

➔ Issue&Trend

시리와 빅스비, 이어서 나올 스마트폰 인공지능의 미래는?

IT 평론가 안병도 (catchrod@hanmail.net)

- I. 애플 시리, 음성인식과 인공지능을 결합하다
- II. 삼성 빅스비, 터치와 카메라까지 인식한다
- III. 게이트박스, 인격을 갖춘 능동형 인공지능으로

최근 IT업계에서 가장 화두가 되고 있는 기술은 바로 인공지능(AI)이다. 전통적인 PC 시대가 가고 스마트폰을 앞세운 모바일 시대가 온 것이 얼마 전이다. 이제는 빅데이터와 클라우드를 거쳐 머신러닝을 갖춘 인공지능 기술이 지능형 IT기기 시대를 열고 있는 것이다. 특히 스마트폰에 들어가는 인공지능은 음성인식 기술과 결합해서 사용자경험을 대폭 향상시키고 있다. 이런 인공지능 기술의 대표격인 애플의 시리와 최근 나온 삼성 빅스비를 통해 스마트폰 인공지능의 현재를 분석하고 미래를 예측해보도록 하자.

I. 시리, 음성인식과 인공지능을 결합하다

예전 피쳐폰 시대에도 음성인식 기술은 적용되는 경우가 있었다. 삼성 애니콜 플립업은 전화번호 대신에 음성으로 말하면 인식해서 전화를 걸어주는 기능이 있었다. 이 점을 고려해 광고영상에는 배우 안성기가 다급한 목소리로 “본부! 본부!”하고 외쳐서 전화를 걸어주는 장면이 들어갔다. 그렇지만 이런 음성인식기술은 기술적 한계로 인식률이 떨어졌을 뿐 아니라 인공지능 기술과 결합되어 적용되지 않았다.



애플이 아이폰에 적용한 시리는 최초로 음성인식과 인공지능을 묶어 이용자에게 여러 가지 서비스를 제공하는 기능이다. 흔히 이것을 '음성인식 개인 비서'라고 표현하기도 한다. 파악한 음성명령을 마치 인간 비서처럼 음지능적으로 판단해서 수행한다는 의미이다.

시리는 SRI 인터내셔널 벤처그룹이 투자한 Siri Inc 라는 회사에서 만들어진 프로그램을 애플이 인수하여 사용한 것이다. 원래는 미국 국방연구소가 군사목적으로 40년간 추진한 인공지능개발 프로젝트의 하나이며 2003년부터 약 8년간 300여명의 연구자와 수백만 달러의 투자로 만들어졌다.

시리는 2011년 10월 4일에 아이폰 4S와 같이 공개되었다. 발표 당시에는 영어, 독일어, 프랑스어 지원으로 시작했지만 이후 일본어를 비롯해, 이탈리아어, 독일어, 한국어, 중국어 등이 추가되었다. 단순히 표준적인 발음을 지원하는 것 외에도 캐나다식 영어라든가 스위스식 이탈리아어 같은 변형 발음도 지원하기 시작했다.



시리가 구현한 인공지능의 작동방식은 간단하다. 사용자가 말을 하면 휴대폰이 마이크를 통해 내용을 녹음해 애플의 서버로 전송한다. 이것을 뉘앙스의 기술을 이용해서 텍스트로 변환한다. 이 텍스트 명령을 인공지능으로 분석하여 어떤 작동을 할 것인지 결정한다. 최종적으로 사용자에게 알맞은 대답을 들려주거나 관련 앱을 실행한다.

애플이 시리를 처음 발표했을 때 사용자들은 인공지능의 존재를 크게 평가하지 않았다. 이전에 나왔던 음성인식 기능을 상업적 목적으로 포장한 것으로 해석하는 경우가 많았다. 또한 이 때는 립러닝에 기초한 인공지능의 작동원리상 사용자가 입력한 데이터가 적어서 다양성과 정확성도 떨어졌다. 립러닝은 사용자가 많이 사용할 수록 인공지능이 스스로 학습해서 더욱 나은 결과를 만드는 알고리즘이다.



스마트폰 인공지능이기에 시리는 아이폰의 홈버튼을 길게 누르고 있으면 실행된다. 화면 아래쪽에 소리에 따라서 물결처럼 요동치는 그래픽이 표현되며 음성인식이 작동되고 있다고 표시한다. 음성인식이 끝나거나 일정시간 음성입력이 없으면 사라진다. 또한 음성으로 "시리야(Hey Siri)"라고 불러서 작동시킬 수도 있다. 손을 쓰지 않고도 작동시

킬 수 있는 인터페이스로서 큰 의미가 있다.



시리는 인공지능이기 때문에 사용자가 다소 모호하게 말하는 음성을 정확한 명령으로 인식한다. 또한 이것을 스마트폰에서 가능한 기능으로 구현한다. 단순한 관련 앱 실행이나 시스템 설정까지도 가능하다. 전화걸기, 음악재생, 문자전송, 일정관리 같은 기본 기능 외에도 영화정보, 식당의 리뷰와 가격대 검색도 지원한다. 아이튠즈를 이용한 콘텐츠 구입과 홈킷을 이용한 가정내 사물인터넷 제어도 가능하다.

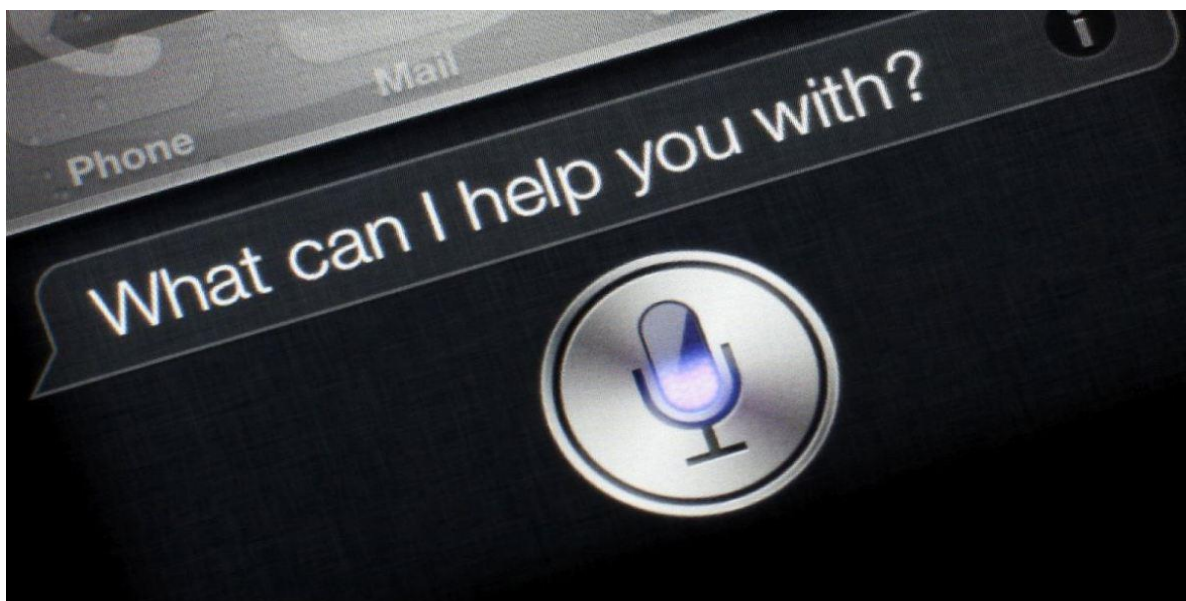
특정한 목적을 가진 작업 이외에도 대화를 통해 농담이나 노래부르기 같은 일상적인 대화도 어느 정도까지 가능하다. 다만 애플 측에서 사용자들이 지속적이고 감정적인 소통을 하지 못하도록 감정과 반응을 제한시켜 놓았다고 한다. 이 부분에서는 애플이 시리와 같은 인공지능의 급속한 발달이 오히려 사용자에게 거부감을 주거나 과몰입을 유발시키는 것을 경계한다고 해석할 수 있다.

애플에서 보자면 시리는 애플 제품군에 탑재되어 매력을 높이는 플랫폼으로서 적절한 기능을 통해 생활에 도움을 주는 정도까지만 간다는 전략이다. 아이폰 외에도 아이패드, 맥에도 시리가 적용되었으며 이들은 모두 플랫폼에 포함된 무료서비스이다. 프로페셔널한 기능을 가진 별도버전 시리를 판매하지 않으며 기기와 떼어내서 서비스하지 않는 것으로 이런 전략성을 확인할 수 있다.



시리는 사용자의 많은 호응을 받으며 이후 경쟁사와 인공지능 서비스 경쟁을 촉발시킨 계기가 되었다. 우선 시리는 이전 세대 음성 인식과 달리 정해진 키워드가 아니면 인식하지 못하는 한계를 넘었다. 자연어 처리를 통해 내용의 핵심 키워드가 없어도 문맥을 파악해 핵심 의미를 이해한다. 예를 들어 '달력'이나 '일정' 같은 단어가 없이 "내가 오늘 해야되는 건 뭐지?" 라고 묻더라도 달력에 잡힌 오늘의 일정을 알려 준다.

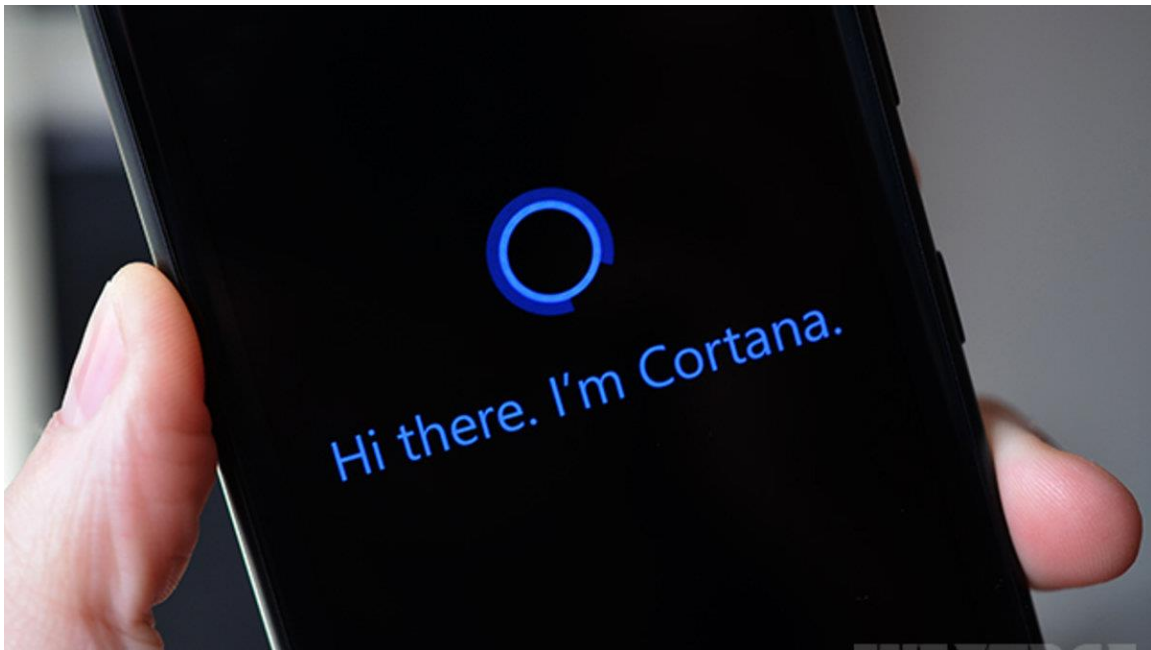
또한 한 문장의 문맥을 넘어 전체 대화 흐름이 가진 문맥도 이해한다. 사용자가 일정을 먼저 확인하고 관련된 인물을 찾아내서 특정한 이름이 아니라 단지 '그 사람'에게 전화해달라고 해도 정확히 전화를 걸어 준다. 다만 이렇게 복잡한 맥락인식은 오류가 날 가능성도 조금씩 늘어난다.



우수한 서비스이지만 문제점이 없는 건 아니다. 우선 주된 입력수단인 음성 인식률이 충분히 높지 않다. 사람과 달리 시리는 말하는 속도와 성량, 발음에 따라서 인식률이 크게 차이난다. 딥러닝으로 인해 개선되긴 했지만 사람들이 기대하듯이 사람에게 말하는 것처럼 편안하게 해도 전부 알아듣는 수준은 아니다. 음성명령을 쓰겠다는 의식하에 또박또박 명확하게 발음하지 않으면 자잘한 부분에서 잘못 인식하는 경우가 많다.

주변 소음이 크거나 음악이 들릴 때 이와 구분해서 정확히 명령만 가려내는 부분도 부족하다. 사람은 상대가 이야기하는 음성과 주변 배경음을 명확히 구분하기 위해 양쪽 귀를 이용하고 맥락까지 뇌로 판단하는 데 비해 아직 스마트폰의 마이크 성능이나 처리능력이 그에 미치지 못한다. 아이폰에서도 자체적으로 주위 소음을 구분하는 기술을 적용했지만 아직 사람정도 수준이 아니다. 또한 음성입력 방식 자체의 문제점으로 소리를 크게 내지 못하는 도서관이나 지하철 같은 환경에서 사용이 어렵다는 점도 단점이 된다.

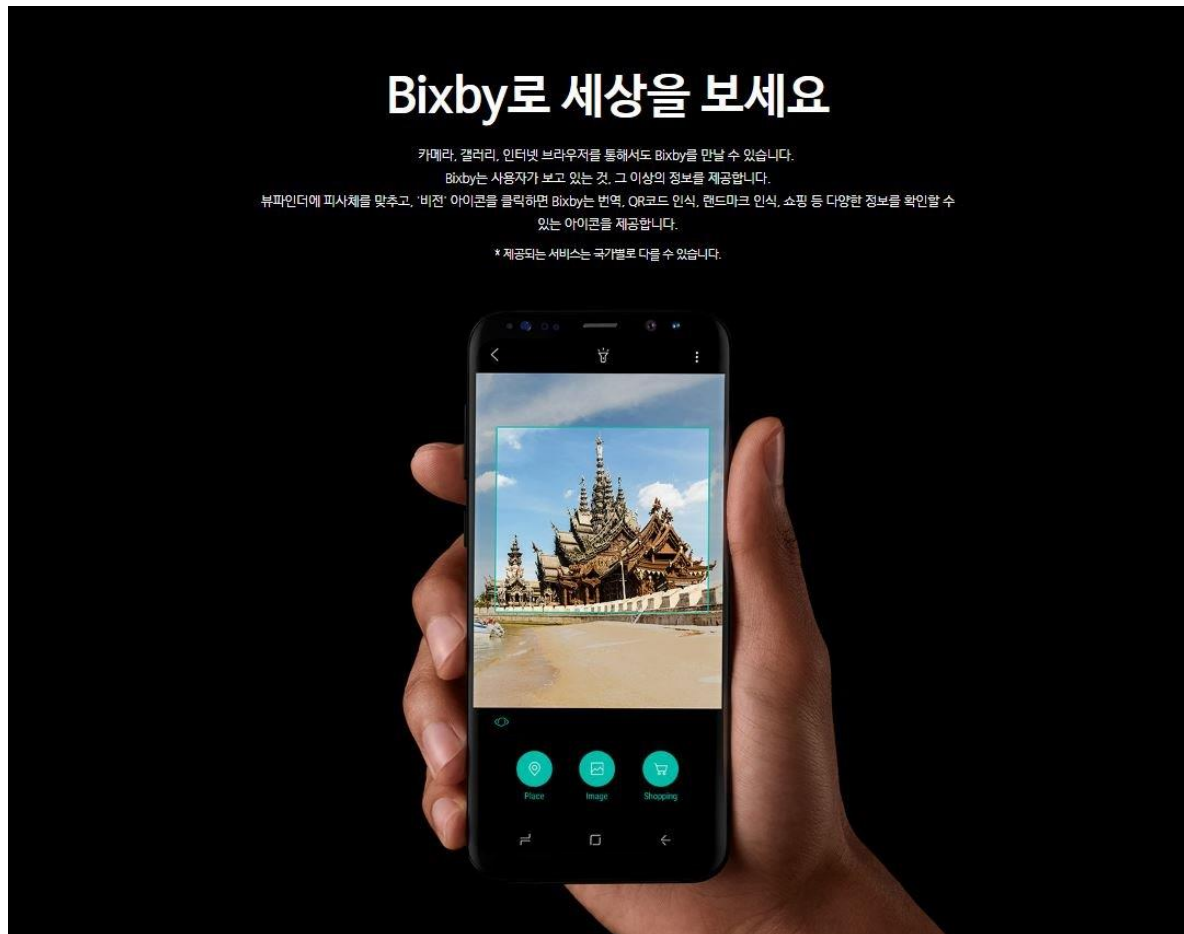
II. 삼성 빅스비, 터치와 카메라까지 인식한다



시리와 비슷한 역할을 하는 스마트폰 인공지능 서비스로 삼성전자의 S 보이스, LG전자의 Q보이스, 마이크로소프트(MS)의 코타나등이 나왔다. 구글도 구글 어시스턴트를 개발해 안드로이드 기기에 서비스하기 시작했다. 시리 이후로 나온 인공지능 대부분은 나름대로의 장점을 강조하지만 고정적으로 쓰이는 제품의 보급률이 낮다는 단점이 있다. 스

마트폰으로 범위를 좁힌다면 안드로이드 계열 휴대폰 가운데 삼성 갤럭시 시리즈를 제외하면 단일 회사 점유율에서 아이폰과 비견될 만한 제품이 거의 없다.

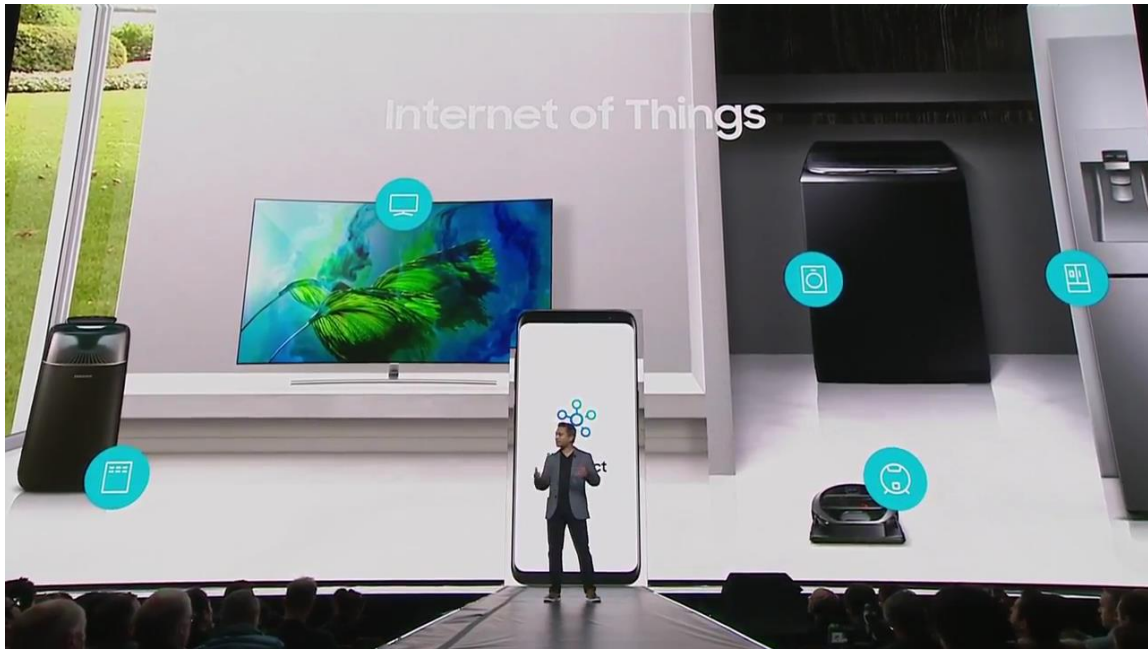
아마존이나 SKT, KT 등은 아예 디스플레이를 배제하고 음성으로 작동하는 셋탑박스 형태를 출시했다. 인공지능을 하드웨어와 함께 묶어서 보급한다는 이런 전략은 스마트폰 같은 독자적인 범용 제품을 생산하지 못하는 업체가 인공지능 시장에 뛰어들기 위한 전략적 선택이다.



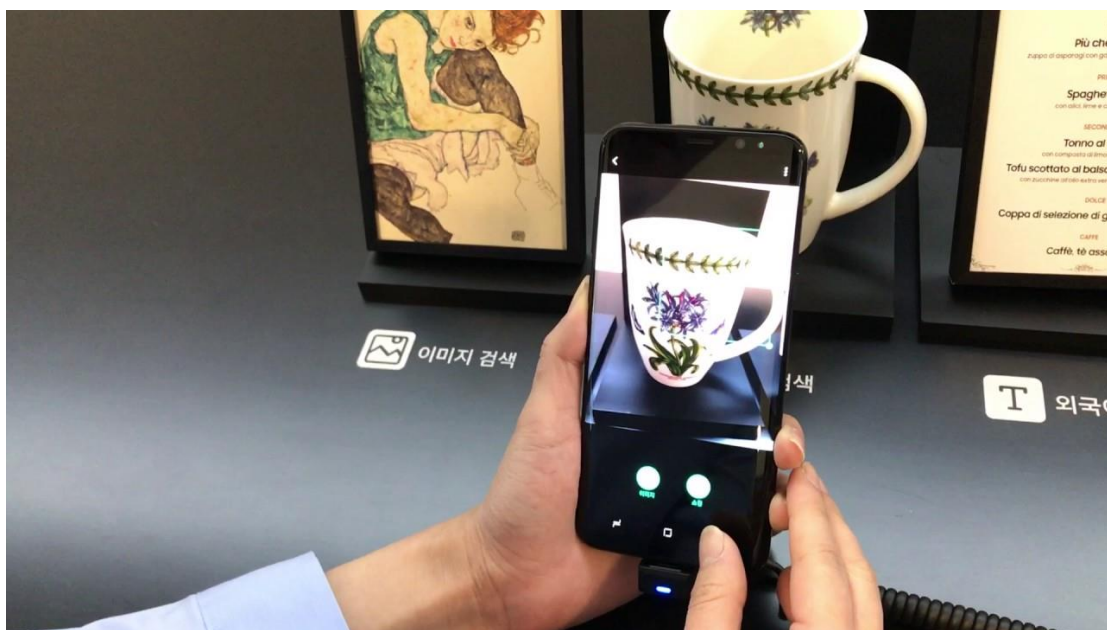
여기서 주목되는 것은 현재 스마트폰 시장에서 애플의 큰 경쟁자인 삼성전자이다. 삼성전자는 새로운 인공지능 서비스 '빅스비(Bixby)'를 개발해서 본격적으로 경쟁에 뛰어들었다. 삼성전자는 2017년 3월 30일에 딥러닝 기반의 음성인식 기반 개인 비서앱으로 빅스비를 선보였다. 빅스비는 우선 갤럭시 S8 시리즈에 기본 탑재되었는데 왼쪽 측면 볼륨 버튼 아래 빅스비 전용 버튼을 만들어줄 정도로 중시하고 있다.

빅스비가 다른 경쟁 서비스와 다른 점은 터치, 음성, 카메라 촬영 같은 여러 입력 수단을 함께 이용할 수 있다는 점이다. 이것은 단순한 '음성' 비서 시스템에 머무르는 경쟁 서비스보다 앞으로 나아간 것이다. 사실 인공지능 서비스에서 전단계의 입력을 위한 수

단은 음성이 아니어도 상관없다. 텍스트면 가장 정확한 편이고 음성 외에 그래픽이나 터치 등 스마트폰에서 이용할 수 있는 모든 것을 이용할 가능성이 열려있다.



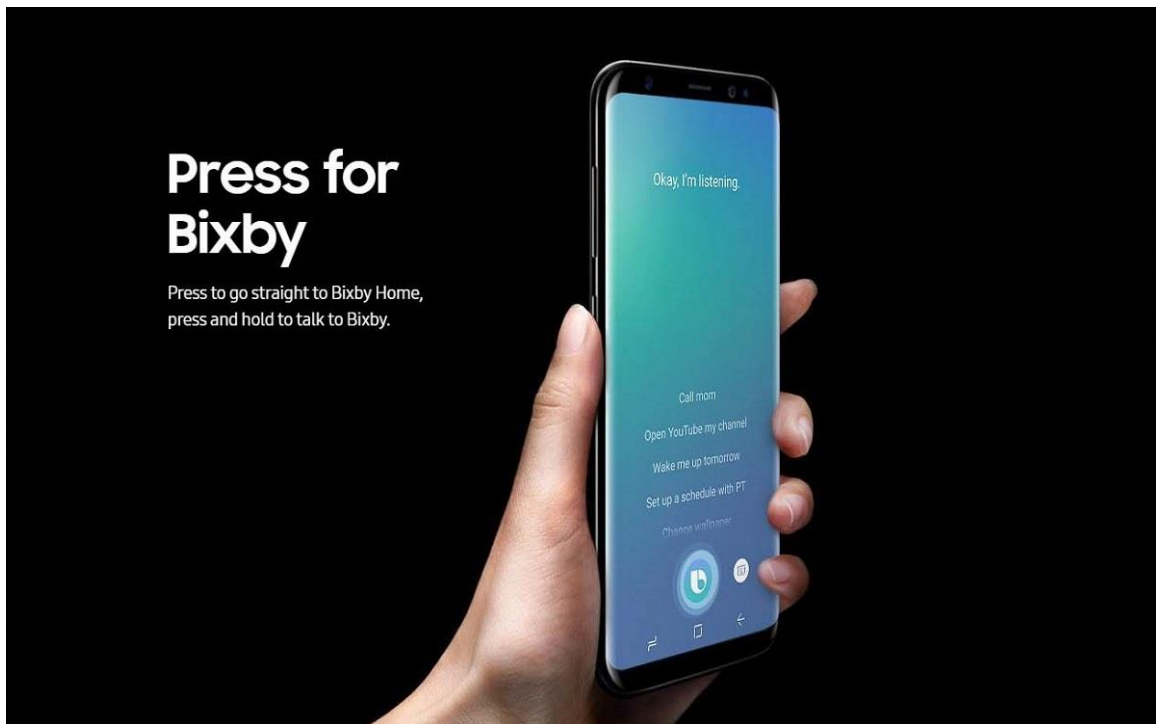
실제로 빅스비에서 보여준 입력 다양성은 충분히 주목할 만하다. 예컨대 빅스비로는 보이스(Voice), 비전(Vision), 리마인더(Reminder), 홈(Home) 기능을 이용할 수 있다. 여기서 "빅스비, 어제 찍은 사진을 보여줘"라고 말하면 곧바로 사진들을 펼쳐서 보여준다. 출력된 화면에서 사용자가 원하는 사진을 손가락으로 고른 뒤 "빅스비, 서울 폴더 만들어서 넣어줘"라고 명령하면 '서울' 폴더를 만들어 선택한 사진을 모아서 넣어준다. 음성만이 아니라 스마트폰 화면에서 터치로 선택된 사진을 객체로 바로 인식한다는 것이다.



카메라를 통한 화상으로 사물을 인식하는 기능도 이런 맥락에서 볼 수 있다. 갤럭시

S8에서 카메라를 켜고 아래 쪽 빅스비 비전 아이콘을 누른 뒤 음료수인 콜라(COLA)병을 찍어보자. 텍스트, 이미지, 쇼핑이라는 버튼 세 개가 등장한다. 이 가운데 텍스트를 선택하면 사진 속에서 'COLA'라는 글자만 빼내온다. 이미지를 누르면 방금 찍은 콜라와 비슷한 사진을 인터넷에서 검색해서 보여준다. 쇼핑을 터치하면 방금 찍은 콜라를 살 수 있는 쇼핑몰로 연결한다.

사용자의 대화에 대한 대답도 상당히 우수하다. 간단한 인사에 대한 응대는 물론이고 날씨와 미세먼지 농도를 말해주며 사용자가 감정을 토로하는 말에 대한 반응도 감성적이다. 다만 정확하고 상냥한 말투에 비해 아직은 목소리가 음절마다 끊어 말하는 기계적인 느낌이 강한 것이 아쉽다.



빅스비는 다른 경쟁 서비스보다 복문 수행능력이 우수하다고 평가받는다. "A하고, B해줘, 그리고 C한 뒤에 D해줘."라는 복문 형식의 명령을 하면 기존 AI 비서들은 1개나 2개 정도의 명령어만 이해한다. 반면 빅스비는 "갤러리에서 여행 이야기의 최근 사진 두장을 찾아서 카톡으로 김철수에게 보내줘"라는 식으로 복잡한 명령을 한 번에 알아듣고 그대로 수행한다. 오히려 사람이라면 지시가 너무 길어 중간에 빼먹을 뻔한 복문 명령을 전부 알아듣고 실행한다.

물론 빅스비도 결점은 있다. 우선 이용 데이터가 쌓여야 점점 우수해지는 딥러닝 기술 특성과 발표 초기라는 특성 때문에 아직은 종종 착오가 일어난다. 영상을 이용한 이미

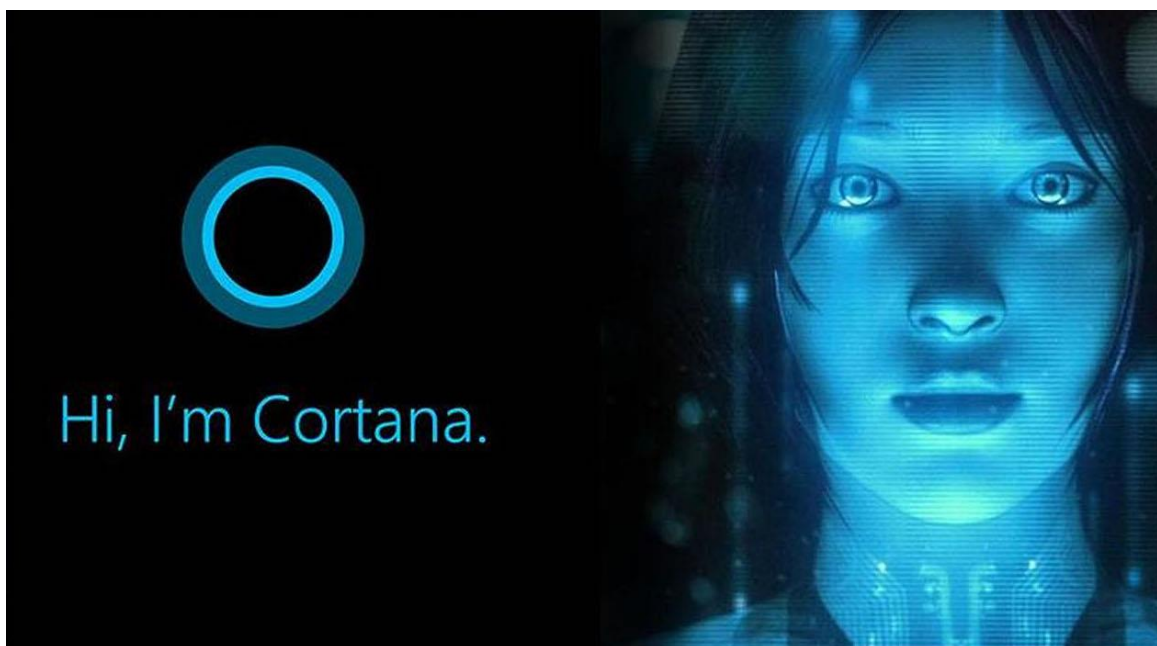
지 검색에서 제대로 정확한 제품명을 찾지 못하거나 쇼핑몰에서 원하는 물건이 아닌 유사한 물건만 찾아오기도 한다.



빅스비는 현재 갤럭시 S8과 갤럭시 S8+에서 서비스되며 이후 출시 예정인 갤럭시 노트8에도 적용된다. 또한 삼성전자 측은 구형인 S6/S7에도 빅스비 업데이트를 검토하겠다는 입장이다.

III. 게이트박스, 인격을 갖춘 능동형 인공지능으로

그렇다면 이런 스마트폰 인공지능 서비스는 앞으로 어떤 방향으로 발전할까? 단지 완성도를 높이고 기능을 추가하는 것 외에 근본적인 혁신은 가능할까? 이런 의문이 제기될 수 있다. 이것에 대한 해답을 생각해보자. 우선 인공지능이 가진 근본적인 성향을 살펴보자.



인공지능은 크게 분류한다면 '강인공지능(Strong AI)'과 '약인공지능(Weak AI)'으로 나뉜

다. 약인공지능은 주로 기능적인 일을 수동적으로 수행한다. 예를 들어 운전, 검색, 사진 판독, 사람 얼굴 인식 같은 특정 분야의 일만 할 수 있다. 스스로 무엇을 할지 판단하지 않고 인간의 명령에 따라서 주어진 틀 안에서만 일하기 때문에 편리하면서도 위험이 적다. IBM 왓슨, 구글 알파고 등 지금까지 나온 인공지능은 모두 약인공지능이다.

강인공지능은 특정한 조건에서만 적용할 수 있고 주로 명령을 받아서 작동하는 약인공지능과 다르다. 모든 상황에 일반적으로 두루 적용할 수 있고 명령 없이도 동작한다. 머신러닝 등을 통해 처음부터 의도되지 않은 운전이나 퀴즈풀기를 다른 사람이 하는 것을 보고 배워서 수행할 수 있다. 인간이 편의점 직원 일을 하다가 공장에서 기계를 돌리기도 하고, 경찰이 되어 보안업무를 수행할 수 있는 것과 마찬가지로 모든 일을 배워서 할 수 있는 것이다.

또한 자의식을 가지기도 한다. 사람처럼 명령받지 않은 일도 스스로 필요하다고 생각하면 수행한다. 또한 스스로의 판단에 의해 외부 명령을 거부할 수도 있다. 자기의 존재를 인지하고 지키려는 의지를 포함해서 같은 명령에 대해서도 조건이나 대상에 따라서 반응을 달리할 수도 있다. 일반적으로 헐리우드의 SF영화 등에서 등장하는 로봇이 바로 강인공지능을 가졌다. 따라서 정말로 사람을 대체할 수 있을 정도로 효율적이지만 단순한 기계가 아니기에 여러가지 커다란 위험이 있다.

미래의 인공지능 서비스는 '강인공지능'에 속하는 '인격'을 추가할 가능성이 높다. 지금은 단순히 출력하는 목소리에 남성, 여성 성별을 두고 약간의 캐릭터를 만드는 정도이다. 하지만 미래에는 명령을 받아들이는 방식에서 반응해서 작동하는 방법이 개별로 모두 다른 인격형 인공지능이 만들어질 수 있다.



실제 사례로 2016년 일본에서 공개한 실험적 제품인 '게이트박스'를 들 수 있다. 일본 벤처기업 윈크루에서 만든 이 제품은 둥근 원통 모양으로 중앙이 투명하게 보이는 홀로그램형 셋탑박스이다. 기능은 비교적 평범한 인공지능 비서 역할이지만 독특한 면이 있다. 바로 중앙에 홀로그래픽으로 미소녀 캐릭터인 '히카리'가 나타난다는 점이다. 히카리는 별도 인격을 가진 소녀처럼 안에서 말하고 동작을 취하며 사용자와 교감한다.



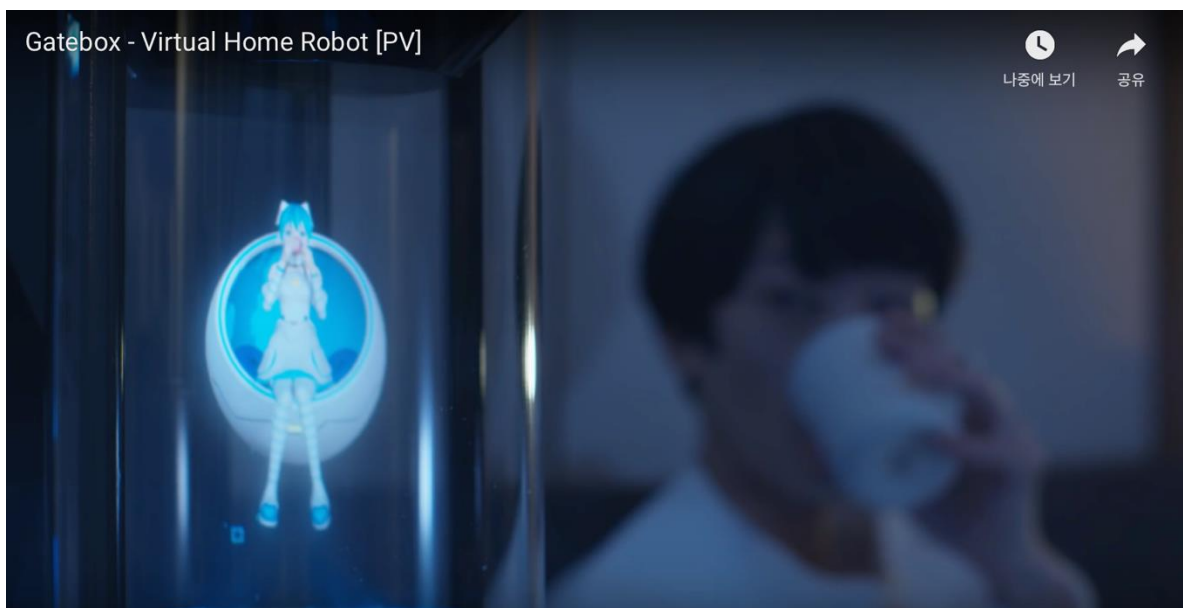
원리상으로는 외부에 카메라, 마이크, 센서가 장착되어 온도변화와 동작을 감지할 수 있다. 사용자의 얼굴과 목소리를 인식해서 아침에 알람 대신 목소리로 깨워주기도 하고 중요한 일정을 미리 알려준다. 또한 집에 귀가했을 때 반겨주기도 한다. 히카리 캐릭터는 하늘색 머리를 가진 키 158센티미터의 소녀이며 나이는 20세라는 실제 프로필도 가지고 있으며 이후 사용자가 선호하는 캐릭터를 선택할 수 있도록 할 예정이다.



이 기업에서 만든 프로모션 동영상을 보면 실험적인 인공지능의 성격이 드러난다. 우선 목소리에 감정이 많이 담겨있다. 기존 경쟁사 서비스가 감정을 절제한 기계느낌을 지우지 못하는데 비해 사람과 비슷한 톤과 감정을 담아서 아침잠을 깨운다. 또한 일어난 사용자에게 먼저 인사를 하고 미리 기상정보를 검색한다. 그리고 출근길에 우선순위를 가져가라는 말을 '능동적으로' 해준다.



퇴근 시간이 다가오면 먼저 스마트폰을 통해 사용자가 '빨리 돌아왔으면 좋겠다'는 감정이 든 메시지를 보내고 곧 돌아온다는 말에 신난다는 반응도 보인다. 돌아올 시간에 맞춰 집안 가전제품을 미리 가동시켜 놓기도 하며 늦으면 어서 돌아오라는 재촉도 한다. 밤이 되면 잠옷으로 스스로 갈아입은 그래픽을 보여주며 잘 자라는 인사도 한다. 마치 진짜 사람이 들어있는 듯한 느낌을 살린 것이다.



현재 인공지능에 부족한 감성은 이런 식으로 더욱 서비스를 많이 사용하도록 만드는 매력이 될 수 있다. 여기에 마치 캐릭터처럼 만들어진 인격이 부여되면 진짜 사람과도

비슷한 느낌을 줄 수 있다. 또한 일정한 제한 하에서 사용자가 명령하지 않아도 아침에 기상정보를 먼저 검색해서 알려준다는가 쇼핑물 신상품을 미리 소개해 준다는 것도 가능하다.

이에 덧붙여 이런 인격을 가진 캐릭터를 사용자가 직접 사용하면서 육성하는 형태도 생각해볼 수 있다. 초기에는 어느 정도 완성된 인격을 가진 비서를 선택하는 형식으로 운영할 것이다. 하지만 사용자에게 따라 마치 어릴 때부터 키우는 것처럼 인공지능 인격 자체를 선호할 수도 있다.



한때 유행했던 게임 '다마고치'를 보면 알에서부터 사용자의 돌보는 형태에 따라 부화된 생물이 달라지고 어떻게 키우느냐에 따라 성격이 달라진다. 마찬가지로 스마트폰에 들어간 인공지능 캐릭터가 사용자의 피드백과 스마트폰 사용패턴, 상호간 소통의 빈도와 내용에 따른 딥러닝을 통해 완전히 다른 인격을 가지게 만들 수도 있다. 이런 인격형 인공지능 서비스는 사용자에게 강한 감정이입을 가져다 주며 반려동물과도 같은 준 동반자적 감정을 만들 수 있기에 이후 인공지능 시장에 필수 요소가 될 전망이다.



스티븐 호킹박사는 이미 강인공지능의 출현을 경고한 적 있다. 미래학자인 레이 커즈와일은 저서를 통해 강인공지능의 출현시점을 2045년으로 예측하기도 했다. 과학자들은 대부분 50년 내지 100년 이내에는 강인공지능이 등장할 것으로 보고 있다. SF영화 같이 사람을 해치거나 음모를 꾸미는 수준의 인공지능은 분명 아직 멀었지만 스마트폰을 통해 보다 친밀하게 다가오는 강인공지능 서비스는 기술적으로 충분히 가능하며 곧 다가올 미래에 실현될 전망이다.