

‘말문을 연’ 대화형 AI 콘텐츠의 미래

-네이버 AI 스피커 ‘클로바’를 중심으로

글 옥상훈 | 네이버 에반젤리스트 | okgosu.exp@navercorp.com

인터페이스의 변화: 모바일에서 음성으로

1980년대에 재미있게 봤던 TV프로그램 중, <전격Z작전>이라는 미드가 있었다. 거기에 나오는 ‘키트’라는 자동차는 무척 인상적이었다. 키트는 알아서 운전(자율주행)도 하고, 무엇보다도 사람 말을 알아듣고 자신도 말을 할 줄 아는 신기한 자동차였다. 그런데 지금 자동차와 IT업계에서는 그러한 영화속 기술을 현실로 만들어가고 있다. 우리는 운전하며 말로 내비게이션을 조작하는 것은 물론이요, 말하는 것만으로 자동차에 시동을 걸거나 문을 여는 활동이 가능한 시대에 살고 있다. 이메일을 예로 들어 보자. 과거에는 PC에 앉아서 메일을 확인할 수밖에 없었다. 하지만, 지금은 손에 들고 있는 스마트폰에서 바로 메일을 확인하고 답장을 할 수 있다. 인터페이스 체계가 PC에서 모바일로 큰 전환을 한번 겪었듯, 현재 곳곳에서는 인터페이스를 모바일에서 음성으로 바꾸려는 시도가 늘어나고 있다. 현재 우리는 음성 명령으로 TV를 켜고, 리모컨을 찾고, 불을 켜고 끌 수 있다. 또한 음악을 재생하고, 날씨나 스케줄을 확인할 수도 있다. 심지어 새로 분양 중인 아파트의 경우, 집에서 나가기 전 엘리베이터를 부르거나 집 안의 가스밸브를 잠그는 일까지도 음성으로 가능하다.

음성명령 인터페이스는 사람이 일상에서 나누는 대화처럼 사용하면 되기에, 일단 사용법을 한 번만 배우고 나면 기존의 PC 웹이나 모바일 앱보다도 훨씬 조작하기 쉽고 직관적이라는 특징이 있다. 예를 들어, PC에서 아이유의 ‘가을아침’이라는 노래를 재생하려면 웹브라우저에서 아이유 노래를 검색한

후 재생버튼을 찾아 눌러서 재생을 해야 한다. 모바일 앱 환경이라면 뮤직 앱을 실행한 후에 다시 아이유 노래를 검색하고 재생버튼을 눌러서 재생한다. 하지만 음성명령 인터페이스를 사용한다면, 기기에 ‘아이유의 가을아침 틀어줘’라고 한 번 말하면 끝난다. 또한 PC나 모바일의 경우 음악을 찾고 재생하는 동안 다른 작업을 할 수는 없지만, 음성의 경우에는 음악을 틀어 달라고 한 후에 불을 꺼 달라고 말한다거나, 알람을 설정한다거나 하는 멀티태스킹 작업이 가능하다. 이처럼 PC웹에서 모바일의 UX가 크게 변화했듯 음성의 UX 또한 매우 다르며, 접근도 달리 해야 하는 상황이다. 하지만 기술 보급이라는 면에서 생각하면 전망은 무척 긍정적이다. 이를테면 어르신들은 컴퓨터에서 인터넷 브라우저를 열고 닫는 일조차도 어려워하지만, 일상생활에서 쓰는 말로 AI 기기를 다루는 법을 익힌다면 이를 훨씬 쉽게 사용할 수 있을 것이다.

AI 스피커 대중화 및 콘텐츠 현황

10여 년 전에는 낯선 기기였던 스마트폰이 이제는 전 국민이 1대 이상 들고 있을 만큼 보급된 것처럼, AI 스피커의 도입현황 또한 스마트폰이 처음 도입되던 초기 추세와 유사한 증가세를 보이고 있다. 각종 리포트나 지표를 보면 AI 스피커가 스마트폰처럼 대중화될 미래는 확실하게 점쳐진다. 불과 2년 전에는 미국 가정의 1% 정도만 AI 스피커를 사용했지만, 2018년 말이면 무려 50% 정도로 보급률이 올라갈 것으로 예상된다.^[주1] 2018년 10월 구글이 구글홈 스피커를 국내에서 판매하기 시작한 이래 SKT, 네이버, 카카오에서 더 다양한 디바이스

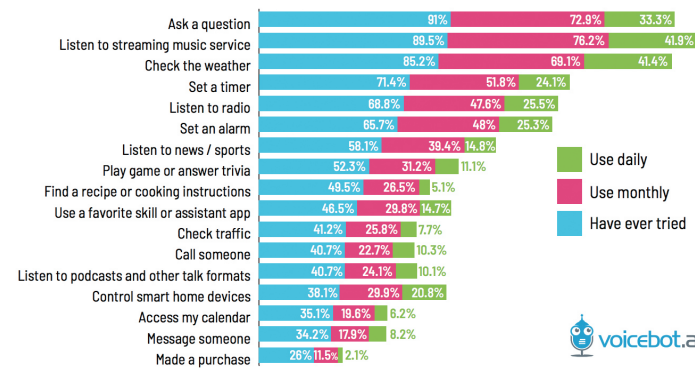


<전격Z작전>은 이제 미래의 일이 아니라 현재 또는 근미래의 일이 되고 있다.

를 출시한다고 간주하면 앞으로의 국내 시장도 경쟁이 매우 치열할 것으로 전망된다. 국내 시장조사 결과에 따르면, 국내 가정에서의 AI 스피커 보급률은 약 15%(300만 대) 정도로 예상했다.^[주2] 이런 추세로 나아가면 내년쯤 AI 스피커는 어느새 보편화된 디바이스로 자리매김할 것이라고 예상된다.

플랫폼이 바뀌더라도 사용자의 니즈는 변하지 않는다는 사실은 항상 들어맞았다. PC에서 메신저가 모바일로 넘어와서 하나의 플랫폼이 되었다는 것은 인간의 커뮤니케이션 욕구가 변치 않는다는 점을 방증한다. AI 스피커의 콘텐츠 현황을 들여다보면, 최근 오디오북이 연평균 20% 성장세를 보이고 있다. 특히 오디오북 검색량의 경우 2년 전에 비해 2배 증가한 8만 건을 돌파했다.^[주3] 이는 당연히 소리 콘텐츠에 대한 소비욕이 스피커에서도 반영된 것이다. voicebot.ai 리포트에 따르면, AI 스피커에서 많이 쓰이고 있는 콘텐츠는 1위가 뮤직, 2위가 날씨, 3위가 단순 질문의 순으로 나타났다. 상대적으로 구매를 하는 비중은 2.1%로 매우 낮은 것으로 파악된다.^[주4] 이러한 양상은 국내 스피커도 크게 다르지 않았다. 하지만 미세한 트렌드의 변화로 최근 들어 기본적인 음악 재생 외에도 다소 부족한 감성과 재미에 대한 욕구가 늘어나고 있다. 비즈니스인사이드에 따르면 Alexa에서 ‘나랑 결혼하자’ 등 감성대화를 건넨 사용자가 2017년 기준

Smart Speaker Use Case Frequency January 2018



[주1] voicebot.ai. <Smart Speaker Consumer Adoption Report>, 2018.3.

[주2] 고현실 기자, <한국 AI 스피커 시장 급성장... "올해 '세계 톱5' 전망">, 연합뉴스, 2018.07.12. 기사.

[주3] 조운주 기자, <콘텐츠 넘친 오디오북 다시 뜬다, 파이낸셜뉴스, 2018.10.21. 기사.

[주4] voicebot.ai. <Smart Speaker Use Case Frequency Report>, 2018.1.

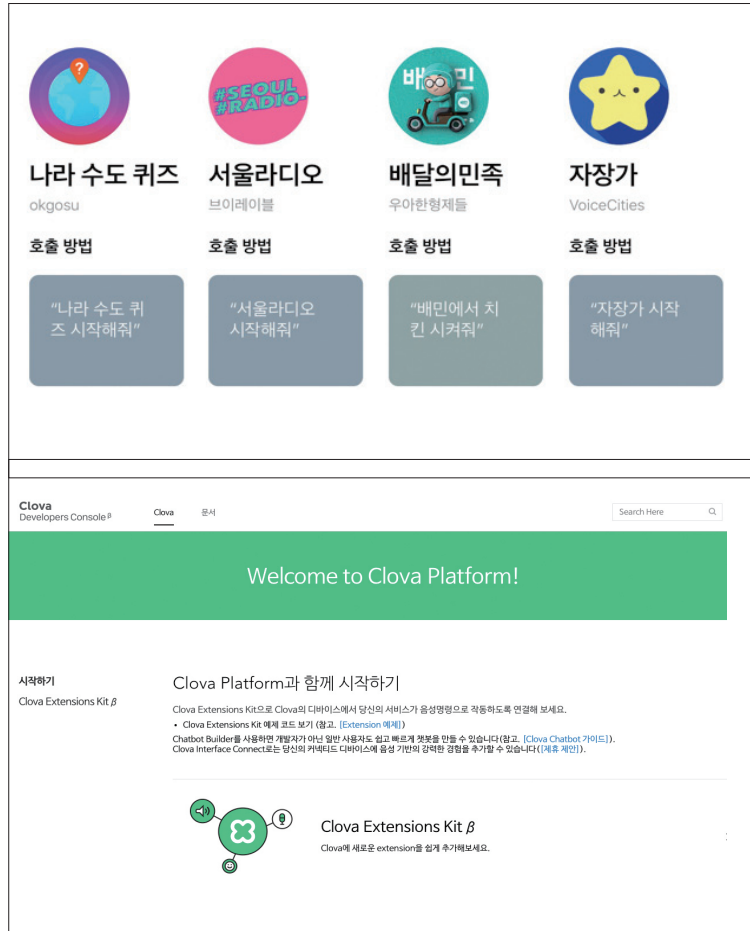
100만 명 이상으로 집계되었으며, S사 스피커 사용자 패턴 분석 결과, ‘좋아해’, ‘심심해’ 같은 감성대화 비중이 4.1%로 늘어난 반면 음악재생은 60%에서 42%로 감소하였다고 한다. 또한 장마처럼 비가 오는 날씨가 감성 대화가 평소대비 8% 증가, 폭염에는 ‘짜증나’ 같은 부정적 대화가 22% 증가했다는 발표가 있다. ^[주5]

AI 스피커 ‘클로바’의 3rd 파티 스킬 소개

클로바에서 음악 재생이나 날씨, 단순 질문을 제외한 분야의 3rd 파티 스킬들을 살펴보면, 특정 부류의 스킬들이 꾸준히 사용 증가세를 보이고 있다. 그 중에서 카테고리 별로 몇 가지만 선별하여 소개해보고자 한다.

우선, AI 스피커와 그야말로 상호작용이라 할 수 있을 간단한 놀이를 즐기는 유저들이 증가하는 추세다. 퀴즈라든가, 시간을 죽이기 위한 여러 활동을 위한 명령어가 클로바에서 인기를 끌고 있다. 퀴즈를 예로 들면, 클로바에 “수수께끼 시작해줘” 또는 “나라 수도 퀴즈 시작해줘”라고 이야기하면 수수께끼나 각국의 수도 맞추기 퀴즈를 즐길 수 있다. 그 외에도 과일 초성 퀴즈, 구구단 게임, 일당백 퀴즈, 서바이벌 퀴즈, 업다운 등 다양한 콘텐츠가 존재한다. 심심풀이 스킬의 경우, 서울시의 다양한 사연을 읽어 주는 ‘서울라디오’나 연애 타로 점을 봐주는 ‘라마마의 연애운’, 마음을 치유해주는 ‘왈이의 마음식당’ 등이 준비되어 있다. 클로바에서 “서울라디오 시작해줘”, “라마마 시작해줘”, “마음식당 시작해줘”라고 하면 위의 스킬들이 실행되고 콘텐츠를 즐길 수 있다.

또, 생활 편의를 위한 스킬들도 준비되어 있다. 우리가 일상에서 자주 활용하는 ‘배달의 민족’과 제휴한 음성 배달 주문 서비스, 배송 확인을 할 수 있는 ‘배송지킴이’가 있다. 예를 들어 “배민에서 치킨 시켜줘”라고 말할 경우, 이전에 주문한 이력이 있다면 자동 주문이 가능하다. 그리고 “배송지킴이 시작해줘”라고 하면, 본인의 핸드폰번호로 등록된 인터넷 구매 물품의 배송 현황을 확인할 수 있다. 그 외에도 ‘진에어’의 비행정보 조회, ‘현대백화점 쇼핑 알리미’의 인기 상품 추천 및 가격 검색,



Clova 주요 3rd 파티 스킬들과 Clova 개발자 콘솔

[주5] 맹하경 기자, <“결혼해줘” 청혼까지 받는 ... 감성 동반자로, 한국일보, ... 기사.
* 인터랙션 모델에 대해 여기서 다 설명하려면 지면이 모자라므로 참고할 만한 Clova 개발 가이드 및 자료 링크로 대신한다.
Clova Dev Days 발표 자료
: <https://www.slideshare.net/ClovaPlatform/>
Clova Dev Days 동영상: <https://www.youtube.com/channel/UCUuTYvmFcwOg06KnH5xx2sQ>
개발가이드: https://developers.naver.com/console/clova/guide/CEK/CEK_Overview.md
배포 가이드: https://developers.naver.com/console/clova/guide/DevConsole/Guides/CEK/Deploy_Extension.md

‘미래에셋대우’나 ‘신한금융투자 알파’의 국내외 증시 현황 안내 및 종목 추천 등이 있다.

한편, 요즘 젊은 세대에게 인기를 끌고 있는 ASMR 사운드도 다양하게 구비되어 있다. 클로바에 수록된 ASMR로는 자장가, 빗소리, 카페소리, 파도소리, 천둥소리 등이 있으며, “자장가 시작해줘”, “빗소리 시작해줘” 라고 말하면 해당 ASMR을 재생할 수 있다.

AI 콘텐츠(스킬)은 어떻게 만드나?

AI 콘텐츠를 만들기 위해서는 각 플랫폼에서 제공하는 툴킷에 대화시나리오를 가공하여 입력하고, 모바일 앱처럼 심사요청을 하여 승인을 받으면 그때부터 서비스할 수 있다. 물론 AI 스피커가 있으면 좋겠지만, 없더라도 플랫폼사가 제공하는 앱(예: Clova 앱)을 설치하면 음성으로 명령하고 테스트 하는 것이 가능하다. 이 부분에서는 챗봇을 제작할 수 있는 Clova 개발자 콘솔(<https://developers.naver.com/console/clova>)을 이용하여 스킬을 만드는 방법을 간략히 소개하고자 한다.

먼저 클로바 스피커와 대화 시나리오를 준비하고, 클로바 앱을 설치하여 준비한다. 그런 후 자신이 서비스할 유형을 결정한다. 만약 FAQ처럼 단순한 질문과 답변만으로 구성된 대화라면 ‘Chatbot extension’을, 사용자의 대화에 대한 답변이 서버의 가공을 거쳐야 한다면 ‘Custom extension’을 선택한다. 그 후, 개발자 콘솔에 기본 정보 및 대화 시나리오를 등록한다. 개발자 콘솔에는 스킬의 기본 정보(스킬이름, 호출 명, 스킬설명)를 입력해 넣는다. 유저가 처음 스킬을 사용할 때 스피커에 말하는 일종의 주문인 셈이다. 그리고 대화 시나리오는 ‘인터랙션 모델’ 화면에 입력한다. 인터랙션 모델은 사용자로부터 입력이 될 만한 문장들과 용어들에 대한 집합이다. 머신러닝 기술이 적용되어, 비슷한 유형의 문장은 대표로 몇 가지만 등록해도 알아듣게 해준다. 마지막으로, 서비스를 배포한다. 개발 중에는 미리 테스트로 등록된 사람만 스킬을 써볼 수 있으며, 클로바 플랫폼이 적용된 스피커 전체 사용자들이 쓸 수 있도록 만들기 위해서는 앱스토

어처럼 본사에 심사 요청을 하면 된다. 스킬 심사는 매주 1회 진행되며, 여기서 승인이 되면 클로바에서 누구나 등록된 호출 명으로 스킬을 사용할 수 있게 된다.

AI 콘텐츠의 향후 전망

다루기 편리하고, 제작도 용이한 AI 콘텐츠의 향후 전망은 앞서 언급한 바처럼 무척이나 밝다. 우선 AI 콘텐츠가 사용자 요청에 단순히 응답 또는 재생만 되는 게 아니라, 사용자와 대화하면서 콘텐츠가 달라지거나, 화면을 보면서 대화하는 등 인터랙티브한 콘텐츠들이 등장할 것이다. 특히 디바이스가 장착된 AI 스피커가 출시되면 요리 레시피를 확인하거나, 주문하려는 상품을 보거나, 그림동화의 그림을 보는 등 화면을 보며 인터랙티브하게 사용할 수 있는 콘텐츠들이 보다 다양하게 만들어질 것이다.

또한, 현재 사용자의 목소리를 학습해 그 목소리를 들으면 누군지 파악하는 화자인증 기술들이 이미 적용되었거나 적용을 앞두고 있다. AI 스피커가 나의 목소리를 알아듣게 된다면 스피커가 사람을 구분하여 알아듣고, 이에 따라 개인화 및 커머스 영역으로 서비스가 확산될 것이다. 즉, 아빠가 내 스케줄을 조회했을 때의 결과와 엄마가 조회했을 때 결과가 다르게 나타날 수 있으며, 아이가 음성 명령으로 물건을 잘못 주문·결제하는 상황을 사전에 막을 수도 있게 될 것이다.

아울러 일원화된 AI 스피커의 목소리 또한 다양해질 전망이다. 현재는 한두 가지의 남자, 여자 목소리만 스피커에서 지원하고 있지만, 음성합성기술이 보다 고도화되고 확산되면 유명인, 캐릭터 목소리를 낼 수 있게 된다. 이에 따라 그 캐릭터에 맞는 재미있는 스킬들이 나올 것으로 기대된다. 마지막으로, 디바이스 플랫폼 종류 확산이 예상된다. 음성으로 명령할 수 있는 디바이스들이 스피커에 국한되지 않고, 다양한 플랫폼(이러폰, 생활가전제품군, 자동차 등)으로 확산될 것이다. 청소기에게 청소를 명령하거나, 냉장고에게 요리 레시피를 물어보는 등의 활동도 가능해질 것이다. ①