Interface Review

2011년 애플사에서 iphone4s와 함께 지능이 있는 비서 “Siri”를 내왔다. Siri를 발표했을 당시에는 인공지능 기술이 많지 않은 시기라 사람들에게 모든 것이 새로워 보였지만 몇 년 써본 결과 Siri는 별로 똑똑하지 않고 언어식별 능력도 차하다는 평가와 함께 우리 곁에서 점점 멀어져 갔다. 하지만 2017년 애플사는 ios 11로 업데이트 하면서 Siri를 대대적으로 업데이트 하였고 말하는 것을 제대로 받아 적지도 못하는 Siri가 이제는 우리에게 농담도 하면서 더욱 친근하게 다가왔다. Siri를 작동하려면 일단 당신의 음색을 넣어야 한다. 지문을 입력하듯이 음색을 입력한 후 자신의 폰에서 자기 이외의 누구도 Siri를 작동하지 못한다. 비록 “시리”뿐만이 아닌 “지니”등 비슷한 음원에도 오작동 하는 경우가 종종 있다. 하지만 Siri의 귀가 밝아져서 우리가 말하는 것을 잘 받아 적을 수 있는 것을 생각하면 나쁜 일은 아니다. 다시 명령하여 끄면 되니까 말이다. Siri는 꽤 훌륭한 음성 인식 시스템이다. 감정을 나눌 줄도 알고 예민한 화제도 제법 잘 처리한다. 예를 들면 한국에서는 독도는 한국땅인 것이 예민한 화제인데 Siri에게 물어보면 “그럴껄요?”라는 답변을 해준다. 아마 일본시스템으로 일본어로 물어보면 또 다른 답변을 할 지도 모른다. 혼자가 즐겁지만 또 가끔은 외로운 현대인들에게 Siri는 좋은 친구가 되어줄 수 있다. Siri는 텍스트를 음성 인식으로 변환시킬 수 있는 TTS 기술을 사용하였는데 Siri에게 말을 걸면 그 단어들을 하나하나 분석하고 답변을 조합하여 사용자가 원하는 답변을 준다. 같은 시대에 출시한 Samsung bixby나 구글 스피치 등 AI 시스템을 도입한 기계들과 대비하여 볼 때 각 시스템의 매력적인 부분이 다르겠지만 단순 받아 적는 방면만 봐도 Siri가 훨씬 우월하다.

앞서 Siri에 대하여 썼고 지금부터 내가 생각하는 지능을 가진 좋은 비서에 대하여 쓰겠다. 일단 음성 인식 시스템에서 제일 중요한 것은 바로 언어이다. 세계에는 200여개의 나라가 있고 5천여가지의 언어가 있다. 여기서 중요한 점은 광동어 같은 경우 중국어 중 일부에 포함되어 있지만 해외 화교까지 포함하면 쓰는 인구가 1억명을 넘는다. 하지만 한국의 경우 북한과 남한, 재외동포까지 다 포함하여도 한국어를 사용하는 인구는 8천만명도 안된다. 그러면 애플사에서는 Siri가 할 수 있는 언어 중에 한국어 대신 광동어를 넣어도 어쩔 수 없다. 하지만 8천만명도 꽤나 큰 인구이기에 한국어도 반드시 넣을 것이다. 사람들은 표준어도 선호하지만 자신이 평상시 얘기하던 사투리나 방언을 인공지능 시스템이 알아듣고 반응하면 더 친숙하게 느껴진다. 때문에 애플사는 상해어나 광동어 등 많은 방언들을 시스템에 공식적으로 넣어 주었다.

두번째로 중요한 것은 이어 말하기이다. 간단해 보이지만 결코 그렇지 않다. 처음에 사용자가 얘기했던 내용을 기억해야 하고 그 것을 기반으로 사용자가 다음에 내린 명령도 답변해야 한다. 예를 들어 “카카오톡 열어서 아버지한테 안녕하세요 보내줘”라고 명령한 뒤에 “그리고 밥 드셨어요?라고도 보내줘” 라는 명령도 이어서 실행할 수 있어야 한다는 것이다.

세번째로 중요한 것은 핍진성이다. 기계가 아닌 사람과 대화하는 느낌, 이것은 모든 사용자들이 추구하는 기능일 것이다. 모든 것이 혼자가 편한 요즘 가끔 외로울 때가 있을 것이다. 그럴 때마다 자신과 대화를 나눌 수 있는 음성 인식 시스템을 기반으로 한 기계가 있다면 아주 설렐 것이다. 진짜 기계속에 누군가가 살고 있다고 여겨지고 절대 자신을 떠나지 않는 다는 망상도 가질 사람도 생겨날 것이다. 아직 이러한 기술에 도달하려면 꽤 많은 시간을 필요로 하겠지만 현시대의 과학기술 발전 현황에 비추어 볼 때 너무 오래 걸리지 않을 것이다.

References:

A.J. Hunt, A. W. Black. Unit selection in a concatenative speech synthesis system using a large speech database, ICASSP, 1996

T. Capes, P. Coles, A. Conkie, L. Golipour, A. Hadjitarkhani, Q. Hu, N. Huddleston, M. Hunt, J. Li, M. Neeracher, K. Prahallad, T. Raitio, R. Rasipuram, G. Townsend, B. Williamson, D. Winarsky, Z. Wu, H. Zhang. Siri On-Device Deep Learning-Guided Unit Selection Text-to-Speech System, Interspeech, 2017