세상을 바꿀 변화의 시작, 음성 인터페이스와 스마트 스피커

지금으로부터 10년 전인 2007년, 미국 샌프란시스코에서 중요한 세상의 변화가 시작되었다. "오늘 세상을 바꿀 세 가지 디바이스(device)를 선보일 것입니다. 와이드 스크린에 터치 인터페이스로 동작하는 아이팟(iPod), 혁신적인 휴대 전화, 그리고 획기적인 인터넷 커뮤니케이터. 아이팟, 전화기, 인터넷 커뮤니케이터… 눈치 채셨나요? 네, 이건 세 가지 디바이스가 아닙니다. 단 하나의 디바이스, 바로 아이폰(iPhone)입니다."

2007년 맥월드(MacWorld) 키노트에서 스티브 잡스가 얘기한 것처럼, 애플은 휴대폰을 재창조했고 세상을 완전히 변화시켰다. 이 '작은 디바이스' 하나만 있으면 언제 어디서나 전화와 음악듣기는 물론, 인터넷과 연결된 수많은 서비스들을 편리하게 이용할 수 있게 됐다.

글 | 이석영 zodiac.lee@kakaocorp.com

현재 카카오에서 카카오의 인공지능 플랫폼인 카카오I와 이를 활용한 카카오의 첫 번째 스마트 스피커 카카오미니 프로젝트를 총괄하고 있다. 닷컴 버블 전인 1999년 부터 IT 서비스 기획을 계속 해오고 있으며, 웹서비스와 스마트폰이 세상을 크게 변화시켰던 현장의 중심인 네이버와 카카오에서 일을 해왔다. AI시대를 맞아, 사람들의 삶을 한번 더 진화시키기 위한 방법을 찾는것이 가장 큰 관심사이다.

스마트폰, 세상을 바꿔버린 슈퍼 디바이스

무엇보다 스티브 잡스가 아이폰을 소개하며 강조한 것처럼, 스마트폰은 다른 휴대용 기기를 통합했다. 오늘날 이 특징은 너무나 당연하게 여겨지고 있지만, 2000년대 초반만 해도 그렇지 않았다. 외출할 때 마다 사람들은 가방에 휴대폰, MP3 플레이어, 디지털 카메라, 휴대용 게임기, PMP(DMB), 전자사전 등의 기기를 넣고 다녀야 했고, 필요할 때 마다 그 중 하나를 꺼내어 이용하곤 했다.

단순히 생활 양식이 바뀐게 아니라, 앞서 언급된 디바이스들은 실제로 세상에서 자취를 감추거나 다른 형태로 바뀌었다. 한 때 필수품 취급으로 받았던 휴대용 MP3 플레이어는 거의 사라졌고, 디지털 카메라 시장은 고급형 모델 중심으로 재편되었다. 스마트폰은 이제 세상에서 가장 많은 사람들이 사용하는 휴대용 게임기가 되었고, 동영상앱을 실행하여 스마트폰으로 영화 한 편을 보는 시대가 되었다.

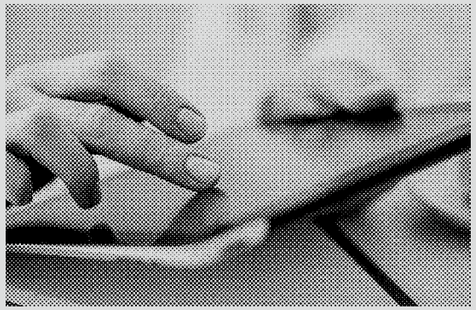
스마트폰의 등장으로 세상에서 자취를 감춘 디바이스는 휴대용 기기에 국한되지 않는다. 자동차에서는 거치용 내비게이션이 점점 사라지고 있고, 집에서는 종합 콤포넌트라고 불리던 오디오 데크가 사라졌다. 집에 유선전화를 개통하는 경우가 크게 줄었으며, 방마다 하나씩 있었던 탁상용 알람 시계 역시 어느 순간 찾아볼 수 없게 되었다.

이 모든 것을 사용하기 위한 단 하나의 인터페이스, '터치'

모든 개별 디바이스의 기능을 하나로 통합한 스마트폰은 "터치 인터페이스"라는 혁신적인 기술이 있었기에 존재 가능했다. 아이폰 이전에도 스마트폰이라고 할 만한 디바이스가 있었으나, 대부분 쿼티(QWERTY) 키보드를 장착하거나, 전용 스타일러스(stylus) 펜을 이용하는 형태였다. 스티브 잡스는 이를 대단히 못마땅 하게 생각했다. "신(神)은 우리에게 이미 스타일러스를 주셨어. 그것도 열 개나." 스티브 잡스가 아이폰 개발팀에게 자기 손을 흔들어 보이며, 자연스러운 터치 인터페이스의 개발이 가장 중요하다고 말한 이야기는 널리 알려진 일화다. 2007년 키노트에서 잡스는 아이폰의 가장 중요한 핵심 기능으로 손가락만으로 동작 가능한 '멀티 터치 인터페이스(multi touch interface)'를 첫 번째로 소개했다. 기존 쿼티 키보드와 스타일러스 펜이 가진 끔찍한 사용성을 함께 언급하면서.

실제로 터치 인터페이스는 매우 쉽고 훌륭하다. 미세한 손가락 움직임에도 반응하는 정전식 디스플레이 장치는 기존 입력 인터페이스 장치들(키패드, 마우스, 포인팅 장치 및 각종 버튼들)이 가지는 조작성의 한계를 사실상 완전히 없앨 수 있기 때문에, 스마트폰에 담겨있는 서비스와 기능들을 거의 무한대에 가까운 방식으로 이용할 수 있게 해준다. 게다가 손가락으로 화면에 표시된 무언가를 눌러 반응을 보는 것은 학습비용이 매우 낮을 뿐만 아니라 자연스러운 형태의 인터페이스이다. 두 세살짜리 아이가 아이폰을 쉽게 조작하는 모습을 보는 것은 이제 별로 놀라운 일도 아니다.

[그림 1] 터치 인터페이스



스마트폰으로의 과도한 통합이 초래한 불편함

너무 많은 기능들이 스마트폰에 담기고 이를 오직 "터치 인터페이스"로만 사용할 수 있게 되면서 불편해진 것도 있다. 대표적인 사례가 가정에서의 음악 감상이다. 과거에는 음악 감상을 위해 테이프나 CD를 데크에 넣고 플레이(Play) 버튼만 누르면 원하는 음악을 좋은 음질로 즉시 들을 수 있었다. 스마트폰과 음악 스트리밍(streaming) 서비스를 통해 언제 어디서나 다양한 노래를 들을 수 있게 되어 편리해진 '수혜'를 모든 사람이 쉽게 누릴 수 있는 것은 아니다. 모든 서비스와 산업이 스마트폰에서 소비되는 현대에, 누구나 쉽게 접할 수 있었던 음악감상을 할 수 없게 된 새로운 소외계층이 생겨났다. 아직도 적지 않은 50대 이상 장년층과 노년층에게 있어 스마트폰과 스트리밍 서비스를 통한 음악 감상은 매우 어려운 과업이다. 음악을 좋은 음질로 듣기 위해, 스마트폰을 블루투스(bluetooth) 스피커와 연결해야 하는데 이는 더더욱 어렵다.

터치 인터페이스 역시 만능은 아니다. 스마트폰은 물리적으로 화면의 크기가 제한되어 있으므로, 한 번에 제공할 수 있는 인터페이스의 정보량이 많지 않다. 그러다 보니 수많은 서비스와 기능들을 이용하기 위해서는 부득이하게 여러 번의 단계를 거쳐 서비스를 이용하도록 설계할 수 밖에 없다. 어떤 서비스를 사용하려고 해도, 보안 잠금을 해제한 후 해당 서비스 앱을 실행하여 몇 번의 터치를 거쳐야만 원하는 기능을 실행할 수 있다. 게다가 스마트폰에는 사용해야 하는 기능이 너무 많다. 터치 인터페이스 자체는 학습비용이 낮지만, 스마트폰에 익숙한 젊은 사람들 조차 앱과 스마트폰에 내재된 기능을 제대로 찾지 못한다. 휴대폰의 설정을 바꾸기 위해 여러 번의 시행착오를 겪는 젊은

사람들의 모습을 쉽게 찾아볼 수 있다.

또한 터치 인터페이스를 사용하기 위해서는 눈(시각)과 손을 필요로 하는데, 이는 태생적으로 멀티 태스킹을 할 수 없도록 만든다. 눈과 손을 온전히 스마트폰을 위해 사용해야만 원하는 기능을 얻을 수 있고 이 과정에서 다른 일들을 동시에 하는 것은 매우 어렵다. 이는 불편함을 넘어 때로는 사용자를 위험에 빠뜨린다. 보행 중이나 운전 중에 스마트폰을 사용하는 것은 대단히 위험하다. 다수의 경우가 실제로 사고로 연결되기도 한다. 스마트폰에 많은 편리한 기능이 담겨 있기 때문에, 사용자들은 앞서 말한 위험에도 불구하고 이동하며 스마트폰을 보는 경우가 많다. 이 역시 슈퍼 디바이스로써 스마트폰의 존재와 이를 터치 인터페이스로 사용해야만 하는 결합이 만들어낸 새로운 종류의 사회적 이슈이다.

가장 자연스럽고 효과적인 인터페이스, '음성 대화'

터치 인터페이스가 스마트폰과 함께 10여년간 가장 훌륭한 인터페이스가 될 수 있었던 것은 앞서 얘기한 것 처럼 자연스러움 덕분이었다. 그러나 터치 인터페이스는 '터치 스크린(touch screen)'이라는 최소한의 물리적 장치를 수반하고, 스크린 위로의 터치라는 물리적 행동을 필요로 하는 한계를 갖고 있다.
이러한 관점에서 봤을 때, 음성을 통한 대화형 인터페이스는 가장 쉽고 자연스럽게 복합적인 기능을 사용할 수 있는 방법일 것이다. 일단 음성 대화 인터페이스는 단계를 거칠 필요가 없이, 모든 서비스 이용을 한 번에 할 수 있게 해준다. 원하는 음악을 듣거나, 뉴스와 날씨 정보를 확인하거나, 알람을 맞추거나, 전화를 걸거나, 심지어 이번에 카카오미니를 통해 제공되는 기능인 카카오톡 보내기 조차도 한 번의 음성 명령으로 즉시 실행된다. 입과 귀를 사용하는 음성 대화 인터페이스는 터치 인터페이스와 달리 인터페이스 장치를 정확히 인지하지 않고도 사용할 수 있어 멀티 태스킹(multi tasking)에 훨씬 적합하다. 음성 대화 인터페이스는 앞서 얘기된 운전 중이나 보행 중에도 위험성 없이 사용할 수 있다. 아침시간 집에서 바쁘게 출근 준비를 하면서도 음성 대화 인터페이스로 날씨, 뉴스, 주가 등을 편하게 확인할 수 있다.

음성 대화를 통한 인터페이스의 가장 큰 장점은 학습비용이 터치 인터페이스보다도 낮다는 점이다. 대화라는 것은 모든 사람이 태어나서부터 배우고 이미 방법을 알고 있는 소통 방식이다. 기존 인터페이스들이 HCI(human-computer interface)라는 학문적 기반에서 발전해 왔고, 이는 사람처럼 대화를 할 수 없는 기계를 사용하기 위해 만들어진 방법임을 생각해 본다면, 음성 대화로 컴퓨터를 조작하는 행위는 궁극의 인터페이스의 한 형태라고도 볼 수 있을 것이다.

물론, 아직 음성 대화를 통한 인터페이스는 완벽하지 않다. 이것이 완전해지기 위해서는 모든 자연스러운 대화를 이해하고 응답되어야 한다. 현재의 기술로 이를 완전하게 구현하는 것은 쉽지 않다. 그러나, 음성인식과 AI기술이 빠르게 발전하고 있기 때문에 자연스러운 모든 대화를 이해하는 컴퓨터 혹은 서비스가 등장하는 것은 어쩌면 그리 멀지 않은 미래의 이야기일 수도 있다.

[그림 2] 영화 속에서 음성 대화 인터페이스가 활용되는 예시 : HER

왜 스마트폰이 아닌 스마트 스피커인가?

사실, 음성 대화 인터페이스로 터치 인터페이스가 가진 한계를 극복하고자 했던 시도는 스마트폰 제조사를 중심으로 이미 몇 년 전 부터 진행되어 왔다. 그러나 애플 시리(Siri)와 구글 어시스턴트(Google Assistant)를 통한 스마트폰 기반의 음성 대화 인터페이스는 시장에 제대로 정착하지 못했다. 음성 인터페이스는 효용성이 떨어지는 기술로 평가받으며, 훨씬 더 먼 미래에나 일상에서 사용될 수 있는 것처럼 보였다. 아마존이 에코(Echo)를 발표하여 사람들의 일상이 음성 인터페이스로 실제로 바뀔 수 있음을 보여주기 전 까지는.

아마존은 "음성 대화 인터페이스"가 가지는 특징과 가치를 제대로 이해하고 있었고, 이것이 에코가 성공을 거둘 수 있었던 가장 큰 요인일 것이다. 아마존은 전원에 상시 연결된 가정용 스피커 디바이스를 통해 알렉사(Alexa) 서비스를 제공했다. 사용자들은 알렉사를 통해 기존 스마트폰 음성 비서와는 다른 두 가지 중요한 사용자 경험을 제공할 수 있었다. 첫번째는 스피커를 24시간 음성 입력 대기 상태로 만듦으로써, 사용자가 디바이스를 사용하기 위한 별다른 준비를 하지 않아도 된다는 것이다. 사용자는 아무 때나 '알렉사'를 부르는 것만으로 서비스를 바로 이용할 수 있게 됐다. 이 과정은 서비스 이용 단계를 단 한 번으로 끝낼 수 있는 음성 인터페이스의 본질적 사용자 가치를 제대로 구현하기 위한 중요한 요소였다. 두번째는 음성 대화만으로 모든 서비스를 완전하게 이용할 수 있도록 만든 점이다. 이를 통해 사용자는 눈과 손을 자유롭게 쓸 수 있을 뿐 아니라 특별한 학습이 필요 없이 자연스러운 대화를 통해 서비스를 이용할 수 있었고, 이를통해 음성 인터페이스의 진정한 편리함을 완전하게 경험할 수 있었다.

스마트폰의 음성 비서는 알렉사가 보인 음성 인터페이스의 경험을 제대로 구현하는데 한계가 있었다. 음성 웨이크업 기능이 있지만, 보조적인 수단이었고 옵션을 꺼 두는 경우가 많아 신뢰도가 낮았다. 또한 스마트폰의 음성 비서는 터치 인터페이스의 병행 사용을 유도했는데 이로 인해 음성 대화는 스마트폰의 오롯하게 활용하는 수단이 되지 못하고, 보조 인터페이스로의 위상을 벗어나지 못하게 된다. 10년간 학습되어 온 '스마트폰 조작=터치 인터페이스 사용'이라는 명제에서 탈피해, 음성 대화로 스마트폰을 조작하는 것은 스마트폰 사용자에게는 낯선 경험으로 인식될 수밖에 없는 환경이 스마트폰의 음성 비서를 보조적 도구로 머물게 만든 이유이기도 했다.

스마트폰으로 인해 사라졌던 디바이스들의 부활

아마존이 발표하는 에코의 새로운 라인업(line up)을 보면, 스피커가 음성 대화 인터페이스 구현에 적합했기 때문에 선택된 것만은 아닌 듯 보인다. 아마존이 새로운 에코 라인업을 통해 선보이는 가장 중요한 기능 중 하나는 전화(Echo Show)와 알람 시계(Echo Spot)이다. 전화 기능과 알람 기능은 기본 에코 디바이스에서도 제공되는 기능이지만, 에코 쇼와 에코 스팟은 이 두 가지 기능을 디바이스의 형태적인 측면으로도 강조하고 있다.

아이러니하게도, 전화기와 알람시계는 에코가 대체한 '오디오 데크'와 더불어 스마트폰의 확산으로 자취를 감춘 가정용 디바이스였고 스마트폰 이전 시대에는 독립적인 기기로서 편리하게 사용되던 것들이다. 아마존의 에코 라인업은 스마트폰 시대에 없어져 버린 가정용 디바이스들을 통합하여 새로운 슈퍼 디바이스로 부활시키려는듯 보인다. 에코와 같은 스마트 스피커를 이용하면 집안에서 스마트폰을 쓰는 것보다 훨씬 편리하게 음악 감상이나 알람 설정을 할 수 있다. 가전 기기 제어와 각종 정보 확인, 커뮤니케이션, 쇼핑, 음식 주문하기도 에코를 통해 이용할 수 있다. 물론 아직은 스마트폰을 이용할 때 더 편리하게 이용할 수 있는 서비스들이 훨씬 많다. 스마트폰의 등장이 개인용 컴퓨터를 완벽하게 대체하지 않았던 것처럼, 스마트 스피커가 또 다른 슈퍼 디바이스가 된다고 해도, 스마트폰 역시 계속 사용될 것이다.

그러나 스마트 스피커는 그동안 스마트폰에게 부여된 과도한 역할 중 가정에서의 IT 서비스 사용 경험을 음성인터페이스라는 편리한 UX와 함께 많은 부분 대체할 수 있을 것이다.

변화는 이제 막 시작되었다.

스마트폰이 등장하고 10년 동안 세상이 더 편리해진것 처럼, 음성 대화 인터페이스와 새로운 가정용 스마트 디바이스의 출현은 앞으로 오랜 시간에 걸쳐 삶의 많은 부분을 변화 시키고 새로운 가치를 만들어 낼 것이다.그리고 언제나 혁신적인 생활 플랫폼을 만들어 왔던 카카오 역시, 카카오의 인공지능 플랫폼인 카카오I와 스마트 스피커 카카오미니를 시작으로 이 거대한 변화를 함께 만들어 나갈 것이다.