

# 출제예감 - 02 음성인식 비서

## Speech Recognition

정보관리기술사 최창환  
(buksamfight@naver.com)

### ZERO UI의 실체화를 통한 AI Speaker로의 진화!

<p><b>Concept</b></p>	<p><b>(정의)</b></p> <ul style="list-style-type: none"> <li>- 음성, 이미지등을 기계(컴퓨터등)로 식별하여, 사용자와 교감하여 편의성을 제공 해 주는 기술</li> <li>- 사용자의 음성을 인식해 사용자의 명령에 따라 음악을 재생하거나 스마트폰, 가전 제품이나 자동차 등 IoT로 제어할 수 있는 기기 등을 작동하는데 사용</li> </ul>
<p><b>KeyWord</b></p>	<p>음성인식, 텍스트 인식, UI/UX, 음성합성, 딥러닝 , AI 플랫폼 , AI Speaker</p>

#### 스피커는 왜 AI를 품었을까

기술사 시험에서 우리는 다양한 관점으로 학습을 하고 있습니다. 전통적 도메인, 최신 트렌드, 사회적 이슈등을 지식과 경험을 통해 답안을 만들기 위해 노력 하고 있습니다. 최근 진행한 2017 년 상반기 트렌드 특강을 준비 하다 보니, 인공지능을 기반으로 한 다양한 기술들과 사회적 이슈를 접할 수 있었습니다. 그 중 시장성과 미래 가치적 측면에서 가장 빈도수가 높게 노출 되는 기술이 'AI Speaker' 였습니다. 그래서 준비 했습니다^^ 한번 보도록 하시지요.

사용자조작화면(UI)이 새롭게 만들어질 때 거대한 플랫폼이 동시에 만들어 지게 됩니다. UI는 사용자가 기계와 쉽게 대화를 나눌 수 있게 도와주는 중개 역할을 하기 때문이죠. PC 시대에는 마우스와 그래픽 중심의 UI 플랫폼이 기반이 되었고, 스마트폰이 등장하면서 화면을 클릭할 수 있는 터치 기반 UI 플랫폼이 등장 하였습니다. 그리고 이제 기업은 차세대 UI 플랫폼으로 텍스트 기반 입력이 아닌 '음성'에 주목하고 있습니다.

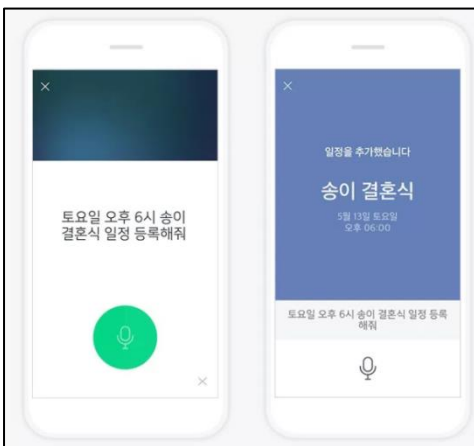


Figure 1 음성인식 사례, 네이버 클로바

음성 기반 플랫폼이 만들어지면 스마트홈으로 가는 사물인터넷(IoT) 시장을 손쉽게 선점할 수 있고 이 핵심엔 현재 스피커가 자리잡게 되었습니다. 음성 기반 플랫폼을 이용하면 손을 이용하지 않고도 편리하게 기기를 관리하거나 제어 할 수 있게 되는데, 이 때 AI 스피커는 인공지능 알고리즘을 이용해 사용자와 음성으로 의사소통을 하게 됩니다. AI 스피커를 이용하면 음성인식을 통해 집안의 기기를

목소리만으로 간편하게 제어하는 식으로 손쉽게 스마트홈 환경을 구축할 수 있게 되는 것입니다. 즉 음성인식

플랫폼하에 모든 기기를 제어 할 수 있는 환경을 제공 하게 되는 것입니다.

또한 터치 기반과 달리 음성 기반 조작은 쉽게 배우고 사용할 수 있다. 우리가 스마트폰에서 ‘시리’나 ‘S 보이스’ 등을 이용해 기기를 제어하는 게 낯설지 않듯, 음성인식 기반 UI 플랫폼은 우리 생활 가까이에 숨어 있습니다. 억지로 배우지 않아도 자연스럽게 배울 수 있는 UI 를 여러분은 얼마나 사용해 보셨나요? 그렇기 때문에 더 가치를 부여 하고 향후 2025년 까지 약 200 조의 산업 생태계를 구성 할 것이라 예측하고 있는 것입니다.

언어와 음성에 대한 연구는 인류가 태동하면서부터 수행 한 학문 중에 하나 입니다. 그 중 ‘음성’을 이용하는 것은 인간과 컴퓨터의 상호작용에 가장 이상적인 인터페이스로, 음성이야 말로 인간이 가진 가장 본질적인 커뮤니케이션 도구가 아닌가 생각이 됩니다.

### 음성인식 기반 기술의 발전 과정

언어와 음성에 대한 연구는 앞서 언급한 바와 같이 1950 년대로 거슬러 올라갈 수 있습니다. 초기에는 개별 숫자, 음절, 모음을 인식하는 시스템을 만들기 위한 노력이 이루어졌으며, 1963년 IBM 이 총 16 개의 단어를 인식할 수 있는 “Shoebbox”라는 기기를 소개 하였고, 1971년부터 시작된 미국 국방부 산하 국방첨단연구사업국(DARPA)의 음성이해연구 프로그램 (Speech Understanding Research)을 통해 비약적으로 발전하는 계기를 얻었습니다(한국콘텐츠진흥원, 2011. 11). 고립단어(Isolated word)를 인식하는데 한정되었던 음성인식기술은 1980 년대에 들어서야 연결단어(Connected word)를 처리할 수 있게 되었으며 인식할 수 있는 단어도 10,000 개 수준으로 크게 늘어나게 됩니다.(Furui, 2005). 또한 미국의 카네기멜론대학(CMU)이 불특정화자의 목소리를 인식할 수 있는 음성인식 소프트웨어 “SPHINX”를 소개하고, 은닉 마르코프 모델(Hidden Markov Model) 로 대표되는 통계적 프레임워크가 이용되기 시작했으나 컴퓨터의 처리 속도 및 저장 공간 문제로 음성인식 기술의 사용화는 요원했으나, 최근에는 이러한 문제들이 클라우드 컴퓨팅 기반하에 ICT 사용 환경이 크게 변화하면서, 실시간으로 유사 음성을 검색하고 대용량의 음성 데이터를 빠른 속도로 분석할 수 있게 되었습니다.

세대	설명
<b>1세대(1952~1968)</b>	- 숫자, 음절, 모음, 음소 인식시스템 개발 시도
<b>2세대(1968~1980)</b>	- 동적시간워핑(Dynamic Time Warping) - 고립단어(Isolated word) 인식 시스템
<b>3세대(1980~1990)</b>	- 연결단어(Connected word)인식 - 통계학적 프레임워크 적용 - 은닉 마르코프 모델, n-gram , Neural net - DARPA 프로그램: SPHINX(CMU), BYBLOS(BBN), DECIPHER(SRI)등
<b>3.5세대(1990~2000)</b>	- 오류최소화(discriminative)를 위한 MCE(Minimum Classification Error), MMI(Maximum Mutual Information) 등의 접근방식 시도 - 노이즈, 마이크로폰, 전송채널, 반향 등에 Robust한 음성인식
<b>4세대(2000~현재)</b>	- DAPRA 프로그램: EARS(Effective Affordable Reusable Speech-to-Text) 프로그램, 중요정보의 추적/추출/요약/번역 - 복합(Multimodal)음성인식: 영상 - 음성 정보 동시 인식

[표 1] 음성인식 기술의 세대별 구분

## 음성인식 비서의 핵심 기술

음성인식 기술은 최근 오픈 소스로 공개 되면서 더욱 자유롭게 활용 될 것으로 예상 되고 있고, 미시간 대학교 연구소 Clarity Lab에서는 Figure2의 아키텍처 기반의 Sirius를 공개 하였습니다.

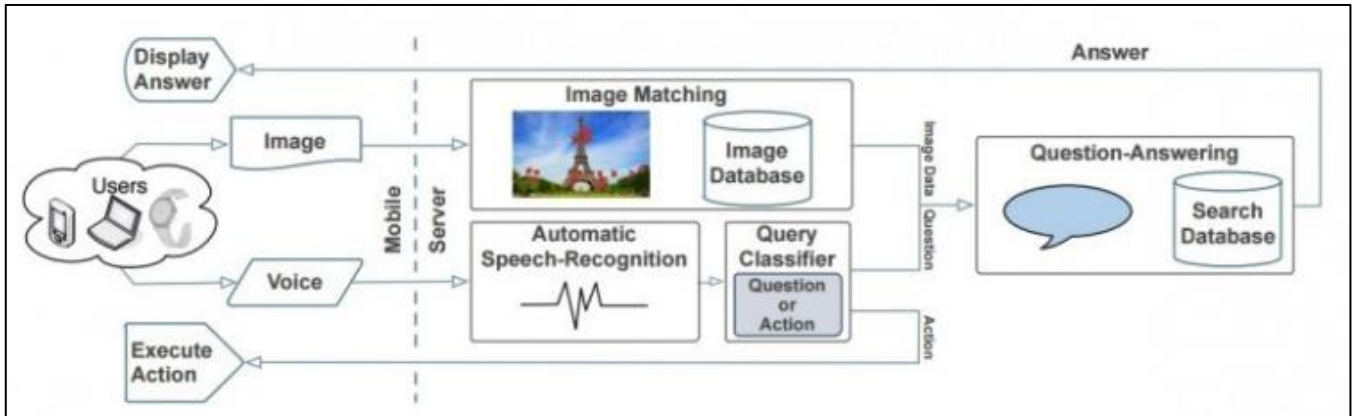


Figure 2 End to End Diagram of the 'Sirius' Pipeline

사용자의 voice 나 상황을 인지하여, 노이즈를 제거 후, 특징을 검출 하고 이에 적합한 ‘답’을 찾아 제공하는 Process Flow 를 기반으로 서비스를 제공 하게 되고, 더 나아가서는 상황을 파악하여 음성인식 비서 스스로 사용자가 원하는 것을 공유하게 진화하게 될 것입니다. 실제 예로, 미국에서 부부가 싸우는 상황을 음성비서가 인지하여 911 에 전화를 걸어 위험한 상황을 해결하기도 했습니다.

구분	설명
음성인식	- 개인 맞춤형 연속어 음성인식 기술 - 실시간 음성인식, 최고의 인식율 (HMM, CNN, RNN 등)
텍스트분석	- 고성능/고정밀/고품질 자연어 처리 기술 - 딥러닝 기반 감성인식 및 자동 분류 기술 (RNN, LSTM 등)
대화처리	- 대화 모델링, 시나리오 구축 - 기존 대화의 챗봇 학습 (빅데이터, 클라우드 등)
음성합성	- 파라메트리 음성 합성(DNN TTS) - 감성 합성(Augmented Sentiment Talk)
Question & Answer	- 질문분석, 광범위한 지식베이스 구축 - 딥러닝 외 다양한 기계학습

[표 2] 음성인식 기술의 핵심기술

음성인식 비서는 통합 대화엔진 등을 통해서 음성인식, 텍스트 분석을 수행하게 되고, 질의응답에 대한 의도파악을 통해 응답정확도에 대한 사용자 Feed-Back 을 지속적으로 반영해서 개선하는 강화 학습체계를 기반으로 동작 하게 됩니다. 이러한 상황에서 글로벌 메이저 기업들은 플랫폼화를 통해 보급 확대로 시장 선점을 도모하고 있는 실정 인데요. 애플·아마존·구글 등 IT 시장을 주도하고 있는 글로벌 기업들은 AI 비서 개발은 물론 서비스의 플랫폼을 통한 보급 확대로 기회를 엿 보고 있고, 최근엔 삼성전자·네이버 등 국내 유수의 기업들도 경쟁에 뛰어들었습니다. 국내외 기업현황은 아래 Figure3,4 에서 확인 하시면 되겠습니다.

Figure 3 해외 음성인식 서비스와 스피커 현황 요약, 각 사 홈페이지 및 언론자료







	Amazon	Google	Apple	Microsoft	Baidu	vinclu
						
SW (출시연도)	Alexa (2014)	Assistant (2016)	Siri (2011)	Cortana (2014)	Duer (2015)	-
전용기기 (출시연도)	Echo (2014)	Home (2016)	-	Invoke (2017)	Little Fish (2017)	Gatebox (2016)
지원언어 (한국어)	2개 (지원안함)	8개 (지원예정)	20여개 (지원함)	8개 (지원예정)	1개 (지원안함)	1개 (지원안함)
클라우드 사용여부	사용함	사용함	사용안함	사용함	확인불가	확인불가
강점	만개이상의 Skills 보유	구글 이용자의 빅데이터 보유	개인정보 보호 및 보안 우수	Windows 유저를 잠재적 이용자로 보유	해외기업의 진입장벽 높은 중국시장에서 유리	최초의 홀로그램 AI상용화

Figure 4 국내 주요기업의 AI 비서 서비스 현황

기업	서비스	종류	특징
SKT	누구 (NUGU)	AI 스피커	- 국내 최초 음성인식 AI 비서 스피커 - 조명, 연결된 앱과 기기, IPTV 제어
KT	기가지니 (GiGA Genie)	AI TV/스피커	- AI TV와 스피커 결합으로 홈IoT 환경 제공 - 기가지니 카메라 별도 설치를 통해 홈캠 활용 가능
삼성전자	빅스비 (Bixby)	AI 비서 UI	- 음성인식, 번역, 콘텐츠(음악, 영화 등)추천 - AI 대화형 엔진 네이버아이(i)와 다양한 AI 기 술을 접목
네이버	클로바(Clova) 베타	AI 비서 앱	- 음성인식, 번역, 콘텐츠(음악, 영화 등)추천 - AI 대화형 엔진 네이버아이(i)와 다양한 AI 기 술을 접목
카카오	카카오 미니	AI 비서 앱	- 음성 언어를 컴퓨터가 해석하여 문자로 변환해주는 시스템인 "뉴톤"과 문자를 음성 언어로 변환해주는 시스템인 "뉴톤 톡"의 기술 기반

각 사는 스피커뿐 아니라 음성인식 AI 기술력을 확장하여 AI 시장 우위를 점하는 것을 목표로 삼고 있습니다. Amazon 은 Echo 의 라인업을 다양화하고 있으며, 현 시점에서는 가장 높은 시장점유율로 유리한 고지에 있는 상태 입니다. 기존에 충성도 높은 고객층을 보유하고 있는 Apple, Microsoft 와 삼성은 AI 비서 서비스를 향후 자사

제품에 적용할 것을 계획하고 있는데, 이를 통해 고객의 편의성을 향상시키고 이용데이터 축적을 통해 자사 AI 비서 서비스의 품질을 안정화 시킬 수 있을 것으로 보이며, Google 과 국내 포털사는 다수의 경쟁력 있는 온라인서비스, 축적된 정보량, 검색기술 등을 기반으로 AI 비서 서비스에서 이동통신사의 홈 IoT 중심의 AI 전략과 차별화될 것으로 생각이 됩니다. 마지막으로 이동통신사의 경우, 음성인식 AI 서비스를 통해 홈 시스템과 가전제품을 제어하는 홈 IoT 중심으로 확장될 것이고, 온라인 서비스자회사(인터넷 쇼핑몰, 음악제공서비스 등)의 콘텐츠를 활용하여 온라인상에서도 경쟁력을 확보할 것으로 예상됩니다.

### 음성인식 비서 기술 선점을 위한 고려 사항

향후 음성인식 AI 개발 툴 공개로 음성인식 AI 관련 응용 서비스들이 활발히 개발될 것으로 기대됩니다. 이는 국내 음성인식 AI 시장을 활성화되는데 기폭제가 될 것이고, Amazon 처럼 자사 이커머스 플랫폼과 시너지를 추구하는 것 외에는 아직 명확한 비즈니스 모델이 없지만 현 시점에서는 시장 규모를 키우기 위한 이용자 확보가 당면 과제로 생각 됩니다. AI 스피커 시장의 본격화에 앞서 짚고 넘어가야 할 중요한 이슈 중 하나는 **이용자의 프라이버시 보호와 관련된 사항**입니다. AI 스피커를 통해 수집된 데이터는 **무선인터넷을 통해 클라우드 서버로 전송되기 때문에 데이터 도난이나 해킹에 대한 우려도 존재**하구요. AI 스피커를 출시한 기업들은 수집된 개인정보는 비밀이 유지될 것이라 약속하지만, 구체적으로 어떤 방식으로 이 개인정보가 보호되는지에 대해서는 알려져 있지 않고 있습니다. 따라서 각 사는 이에 대해 구체적이고 투명한 대책을 마련하고 잠재적 이용자들의 신뢰를 확보하는 것이 우선시 되어야 할 것입니다.

“끝”

## Contents connect communications!!

아이리포에 오시면 더 많은 지식을 가져가실 수 있습니다.

아이리포 온라인 : <http://www.ilifo.co.kr>

아이리포 지덤시리즈 : <http://www.jidum.com>

아이리포 IT지식창고 : <https://www.ilifo.co.kr/boards/knowledge>

아이리포 기술사/감리사 카페 : <http://cafe.naver.com/itlf>

서울시 마포구 상암동 1610번지, DDMC 3층 아이리포 교육센터

TEL: 02-303-9997 | MAIL: edu@ilifo.co.kr