

음성 인식 서비스



데이터분석본부 호남지원 선임연구원 이은지 Tel: 062-951-7704, e-mail: eunji_lee@kisti.re.kr

KEY FINDING

1. 음성 인식 기술은 인간의 언어를 사용해 기계와 상호작용할 수 있는 기능을 제공하며, 최근에는 생체 인증 및 문서화 기술이 주목받고 있다.
2. 음성 인식 기술은 종단간 방식의 적용과 자연어처리, 자연어이해 기술의 발달로 정확도가 견고해지면서, 다양한 산업별 응용서비스 확장이 기대된다.
3. 음성 인식 서비스의 세계시장은 2021년 82억7,200만 달러 규모이며, 연평균 21.6 % 성장으로 2026년 219억9,500만 달러가 될 것으로 전망된다.
4. 음성 인식 서비스는 기반기술, 플랫폼, 디바이스, 응용서비스 등 융복합형 산업을 형성하고 있으며, 커넥티드카, 웨어러블 기기, 로봇 시장의 성장은 기존의 기기 제어, 명령 실행 등을 지원하는 음성 인식 서비스 시장을 지속적으로 견인할 것이다.
5. 최근 비대면 업무 증가는 음성 기록, 문서화 서비스의 수요를 크게 증가시킬 것으로 예상되며, 단순 반복 업무 자동화와 개인 식별 및 인증을 기반으로 한 금융, 정부, 의료, 기업 서비스 등 응용서비스 시장이 성장할 것으로 전망된다.

1) 음성 인식 서비스의 개요

| 음성 인식 기술의 개념

음성 인식 기술이란 일반적으로 컴퓨터가 입력받은 인간의 음성 언어를 인식해 문자로 변환하는 기술을 말하며, 인식된 결과에 대하여 음성 이해(Speech Understanding)과정을 통해 음성의 의미에 부합하는 명령을 수행하는 기술을 포함한다.

최초의 음성 인식 기술은 컴퓨터로 입력된 음성 파형이 어떤 음과 가까운지, 어느 단어를 나타내는지를 인식하는 방식이었다. 컴퓨팅 성능의 향상으로 언어 모델링 개념이 도입되면서 대용량의 언어 모델을 기반으로 언어적 패턴의 특징을 학습해 입력되는 음성의 특징을 기반으로 음성을 인식하는 단계적 방식으로 발전하였다.

최근에는 종단간(End-to-End) 구조 기반의 학습 방식이 주목받고 있다. 종단간 방식은 인간 처럼 음성을 들은 후 바로 단어와 문장으로 인식하는 방법으로 단계적으로 처리하는 기존의 방식에 비해

더 많은 데이터와 연산과정이 필요한 난점을 가지고 있지만, 단계마다 필요한 작업들을 생략하고 시스템을 하나로 통합해 운영하기 때문에 시스템 복잡도를 낮출 수 있어 음성 인식의 정확성 및 속도의 향상을 위한 핵심기술이 될 것으로 평가받고 있다.

I 음성 인식 서비스의 현황

음성 인식 기술은 타 산업의 제품 및 서비스에 적용을 통해 새로운 비즈니스 창출이 가능한 기반기술로서 사용자의 편의성을 증진시

그림 1 단계적 모델과 종단간 모델 개념도

〈단계적 모델(기준)〉

Audio → Compute features → Phonemene recognizer → Final recognizer → Output

〈종단간 모델〉

Audio → Learning algorithm → Transcript

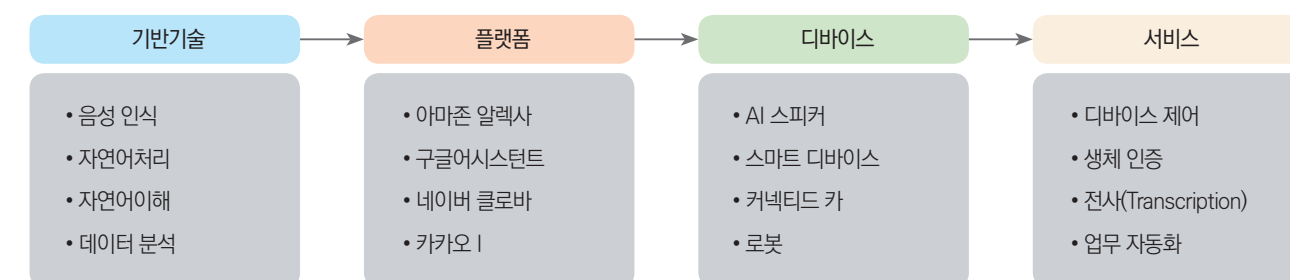
출처 : Andrew Ng, Machine Learning Yearning

표 1 음성 인식과 터치/텍스트 기반 기술 비교

| 구 분 | 음성 인식 | 터치/텍스트 인식 |
|-----------|-------------------------|----------------------|
| 입력속도 | 150 단어/분 | 40 단어/분 |
| 입력방식 | Random Access | Hierarchical GUI |
| 정보 입출력 방식 | 마이크, 스피커 | 키보드, 디스플레이 |
| 커뮤니케이션 방식 | 컨텍스트 기반 양방향 | 단방향 |
| 멀티태스킹 | Hands-free기능으로 멀티태스킹 가능 | 터치나 텍스트 입력시 멀티태스킹 불가 |

출처: 음성인식 AI 비서 시장의 현황과 시사점, KISDI, 2017

그림 2 단계적 모델과 종단간 모델 개념도



키기 위해 지속적인 연구가 필요한 인간-컴퓨터 상호작용(Human-Computer Interaction) 기술 중 하나로 평가받고 있다. 그리고 음성 인식 기술은 키보드나 기타 입력장치를 통한 입력이 아닌 사람의 음성을 통한 입력이 가능하므로 손발이 자유롭지 못한 상황에서도 정보를 전달할 수 있어 사용자가 다른 활동을 하면서도 서비스를 이용할 수 있는 점에서 안전성과 편의성을 지향하는 다양한 서비스에 도입되고 있다.

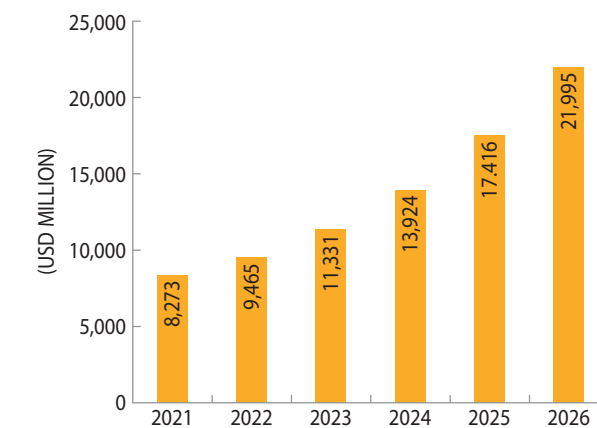
초기 음성 인식 기반 서비스는 AI 스피커, 스마트 홈, 스마트 카와 같이 음성 인터페이스를 통해 디바이스를 제어하는 분야를 중심으로 발전하였으며, 현재는 단순하고 반복적인 업무의 자동화를 위해 음성 인식 기술을 적용하고자 하는 수요가 크게 증가함에 따라 콜센터 문의 및 서비스 응대, 음성 기록의 자동 텍스트 변환 등 업무 자동화 분야로 영역이 확장되고 있다.

2) 시장동향 및 전망

I 시장 규모

각종 산업분야에서 AI, IoT, AR 등의 첨단기술을 서비스에 접목하려는 시도가 계속되는 가운데 음성 인식 기술을 활용한 서비스가 증가하고 있다. 음성 인식 서비스의 세계 시장규모는 2021년 82억

그림 3 세계 음성 인식 시장규모('21-'26)



출처 : Markets&Markets, Speech & Voice Recognition Market, 2021

그림 4 지역별 음성 인식 시장규모('21-'26)

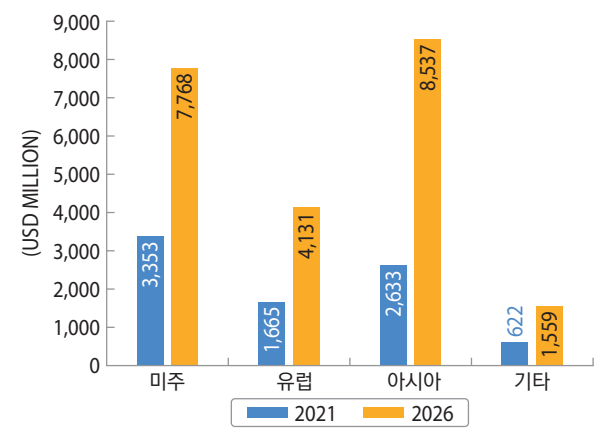


표 4 지역별 음성 인식 서비스 시장 현황 및 전망(2021-2026)

(단위: 백만달러)

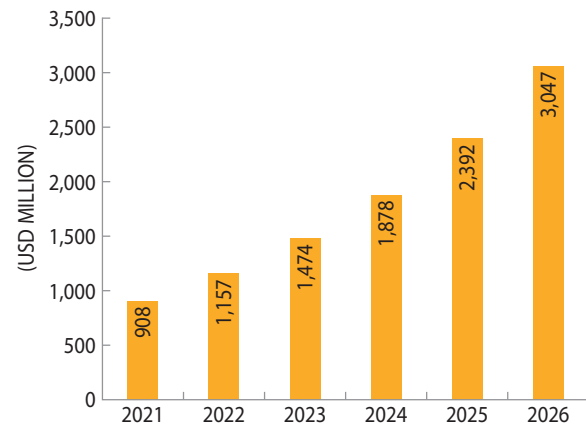
| 구분 | 2021년 | 2022년 | 2023년 | 2024년 | 2025년 | 2026년 | CAGR(%) |
|-----|-------|-------|--------|--------|--------|--------|---------|
| 미주 | 3,353 | 3,732 | 4,346 | 5,195 | 6,321 | 7,768 | 18.3 |
| 유럽 | 1,665 | 1,882 | 2,224 | 2,696 | 3,323 | 4,131 | 19.9 |
| 아시아 | 2,633 | 3,142 | 3,919 | 5,011 | 6,514 | 8,537 | 26.5 |
| 기타 | 622 | 709 | 842 | 1,022 | 1,258 | 1,559 | 20.2 |
| 합계 | 8,273 | 9,465 | 11,331 | 13,924 | 17,416 | 21,995 | 21.6 |

출처 : Markets&Markets, Speech & Voice Recognition Market, 2021

국내의 음성 인식 시장규모는 2021년 9억800만 달러이며, 연평균 27.4%로 성장해 2026년 30억4800만 달러가 될 것으로 전망된다. 이는 COVID-19의 영향으로 조성된 비대면 문화의 확산에 따라 언택트 스토어, 금융권 컨택센터, 상담 및 조사가 필요한 공공·의료

서비스 분야로 음성 인식 솔루션의 적용 범위가 확대되었으며, 팬데믹이 종식되더라도 편리한 언택트 소비 경험이 하나의 문화로 정착되어 지속적으로 음성 인식 응용 서비스 수요확대가 이어질 것으로 전망된다.

그림 5 국내 음성 인식 서비스 시장규모('21-'26)



출처 : Markets&Markets, Speech & Voice Recognition Market, 2021

그림 6 아시아 국가별 음성 인식 시장규모('21-'26)

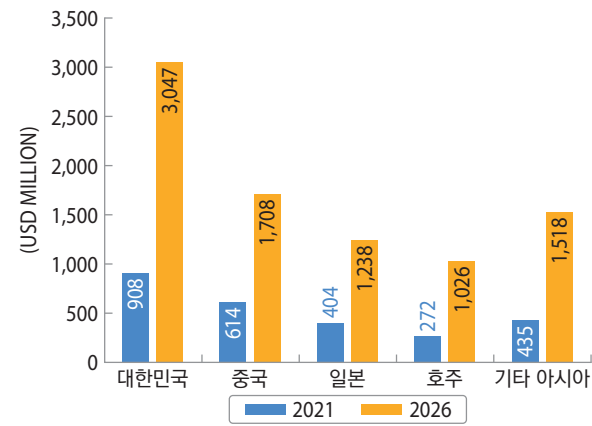


표 5 아시아 지역별 음성 인식 서비스 시장 현황 및 전망(2021-2026)

(단위: 백만달러)

| 구분 | 2021년 | 2022년 | 2023년 | 2024년 | 2025년 | 2026년 | CAGR(%) |
|--------|-------|-------|-------|-------|-------|-------|---------|
| 대한민국 | 908 | 1,157 | 1,474 | 1,878 | 2,392 | 3,048 | 27.4 |
| 중국 | 614 | 753 | 924 | 1,134 | 1,392 | 1,708 | 22.7 |
| 일본 | 404 | 505 | 632 | 791 | 989 | 1,238 | 25.1 |
| 호주 | 272 | 355 | 463 | 603 | 789 | 1,026 | 30.4 |
| 기타 아시아 | 435 | 559 | 717 | 921 | 1,182 | 1,518 | 28.4 |

출처 : Markets&Markets, Speech & Voice Recognition Market, 2021

적용 분야별 시장규모

음성 인식 서비스는 소비자용 서비스, 자동차, 기업, 금융, 의료, 군사, 교육, 기타 분야로 구분할 수 있다. 그 중에서 소비자용 서비스가 가장 높은 점유율을 보이며, 2021년 36억1,900만 달러이고 연평균 27.6%로 성장해 2026년 122억5,000만 달러에 이를 것으로 전망된다. 소비자용 서비스가 높은 비중을 차지하는 이유는 모바일, 랩톱, 태블릿 및 스마트 홈 장치와 같은 소비자 제품의 음성 인식 기술 보급률이 높기 때문이다. 향후 웨어러블 디바이스에 대한 수요 증

가 및 개인화된 장치에 대한 인증·보안 문제와 같은 요인은 음성 인식 시장의 성장을 촉진할 것으로 예상된다.



표 6 음성 인식 서비스 분야별 활용 사례

| 구분 | 활용 사례 |
|----------|---|
| 소비자용 서비스 | 모바일, 웨어러블 디바이스, IoT 홈 서비스 등 음성 인터페이스 기반 디바이스 제어 |
| 의료 | 디지털 헬스케어 분야(전자의무기록 자동작성, 환자기록 검색 등) |
| 기업 | 고객센터(예약서비스 등 단순·반복 업무응대), 회의록 작성 등 |
| 금융 | 금융/은행/보험 상담업무, 사용자 인증 등 |
| 정부 | 음성 인식 모바일 행정 서비스(민원·상담 업무 등), 민원인 신원인증 등 |
| 자동차 | 주행 중 기타장치 제어, 사용자 음성 기반 인증 등 |
| 교육 | e-learning, 어학교육 콘텐츠 등 |
| 유통업 | 음성 쇼핑 서비스 (음성인식 기반 간편구매 서비스, 쇼핑제한 등) |

출처 : Markets&Markets, Speech & Voice Recognition Market, 2021, KISTI 재구성

표 7 지역별 음성 인식 서비스 시장 현황 및 전망(2021-2026)

(단위: 백만달러)

| 구분 | 2021년 | 2022년 | 2023년 | 2024년 | 2025년 | 2026년 | CAGR(%) (2021-2026) |
|----------|-------|-------|-------|--------|--------|--------|---------------------|
| 소비자용 서비스 | 3,508 | 4,256 | 5,370 | 6,924 | 9,048 | 11,888 | 27.6 |
| 의료 | 1,263 | 1,322 | 1,433 | 1,578 | 1,745 | 1,917 | 8.7 |
| 기업 | 1,002 | 1,108 | 1,279 | 1,511 | 1,814 | 2,191 | 17.0 |
| 금융 | 496 | 509 | 539 | 576 | 613 | 639 | 5.2 |
| 정부 | 169 | 178 | 195 | 217 | 242 | 270 | 9.8 |
| 자동차 | 167 | 185 | 213 | 252 | 302 | 365 | 17.0 |
| 교육 | 154 | 170 | 195 | 229 | 274 | 329 | 16.3 |
| 군사 | 153 | 157 | 167 | 179 | 191 | 201 | 5.6 |
| 법률 | 99 | 111 | 131 | 157 | 193 | 237 | 19.1 |
| 유통업 | 89 | 98 | 113 | 134 | 161 | 195 | 17.2 |
| 기타 | 19 | 20 | 21 | 23 | 25 | 27 | 7.4 |
| 합계 | 7,119 | 8,114 | 9,656 | 11,780 | 14,608 | 18,259 | 20.7 |

출처 : Markets&Markets, Speech & Voice Recognition Market, 2021

국내외 주요기업 현황

음성 인식 서비스 세계 시장에서는 애플, MS, 구글, 아마존 등의 빅테크 기업이 시장을 주도하고 있으며, 기업별 자사 AI 플랫폼을 중심으로 각종 AI 서비스를 확장하는 형태로 서비스를 제공하고 있다.

최근에는 빅테크 기업의 활발한 스타트업 인수합병(M&A)과 적극적인 R&D 투자를 기반으로 AI 플랫폼 및 서비스 영역을 확장하고 있다. 최근 클라우드 기반으로 AI 플랫폼을 운영하는 빅테크 기업들은 기술력 보유 스타트업에 대한 M&A를 수익증대 방법 중 하나로 여기고 있으며, M&A는 새로운 서비스를 위한 초기 투자비용이나 기술개발 비용을 지출하지 않고도 즉시 사업을 시작할 수 있고, 쉽게 경쟁

우위를 확보할 수 있기 때문에 시장진출 전략으로 활용되고 있다. 국내 시장에서는 네이버, 카카오가 AI 플랫폼을 출시하고 주도해 왔으며, 최근에는 국내 AI 플랫폼 분야 스타트업의 성장으로 의료분야,

콜센터 상담분야 등 자체 음성 인식 기술을 적용해 다양한 응용서비스분야로 영역을 확장하고 있다.

표 8 주요기업별 AI 플랫폼 및 음성 인식 기술 적용 분야

| 주요기업 | AI 플랫폼 · 음성 인식 기술 | 적용 분야 |
|--------|---|--|
| 애플 | • Siri • Apple OS 탑재 기기 명령 실행, 기기 제어, 음성 인식, 화자 인식 등 | • 아이폰, 아이패드, 맥PC, 애플워치 • 스마트 스피커 |
| MS | • Azure AI • 명령 실행, 음성 인식(STT), 음성 합성(TTS), 화자 인식 등 | • 윈도우 PC • 모바일 |
| 구글 | • Google Assistant • Google Now의 정보 분석 기능과 음성 인식 기술의 접목으로 개인화된 AI 비서 중 높은 서비스 품질로 평가됨 | • 스마트 스피커 • 스마트 TV • 카 인포테인먼트 |
| 아마존 | • Alexa • 종단간 모델 적용으로 음성 인식 정확도 개선, 음성 합성, 명령 실행 등 | • 음성 쇼핑 • 스마트 스피커 • 스마트 TV • 카 인포테인먼트 |
| 네이버 | • CLOVA Speech(단계적 모델), CLOVA NEST(종단간 모델) • 음성 인식, 음성 합성, 화자 인식 등 한국어 음성 인식 중 가장 우수한 성능으로 평가됨 | • 회의록 자동작성 • 환자상태 케어콜 |
| 카카오 | • 카카오 I • 음성 인식, 음성 합성 등 | • 자동차 음성인식 • 스마트 스피커 |
| 솔트룩스 | • AI Suite • 음성 인식, 음성 합성, 시맨틱 검색, 자연어처리 등 솔루션 제공 | • AI 상담시스템 • 콜봇 서비스 |
| 마인즈랩 | • maum ai • 음성 인식, 음성 합성 등 인공지능 클라우드 서비스 제공 | • AI 콜센터 • 회의록 자동작성 |
| 셀바스 AI | • Selvy • 딥러닝 기반 패턴인식 솔루션 및 서비스 개발 | • AI 의료 음성인식 • AI 고객센터 • AI 조서작성 • 안내방송 |

3) 분석자 인사이트

정부의 정책과 지원에 힘입어 AI 산업 분야는 빠르게 성장했고, 상용화 단계에 이르렀다. 이러한 시장 수요의 성장에 따라 음성 인식 분야 역시 응용 솔루션 등 최신 기술의 개발이 가속화할 것으로 전망된다. 음성 인식 기술은 인간의 언어를 사용하여 기계와 상호 작용할 수 있는 기능을 제공해 왔고, 최근에는 가장 빠른 생체 인식 및 문서화 기술로서 주목받고 있다.

음성 인식 기술은 종단간 방식의 적용과 자연어처리(NLP), 자연어 이해(NLU) 기술의 발달로 정확도가 견고해지면서, 다양한 산업별 서

비스가 점차 확대되어 음성 인식 서비스의 대중화를 예상할 수 있다. 커넥티드카, 웨어러블 기기, 로봇 시장의 성장은 기존의 기기 제어, 명령 실행 등을 지원하는 음성 인식 서비스 시장을 지속적으로 견인할 것이며, 최근 비대면 업무 증가는 음성 기록 서비스의 수요를 크게 증가시킬 것이다. 이 외에도 음성 인식 기술을 기반으로 한 업무 자동화(전사, 고객센터 등), 개인 식별 및 인증은 금융, 정부, 의료, 기업 서비스 등 다양한 응용서비스 시장이 성장할 것으로 예상된다.

음성 인식 시장 진입을 위한 촉진요인, 저해요인 등의 주요 요인은 <표 9>와 같이 정리할 수 있다.

표 9 시장 영향요인 분석

촉진 요인

- 종단간(End-to-end) 방식 적용 등 음성 인식 정확도 향상 기대
- 음성 인터페이스 기반 스마트 애플리케이션 소비자 인식 확대
- 음성 인식 기술을 활용한 업무 자동화에 대한 관심 증가

저해 요인

- 기술개발에 대한 막대한 투자비용으로 소규모 업체의 시장 진입 어려움
- 개인 정보 보호 및 보안에 대한 우려
- 노이즈 환경에서의 음성 인식 성능 저하
- 클라우드 기반 AI 플랫폼 서비스 기업의 독과점

기회 요인


- 로봇, 커넥티드카 분야에서 높은 성장 잠재력
- 음성 기반 생체 인식 시스템에 대한 수요 증가
- 첨단 기술 제품에 대한 소비자 수용 및 선호도 증가

위험 요인

- 한국어 음성 인식 성능 한계
- 음성 인식 학습 데이터에 대한 규제
- 서비스 환경 등 음성 샘플의 품질 변화로 인한 오류 증가 가능성

음성 인식 시장은 기반 기술에 종속성 때문에 기반 기술을 가지고 있지 않은 소규모 기업의 시장 진출은 어려울 수 있다. 하지만 공개된 API 활용 또는 AI 플랫폼 기업과의 협력을 통해 새로운 서비스를 발굴할 수 있다.

다각적 관점에서 차별화된 서비스의 개발을 위해 기업에서 발생되는 다양한 서비스 도메인별 음성 시나리오 데이터의 수집과 라벨링된

AI 학습데이터 구축은 서비스 경쟁력 확보 및 시장진출에 중요한 역할을 할 것으로 판단된다. 음성 인식 서비스는 콘텐츠, 플랫폼, 네트워크, 디바이스 등의 기술을 포함하는 융복합형 산업의 특성을 가지고 있기 때문에 시장 트렌드에 주목하고 꾸준한 분석을 통해 새로운 사업 기회를 발굴하려는 노력이 필요할 것이다. 

ASTI MARKET INSIGHT



본원 (우)34141 대전광역시 유성구 대학로 245 한국과학기술정보연구원
T. 042) 869-1004, 1237 F. 042) 869-1091

분원 (우)02456 서울특별시 동대문구 회기로 66 한국과학기술정보연구원
T. 02)3299-6114 F. 02)3299-6244

