

# 음성의 진화: 하이퍼보이스(Hypervoice) 기술



현욱 한국전자통신연구원 표준연구센터 책임연구원

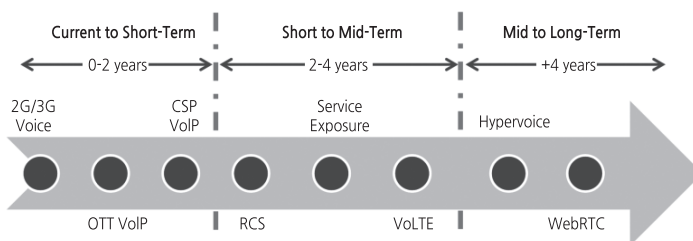
## 1. 머리말

하이퍼보이스는 음성 데이터와 텍스트 간 연결이 가능한 색인기능을 제공함으로써 멀티미디어 데이터에 대한 접근성을 향상시키며, 다양한 매시업(mesh-up) 서비스를 만들어 낼 수 있는 응용 기술이다. 기존의 하이퍼텍스트(Hypertext) 기술이 문서/파일 단위의 이동을 주목적으로 하였다면, 하이퍼보이스는 음성 데이터의 원활한 탐색 지원을 주요 목표로 한다. 이와 같이 미디어 내에서의 원활한 탐색(navigation) 기능은 엔터프라이즈급 UC(Unified Communication) 서비스, VoD, 오디오

오북 등에도 쉽게 접목이 가능하다는 장점이 있다.

2013년 가트너 보고서에 따르면, 음성 혁명의 다음 레벨로 각종 미디어의 매시업이 이루어질 것이며, 이러한 음성정보를 인덱싱하고 검색하게 함으로써 사용자에게 좀 더 나은 UX를 제공하는 하이퍼보이스가 각광받을 것으로 예측했다. 그 후 북미를 중심으로 다양한 솔루션/서비스가 제공 중에 있다. 또한, 최근 각광받고 있는 인공지능 기술 기반의 음성인식기술의 성능이 비약적으로 상승함에 따라 앞으로 더욱 다양한 하이퍼보이스 서비스들이 출시될 것으로 예상된다.

2014년에 발간된 Global Industry Analysis에 따

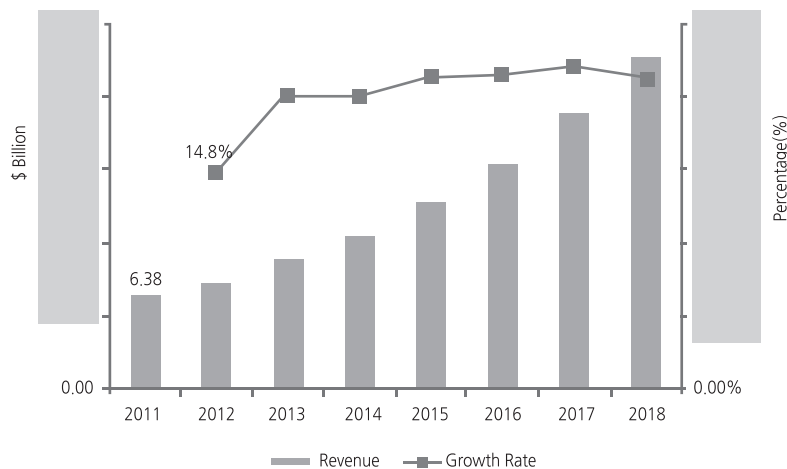


CSP = communications service provider, VoIP = voice over Internet Protocol; VoLTE = voice over Long Term Evolution; OTT = over-the-top; WebRTC = Web real-time communication

※ 출처: Gartner Market Trends 2013: The Future Evolution of CSP Voice Services

[그림 1] 음성통신 서비스 시장의 진화 방향 예측

EMEA UNIFIED COMMUNICATIONS MARKET, 2012 - 2018(USD BILLION)



Source: Transparency Market Research

[그림 2] UC(Unified Communications) 시장 예측

르면, 하이퍼보이스 기술의 주요 대상 중 하나인 UC 시장이 2013년 14.8%의 대폭적인 성장에 이어 2018년에는 61.9억 달러 성장할 것으로 예측하였다.

본고는 하이퍼보이스 서비스의 핵심 요소기술과 국내/국제 표준화 현황을 간략히 기술하고, 하이퍼보이스 서비스를 전망하고자 한다.

## 2. 핵심 요소 기술

하이퍼보이스 서비스 확산을 위한 주요 요소 기술로는 음성인식 기술, 하이퍼보이스 마크업 및 변환기술, 매시업 응용서비스 기술이 있으며, 이 세 가지 요소 기술들의 적절한 조합이 서비스 성패를 가를 것으로 생각된다.

### 2.1 인공지능 기반 음성인식 기술

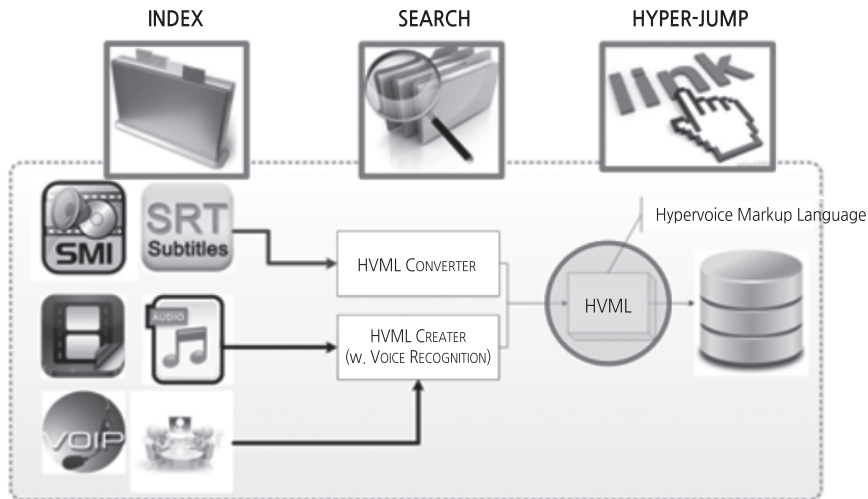
하이퍼보이스 서비스의 품질은 음성인식 기술의 성능과 밀접하게 관련되어 있다. 즉, 음성인식 기술의 수준에 따라 서비스에 대한 평가가 엇갈리게 될

것이다. 이에 전 세계 유수의 IT 기업에서는 Deep Learning 기반의 음성인식 기술을 확보하고 시장을 선점하기 위한 서비스 출시에 박차를 가하고 있으며, 대다수의 API(Application Programming Interface)는 Deep Learning 기술을 이용하여 음성인식의 정확도를 높이고 있다.

2016년 3월에 구글은 딥러닝 기술을 이용, 단문 모드/연속모드를 지원하는 Cloud Speech API를 공개하였으며, 지원언어 또한 지속적인 확장 중에 있다. 국내의 네이버 또한 2016년 1월에 음성인식/음성합성/기계번역 등에 관련된 API를 공개하는 등 후속 기술들이 등장하고 있다. 이 외에도, 아마존, IBM, 애플, 다음 등도 자체 기술을 확보하고 API를 공개하며 기술지원을 통해 3<sup>rd</sup> Party 제조사 등의 우군을 확보하는 등 생태계를 키우기 위한 세력 규합이 이뤄지고 있는 상태다.

### 2.2 하이퍼보이스 마크업 표준 및 변환 기술

하이퍼보이스 마크업 언어(HVML, HyperVoice



[그림 3] 하이퍼보이스의 활용 사례

<표 1> HVML(HyperVoice Markup Language) 관련 W3C 표준

HVML	설 명	지원 브라우저
WebVTT	WebVTT(Web Video Text Tracks Format)은 HTML<track> 요소를 통해 외부 텍스트 트랙 리소스를 마크업함으로써, 비디오 내용에 캡션이나 자막을 제공하고 텍스트 비디오 설명, 내용 탐색을 위한 정보를 미디어와 시간이 일치하도록 메타 데이터를 제공하는 기술	IE10, 크롬, 사파리, 오페라
TTML	Timed Text Markup Language(Timed Text Markup Language)는 저작 시스템 간의 상호 교환을 위해 시간이 지정된 텍스트 미디어를 나타내는 기술로, HTML<track>, <text>, <textstream>를 통해 사용되며, HbbTV, DVB, DASH, ARIB, CFF 등의 TV 셋톱 등에서 많이 사용되고 있다.	DVB 셋톱 등 방송 산업

Markup Language)는 텍스트 데이터와 보이스 데이터 간 맵핑 관계를 제공하기 위한 마크업 언어를 총칭한다. 2000년대 초 모토로라에서 VoxML(Voice Markup Language) 기술을 개발하여 SDK를 제공하는 등 서비스 확산을 도모 하였으나, 당시 음성인식 기술의 미성숙과 자사 라이선스 지위 강화를 위해 국제 표준화에 소극적이었던 문제 등으로 실패했다.

하이퍼보이스 마크업 언어와 관련하여, W3C(World Wide Web Consortium)에서 표준화되고 있는 HVML 관련 표준들로는 <표 1>과 같은 TimedText, WebVTT 등이 있다.

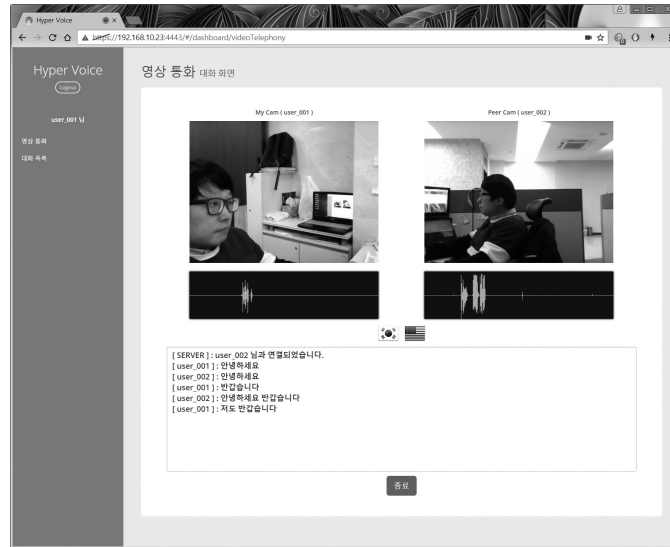
한동안 WebVTT와 TTML의 경쟁이 있었으나, 최

근에는 WebVTT가 주류로 자리매김하였으며 W3C의 Web Media Text Tracks 그룹에서 WebVTT 규격, 기존 포맷(TTML, SRT 등)의 변환을 위한 맵핑 작업이 진행 중에 있다. 이를 기반으로 [그림 3]과 같이 다양한 변환 툴이 등장할 것으로 예상된다.

## 2.3 하이퍼보이스 기반 매시업 서비스를 위한 Web API 기술

하이퍼보이스 기반 매시업 서비스를 구현하기 위해 사용될 수 있는 주요 Web API들은 <표 2>와 같다.

[그림 4]는 하이퍼보이스 서비스에 대한 PoC(Proof of Concept)를 위해 ETRI에서 개발 중인 프로토타입 화면으로, 위 3가지의 API 조합을 통해 원



[그림 4] 웹 기반 하이퍼보이스 프로토타입

<표 2> 하이퍼보이스 기반 매시업 서비스를 위해 유용한 Web API 기술들

API	설 명	지원 브라우저
WebRTC API	WebRTC API는 웹 브라우저 간 직접 영상통화 서비스를 구현하기 위한 API 기술	크롬, 파이어폭스, 오페라, 안드로이드, iOS 등
WebAudio API	웹 브라우저상에서 오디오 데이터를 직접 제어하기 위한 API 기술	크롬, IE9, 파이어폭스, 오페라, 사파리, 안드로이드, 타이젠 등
WebSpeech API	브라우저에 내장된 음성인식 기능을 이용하기 위한 API 기술	크롬

격영상회의에서 생성된 회의록과 녹음된 미디어간의 맵핑을 제공하는 HVMML 생성기능을 보여준다.

### 3. 국내 · 국제 표준화 현황 및 전망

#### 3.1 국제표준화 현황

W3C에서는 2004년도부터 VoiceXML, SCXML, SISR, SRGS, SSML, CCXML 등의 TTS(Text-to-speech) 표준을 개발하였으며, 최근에는 TTML(Timed Text Markup Language)와 WebVTT(Web Video Text Tracks)에 대한 표준화가 활발히 진행 중에 있다. 북미의 에릭슨, 인터디지털 등을 중심으로 2012년도에 구성된 하이퍼보이스

컨소시엄에서는 백서 발간, 기술 마케팅 등 프로모션 활동을 진행 중이며, 최근 들어 인공지능 기술의 급성장에 따른 음성 인식률의 비약적인 향상과 더불어 관련 활동에 탄력을 받을 것으로 예상된다.

#### 3.2 국내표준화 현황

TTA 차세대PC PG에서는 제조사, 이통사, 소방방재청, 경찰청, 국책연구원 등을 주축으로 전담반을 운용해 음성만으로 긴급통화를 연결하도록 하는 ‘스마트폰 음성인식 긴급전화 서비스 표준’을 2014년에 제정하였다. 2015년 TTA 지능형로봇 PG에서는 로봇에 탑재된 음성인식 엔진의 객관적인 성능평가를 위한 ‘서비스 로봇 음성인식 성능평가를 위



[그림 5] 하이퍼보이스 서비스 활용 분야

한 API' 표준을 제정하였다. 최근 인공지능을 기반으로 하는 음성인식 기술의 가파른 성장과 더불어 앞으로 음성인식 정보의 활용과 새로운 서비스 창출을 위한 표준들이 지속적으로 개발될 것으로 예상된다.

#### 4. 하이퍼보이스 서비스 및 활용분야

하이퍼보이스가 가장 먼저 적용될 수 있는 분야로는 원격회의, 개인음성비서, 멀티미디어 지점 검색 등이 있으며, 로봇 및 웨어러블 컴퓨팅 기기와의 통신 서비스에도 적용이 가능하다.

##### 4.1 원격 교육/웹 세미나

현재 대다수의 원격 교육은 VoD와 교재가 병합된 형태로 서비스가 되고 있으나, 미디어 데이터 검색기능은 제한적이다. 하이퍼보이스 기술 접목을 통해, 사용자가 콘텐츠 내에서 원하는 내용에 바로 접근할 수 있도록 함으로써 사용자 편의성을 증대시킬 수 있고, 제품의 경쟁력 향상에도 직접적인 영향을 미칠 수 있다.

##### 4.2 멀티미디어 라이브러리

현재 멀티미디어들에 대한 검색은 제목이나 미리 지정된 메타데이터 방식의 검색만 가능하거나 자막 데이터에 대한 검색도 일부 가능하나, 미디어 검

색을 제공하는 경우는 흔치 않다. 하이퍼보이스 기술을 접목하게 되면, 콘텐츠 내부의 데이터에 대해서도 검색이 가능하므로 사용자의 콘텐츠 접근성을 향상시킬 수 있는 기능을 제공할 수 있다.

##### 4.3 UC

UC 환경에서 단순한 Call Recording을 넘어서, 음성데이터의 자료화가 가능해지게 함으로써 업무 생산성을 높일 수 있다. 북미 업체에서는 보이스와 텍스트 간 인덱싱을 통해 음성, 영상 통신을 비롯한 다자 간 컨퍼런싱, 전화 녹음, Voice-to-text 변환, 음성 분석 등에 관련된 CaaS(Communication-as-a-service) 개념의 클라우드 플랫폼으로 서비스를 제공 중에 있다.

#### 5. 맺음말

하이퍼보이스 기술은 음성인식, 딥러닝, 머신러닝 등의 기반 기술을 바탕으로 다양한 응용 서비스를 창출할 수 있는 일종의 프레임워크 기술에 해당된다. 이를 위해서는 하이퍼보이스와 보이스 데이터 간 연계 관계를 기술하는 HVML 마크업 언어의 표준화와 더불어 각종 매시업 서비스에 대한 아이디어를 기반으로 한 비즈니스 모델, 서비스 시나리오, 관련 응용 표준들의 개발이 선행되어야 한다. 이러한 개발이 이루어진다면 음성 서비스에 혁명

(Voice Revolution)적 변화를 일으킬 것이며 따라서 관련 핵심기술을 확보하는 것이 매우 중요하다.



※이 논문은 2015년도 정부(미래창조과학부)의 재원으로 정보통신 기술진흥센터의 지원을 받아 수행된 연구임[No.R0127-16-1053, 4K/8K UHD 영상컨텐츠 분산 스트리밍 프로토콜 표준개발].

## [참고문헌]

- [1] ICT 표준화전략맵 2017, [http://www.tta.or.kr/data/reporthosulist.jsp?kind\\_num=5](http://www.tta.or.kr/data/reporthosulist.jsp?kind_num=5)
- [2] Hypervoice consortium, <http://www.hypervoice.org/>
- [3] 다음 뉴튼 API, <https://developers.daum.net/services/apis/newtone>
- [4] 네이버 음성인식 API, <https://developers.naver.com/products/vrecog>
- [5] 구글 클라우드 스피치 API, <https://cloud.google.com/speech/>
- [6] W3C, Web Speech API Specification, <https://dvcs.w3.org/hg/speech-api/raw-file/tip/webspeechapi.html>
- [7] Gartner Report, 'Market Trends: The Future Evolution of CSP Voice Service', 14, Nov 2013,
- [8] Unified Communications Insight, 'Hypervoice: the new 'bionic' Voice in Unified Communications', 2013
- [9] Hypervoice consortium, 'Hypertext to Hypervoice – Linking what you say to what you do', 2012
- [10] <https://w3c.github.io/webvtt/>
- [11] <https://aws.amazon.com/ko/amazon-ai/>
- [12] [https://en.wikipedia.org/wiki/HTML5\\_Audio](https://en.wikipedia.org/wiki/HTML5_Audio)
- [13] [http://www.voicexml.org/wp-content/uploads/sites/2/static/Review/Jul2003/features/Jul2003\\_motorola\\_voxgateway.html](http://www.voicexml.org/wp-content/uploads/sites/2/static/Review/Jul2003/features/Jul2003_motorola_voxgateway.html)
- [14] [https://en.wikipedia.org/wiki/HTML5\\_Audio](https://en.wikipedia.org/wiki/HTML5_Audio)



## 미션 크리티컬 Mission Critical, MC

업무 수행을 위하여 가장 중요한(필수 불가결한) 요소.

미션 크리티컬 요소가 정상적으로 작동되지 않거나 파괴되면 업무수행 전체에 치명적인 영향을 미쳐, 조직이나 사회에 재앙을 가져올 수 있다. 예를 들어 온라인 (online) 비즈니스 회사의 통신 시스템이나 재난 통신망, 항공기 운항의 관제 시스템, 그리고 IT 정보 제공 회사의 데이터베이스 시스템 등이 이에 해당된다. 미션 크리티컬 시스템은 완벽한 동작을 위하여 보안시스템도 철저하게 갖추어야 한다.