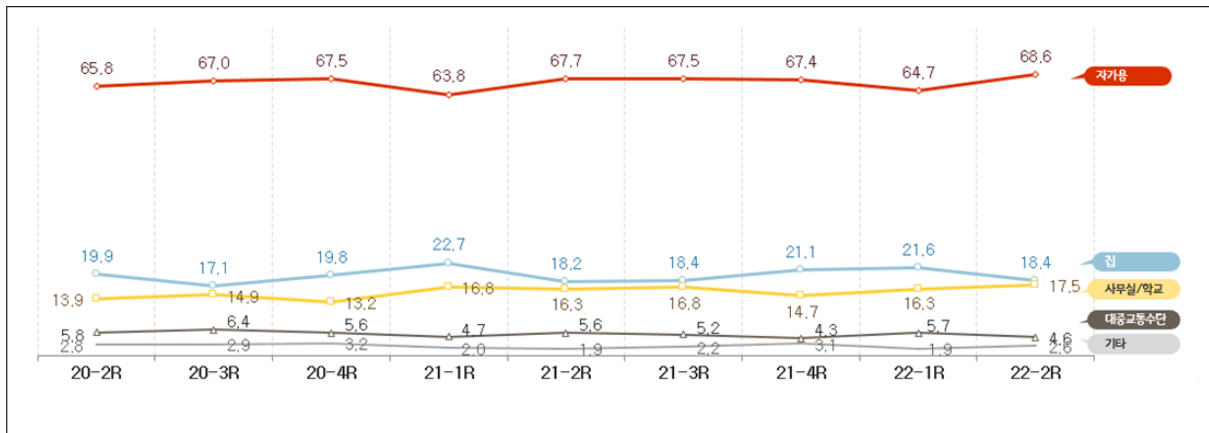
[표 8-1] 라디오 유튜브 채널 현황 (2024.6)

8.2.2 커넥티드카 (국내 브랜드 완성차)

자동차에서 라디오 청취는 전체 라디오 청취의 70%에 달할 정도로 대부분의 라디오 청취가 이루어지는 공간이다. 현재는 자동차의 라디오 주파수 수신 장치를 통해 청취가 이루어지고 있지만, 현재 자동차 업계는 전기차 및 AI 혁신과 맞물려 빠른 속도로 커넥티드카 전환이라는 변화가 이루어지고 있다는 점을 고려하면 머지 않은 미래에 차량 내 라디오 청취 방식에 있어서도 큰 변화가 예상된다. 커넥티드카란 운전자와 자동차를 정보통신기술(ICT)로 연결해 양방향 소통이 가능한 차량을 의미하며, 운전자는 차량에 기본 탑재되는 IP기반의 인포테인먼트 서비스 (내비게이션, 실시간 교통정보, 뉴스, 게임, 음악 등)를 즐길 수 있다.

국내 차세대 커넥티드카 라디오 서비스에 대한 논의는 현대차와 주요 지상파 라디오 방송사들이 함께 수 년 전부터 논의해 왔다. 작년 하반기에 현대, 기아차를 대상으로 라

디오 하이브리드 서비스를 시작한 이후 별다른 논의가 없다가 최근 현대차에서 차세대 커넥티드카에 대한 전략적인 방향이 어느 정도 구체화되면서 다시 논의가 시작되고 있는 상황이다.



(그림 8-1) 라디오 청취 장소 (한국리서치 MRS청취율 조사)

국내 주요 라디오 방송사와 현대차가 공동으로 개발한 차량내 라디오 하이브리드 서비스란 자동차에서 라디오를 청취할 때 음성은 주파수를 통해서 수신한 음성을 듣고, 온라인을 통해서 수신한 라디오 프로그램 편성 정보 및 프로그램 대표 이미지 등의 부가 정보를 차량 내 대시보드 모니터를 통해 제공받는 서비스다. 2023년 9월부터 하이브리드 라디오 서비스를 시작했으며, 현재 300만대 이상의 차량에서 서비스를 이용하고 있다.

하이브리드 라디오 방식은 기본적으로 주파수를 통해 청취가 이루어지며, 편성 정보는 API를 통해 주기적으로 업데이트 하는 방식이므로 차량 내 실제 라디오 청취량을 알 수 없다. 커넥티드카에서 최종적으로는 100% IP기반의 온라인 라디오 서비스가 구현될 가능성이 크다고 할 수 있음에도 불구하고 이러한 중간 형태의 서비스를 거치는 이유는 자동차 업계가 이니셔티브를 쥐고 있는 상황에서 라디오가 개별 앱이나 오브젝트로 전략하지 않기 위해서 선제적으로 ‘라디오 섹션’으로 들어가야 한다는 점이 크게 작용했으며, 커넥티드카 전환 시대에 운전자들에게 차별화된 라디오 청취 경험을 미리 제공하고, 라디오 매체에 대한 니즈를 유지하는 데 의의가 컸다고 할 수 있다.

현대차는 최근 차량 내 데이터 사용 비용에 대한 정책과 함께 차세대 차량용 OS에 대한 전략 방향을 정하면서 새롭게 차세대 인포테인먼트 전략을 수립하고 라디오 방송사와 협력을 준비하는 움직임이 있다. 차량 내 라디오 청취 방식은 차량 내 라디오 청취 비율을 고려하면 IP기반의 온라인 스트리밍인지 주파수인지에 따라 한 쪽이 과반이 넘어가는 지배적인 방식이 결정되기 때문에 라디오 산업에 전반적으로 영향을 미칠 수 있는 사안이라고 할 수 있다.

8.2.3 커넥티드카 (해외 브랜드 완성차)

흔히 외제차라 부르는 해외 브랜드 완성차 업계도 활발하게 차세대 커넥티드카 모델에 대한 준비와 여러가지 시도들을 하고 있으며, 라디오 온라인 스트리밍 서비스의 관점에서 크게 두 가지 형태로 구분할 수 있다.

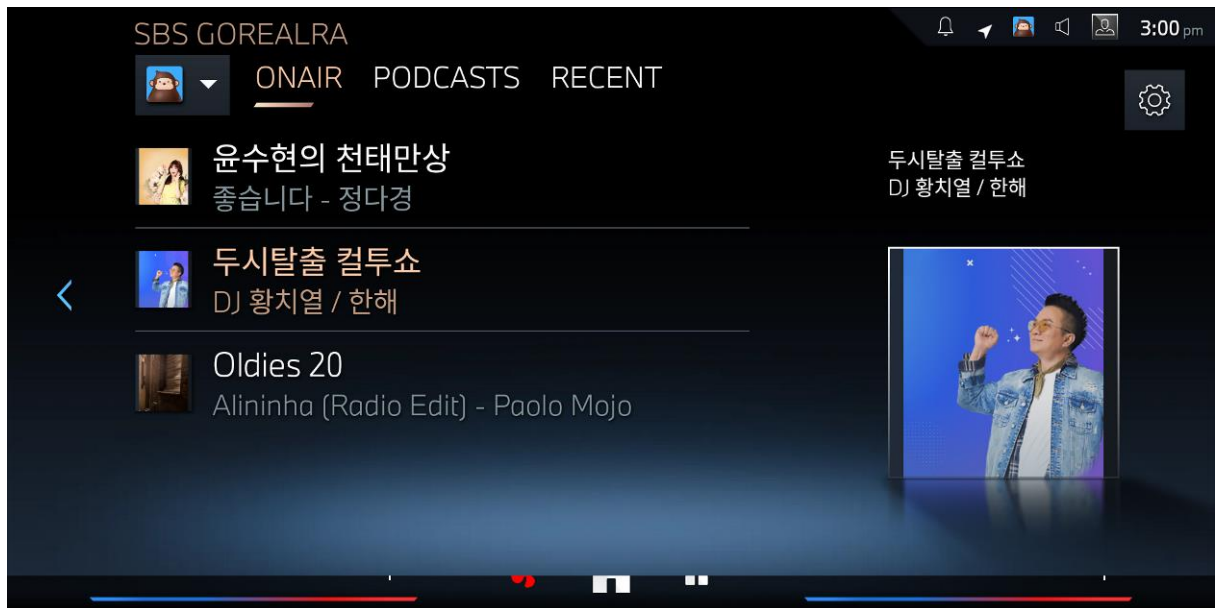
8.2.3.1 미들웨어를 통한 서비스 제공

완성차 업체와 라디오 방송사 중간에 미들웨어 서비스 제공 업체가 온에어 스트리밍 품질 및 방송 편성 정보 등의 제공을 담당하는 방식으로 서비스 제공하는 방식을 의미한다. 이 방식은 하나의 완성차 업체가 개별 라디오 방송 채널을 관리하는 어려움과 라디오 방송사 입장에서 여러 완성차 업체에 개별적으로 서비스를 제공해야 하는 번거로움을 줄일 수 있다는 장점이 있다. 북미와 일본에서는 엑스페리(X-peri) 라는 업체가 DTS autostage 라는 미들웨어를 통해 라디오 온라인 스트리밍 서비스를 메르세데스 벤츠, BMW 등에 제공하고 있으며 국내에서도 몇 군데의 라디오 방송사들이 엑스페리를 통해 메르세데스 벤츠에 라디오 온라인 서비스를 제공하고 있다.

라디오 방송사는 미들웨어 업체에 온라인 스트리밍 정보 및 라디오 편성 정보를 제공하고, 미들웨어 업체는 차량에 설치된 미들웨어를 통해 수집되는 라디오 청취 데이터를 이용하여 각 사별로 차량내 라디오 사용 통계를 방송사에 제공한다. 방송사 입장에서는 스트리밍과 편성 정보만 제공하기 때문에 데이터 및 부가 서비스에 있어 자유도가 낮다고 할 수 있다.

8.2.3.2 차량용 앱을 통한 서비스 제공

차세대 자동차는 운영 체제를 갖춘 하나의 움직이는 컴퓨터라고 할 수 있다. 완성차 업체별로 차이가 있지만 OS에 있어 크게 리눅스 기반의 커스터마이징 된 자체 OS를 사용하는 쪽과 안드로이드 오토모티브 OS를 사용하는 쪽이 있으며, 안드로이드 오토모티브는 개별 차량용 OS 보다 앱 마켓 생태계를 유지하기가 용이한 측면이 있어 새로운 모델에 있어서 점점 안드로이드 오토모티브를 채택하는 쪽이 증가하고 있다. 유럽 쪽에는 이미 유럽에 기반을 둔 다양한 완성차 업체를 고객으로 두고 차량용 앱을 소싱하는 업체들이 있다. 포르투갈의 포레시아(Faurecia)는 그 중 하나로 운전자에게 편의를 제공하는 다양한 차량용 애플리케이션을 소싱하고 있다. SBS 라디오는 올해 1월 포레시아의 앱 마켓인 앱토이드(Aptoid)를 통해 한국에 출시되는 일부 BMW 신형 모델에 차량용 고릴라 서비스를 제공하고 있다. 한번 개발된 차량용 고릴라 앱은 안드로이드 오토모티브 OS를 탑재한 자동차라면 간단한 테스트 과정을 거쳐 손쉽게 다른 브랜드의 차량으로 확장할 수 있는 장점이 있다. 방송사 입장에서는 자체적으로 개발하는 서비스이기 때문에 자사만의 부가서비스 및 데이터 측면에서 유리한 방식이라 할 수 있다.



(그림 8-2) SBS 고릴라 차량용 앱 화면 예시

8.3 온라인 라디오 스트리밍 청취 통계의 장점 및 활용

지금까지 간략하게 라디오 온라인 스트리밍 서비스가 이루어지는 다양한 형태들을 살펴보았다. 현재도 추가적으로 다양한 디바이스와 플랫폼을 통해 온라인으로 라디오 콘텐츠를 제공하는 다양한 논의가 진행하고 있으며, 이는 라디오 매체가 존재하고 기술 발전이 이루어지는 한 지속적으로 시도될 것이며 결과적으로 온라인 청취는 증가하는 방향으로 변화할 수밖에 없다.

그러나 라디오 청취 환경이 점점 다변화하고 복잡해짐에 따라 기존의 청취율 조사 방식으로는 콘텐츠 평가 뿐 아니라 청취자에 대한 이해와 파악이 점점 어려워지고 있다. 이에 비해 온라인 라디오 청취는 주파수를 통한 라디오 청취와 달리 개개인의 청취 행위에 대해 어플리케이션 로그 기록을 남길 수 있고, 이를 통해 정확한 라디오 청취 통계 수립이 가능하다.

어플리케이션 로그 데이터를 이용한 데이터 분석 방식은 3,000명의 샘플을 통해 기억에 의존하는 설문 조사 방식에 기초한 기존의 라디오 청취율 조사 방식에 비해 수백만에 이르는 온라인 라디오 청취자에 대한 전수 조사가 가능하고, 로그를 통해 남겨진 정확한 기록을 이용해서 분석하고, 3달에 한 번씩 실시하며 2주에 걸친 조사 기간으로 인해 발생하는 ‘시간적 지연’ 대신 실시간으로 분석이 가능한 장점을 가지고 있다.

어플리케이션 로그 데이터를 이용한 데이터 분석이 가지는 이러한 장점에도 불구하고 ‘데이터 오너십’ 문제로 인해 자사의 어플리케이션 로그 데이터에만 접근이 가능한 점으로 인해 방송사간 비교가 불가능하며, 전체 라디오 온라인 스트리밍 청취에 대한 측정 통계를 생성할 수 없다는 한계점이 있었다. 올해 마련된 라디오 온라인 스트리밍 측정 방식에 대한 표준화는 ‘데이터 오너십’에 대한 문제를 피하면서 전체 온라인 라디오 청취에 대한 통계 생성이 가능한 대안이 될 수 있다.

앞서 살펴본 바와 같이 라디오 온라인 청취는 조금씩 지속적으로 증가하고 있으며, 차량에서의 청취 방식의 변화에 따라 급격하게 증가할 수도 있다. 지금과 같이 다양한 라

디오 콘텐츠 소비가 이루어지는 환경에서 기존과 같은 청취율 조사 방식으로는 라디오 청취자의 청취 행태에 변화에 대해 정확하게 파악하기에는 한계가 있으므로 라디오 온라인 스트리밍 측정 방식에 대한 표준을 이용한 측정 통계를 적극 활용할 필요가 있으며, 향후 증가하는 라디오 온라인 청취에 대해서도 플랫폼에 관계없이 온라인 스트리밍 통계를 생성 가능한 형태로 확장해 나가는 것이 필요하다.

9 FAST 서비스 광고 삽입 기술

9.1 개요

FAST (Free Ad-Supported Streaming TV)는 단어 뜻 그대로 광고 기반 무료 스트리밍 TV를 의미한다. 광고를 보는 대신에 무료로 볼 수 있는 실시간 채널 서비스다. 온라인 영상 시청 서비스는 기존에는 VOD (Video On Demand) 방식으로 콘텐츠를 소비하는 구독형 OTT가 많은 비중을 차지했다면, 최근에는 전통적인 리니어 방식의 TV 포맷이 다시 관심을 얻고 있다. 전통적인 리니어 방식의 서비스인 FAST는 국내 시장보다는 미국 등 해외 시장에서 점유율을 높이고 있으며, 이제 주요 스마트TV나 셋톱박스에 내장되어 출시되고 있다. 이 장에서는 방송사에서 FAST 채널을 서비스하기 위한 광고 삽입 기술에 대해 정리하고자 한다.

9.2 FAST 서비스 특징

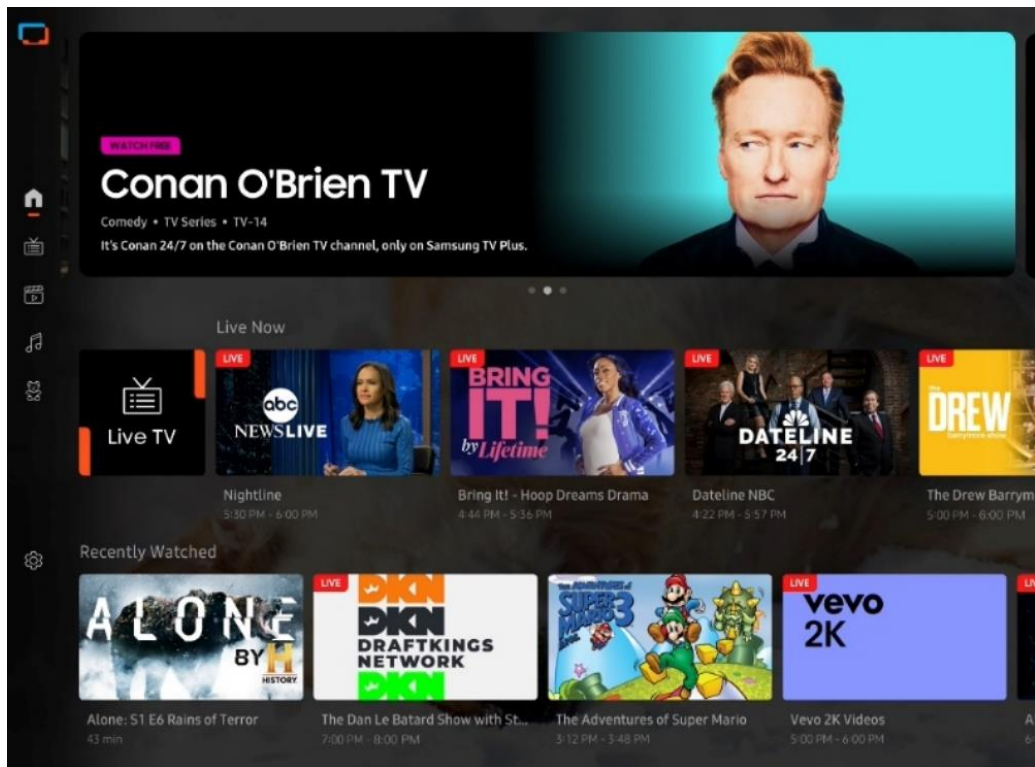
시청자 입장에서 FAST 채널은 무료로 볼 수 있다는 장점이 있다. 기존 전통적인 리니어 방식의 TV 채널(지상파TV 등)과 동일한 시청 형태로 광고를 시청함으로써 무료로 콘텐츠를 볼 수 있다. 이미 지상파TV 등으로 익숙해진 시청 방식으로 서비스함으로써 시청자가 편안하게 느껴질 수 있다는 점도 장점이다. 특별히 서비스에 가입하지 않아도 스마트TV나 셋톱박스에 내장된 서비스를 통해 쉽게 접근이 가능하다는 점도 장점으로 꼽을 수 있다.

방송사 입장에서 FAST 채널을 통해 구작 콘텐츠에 대한 활용도를 높이는 등 다양한 사업 기회를 얻을 수 있다. 다양한 구작 콘텐츠를 FAST 채널로 편성하면, 직접적으로는 더 많은 광고 슬롯을 확보하고 시청자에게 온라인 광고를 집행할 기회가 늘어나게 된다. 또한, 간접적인 효과로 구작 콘텐츠를 FAST로 시청하는 시청자들에 대한 구독형 VOD 서비스로의 유입을 기대할 수 있다.

국내 TV 제조사인 삼성전자와 LG전자도 자사 스마트TV를 통해 FAST 채널을 제공할 수 있는 플랫폼인 삼성TV플러스와 LG채널을 운영하고 있으며, 많은 방송사가 이 플랫폼으로 다양한 FAST 채널을 서비스하고 있다. 이 FAST 플랫폼은 스마트TV나 셋톱박스과 같은 고정형 장치만이 아니라 모바일 앱을 지원하여 더 많은 시청자 확보가 가능하다.

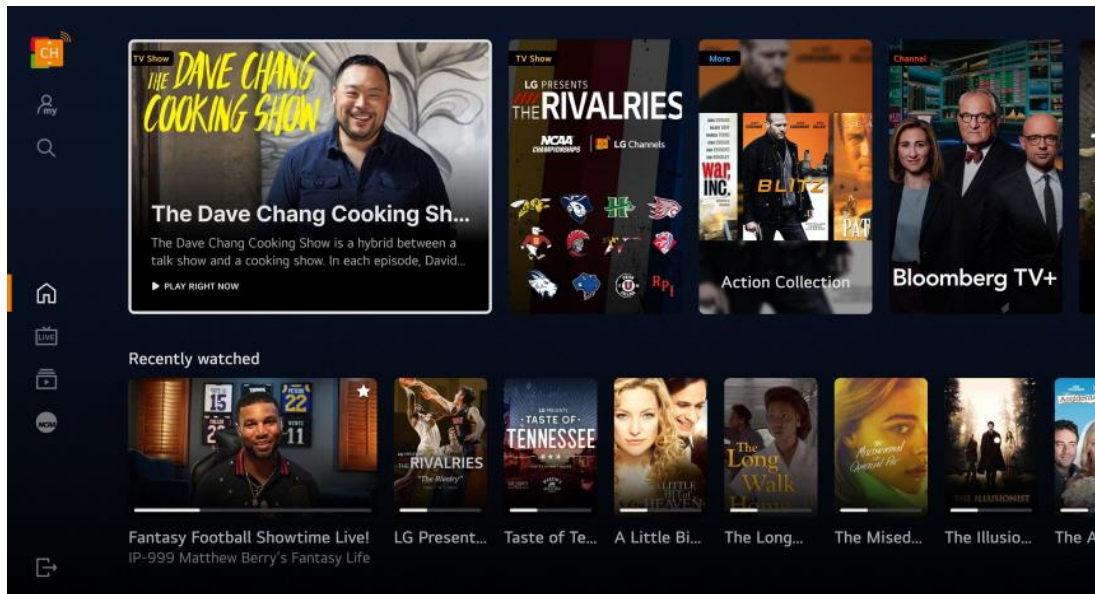


(그림 9-1) 주요 FAST 서비스 플랫폼



(그림 9-2) 삼성TV플러스 서비스 화면

[출처] <https://news.samsung.com/global/samsung-tv-plus-update-makes-browsing-and-viewing-content-easier-and-faster>



(그림 9-3) LG채널 서비스 화면

[출처] <https://www.lgcorp.com/media/release/26787>

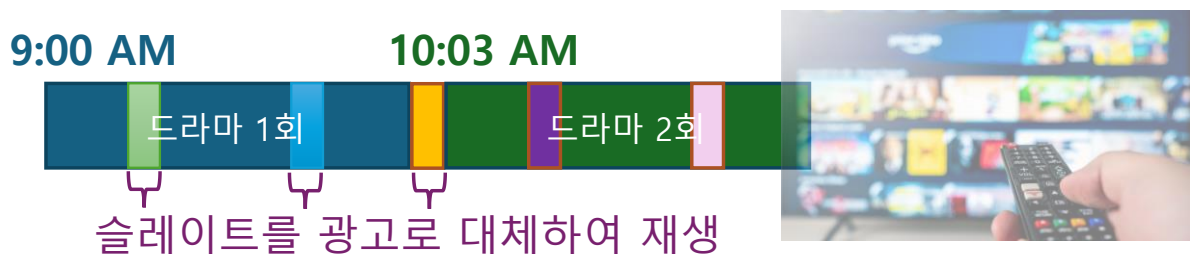
9.3 FAST 채널에서의 광고 방식

FAST 채널은 온라인 스트리밍으로 서비스되는 특성 상, 기존 리니어 방식의 TV와는 다른 방식으로 광고를 재생한다. 기존 리니어 TV에서는 콘텐츠 자체에 광고를 편성하여 방송하고 모든 시청자가 동일한 편성된 광고를 시청하는 방식이다. FAST 채널에서는 광고를 재생하기 위해 방송사가 방송 콘텐츠의 중간에 광고가 삽입될 지점(이하 광고 구간)만 미리 지정하는 방식을 사용한다. 광고 구간으로 지정된 부분이 재생될 때 시청자에게는 광고가 대신 재생되는 방식이다.

콘텐츠가 광고로 대체되지 않게 하기 위해 광고 구간으로 지정되는 지점은 슬레이트 영상으로 편성하게 된다. 광고를 직접 편성하지 않고 이런 방식으로 슬레이트를 편성한 후 광고 구간으로 지정하여 슬레이트 영상을 재생하는 시점에 광고로 대체하여 재생하면, 콘텐츠가 편성되는 시점에서는 광고가 고정되지 않고 시청자가 콘텐츠를 재생하는 시점에 동적으로 재생될 광고를 선택할 수 있게 된다. 즉, 시청자마다 다른 광고가 재생될 수 있고, 이를 통해 시청자에 타겟팅 된 광고를 재생할 수 있다는 점이 방송 광고와 다른 온라인 광고의 특징이다.



(그림 9-4) FAST 채널에서 광고 구간 지정의 예



(그림 9-5) 시청자는 슬레이트 대신 광고로 대체된 영상 시청

(그림 9-4)는 FAST 채널을 편성하는 방식에 대한 예시이다. 드라마 1회~3회까지 편성하고 일정 시간마다 슬레이트 영상을 삽입하여 편성함으로써 광고 구간 지정한다. 슬레이트 영상이 편성될 때마다 실제 드라마 회차의 길이는 길어진다. 예시에서는 1분짜리 광고 구간을 3회 지정하기 위해 1분 슬레이트 영상을 3회 편성함으로써 다음 회의 드라마 편성시간이 3분 늘어난 것을 볼 수 있다.

이렇게 편성된 슬레이트가 포함된 콘텐츠는 온라인 스트리밍을 통해 시청 디바이스에서 재생되는데 이 때 슬레이트 영상 대신 시청자에 맞는 광고가 재생되게 된다. (그림 9-5)

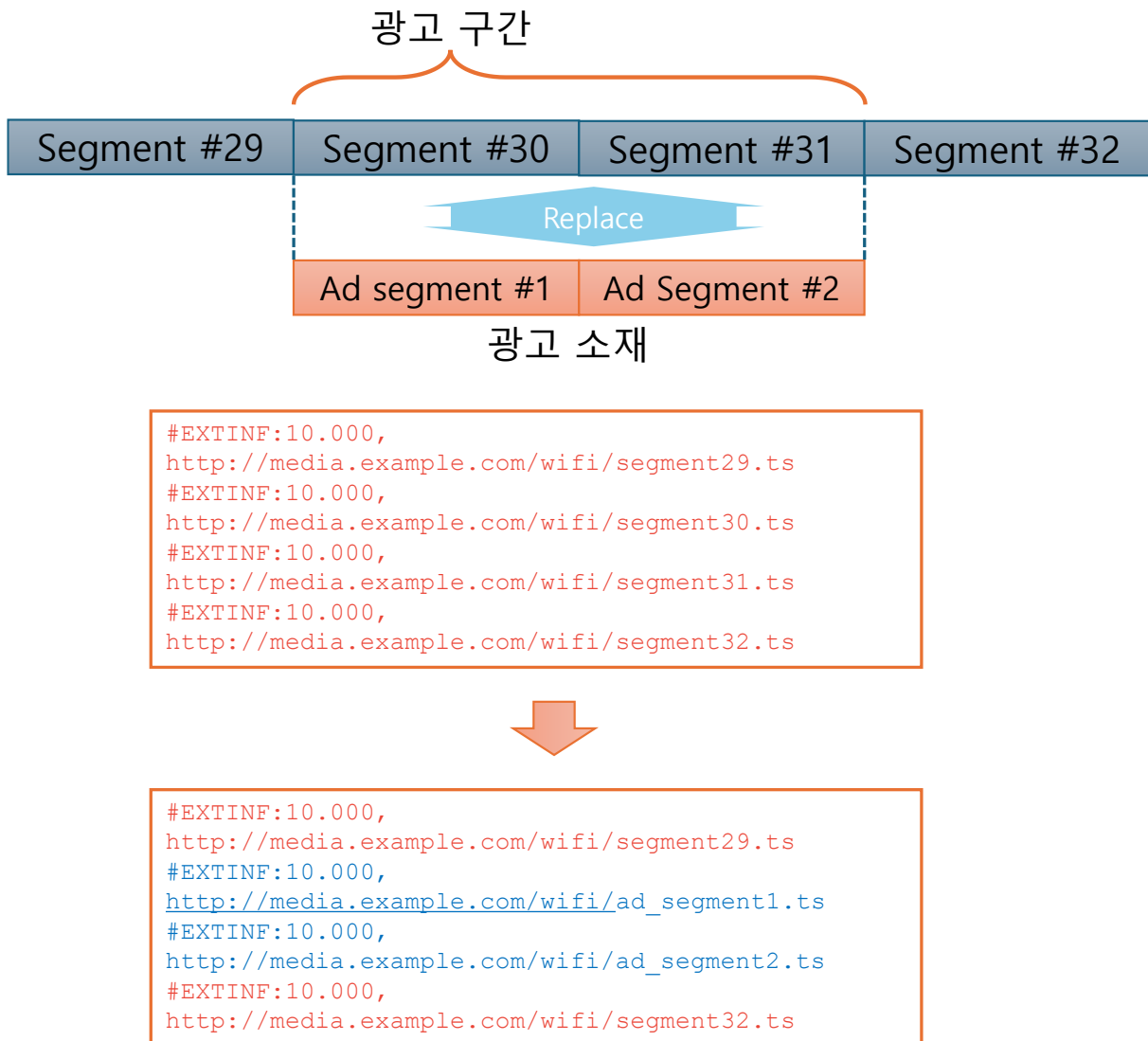
9.4 FAST 채널 광고를 위한 메타데이터

시청자에게 광고 구간으로 지정된 슬레이트 영상 대신에 광고 영상을 보여주기 위해서는 광고 구간인 슬레이트 영상을 인식하여 이를 광고로 대체하여 재생할 수 있는 방법이 필요하다. 슬레이트 영상을 인식하기 위해서 방송사에 미리 광고 구간에 대한 메타데이터를 스트림에 삽입하여 송출하게 된다. 메타데이터 포맷으로 SCTE-35 규격이 주로 사용된다.

SCTE-35 규격은 표준화 단체인 SCTE(Society of Cable Telecommunications Engineers)와 ANSI에서 공동으로 제정한 표준으로 MPEG TS 스트림에서 큐 톤(cue tones) 신호를 삽입할 수 있도록 정의하고 있다. 큐 톤 신호는 원래 오디오 톤으로 구성

된 메시지로, 방송 네트워크에서 지역 방송사의 자동화 장비에서 인식하여 지역 TV 광고나 라디오 광고를 자동으로 삽입하도록 신호를 보내는 역할을 하였다. 디지털 방송으로 전환되면서 오디오 쿼터톤이 아닌 디지털 쿼터톤 신호인 SCTE-35 신호로 대체되고 있다. SCTE-35 방식 역시 MPEG TS 스트림이 포함된 HLS 프로토콜에서 광고 대체구간 정보를 전달하는데 사용될 수 있다.

SCTE-35 규격은 최근 OTT 서비스에서 주로 사용하는 HTTP 기반의 스트리밍 프로토콜로 확장되고 있다. 예를 들어 HLS의 m3u8 플레이리스트 상에서 SCTE-35 규격에 맞추어 Chunk를 나누고 확장 태그를 추가하는 방식으로 HTTP 기반 스트리밍 서비스에서 광고 대체구간 정보를 전달할 수 있다.



(그림 9-6) HLS에서 Chunk 대체를 통한 광고 삽입의 예

(그림 9-6)은 HTTP 기반의 스트리밍 프로토콜인 HLS (HTTP Live Streaming)에서 광고 구간에 해당하는 슬레이트 영상 부분의 Chunk를 광고 영상의 Chunk로 대체하여 시청 디바이스에서 플레이하는 최종적인 플레이리스트(.m3u8) 상에서 광고 영상이 재생되도록 한 예이다.

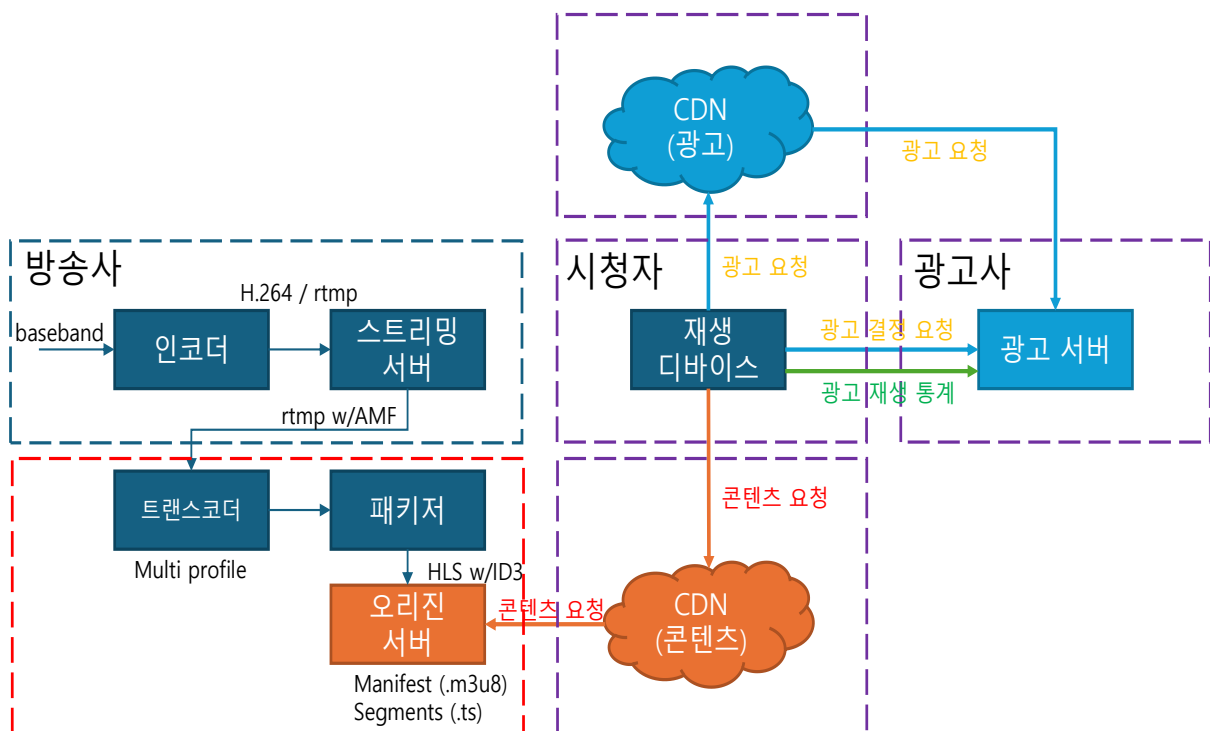
SCTE-35 방식으로 삽입된 광고 구간을 광고 영상으로 대체할 때, 어느 단계에서 광고 영상 대체가 수행되는가에 따라 클라이언트측 동적 광고 삽입(Client-side Dynamic Ad Insertion)과 서버측 동적 광고 삽입(Server-side Dynamic Ad Insertion)으로 나눌 수 있다.

9.4.1 클라이언트측 동적 광고 삽입

FAST 채널을 시청하는 TV나 셋톱박스 같은 시청 디바이스(클라이언트)에서 스트림에 삽입된 메타데이터(SCTE-35)를 인식하여 광고를 대체하는 방식이다. 이를 위해 시청 디바이스는 라이브 스트림을 재생하는 기능만이 아니라 광고 구간 식별을 위해 메타데이터를 인식하는 기능, 광고 서버와 통신하여 대체하여 재생될 광고를 받아오는 기능, 받아온 광고를 정확한 타이밍에 맞춰 실시간으로 대체하여 재생할 수 있는 기능을 갖추어야 한다.

클라이언트측 동적 광고 삽입 방식은 시청 디바이스가 콘텐츠의 재생만이 아니라 광고 서버와 통신 및 광고 재생 등의 부가 기능을 구현하여야 하며 구현 복잡도가 증가한다는 단점이 있다. FAST 채널에서 슬레이트 영상을 대체하여 온라인 광고를 하기 위해 기존에 출시된 광고 대체 기능이 지원되지 않는 구형 시청 디바이스들을 소프트웨어 업데이트를 통해 모두 업그레이드해야 한다는 점도 장벽이 될 수 있다.

또한, 시청 디바이스 전 단계에 광고를 차단하는 기능 (브라우저의 Ad Blocker나 네트워크 공유기에서의 광고 서버 차단 기능 등)이 존재하면, 광고 서버와의 통신이 방해되어 광고가 재생되기 어렵다는 점도 단점으로 꼽힌다.



(그림 9-7) 클라이언트측 광고 삽입에 대한 다이어그램

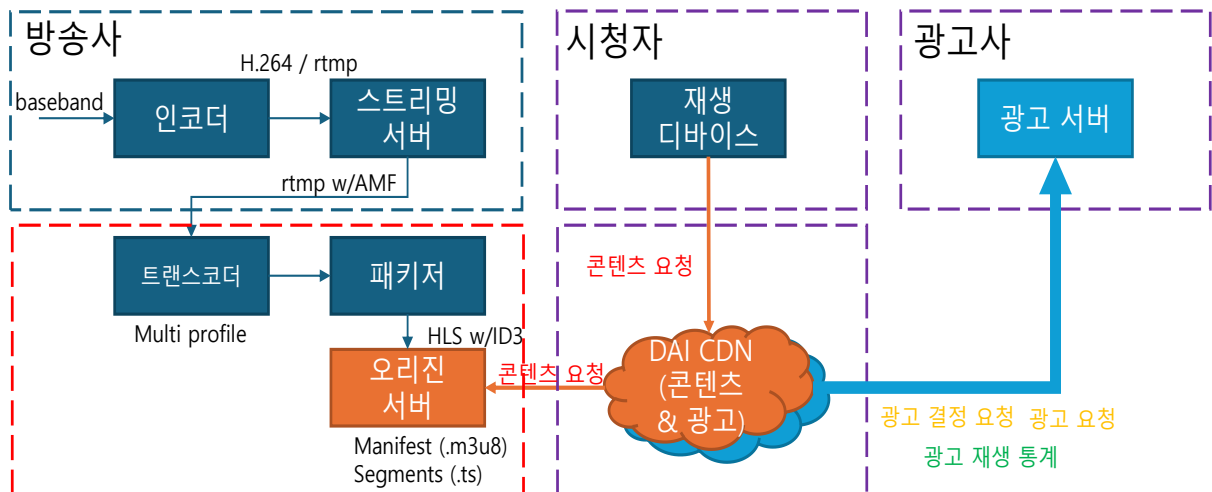
(그림 9-7)은 클라이언트측 광고 삽입 방식을 사용하는 전체적인 시스템 구성도를 나

타낸다. 방송사에서는 원본 신호를 스트리밍하기 위해 동영상 인코딩 과정을 거치고 생성된 스트림에 광고구간을 인식할 수 있도록 광고구간 메타데이터를 삽입한다. 시청 디바이스로 온라인 스트리밍하기 위해 다양한 화질로 트랜스코딩 되고 HTTP 기반 스트리밍 프로토콜로 변환되어 플레이리스트(.m3u8)와 함께 영상의 Chunk(.ts)로 제공된다.

재생 디바이스는 CDN을 통해 스트리밍 되는 콘텐츠를 재생하고, 광고 구간이 인식되면, 광고 서버에 광고를 요청하여 별도의 광고 파일을 다운로드 받아서 콘텐츠의 슬레이트 구간에 재생해 주는 기능이 필요하다.

9.4.2 서버측 동적 광고 삽입

방송사에서 송출된 온라인 스트림은 네트워크의 효율성을 위해 CDN(Contents Delivery Network)을 통해 시청 디바이스로 전달되는데, CDN에서 스트림에 대한 콘텐츠의 Chunk와 Chunk의 목록 정보를 갖는 플레이리스트를 만들게 된다. 이때 슬레이트 영상에 해당하는 Chunk 파일을 광고 영상의 Chunk 파일로 대체하게 되면, CDN 단계에서 실시간으로 광고를 삽입할 수 있게 된다. 즉 시청 디바이스에게는 콘텐츠와 광고가 단일 스트림으로 제공되게 된다. 시청 디바이스에서는 미리 광고로 대체된 스트림을 재생하면 되기 때문에, 광고 삽입에 대한 추가적인 작업 없이 FAST 채널을 재생할 수 있는 장점이 있다. 하지만 CDN에서 광고 영상을 삽입한 스트림은 모든 동일 채널 시청자에게 전달되므로, 시청자마다 특화된 타겟 광고를 재생할 수 없다는 점이 단점이다.

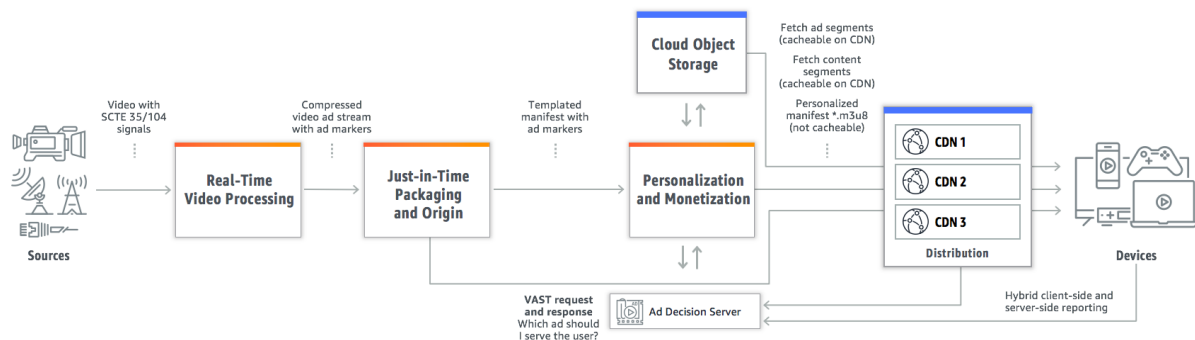


(그림 9-8) 서버측 광고 삽입에 대한 다이어그램

최근 서비스되고 있는 CDN에서는 이런 점을 보완하기 위해 동일 채널을 시청하는 시청자라 할 지라도 시청자마다 별도의 플레이리스트를 구성하여 광고만 다르게 재생할 수 있도록 지원하고 있다. 방송사에서 슬레이트 영상을 편성하고 광고 구간을 메타데이터에 넣어서 CDN으로 송출하면, CDN은 광고 구간 메타데이터를 인식하고 슬레이트 영상의 Chunk만 광고 콘텐츠의 Chunk로 대체하여 플레이리스트를 만든다. 이 때 시청자마다 별도의 타겟팅 된 광고를 광고 서버에서 받아오게 되고 방송 콘텐츠는 동일하지만 광고 구간의 Chunk만 다른 플레이리스트를 생성하여 시청자에게 전송한다. 이 방식을 사용하면

시청 디바이스에서 광고 삽입을 신경 쓸 필요 없다는 장점과 시청자 별로 특화된 타겟팅 광고를 재생할 수 있다는 장점을 동시에 가져갈 수 있다. 또한 시청자에게는 미리 삽입된 광고 스트림을 전달하므로 광고 차단 기능에도 광고가 재생 가능하게 된다.

(그림 9-8)은 서버측 광고 삽입 방식을 사용하는 전체적인 시스템의 구성도를 나타낸다. 클라이언트측 광고 삽입 방식의 구성도와 비교했을 때, 재생 디바이스는 별도로 광고 서버와의 연동이 필요치 않고, 단일 스트림을 재생하는 역할만 수행해도 된다. 대신 CDN에서 광고 서버와 연동하여 광고 구간에 재생될 광고 영상을 받아와서 단일 스트림으로 합치는 작업을 수행해야 한다.



(그림 9-9) AWS Cloud에서 제공하는 서버측 광고 삽입 서비스 단계

[출처] <https://aws.amazon.com/ko/media/tech/what-server-side-ad-insertion-ssai/>

9.5 결론

FAST 서비스는 기존 리니어TV의 장점과 온라인 서비스의 장점을 모두 갖추어 해외 시장을 시작으로 국내에서도 점점 활성화되고 있는 서비스이다. 기본적으로 FAST 채널 시청은 무료이기에 광고 기반으로 서비스하는 것이 일반적이다. FAST 채널에 광고를 삽입하기 위해서는 콘텐츠의 중간에 슬레이트 영상을 삽입하고, 슬레이트 영상을 광고 구간으로 표시하는 메타데이터를 삽입하여 스트리밍한다. 광고 구간 메타데이터를 인식하여 슬레이트 영상을 실제 광고 영상으로 바꿔서 재생하게 된다.

슬레이트 영상을 광고 영상으로 대체하는 역할을 어느 단계에서 수행하는가에 따라 클라이언트측 동적 광고 삽입과 서버측 동적 광고 삽입으로 나눌 수 있다. 클라이언트측 동적 광고 삽입 방식을 사용하려면 스마트TV나 셋톱박스과 같은 시청 디바이스에서 광고 인식, 광고 소재 다운로드 및 재생 등의 기능을 수행해야 하므로 시청 디바이스의 구현 난이도의 증가, 구형 디바이스 지원 불가 등의 이슈가 있을 수 있다. 서버 측 동적 광고 삽입 방식은 CDN(Content Delivery Network)에서 슬레이트 영상의 Chunk를 광고 영상의 Chunk로 대체하여 플레이리스트를 만드는 방식으로 하나의 스트림으로 통합하여 재생 디바이스로 스트리밍 되므로 시청 디바이스에서의 이슈를 해결할 수 있지만, 모든 시청자에게 동일 광고가 재생되어 시청자별 타겟팅 된 광고를 재생하기 어렵다.

최근 동적 광고 삽입을 지원하는 CDN 서비스에서는 이런 문제를 해결하기 위해 시청자 별로 별도의 플레이리스트를 작성하여 광고 구간만 다른 플레이리스트를 시청자에게

개별 서비스하는 방식을 도입하고 있다.

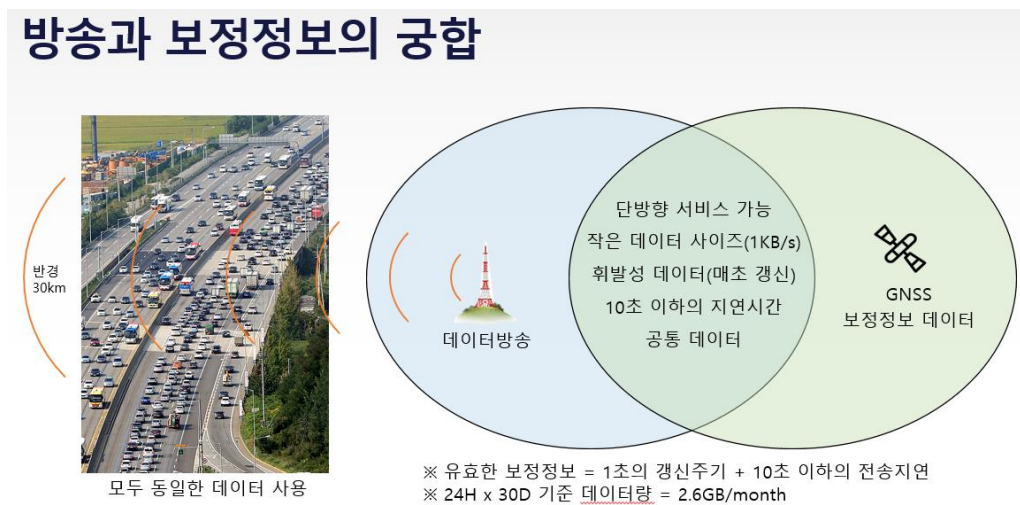
10 다양한 네트워크망을 활용한 고정밀 위치정보 서비스 및 연구 동향

10.1 개요

고정밀 위치정보 서비스는 자율주행, 드론, 물류, 군사 등 다양한 산업의 발전과 더불어 점차 중요한 인프라로 자리 잡고 있다. 특히, GNSS(Global Navigation Satellite System)와 이를 보완하는 정밀 위치 보정 기술은 실시간 높은 정밀도를 요구하는 응용 분야에서 필수적이다. 기존에는 RTK(Real-Time Kinematic)와 같은 정밀 측위 서비스가 주로 이동 통신망을 통해 제공되었으나, 최근에는 ATSC 3.0과 같은 지상파 방송망, DAB/DMB망, 그리고 KASS(Korean Augmentation Satellite System) 등 새로운 네트워크와 보정 기술을 적용하려는 시도가 활발히 이루어지고 있다. 이러한 네트워크망을 활용한 고정밀 위치정보 서비스는 넓은 커버리지와 높은 경제성으로 사용자들에게 실질적인 혜택을 제공할 수 있으며, 다양한 산업에서의 활용 가능성을 넓혀주고 있다. 본 보고서에서는 다양한 네트워크망을 활용한 고정밀 위치정보 서비스에 대해 알아보고, 각 서비스별로 시장에서의 기술 완성도 및 활용도를 중심으로 살펴보고자 한다.

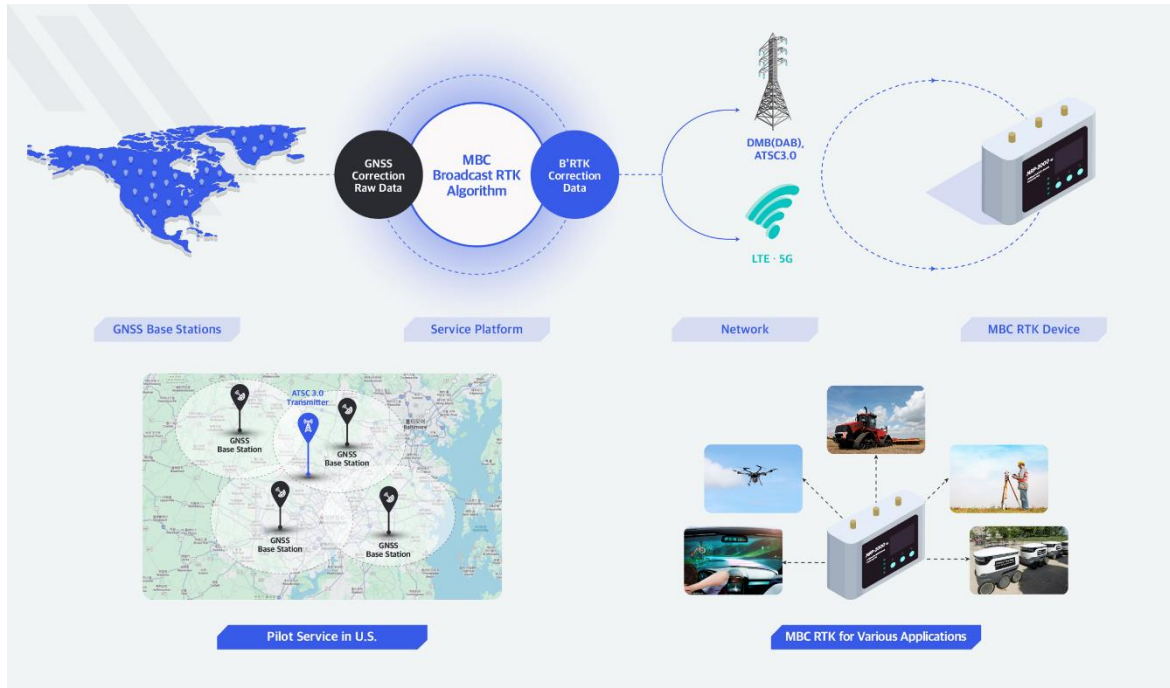
10.2 ATSC 3.0 방송망을 활용한 고정밀 위치정보 서비스

일반 GPS와 달리 cm급 위치정밀도를 갖는 RTK(Real-Time Kinematic)와 같은 정밀측위 시스템은 GNSS 위성신호 외에 별도의 네트워크망을 활용하여 위치 보정신호 전송이 필요하다. 기존에는 주로 이동통신망이나 위성망이 사용되었지만, 최근에는, ATSC 3.0과 같이 이동수신을 지원하는 지상파 방송망을 활용하려는 시도가 활발히 이루어지고 있다. 지상파 방송망은 넓은 네트워크 커버리지, 풍부한 전송 대역, 저렴한 수신 모듈가격 그리고 무료로 가까운 이용요금의 장점을 지니며, 이를 바탕으로 위치 보정신호를 전송하는 매체로 평가받고 있다.



(그림 10-1) 보정 정보 제공을 위한 네트워크로써 방송망의 효용
[출처] 방송미디어 서비스 및 기술 동향, 미래방송포럼 기술보고서

방송망은 양방향 통신이 불가하므로, RTK 보정 정보 제공 시 기존에 사용하던 NTRIP(Networked Transport of RTCM via Internet Protocol)을 활용할 수는 없다. 대신 다양한 GNSS 기준국의 정보를 방송 신호로 전송하고, 사용자가 적절한 보정 정보를 선택할 수 있도록 추가 정보를 제공해야 한다. 이를 위해서 아래 (그림 10-2)와 같이 Broadcast RTK 알고리즘이 필요하며, 이는 방송망으로 RTK보정 신호를 전송하는 별도의 데이터 생성을 포함한다.



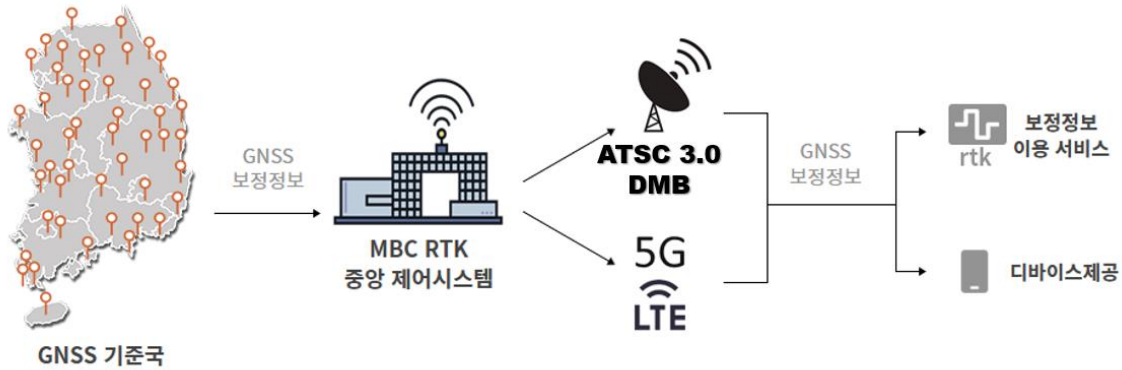
(그림 10-2) 방송망을 활용한 RTK 서비스 전달 방식

[출처] NAB 2024 Show

최근 글로벌 지상파 TV 방송시장은 마지막 대 전환기라고 할 수 있을 정도로 국가별 방송시스템이 바뀌고 있다. 그 중심에는 국내 UHD TV와 미국의 NextGen TV 방송 표준인 ATSC 3.0이 있다. 미국은 기존 ATSC 1.0 방송시스템을 ATSC 3.0으로 활발히 전환 중이고 캐나다도 전환을 위한 실험방송을 추진 중이다. 남미지역에서는 지상파 직접 수신율이 70%가 넘는 브라질이 차세대 방송 표준으로 ATSC 3.0을 결정했고, 인도는 대다수 국민의 휴대폰을 통한 영상 시청 방식을 이동 방송으로 전환하려는 정책(Direct to Mobile)을 시행 중이며, 관련 기술 표준으로서 ATSC 3.0의 채택 가능성이 매우 높은 상황이다. 이러한 차세대 지상파 방송시스템 전환은 대규모 투자비가 들 수밖에 없고, 신규 수익 창출 및 서비스 성공을 위해서, 시청자에게 더 나은 미디어와 함께 새로운 서비스를 제공할 필요가 있다. 이런 상황에서 지상파 방송사들은 새로운 수익 모델을 모색 중이며, 방송망을 활용한 RTK 서비스에 대한 관심이 증가하고 있다.

방송망을 통해 RTK 보정 신호를 제공하는 방식이 매력적인 방법이지만, 방송망의 한계도 존재한다. 상용 서비스를 수행하기 위해서는 신호가 도달하지 못하는 음영 지역을 최소화하여야 하지만, High-Tower/High-Power 가 기본인 방송 신호의 특성 상 산악 지대에서의 수신율 저하는 서비스를 안정적으로 공급하기에 부족함이 있다. 또한 지

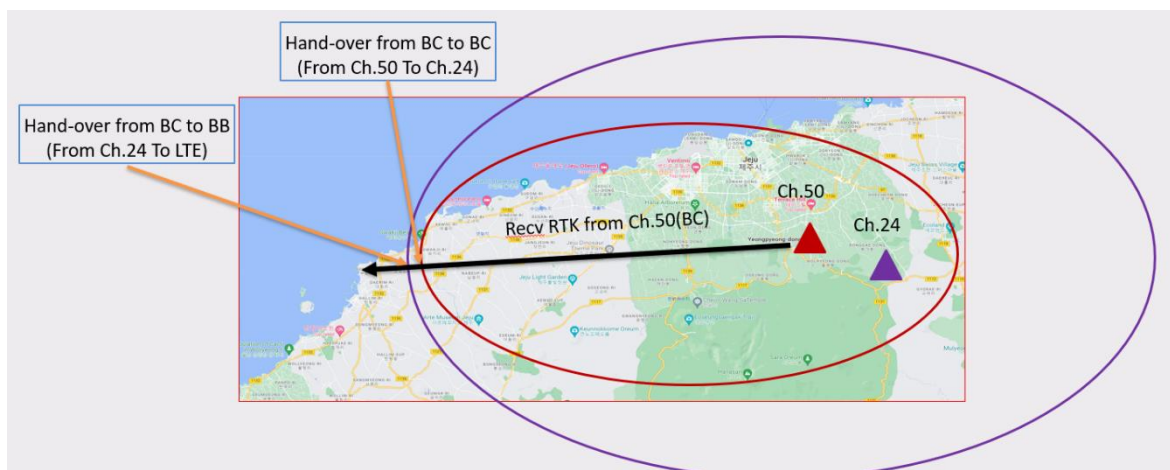
상파 방송사의 열악한 경영 환경 상 방송 중계기 확산이 느려지는 것도 마이너스 요인일 수밖에 없다. 이러한 이유로 방송망과 통신망을 결합한 Hybrid 형식의 RTK 서비스가 발생하게 되었다. 방송망이 가능한 지역에서는 무료인 방송망을 통해서 RTK 서비스를 수신하여 통신비용을 절감하고, 방송망을 활용할 수 없는 지역은 통신망을 활용하여, 서비스 커버리지를 확대할 수 있다.

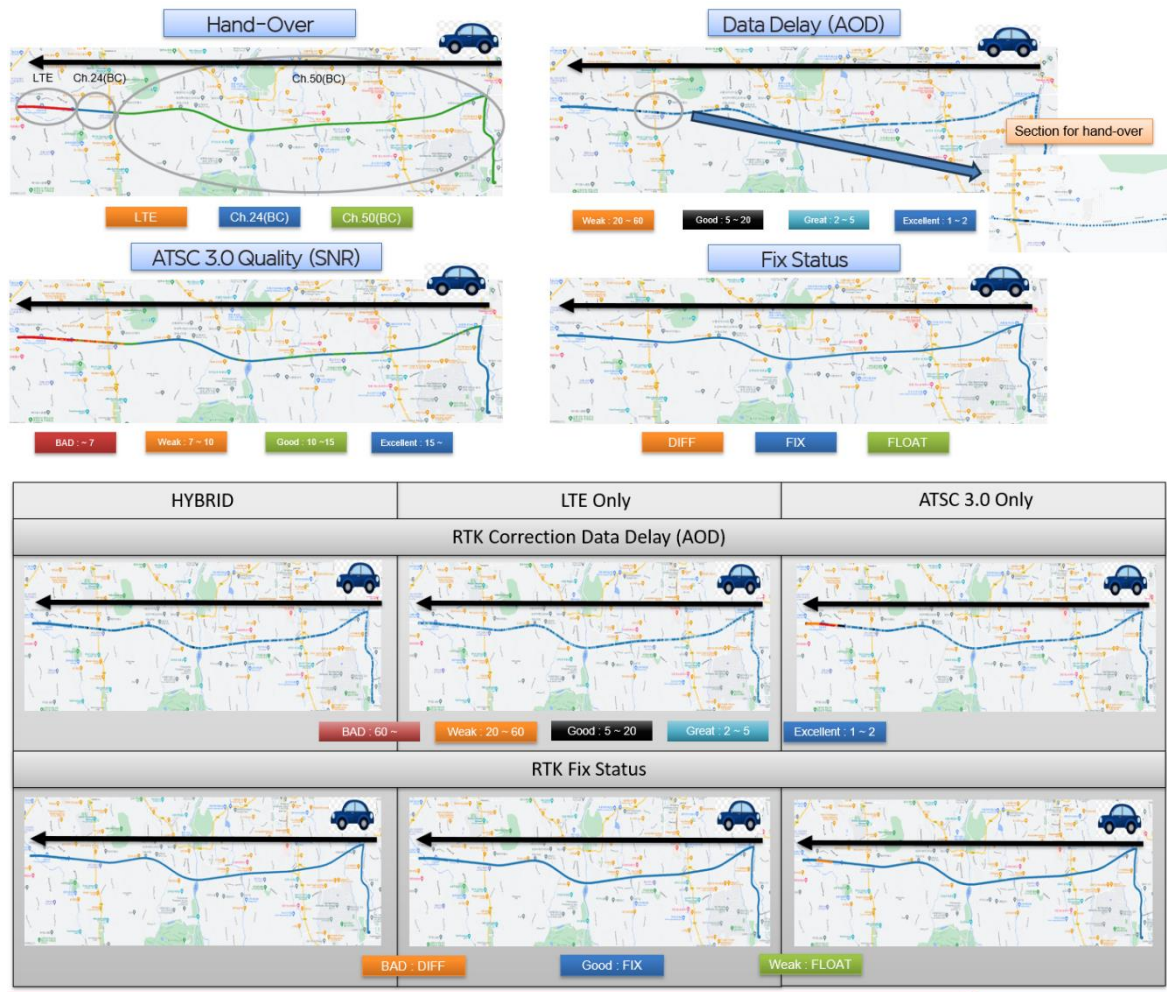


(그림 10-3) 방송/통신 융합형 Hybrid RTK 서비스 전달 방식

[출처] 방송미디어 서비스 및 기술 동향, 미래방송포럼 기술보고서(2023년)

방송망과 통신망을 결합한 RTK 서비스의 성능을 확인하기 위해서 제주테크노파크 및 한국전파진흥협회(RAPA), (주)문화방송(MBC)는 제주 지역에서의 필드 테스트를 실시하였다. 2개의 주파수를 활용하여, Broadcast-To-Broadcast 핸드오버 기능을 검증하였고, 방송망이 도달하지 않는 지역에서 Broadcast-To-Broadband 핸드오버 성능을 검증하였다. 아래 테스트 결과는 방송 신호가 약해짐에 따른 통신망으로의 전환이 적절하게 이뤄졌음을 확인해 주고 있으며, RTK 보정 신호를 전달할 때 통신망을 활용하지 않았을 경우, 특정 지역에서 서비스가 정상적으로 수행되지 못함을 보여준다. 아래 테스트 결과를 확인하면, 방송망과 통신망을 결합하게 되면 기존 서비스 대비 좋은 성능을 가지면서 훨씬 효율적인 서비스를 제공할 수 있음을 검증할 수 있었다





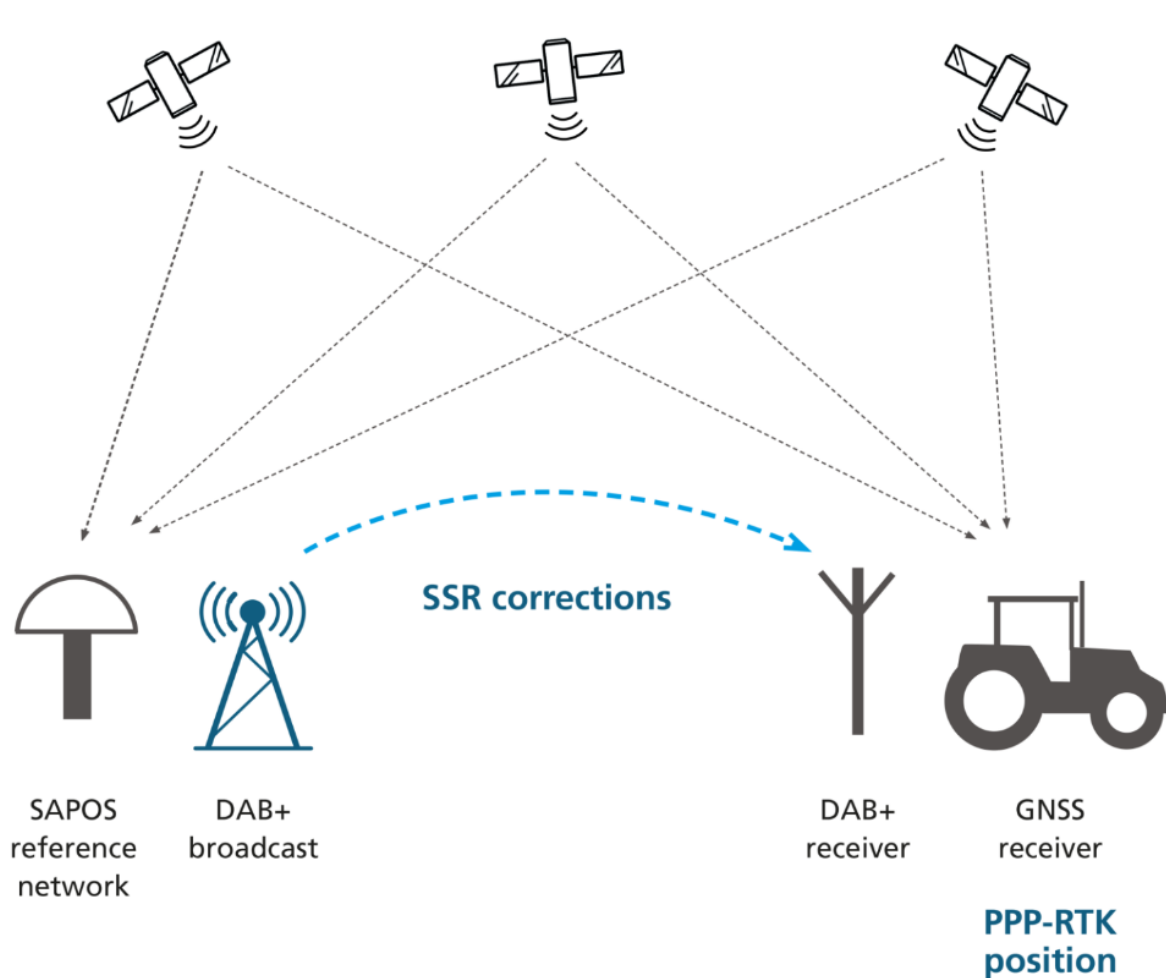
(그림 10-4) 방송망과 통신망을 결합한 RTK 서비스 성능 테스트 결과
[출처] 방송미디어 서비스 및 기술 동향, 미래방송포럼 기술보고서(2023년)

미국 최대의 지상파 방송사인 NexStar 미디어그룹과 Sinclair방송그룹의 합작사인 BitPath사는 방송망과 유선 통신망을 동시에 활용하는 Hybrid RTK 서비스 기술을 도입하여 미 전역에 서비스를 개시하기 위해 인프라 구축을 확대해 나가고 있다. 2024년 볼티모어/워싱턴 DC 지역을 시작으로 덴버 및 라스베이거스, 내슈빌 등 약 5개 지역에서 서비스를 오픈할 예정이며, 2025년도에는 30개 이상 지역으로 확대해 나갈 계획을 가지고 있다. 주요 활용 분야로는 측지/측량을 위한 보정정보 제공 분야와 자율주행/정밀 내비게이션 서비스를 포함한 미래형 자동차 산업에 적용하기 위한 기획을 진행하고 있다.

ATSC 3.0 방송망을 활용한 고정밀 위치정보 서비스는 방송 기술과 RTK GPS의 융합을 통해 새로운 소비자 요구를 충족시킬 수 있는 혁신적인 접근법이다. 방송망의 경제성과 광범위한 신호 도달 범위를 활용하여 고정밀 위치정보 서비스를 제공함으로써, 자율주행차, 드론, 스마트 물류 등 다양한 산업 분야에서의 활용 가능성을 확대할 수 있다. 하지만, RTK 서비스를 위한 방송 인프라 구축, 수신기 보급 등 선결해야 할 과제가 많은 것 역시 사실이다. 앞으로의 기술 발전과 서비스 상용화를 통해 고정밀 위치정보 서비스의 잠재력을 극대화할 필요가 있다.

10.3 DAB/DMB 망을 활용한 고정밀 위치정보 서비스

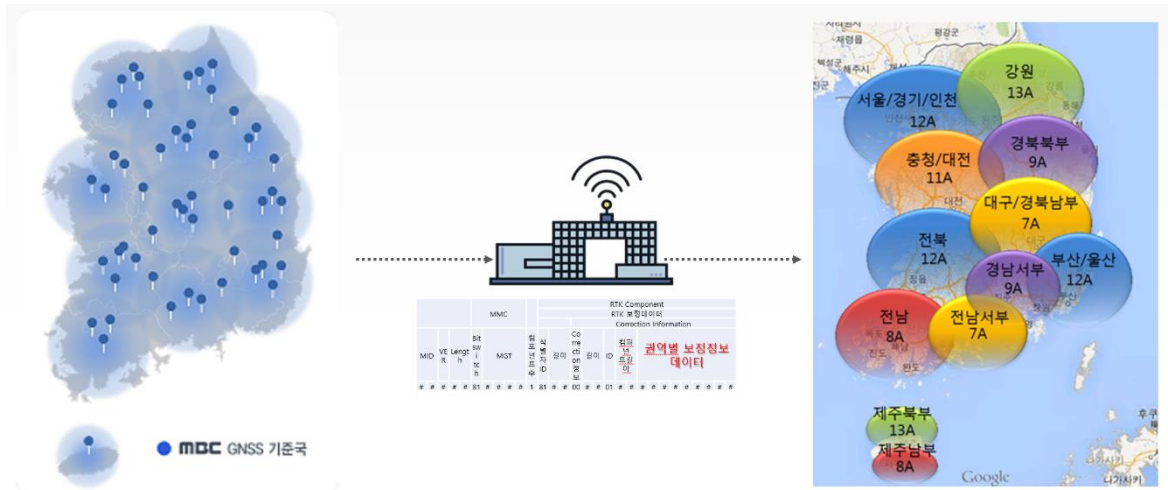
위에서 언급한 ATSC 3.0 방송망이 매력적인 RTK 보정정보 전송 매체이기는 하지만, 유럽/중국 등 전 지역에서 상용화 되어 있지 않다. 글로벌 판매를 목표로 하고 있는 양산 자동차 시장에 방송망을 통한 RTK 서비스를 제공하기 위해서는 유럽에서 사용하고 있는 방송망을 활용한 보정 정보 제공 방법도 고려되어야 한다. 유럽 지역에서는 디지털 라디오에서 사용하고 있는 DAB(Digital Audio Broadcasting) 규격이 이동방송을 위한 표준으로 활용되고 있고, 이미 독일/북유럽 등의 양산 자동차에 적용되어 있다. 이 표준은 또한 국내 이동 방송 규격으로 2005년부터 사용되고 있는 DMB(Digital Multimedia Broadcasting) 표준과 일치한다. Alberding GmbH, inPosition GmbH, Geo++, Fraunhofer IIS 등은 DAB+ 데이터 채널을 활용하여 SSR(State Space Representation) 보정 데이터를 전송하는 시스템을 개발하여 Intergeo 2024에서 전시하였다.



(그림 10-5) DAB+ 방송망을 활용한 RTK 보정정보 전달 방식

[출처] <https://www.iis.fraunhofer.de/en/ff/lv/lok/proj/ssroverdab.html>

국내에서도 KBS 및 MBC에서 DMB 망을 활용한 RTK 보정정보를 제공하는 서비스를 시작하였고, MBC의 경우 OSR(Observation State Representation) 기반의 권역별 보정정보 서비스를 제공하고 있다.



(그림 10-6) DMB 방송망을 활용한 권역별 RTK 서비스 현황

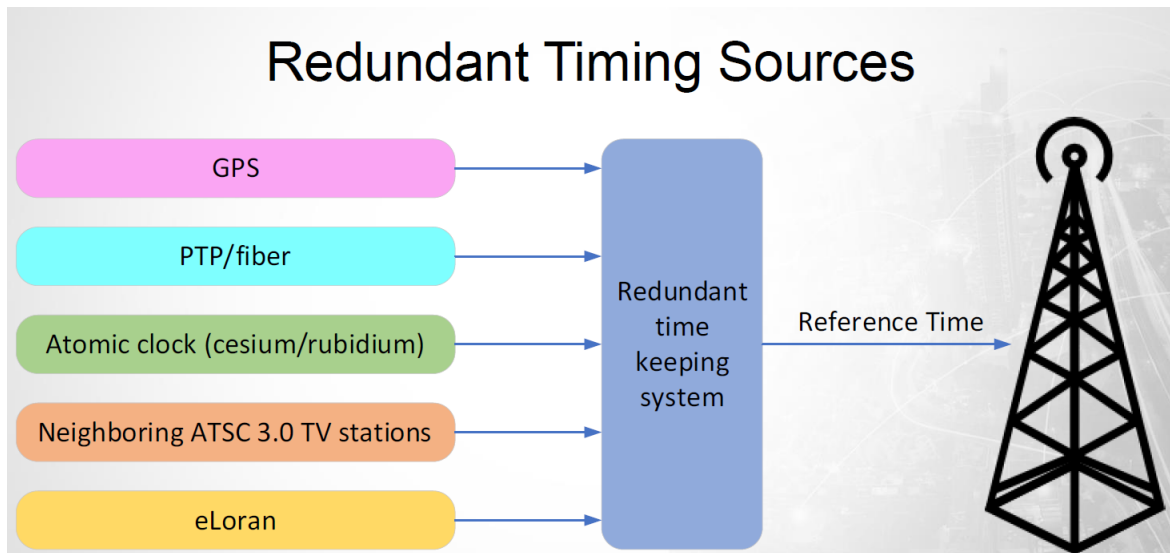
[출처] 위성측위기술과 정밀측위 서비스 활용, 2023

DAB 망의 작은 주파수 대역폭은 대용량 RTK 보정 정보를 제공하기에는 한계가 있다. 이러한 단점을 보완하기 위해서, 데이터 압축 기술을 활용하거나 보정 신호 전송 스케줄을 조정하는 등의 적절한 기술적 진보가 필요하다. 아직 전 세계적으로 초기 단계에 머무르고 있기는 하지만, 향후 ATSC 3.0 방송망과 함께 글로벌 시장에서 방송망을 통한 RTK 서비스의 사용성 향상을 위해, DAB 망을 활용한 보정 정보 제공 기술의 발전도 지켜볼 필요가 있을 것이다.

10.4 Broadcast Positioning System(BPS)

Broadcast Positioning System(이하 BPS)는 ATSC 3.0 방송 신호를 활용하여 수신기에서 시간과 위치 정보를 추정하는 시스템이다. ATSC 3.0은 IP 기반의 방송 표준으로, 국내에서는 지상파 UHD 방송 표준으로 2017년부터 상용 서비스를 제공하고 있다. 이런 ATSC 3.0 신호를 송출하는 TV타워를 활용하면, 기존 GPS나 인터넷 연결이 불가능한 환경에서도 위치 정보 제공이 가능하다. BPS는 단독으로 동작하며, 독립적 시스템으로써 추가적인 GPS나 인터넷 연결 없이 위치 추정을 수행할 수 있다.

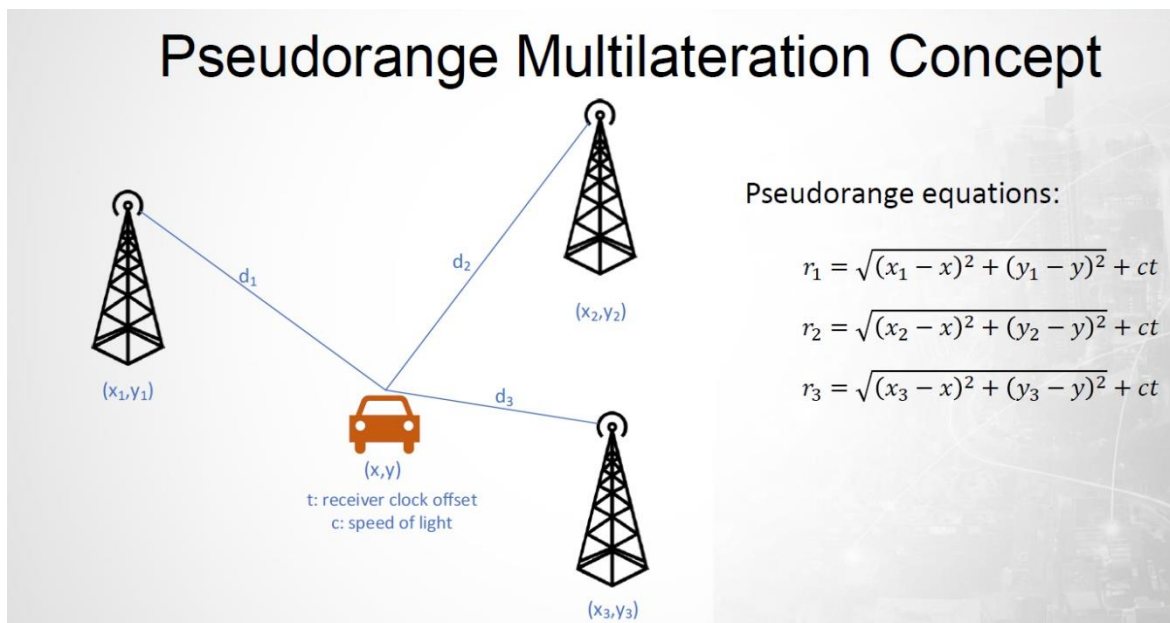
단일 TV 타워를 활용할 경우, 특정 위치에서 고정밀의 시간 정보를 제공할 수 있다. ATSC 3.0 기반의 TV 타워는 GPS와 같은 시간 정보 수신 기능이 없어도 정밀한 시간 동기화를 지원할 수 있는 강점이 있다. 단일 TV 타워는 95%의 신뢰도로 100 나노 초(ns) 이내의 오차를 유지하여 정확한 시간 정보를 제공한다. 이 기능은 특히 실내 위치 기반 서비스나 교통 시스템에서 신호 주파수 간 시간 동기화를 위해 활용될 수 있다.



(그림 10-7) 단일 TV 타워를 활용한 고정밀 시간 정보 제공 방식

[출처] <https://www.nab.org/bps/>

ATSC 3.0 기반 BPS는 하나 이상의 TV 타워를 통해 위치를 추정할 수 있으며, 복수의 TV 타워 신호를 활용하면 더욱 정확한 위치 추정이 가능하다. 네 개의 TV 타워가 위치 정보를 제공할 경우 100미터 이내의 평균 정확도를 확보할 수 있다. 이를 활용하여 GPS 신호가 불안정한 도심 지역, 고층 건물 밀집 지역 및 실내 환경에서 교통, 물류, 재난 구조 등 위치 기반 서비스가 요구되는 다양한 분야에 적용할 수 있다.



(그림 10-8) 복수의 TV 타워를 활용하여 위치 정보를 계산하는 방법

[출처] <https://www.nab.org/bps/>

BPS 서비스는 아래와 같은 장점을 가지고 있다.

- ATSC 3.0을 통해 TV 방송국이 이미 운영되고 있는 만큼, 추가적인 시설 확장 없이 BPS 시스템을 쉽게 배포할 수 있다.

- BPS는 비상 상황에서도 안정적인 서비스를 제공하도록 설계된 방송 송출 시설을 기반으로 한다. 전력이나 인터넷 연결이 차단된 상황에서도 방송 신호는 독립적으로 작동할 수 있어, 재난 및 비상 상황에서도 위치 추적 기능을 지속적으로 제공할 수 있다. 이는 응급 구조와 같은 분야에서 매우 중요한 기능이다.
- BPS는 고출력 및 고탑 송출 시설을 통해 실내에서도 신호를 수신할 수 있다. GPS가 도달하기 어려운 건물 내부나 지하 등에서도 위치 추적이 가능하며, 상업용 건물, 공항, 쇼핑몰, 물류 창고 등 실내 위치 기반 서비스가 필요한 다양한 장소에서 활용될 수 있다.
- BPS는 무료로 사용할 수 있는 위치 정보 시스템으로, 통신사나 GPS 서비스에 추가적인 비용을 지불할 필요가 없다. 이를 통해, 비용 효율적인 위치 기반 서비스를 제공할 수 있어 소규모 사업자와 일반 사용자에게도 접근성이 높다.
- BPS는 무제한의 사용자에게 위치 정보를 제공할 수 있어, 사용자가 증가해도 서비스 품질 저하나 비용 부담이 발생하지 않는다. 이는 위치 기반 서비스를 제공하는 스마트 시티, 교통 시스템, 물류 산업에서 매우 중요한 요소이다.

하지만 BPS는 아래와 같은 단점도 함께 지니고 있다.

- BPS는 GPS에 비해 위치 정확도가 떨어질 수 있다. 특히 실외에서 수십 미터 이내의 정밀도를 요구하는 상황에서는 GPS보다 정확도가 낮아, 고정밀을 요구하는 응용 분야에 한계가 있을 수 있다.
- BPS는 ATSC 3.0 표준을 준수하는 방송 인프라가 갖춰진 지역에서만 사용 가능하며, ATSC 3.0이 배포되지 않은 국가나 지역에서는 BPS를 이용할 수 없다. 따라서 글로벌 통일성이 부족하고, 특정 지역에 종속된 위치 서비스로 제한될 수 있다.
- BPS를 사용하기 위해서는 ATSC 3.0을 수신할 수 있는 수신기 칩셋이 필요하다. 하지만 아직 모든 기기에 BPS를 지원하는 칩셋이 탑재되지 않아, 호환 기기가 제한적일 수 있다. ATSC 3.0 수신이 가능한 기기가 대중화되지 않은 지역에서는 활용도가 낮다.
- BPS는 주로 고정된 송출 타워를 기반으로 작동하기 때문에, 송출 타워가 없는 장소나 이동성이 필요한 응용에는 제약이 있다. 따라서 송출 타워의 수나 밀집도가 낮은 지역에서는 위치 정확도가 떨어지거나, 서비스 제공이 제한될 수 있다.
- BPS가 실내에서도 수신 가능하다는 장점이 있지만, 건축 구조나 재료에 따라 실내에서도 신호 간섭이 발생할 수 있다. 이로 인해 GPS와 달리 실내에서도 안정적인 위치 정보가 제공되지 않는 경우가 발생할 수 있다.

BPS는 저비용, 비상 상황 대응력, 실내 수신 가능성 등 여러 장점을 통해 위치 기반 서비스의 중요한 대안으로 자리 잡을 가능성이 크다. 특히 스마트 시티 구축, 물류 및 재난 구조와 같은 분야에서 BPS는 GPS를 보완하거나 대체할 수 있는 역할을 할 수 있을 것이다. 다만, 한계점 개선을 위한 기술 개발과 글로벌 인프라 확장이 이루어진다면 BPS

의 활용 가치는 더욱 높아질 것으로 기대된다.

10.5 한국형 정밀 GPS 위치보정시스템(KASS)

한국형 정밀 GPS 위치보정시스템(Korean Augmentation Satellite System, 이하 KASS)은 GPS(Global Positioning System)의 위치 정확성을 높이기 위한 위성 기반 보정 시스템으로써, 기존 GPS 신호 이용 시 발생하는 위치 정보의 정확도를 1~3미터 수준으로 향상시킨다.



(그림 10-9) 복수의 TV 타워를 활용하여 위치 정보를 계산하는 방법

[출처] 국토교통부 보도자료, 한국형 항공위성서빗(KASS) 운영 시대 본격 개막

KASS를 활용하게 되면 다음과 같은 효과를 기대할 수 있다.

- 지상에 설치된 항행 안전시설의 위치와 관계없이 유연한 항공로 설정을 가능하게 하여 공역의 혼잡도를 해소할 수 있다.
- 다양한 착륙경로 운영을 지원하여 공항의 수용능력을 증가시키며, 이를 통해 항공 운영의 효율성을 향상시킨다.

국토교통부는 착륙 최저치 개선도가 높은 지방공항(무안, 울산)에서 KASS를 최초로 운영하고, 단계적으로 대상 공항을 확대할 계획을 가지고 있다. 또한, KASS 신호 제공을 위한 항공위성 2호기를 추가 발사하고 운영하여 안정적인 서비스를 제공할 수 있는 기반을 강화시키고, 정밀 위치 정보를 인터넷을 통해 제공하는 KDAS(Korean Differential Augmentation System) 서비스를 시행하여 다양한 위치 기반 산업 분야에서의 신호 사용

편의성을 향상시킬 계획이다.

KASS는 많은 장점을 가지고 있지만 몇 가지 한계점도 존재한다.

- KASS는 한국 내에서만 사용할 수 있는 시스템으로, 국제적인 활용 가능성이 제한될 수 있다.
- 새로운 시스템이기 때문에, 신호 안정성 및 커버리지 확장 등 기술적 과제가 존재한다.
- KASS의 성공적인 운영을 위해서는 지속적인 인프라 구축과 유지 관리가 필요하며, 이는 추가적인 비용을 초래할 수 있다.

KASS는 위치 정확도를 높이고 안전성을 강화하는 데 중요한 역할을 하는 시스템이다. 항공, 해양, 교통, 농업 등 다양한 분야에서 활용될 수 있는 KASS는 앞으로의 기술 발전과 인프라 확장을 통해 더욱 넓은 응용 가능성을 가지고 있다. KASS가 성공적으로 운영된다면, 한국의 위치 기반 서비스는 물론, 글로벌 위치 보정 시장에서도 중요한 역할을 할 것으로 기대된다.

10.6 LEO PNT : 차세대 내비게이션 시스템

기존 GNSS(Global Navigation Satellite System) 신호는 도심, 실내 등에서 약해지거나, 재밍(Jamming)과 스푸핑(Spoofing) 같은 공격에 취약하다. LEO(Low Earth Orbit) 위성은 지구에 훨씬 가까이 위치하여 강력한 신호와 짧은 지연 시간을 제공한다.

항목	LEO PNT	GNSS
궤도 고도	500~2,000km (저궤도)	약 20,000km(중궤도)
신호 강도	강함	약함
위치 정확도	센티미터 수준	오차 범위 5~10미터
보안성	재밍 및 스푸핑에 강함	보안 취약
주요 활용	자율주행, IoT, 군사, 항공	네비게이션, 항공 및 해양 운송

GEOSTATIONARY EARTH ORBIT (GEO)

Orbit: 35,786km
Earth Coverage: Very large (min. 3 satellites required)
Beam footprint size/km: 200-3500
Latency: High (~250 to 500ms)
Active: ~541
Usage:

- Mobile Backhaul
- Weather data & Broadcast TV
- HTS technology allow basic MBB

MEDIUM EARTH ORBIT (MEO)

Orbit: 2,000 to 34,000km
Earth Coverage: Large (min. 6 satellites required)
Beam footprint size/km: 100-1000
Latency: Medium (~180ms)
Active: ~159
Usage:

- PNT (GNSS, GPS)
- Higher throughput (HTS) Backhaul & data connectivity

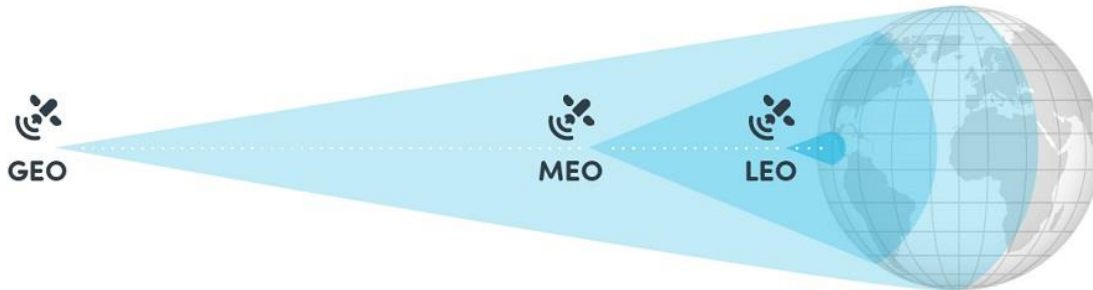
LOW EARTH ORBIT (LEO)

Orbit: 500 to 2000km
Earth Coverage: Small (100-1000 satellites required)
Beam footprint size/km: 100-1000
Latency: Low (~40ms)
Active: >3,700
Usage:

- Higher throughput & lower latency data services
- Direct to phone connectivity
- 3D Positioning

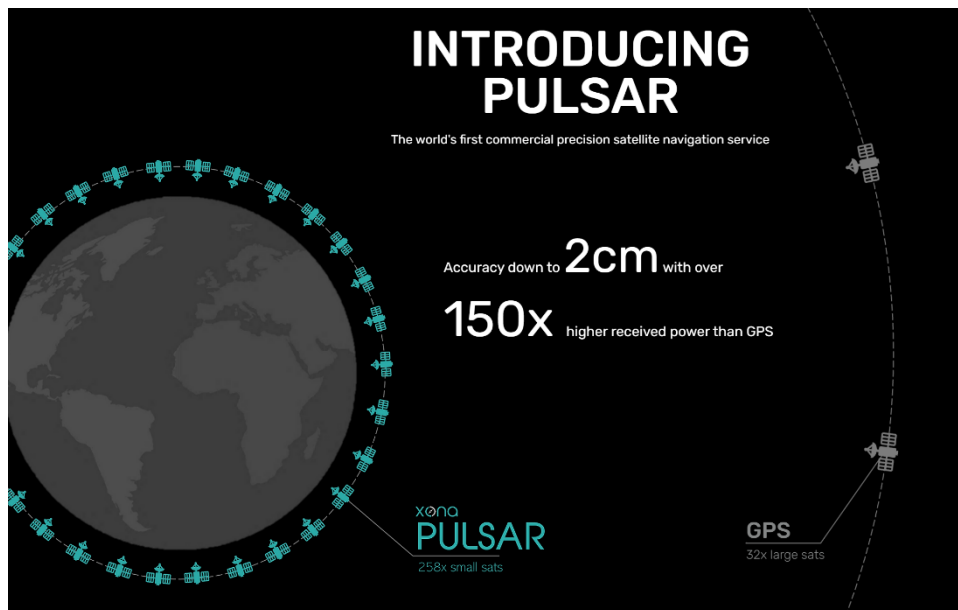
Potential:

- PNT (Relay or Satellite Time and Location [STL])



(그림 10-10) GEO, MEO, LEO 궤도 비교

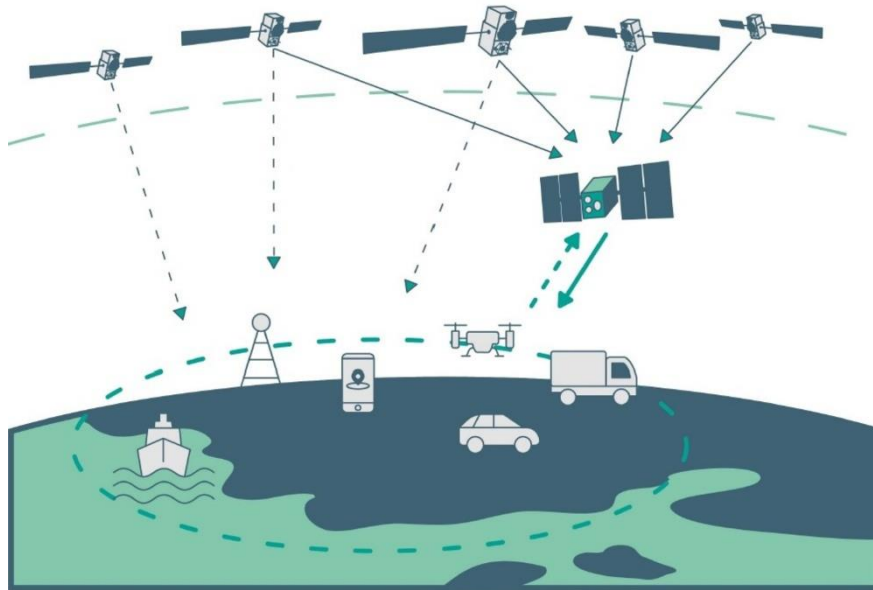
[출처] Satellite Today “Schematic of Orbital Altitudes and Coverage Areas”, 2020



(그림 10-11) Xona Space Systems의 Pulsar 위성

[출처] Xona Space Systems, “Introducing Pulsar”, 2024

스페인의 GMV(GMV Innovating Solutions)는 ESA(European Space Agency)와 협력하여 LEO PNT 시범 프로젝트를 이끌고 있다. 이 프로젝트는 총 5개의 소형 위성으로 구성되며, GNSS(GPS 및 Galileo) 시스템을 보완하기 위한 목표를 가지고 있고, UHF, L, S, C 밴드를 활용한 신호를 통해 기존 GNSS의 기능을 확장하고 신호의 무결성 감시 기능을 실시간으로 제공할 예정이다.



(그림 10-12) XESA의 LEO PNT를 활용한 다층 시스템

[출처] European Space Agency, “Industry invited to bid for low-Earth orbit satnav demo”, 2023

LEO PNT의 센티미터 수준의 정밀도는 자율주행차와 드론이 도심과 복잡한 환경에서 안정적으로 운영할 수 있는 기반 기술이 될 수 있으며, 물류 산업의 효율성을 높이고, 사고 위험을 줄일 수 있다. 또한 군사 작전에서 필수적인 정확한 위치와 타이밍 정보를 제공할 수 있으며, 스푸핑과 재밍에 강인한 특징은 군사 시스템의 안정성을 높일 수 있다. 또다른 장점으로, 해양과 항공기 운항에서 GNSS 신호를 보완할 수 있으며, 극지방과 같이 기존 GNSS 신호가 약한 지역에서도 안정적인 내비게이션 기능을 제공할 수 있다.

10.7 결론

고정밀 위치정보 서비스는 여러 네트워크망을 활용하여 사용자들에게 보다 정확한 위치 정보를 제공하고 있다. ATSC 3.0 방송망과 DAB/DMB망은 저렴한 비용과 넓은 커버리지로 위치 보정 신호를 효율적으로 전달하며, KASS와 같은 위성 기반 시스템은 항공과 해양 운송 등 다양한 분야에서 활용될 수 있다. 또한, LEO PNT와 같은 신기술은 도심과 실내 환경에서의 위치 정확도를 더욱 향상시킬 수 있는 잠재력을 가지고 있다. 다만, 네트워크 인프라 구축과 신호 안정성 향상과 같은 과제도 남아 있는 상황이다. 앞으로의 기술 발전과 인프라 확장을 통해 고정밀 위치정보 서비스는 자율주행, 군사, 교통 등 다양한 분야에서 더욱 널리 활용될 것으로 기대된다.

부 록 Ⅰ-1

(본 부록은 기술보고서를 보충하기 위한 내용으로 기술보고서의 일부는 아님)

시험인증 관련 사항

Ⅰ-2.1 시험인증 대상 여부

해당 사항 없음

Ⅰ-2.2 시험표준 제정 현황

해당 사항 없음

부 록 1-2

(본 부록은 기술보고서를 보충하기 위한 내용으로 기술보고서의 일부는 아님)

참고 문헌

- [1] AX Semantics, What is Natural Language Generation (NLG) & where is it used?, 2021.05.05.,
<https://www.ax-semantics.com/en/blog/natural-language-generation-explained>
- [2] Gnani Marketing, How do Voice Bots Handle Languages & Accents? | Gnani, 2022.07.01.,
<https://www.gnani.ai/resources/blogs/how-do-voice-bots-handle-languages-accents/>
- [3] Fabio Chiusano, Two minutes NLP — 33 important NLP tasks explained, 2021.12.07.,
<https://medium.com/nlplanet/two-minutes-nlp-33-important-nlp-tasks-explained-31e2caad2b1b>
- [4] Google Cloud, Cloud Speech-to-Text,
<https://cloud.google.com/speech-to-text?authuser=00&hl=ko>
- [5] 미디어오늘, 영화 ‘밀수’ 이어 ‘수사반장’ 드라마에도...자막 전성시대, 2024.05.04.,
<https://www.mediatoday.co.kr/news/articleView.html?idxno=317767>
- [6] SK텔레콤, SKT-JTBC스튜디오, AI 영상 자막기술 공동 개발, 2021.10.27.,
<https://news.sktelecom.com/171867>
- [7] 연합뉴스, SKB, AI 기술 활용한 한글 자막 생성 시스템 개발, 2023.08.21.,
<https://www.yna.co.kr/view/AKR20230821017400017>
- [8] 경향신문, 인공지능으로 자막 생성...VOD를 더 편하게, 2023.10.18.,
<https://www.khan.co.kr/ePR/article/202310181935015>
- [9] 한겨레, 한글 자막 안 나오면 답답하죠?...AI로 자막 만든다, 2023.08.21.,
<https://www.hani.co.kr/arti/economy/it/1105092.html>
- [10] 조선일보, 34國 번역가 3만명 AI로 보조... 넷플릭스·디즈니 세계 자막 시장 장악한 한국인, 2022.10.14.,
https://www.chosun.com/economy/tech_it/2022/10/14/QIXYG6BQFNIRJCXECN5UY2TIU/

[11] 엑스엘에이트, [어젯밤 당신이 본 글로벌OTT 자막, '누가' 번역했을까?], 2022.08.26.,
<https://m.post.naver.com/viewer/postView.naver?volumeNo=34201193&memberNo=58756197>

[12] <https://www.xl8.ai/products/mediacat-translate>

[13] 김태영, 간단한 음성 데이터 처리, 2024.07.15.,
<https://medium.com/@ashiofumiirerujiyu/간단한-음성-데이터-처리-72938c3f09f4>

[14] HuggingFace, Introduction to audio data,
https://huggingface.co/learn/audio-course/chapter1/audio_data

[15] MyloveKBS, 원고부터 선곡까지...국내 최초 AI가 진행하는 라디오 '스테이션 X' 첫 방, 2024.09.02.,
https://mylovekbs.kbs.co.kr/index.html?source=mylovekbs&sname=mylovekbs&style=blog&contents_id=70000000402009

[16] 한겨레, 대본도 진행도 혼자서...KBS 심야방송에는 '인공지능 DJ'가 있다, 2024.09.25.,
https://www.hani.co.kr/arti/culture/culture_general/1159723.html

[17] 연합뉴스, KT, AI 기술로 '마왕' 신해철 음성 복원...라디오방송 3편 제작, 2021.09.12.,
<https://www.yna.co.kr/view/AKR20210912013400017>

[18] 연합뉴스, 故 신해철, 'AI 新해철'로 돌아온다...목소리 모델 공개, 2024.05.03.,
<https://www.yna.co.kr/view/AKR20240503064700005>

[19] 연합뉴스, AI와 라디오 진행한 최화정 "사람의 온도는 따라 하지 못해", 2023.8.20.,
<https://www.yna.co.kr/view/AKR20230818113800005>

[20] SDF DIARY, Ep.162 | 'AI 화정'의 비밀, 2023.08.16.,
<https://sdf.or.kr/m/newsletter/10000001232>

[21] SBS, 최화정 목소리 학습한 'AI 화정' 등장...SBS 파워FM, 8월 1일 'AI 뮤직페스티벌' 개최, 2023.07.25.,
https://ent.sbs.co.kr/news/article.do?article_id=E10010274427&plink=COPYPASTE&cooper=SBSENTERNEWS

[22] SBS, '최화정X'AI 화정', '최파타' 더블 진행... 'AI 뮤직 페스티벌' 오늘(1일) 방송, 2023.08.01.,

https://news.sbs.co.kr/news/endPage.do?news_id=N1007290829&plink=COPYPASTE&cooper=SBSNEWSEND

[23] Getting started with Google Ad Manager's Dynamic Ad Insertion

https://services.google.com/fh/files/misc/getting_started_with_dynamic_ad_insertion.pdf

[24] Google IMA SDK

<https://developers.google.com/interactive-media-ads>

[25] Client-Side Ad Insertion (CSAI) vs. Server-Side Ad Insertion (SSAI)

<https://www.wowza.com/blog/csai-vs-ssai-client-side-ad-insertion-server-side-ad-insertion>

[26] Understanding SCTE-35

<https://codesequoia.wordpress.com/2014/02/24/understanding-scte-35/>

[27] Satellite Today “Schematic of Orbital Altitudes and Coverage Areas”, 2020

[28] Xona Space Systems, “Introducing Pulsar”, 2024

[29] European Space Agency, “Industry invited to bid for low-Earth orbit satnav demo”, 2023

[30] Broadcast Positioning System(BPS) Using ATSC 3.0 (Tariq Mondal, Robert D. Weller, Sam Matheny), 2023

[31] 한국형 항공위성서비스(KASS) 운영 시대 본격 개막, 국토교통부 보도자료, 2022

[32] 방송 미디어 서비스 및 기술 동향, 미래방송포럼 방송하이브리드분과 기술보고서, 2023

[33] NAVISP-EL2-069 “SSRoverDAB+”, 2023

부 록 1-3

(본 부록은 기술보고서를 보충하기 위한 내용으로 기술보고서의 일부는 아님)

영문기술보고서 해설서

해당 사항 없음

부 록 1-4

(본 부록은 기술보고서를 보충하기 위한 내용으로 기술보고서의 일부는 아님)

기술보고서의 이력

판수	채택일	표준번호	내용	담당 위원회
제1판	2024.11.29	제정 FBMF-TR-017	-	방송IP하이브리드 분과위원회
오류정정				
제2판				