Towards a Proactive Smart Speaker Responding to User's Desk Activities

Huhn Kim[1], Sohyang Lee[2*]

[1]Department of Mechanical System Design Engineering, Professor, Seoul National University of Science and Technology, Seoul, Korea

[2]Department of Design and Engineering, Student, Seoul National University of Science and Technology, Seoul, Korea

Abstract

Background Users tend to more positively evaluate intelligent agents with higher personification properties. However, the conversations with smart speakers are currently initiated by the user's unilateral utterance of the call word, and the conversation does not take into account information or tastes of the user's situation. This differs from the general communication characteristics between people. In this work, we evaluate the user preference by different levels of active response of the smart speakers that perform proactive conversations by automatically recognizing the user's desk activity.

Methods First, we defined a system concept based on deep learning and rule-based models with data acquired from microphones and sound sensors, human sensitivity sensors, and light sensors to automatically recognize the users' desk activity. Second, we divided the task situations that can be judged by the system into 19 different levels, and we derived specific scenarios by dividing the active level of the interaction into four stages: non-response, simple response, situation prediction and suggestion, and proactive response executing the suggestions. Third, we evaluated which level of active interaction is more preferred through user evaluation for each task situation.

Results In most task situations, situation prediction and suggestion and proactive response interactions have been shown to be preferable to non-response, while simple response interactions have been evaluated negatively. In particular, the participants in the experiment were found to be concerned about context interruption, especially in situations where they were immersed in certain tasks or where there were several people together.

Conclusions Smart speaker's proactive conversation depending on user's context will be very useful if the system's higher recognition accuracy is supported, thereby providing a more extended user experience.

Keywords Smart Speaker, Voice User Interface, Desk Activity, Context Recognition, Proactive Interaction

This work has been conducted with the support of the "Project for Nurturing Advanced Design Professionals" initiated by the Ministry of Trade, Industry and Energy of the Republic of Korea, and was published based on Master's thesis of the corresponding author in Seoultech 2020.

*Corresponding author: Sohyang Lee (kkangyu12@naver.com)

Citation: Kim, H., & Lee, S. (2021). Towards a Proactive Smart Speaker Responding to User's Desk Activities. *Archives of Design Research, 34*(3), 155-171.

http://dx.doi.org/10.15187/adr.2021.08.34.3.155

Received : Feb. 15. 2021 ; **Reviewed :** Apr. 06. 2021 ; **Accepted :** May. 26. 2021
pISSN 1226-8046 **eISSN** 2288-2987

1. 연구의 배경 및 목적

인공지능 기술의 발달로 인해 인공지능은 다양한 제품군에 적용되고 있다. 일상생활에서 인공지능 기술을 쉽게 접할 수 있는 매개물 중 하나로 스마트 스피커를 꼽을 수 있다. 스마트 스피커는 음성 사용자 인터페이스(Voice user interface, VUI)를 기반으로 한 지능형 에이전트(Intelligent agent)가 적용된 제품이다. 이 지능형 에이전트는 사용자의 명령이나 질문에 대한 답을 즉각적으로 제공하는 등 충실한 비서의 역할을 수행한다. 아마존의 알렉사(Alexa), 구글의 구글 어시스턴트(Google assistant), 애플의 시리(Siri), 삼성의 빅스비(Bixby) 등은 지능형 에이전트의 대표적인 예이며, 각 기업은 이들을 적용한 스마트 스피커를 경쟁적으로 출시하고 있다. 스마트 스피커와의 대화는 호출어 발화로부터 시작된다. 사용자가 “오케이. 구글” 또는 “하이, 빅스비” 등 미리 정해진 호출어를 발화하면 대화가 가능한 상태로 진입하는데, 대부분 사용자의 일방적 호출에 의해서만 수동적으로 대화를 시작한다.

나스, 슈에르, 그리고 타우버(Nass, Steuer & Tauber, 1994)는 컴퓨터와 같은 시스템에 인간과 유사한 속성이 나타날 때, 사람들은 이러한 시스템을 사회적 존재로서 인식한다는 CASA(Computers are social actors) 패러다임을 발표하였다. 또한, 리브즈와 나스(Reeves & Nass, 1996)는 사람들이 무생물인 컴퓨터와 인터랙션을 하고 있다는 사실을 인지하면서도 의인화된 신호(Anthropomorphic cue)를 보내는 컴퓨터와의 인터랙션에서 사회적 규칙과 행동을 보인다고 하였다. 이러한 CASA 패러다임에 기반하여 에이전트나 로봇과 같은 인공물에 의인화 속성을 부여하여 사용자의 반응을 관찰한 연구들이 진행되어왔다. 홍은지, 조광수, 그리고 최준호(Hong, Cho & Choi, 2017)에 따르면 인공물을 의인화하는 방법은 크게 세 가지이다. 첫째, 외적 의인화는 인공물에 인간과 유사한 얼굴, 체형, 성별을 부여하는 것이다. 또한, 김헌, 고재영, 김승완, 황호연(Kim, Ko, Kim & Hwang, 2019, 2020)에 따르면 귀에 들리는 음성이나 말투, 심지어 말의 빠르기나 길이도 에이전트에게서 느끼는 개성에 영향을 미친다. 둘째, 내적 의인화는 정서, 감정 등 인간의 내적 상태를 흉내내는 것이다. 셋째, 사회적 의인화는 양방향 의사소통을 의미한다.
이재길, 김기준, 이상원, 신동희(Lee, Kim, Lee & Shin, 2015), 아라우조(Araujo, 2018), 쉬, 양, 마, 루, 카오(Shi, Yan, Ma, Lou & Cao, 2018), 응우옌, 타, 프라이부톡(Nguyen, Ta & Prybutok, 2018)의 연구 결과를 종합하면, 에이전트의 의인화 속성이 높을수록 교감의 정도와 참여도, 신뢰성 등의 사회적 반응이 높게 나타나고, 이는 지속적 사용 의도를 촉진한다고 한다. 이처럼 지능형 에이전트의 디자인 시에 다양한 의인화 방법을 통해 에이전트의 사회성을 높이는 것이 중요하다. 일부 스마트 스피커의 VUI에서도 의인화된 속성을 볼 수 있는데, 미지원 기능임을 안내할 때에는 미안한 말투로, 축하의 말을 건넬 때는 밝은 톤으로 말하는 등 대화 내용에 따라 다른 톤과 매너를 사용하고 있다. 또한, 일상 대화와 농담, 감성 대화를 시도하는 등 인간의 의사소통 특성을 모방하는 모습을 보이지만, 아직은 미흡한 수준이다.

또한, 인간 사이의 의사소통에서는 상대방이 놓여 있는 상황에 대한 정보를 인지하고, 적절한 시점에 상대방에게 말을 걸어 대화를 시도한다. 이는 콘텍스트 인지(Context-awareness)와 연관된 것으로, 데이와 어보우드(Dey & Abowd, 2000)는 사용자에게 태스크와 관련된 정보 및 서비스를 제공하기 위해 맥락(Context)을 이용한다면 이 시스템은 콘텍스트를 인지한다고 하였다. 이춘화와 헬랄(Lee & Helal, 2003)에 의하면 콘텍스트를 인지하는 시스템은 사용자에게 적절한 시간 및 장소에 적합한 서비스를 선택하여 추천한다고 하였으며, 더불어 괴츠, 키슬러, 그리고 파워스(Goetz, Kiesler & Powers, 2003)에 따르면 사용자는 심각하거나 혹은 장난스러운 주변 콘텍스트에 적절히 대응하는 에이전트를 더 선호한다. 이러한 배경을 바탕으로 본 연구에서는 사용자의 일방적인 호출에 의해 대화가 시작되는 한계에서 벗어나 사용자의 콘텍스트를 파악하고, 그에 따른 제안이나 추천 기능을 실행하는 등의 능동적(Proactive) 인터랙션을 시도하는 스마트 스피커를 제안하였으며, 이러한 스마트 스피커의 능동적 개입에 대한 사용자의 반응을 평가하였다.

2. 능동적 인터랙션을 시도하는 스마트 스피커의 콘셉트 정의

2. 1. 데스크 액티비티의 정의

먼저, 능동적 인터랙션을 시도하는 스마트 스피커의 콘셉트를 정의하기 위해 능동적 개입이 사용자로부터 유의미한 반응을 끌어낼 수 있는 사용 상황을 조사하였다. 이를 위해 스마트 스피커가 설치되는 장소에 주목했는데, 킨셀라와 무슐러(Kinsella & Mutchler, 2019)에 따르면 스마트 스피커의 일반적 설치 장소는 Figure 1과 같았다. 거실이 44.4%로 가장 설치 비율이 높았는데, 2018년의 조사 결과와 비교하면 1.5% 감소하였다. 반면, 침실과 화장실, 그리고 홈 오피스와 같이 가정 내 독립된 공간에서의 설치 비율은 모두 증가하였다. 특히 서재와 같은 홈 오피스에서의 설치 비율은 2018년에 비해 3.5% 증가하였다. 이와 더불어 스마트 스피커를 2대 이상 보유하고 있는 가정의 비율 또한 증가했는데, 2018년의 34.3%에서 2019년에는 41.9%로 증가하였다(Figure 2). 이러한 이유로 향후 가정 내 독립된 공간마다 스마트 스피커가 설치될 것이라는 전망이 유력하다.

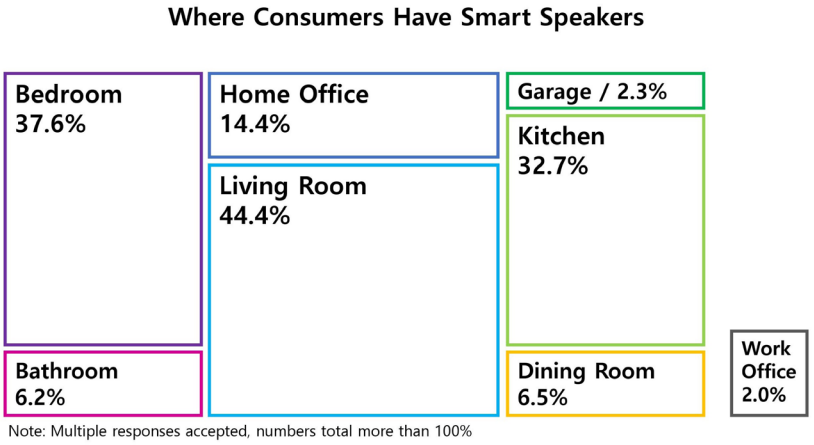

Figure 1 Primary location of smart speakers in 2019
(Smart speaker consumer adoption report 2019)

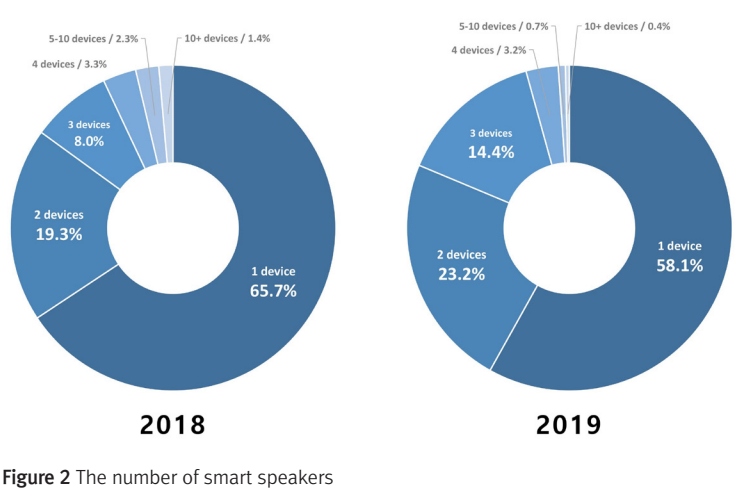

Figure 2 The number of smart speakers
(Smart speaker consumer adoption report 2019)

서재와 같은 공간에서 책상 위에 놓인 스마트 스피커의 경우 침실이나 거실 등과는 달리 발생 가능한 사용자 콘텍스트가 한정되어 있어 예측이 가능하기에 본 연구에서는 이러한 상황에서 스마트 스피커가 시도할 수 있는 능동적 인터랙션에 중점을 두었다. 책상 위에 놓인 스마트 스피커는 책상 앞에 앉아 있는 사용자의 콘텍스트 정보를 인식하고, 그에 필요한 대화를 시도하거나 추천 기능을 제안 또는 실행할 수 있을 것이다. 본 연구에서는 이처럼 사용자가 책상 앞에 앉아서 하는 활동들을 데스크 액티비티(Desk activity)라고 정의하였다. 이를 세분화하기 위해 먼저, 가정에서 수행된 콘텍스트 유추와 관련한 연구 내용 중 책상과 그 주변에서 일어나는 활동들을 추출하였다. 이때, 거실과 몇 개의 독립된 방으로 구분된 일반적인 저택 환경뿐만 아니라 거실과 침실, 서재 등의 공간이 구분되어 있지 않은 원룸 환경 모두를 고려하였다. 추출된 활동은 어피니티 다이어그램(Affinity diagram)을 활용하여 유사한 키워드 그룹으로 분류하였고, 각각의 키워드 그룹을 데스크 액티비티로 구분하였다. 그 결과, 데스크 액티비티는 '자리 복귀', '컴퓨터 작업', '공부', '독서', '미팅 또는 회의', '잠', '간식 섭취', '사색', '자리 떠남', '기타 활동'의 10가지로 세분화되었다. 자리 복귀란 외출했던 사용자가 돌아와 책상 앞에 앉는 행동을 의미하며, 데스크 액티비티에서의 잠은 책상에 엎드리거나 의자에 기대 잠이 든 것을 의미한다. 또, 자리 떠남은 외출 등의 목적으로 사용자가 책상 앞에서 떠나는 행동을 의미하며, 기타 활동은 나머지 9가지의 분류 중 어디에도 속하지 않는 활동들을 의미한다.

2. 2. 능동적 스마트 스피커의 하드웨어 구성

앞서 정의한 데스크 액티비티를 스마트 스피커가 인지하기 위해서는 각종 센서를 이용하여 사용자의 활동을 측정할 수 있어야 한다. 실내 공간에서, 사용자의 활동을 측정하려는 시도로서 크리쉬난, 쿡(Krishnan & Cook, 2014)과 루, 푸(Lu & Fu, 2009)는 모션 센서, 가속도 센서, 진동 센서, 압력 센서, 조도 센서 등을 문, 소파, 테이블, 가전기기 등에 부착하였다. 이러한 방식은 비교적 넓은 공간에서 스마트 홈 환경을 구성하기에 유리한 장점이 있으나, 본 연구에서는 콘텍스트를 데스크 액티비티로 제한하였으므로 인식하려는 사용자의 활동 범위가 책상 주변으로 한정적이다. 따라서 스마트 스피커에 사용자의 데스크 액티비티를 인지할 수 있는 센서들의 장착이 필요하다. Figure 3은 본 연구에서 제안하는 능동적 스마트 스피커의 구현에 필요한 하드웨어 구성을 보여준다.

첫째, 사용자의 데스크 액티비티를 인지하기 위한 핵심적인 요소는 마이크이다. 마이크는 스마트 스피커에 내장된 필수 요소이며, 최근 활발한 연구가 진행되고 있는 머신러닝이나 딥러닝과 같은 알고리즘 기술을 이용하면 녹음된 소리를 듣고 사용자의 활동을 예측할 수 있다. 아가왈, 제인, 쿠무르, 그리고 파텔(Agarwal, Jain, Kumur & Patel, 2018)은 스마트 스피커에 내장된 마이크들에서 인식한 소리를 딥러닝 알고리즘을 이용하여 어깨 회전, 다리 들어올리기, 제자리 걷기 등 10가지의 움직임을 96%의 정확도로 분류하였다. 또한, 스마트폰의 마이크와 가속도 센서를 이용하여 자는 동안의 숨소리 및 움직임을 분석하여 수면 단계(깊은 잠 또는 렘수면 등)를 기록하고 수면의 질을 측정하는 앱 등의 개발사례는 마이크를 이용하여 사용자의 활동을 측정할 수 있음을 뒷받침한다.

둘째, 마이크와 더불어 인체 감응 센서와 조도 센서가 필요하다. 이 센서들은 사용자의 움직임과 사용자 주변의 밝기 데이터를 측정하여 소리 데이터만으로는 알기 힘든 사용자 주변의 환경에 관한 정보를 습득할 수 있다. 그리하여 사용자의 데스크 액티비티를 더욱 정확히 파악하고, 주변 환경에 따른 다양한 서비스 제공을 가능하게 해 준다.

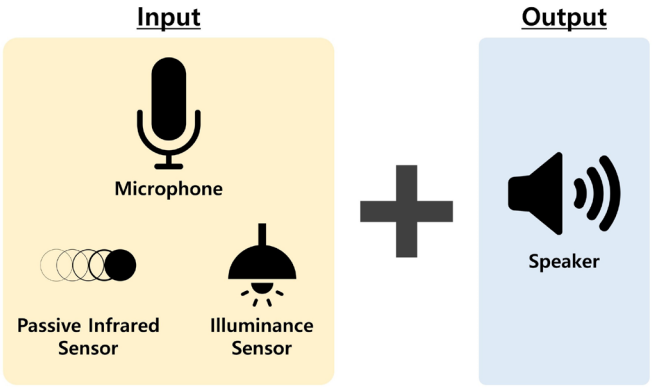

Figure 3 Hardware system for proactive smart speakers

2. 3. 능동적 스마트 스피커의 소프트웨어 시스템 구성

스마트 스피커 하드웨어 시스템을 통해 입력된 데이터를 이용하여 사용자의 데스크 액티비티를 판별하는 소프트웨어 시스템은 두 개의 모델로 구성하였다. 사람은 책 넘기는 소리, 마우스 클릭 소리 등 특정 소리만을 듣고도 독서를 하는지, 컴퓨터를 사용하는지와 같은 대략적인 상황을 유추할 수 있다. 이에 착안하여, 첫 번째 모델은 마이크에서 인식된 사운드 데이터를 딥러닝 알고리즘에 기반하여 소리 종류로 분류해 내는 사운드 인지 모델이다. 두 번째 모델은 첫 번째 모델의 출력값인 소리 종류와 나머지 센서들로부터 측정된 데이터들의 변화량 및 시간 데이터를 입력값으로 하며, 이를 특정 규칙 기반으로 사용자의 데스크 액티비티를 판별하는 모델이다.

본 연구에서는 소프트웨어 시스템의 구성을 결정하는 과정에서 구글의 오디오셋(AudioSet) 데이터 중 데스크 액티비티의 상황 구분에 이용할 수 있는 16가지 레이블에 해당하는 데이터만을 추출하여 이를 구분해내는 딥러닝 모델로 학습시켰다. 그 결과 정확도는 59.82%로 다소 낮게 측정되었는데, 이는 오디오셋 데이터가 유튜브의 동영상에서 추출된 만큼 순수하지 않은 데이터인 점과 데이터의 개수가 적었던 것이 원인이기도 하지만 사운드 분류 기술의 성능이 아직 미흡하기 때문이기도 하다. 이를 보완하기 위하여 두 번째 모델은 규칙 기반 모델로 구성하였다. 규칙 기반 모델의 입력값은 정해진 조건에 따라 출력값이 정해지기에 규칙을 잘 설정하기만 하면 사운드 분류 모델의 낮은 정확도를 보완할 수 있다.

이러한 소프트웨어 시스템과 앞서 정의했던 하드웨어 시스템을 바탕으로 한 전체 시스템의 아키텍처는 Figure 4와 같다. 사용자가 책상 앞에 앉아 어떠한 행동을 하면 스마트 스피커에 내장된 마이크, 인체 감응 센서, 조도 센서가 상황에 대한 정보를 수집한다. 마이크로부터 입력된 사운드 데이터는 사운드 인지 모델을 통해 소리 크기에 대한 데이터와 함께 어떤 종류의 소리인지 판별된다. 인체 감응 센서, 조도 센서는 아두이노와 같은 하드웨어를 통해 데이터가 입력되며 사운드 인지 모델에서 예측한 소리의 크기·종류 데이터와 시간 데이터에 기반한 규칙 기반 모델을 통해 사용자의 데스크 액티비티와 구체적인 태스크 상황을 예측한다. 그리고 예측된 상황에 따라 미리 정해진 시나리오에 기반한 소리를 스마트 스피커가 들려준다.

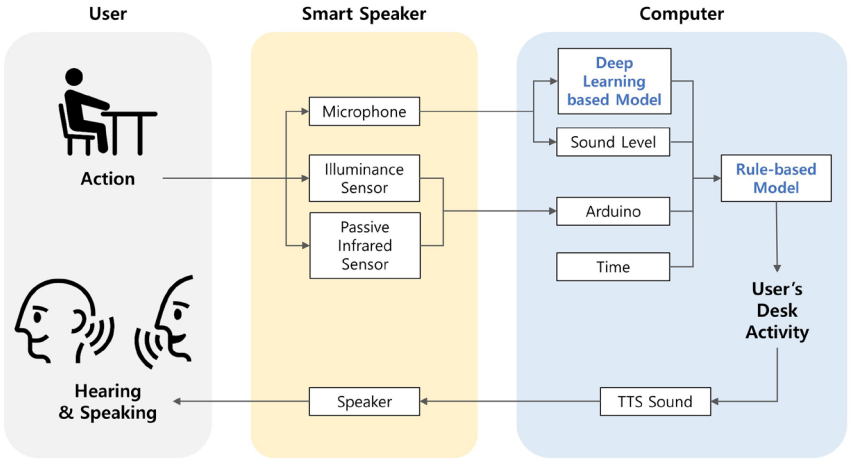

Figure 4 System architecture for proactive smart speakers

3. 능동적 VUI 시나리오 개발

3. 1. FGD를 통한 인터랙션 아이디어 수집

앞서 정의한 10가지 데스크 액티비티를 인지하는 스마트 스피커와의 유용한 사용자 인터랙션 아이디어를 수집하고자 포커스 그룹 토의(Focus group discussion, FGD)를 실시하였다. 스마트 스피커에 대한 배경지식을 지닌 평균 나이 26.0세의 20대 남녀 6명(남: 3, 여: 3)이 참여하였다. 이들 중 3명은 최소 3개월 이상 스마트 스피커를 사용 중인 참여자들이었다. 참여자들은 먼저 각 데스크 액티비티별로 기대하는 인터랙션 발생 상황과 구체적인 인터랙션 방안을 포스트잇에 5분간 작성하였다. 그 후 자신의 아이디어를 공유하고, 서로 의견을 나누면서 아이디어를 발전시켰다. 이러한 방법으로 FGD를 약 90분간 진행하여 총 113개의 아이디어를 도출하였다.

3. 2. 인터랙션을 위한 태스크 상황 구체화

인터랙션이 발생하는 태스크 상황을 구체화하기 위해 수집된 아이디어들 중에서 앞서 정의한 스마트 스피커 시스템(Figure 4)이 인식할 수 있을 것으로 예상되는 아이디어만을 추출하였다. 제안한 시스템은 소리가 주요 정보원 중 하나이므로, 다른 데스크 액티비티와 유사한 소리가 발생하거나 특징적인 소리가 발생하지 않아 사용자의 활동 인식에 어려움이 있을 것으로 판단되는 '독서', '사색', '자리 떠남'과 관련된 아이디어는 제외하였다. 최종 선정된 태스크 상황과 인터랙션 아이디어는 Table 1과 같으며, 본 연구에서 제안한 VUI의 인터랙션 시나리오는 이를 기반으로 작성하였다.

Table 1 19 Task situations

번호	7종류 데스크 액티비티	태스크 상황
1	자리 복귀	장시간(6시간 이상) 자리를 비운 후 자리에 복귀
2		단시간(2시간 이상, 6시간 미만) 자리를 비운 후 자리에 복귀
3	컴퓨터 작업	1시간 동안 거의 움직이지 않고 기계적으로 컴퓨터 작업을 함
4		컴퓨터 작업 중 펜으로 메모를 함
5		컴퓨터 작업 중 자정이 넘음
6	공부	1시간째 공부를 하고 있음
7		어두운 방 안에서 공부를 함
8		공부 중 자정이 넘음
9		공부 중 집중이 되지 않아 움직임이 많음
10	미팅 또는 회의	미팅/회의 중 대화 소리가 큼
11		미팅/회의 중 박수 소리가 남
12		미팅/회의 중 웃음소리가 남
13	잠	낮(10시~18시)에 30분 이상 책상에 엎드려(의자에 기대어) 잠
14		밤(23시~6시)에 30분 이상 책상에 엎드려(의자에 기대어) 잠
15	간식 섭취	간식 섭취 후 트림을 함
16	기타 활동	재채기를 함
17		기침을 함
18		움직임이 많음
19		한숨을 쉼

3. 3. 스마트 스피커 VUI의 능동적 반응 수준 정의

'능동적임'은 주관적인 지표이기에 개개인의 성향이나 상황에 따라 허용할 수 있는 능동적 반응의 수준은 다를 수 있다. 그러므로 VUI 시나리오를 개발하기에 앞서 능동적 대화를 시도하는 VUI 시스템의 능동적 반응에 대해 적절하게 그 수준을 구분할 필요가 있다. 그리고 구분된 능동적 반응의 수준에 따라 서로 다른 시나리오의 제공이 필요하며, 이에 대한 사용자들의 반응을 관찰하는 것이 중요하다.

본 연구에서는 사용자의 행동에 따른 스마트 스피커 VUI의 능동적 반응을 Figure 5와 같이 4단계로 구분하였다. 1단계는 무반응으로, 감지된 사용자의 행동에 대해 어떠한 인터랙션도 시도하지 않으며, 사용자의 호출에 의해서만 인터랙션을 시도하는 기존의 스마트 스피커가 여기에 해당한다. 2단계는 사용자의 행동에 대해 시스템으로부터 측정된 정보에 기반하여 단순 반응 인터랙션을 시도한다. 3단계는 시스템으로부터 측정된 정보에 기반하여 사용자의 태스크 상황을 예측하고, 그에 따른 적합한 사용자 행동을 제안하는 단계이다. 마지막으로 4단계는 3단계와 유사하게 사용자의 태스크 상황을 예측하는 한편, 태스크 상황에 따른 추천 행동을 제안하는 것에 그치지 않고 이를 직접 실행까지 하는 단계이다. 이렇듯 능동적 수준이 증가할수록 VUI는 더 능동적이고 적극적인 성격을 갖는다고 할 수 있다.

능동적 단계	인터랙션 유형
1단계	무반응
2단계	단순 반응
3단계	상황 예측 및 그에 따른 제안
4단계	능동적 대응(제안 사항 실행)

[예시]

능동적 단계	방 안이 어두운 상황에 대한 단계별 대응
1단계	무반응
2단계	방 안이 어둡다는 사실만을 알려줌
3단계	스탠드를 켜는 것을 제안
4단계	스탠드를 켜줌

Figure 5 Four levels of proactive interaction

3. 4. 능동적 반응 수준에 따른 VUI 시나리오 개발

Table 1의 각 태스크 상황별로 Figure 5의 능동적 인터랙션 단계를 적용하여 사용자에게 도움이 될 시나리오를 도출하였다. 각 태스크 상황별로 도출된 2~4단계의 능동적 반응 수준별 인터랙션의 내용은 Table 2와 같았고, 이를 구체화하여 시나리오 스크립트를 작성하였다. 예를 들면, Figure 6은 태스크 상황 2, 4, 13에 해당하는 시나리오 스크립트이며, 나머지 태스크 상황에 대한 시나리오 역시 동일한 톤으로 스크립트를 작성하였다.

Table 2 Proactive interactions for each task situations

태스크	데스크 액티비티	단계	능동적 수준별 인터랙션	태스크	데스크 액티비티	단계	능동적 수준별 인터랙션
1	자리 복귀	2	돌아왔다는 단순반응	11	미팅/ 회의	2	박수 소리가 난다는 단순반응
		3	긴 일정이 있었는지 추측			3	기쁜 일이 있는지 추측
		4	일정이 어땠는지 물어보고, 힐링 음악 재생			4	축하 송 재생
2		2	일찍 돌아왔다는 단순반응	12		2	웃음소리가 난다는 단순반응
		3	보일러와 가습기를 켜는 것을 제안			3	재밌는 일이 있는지 추측
		4	방안의 온·습도 안내 및 보일러와 가습기 작동			4	난센스 퀴즈
3	컴퓨터 작업	2	컴퓨터 작업 중이라는 단순 반응	13	잠	2	자고 있는지 물어봄
		3	스트레칭 추천			3	피곤한 상태인지 추측
		4	스트레칭 방법 안내 및 움직임 감지하여 칭찬의 말			4	알람 재생 및 일정 브리핑
4		2	뭔가 적고 있다는 단순반응	14		2	자고 있는지 물어봄
		3	음성 메모 기능을 추천			3	피곤한 상태인지 추측
		4	음성 메모 기능 실행			4	알람 재생 및 침대에서 편히 잘 것을 제안
5		2	자정이 지난 것을 알려줌	15	간식 섭취	2	트림을 했다는 단순반응
		3	자러 갈 것을 추천			3	속이 더부룩한지 추측
		4	평소 기상 시간에 맞춰 알람 설정			4	트림하는 이유를 소개
6	공부	2	공부 중이라는 단순반응	16	기타 활동	2	재채기를 했다는 단순반응
		3	집중력을 칭찬해 줌			3	실내 공기 환기를 추천
		4	집중에 도움 되는 ASMR 재생			4	실내 공기질이 나쁘다면 공기청정기 작동
7		2	방 안이 어두운 것을 알려줌	17		2	기침을 했다는 단순반응
		3	스탠드를 켜는 것을 제안			3	감기에 걸렸는지 추측
		4	스탠드 작동			4	감기에 걸렸는지 물어보고, 적정 습도로 유지되도록 가습기 작동
8		2	자정이 지난 것을 알려줌	18		2	분주하다는 단순반응
		3	공부할 것이 남았는지 물어봄			3	바쁜 일이 있는지 추측
		4	공부할 것이 남았는지 물어보고, 졸음 깨는 방법 안내			4	주요 뉴스 브리핑
9		2	움직임이 많다는 단순반응	19		2	한숨을 쉬었다는 단순반응
		3	집중이 되지 않는지 추측			3	스마트 스피커에 고민을 얘기해 볼 것을 추천
		4	명언을 인용해 집중하여 공부하도록 격려			4	격려의 말
10	미팅/ 회의	2	대화 소리가 크다고 알려줌				
		3	목소리를 낮출 것을 제안				
		4	목소리를 낮출 것을 제안하고, 잔잔한 음악 재생				

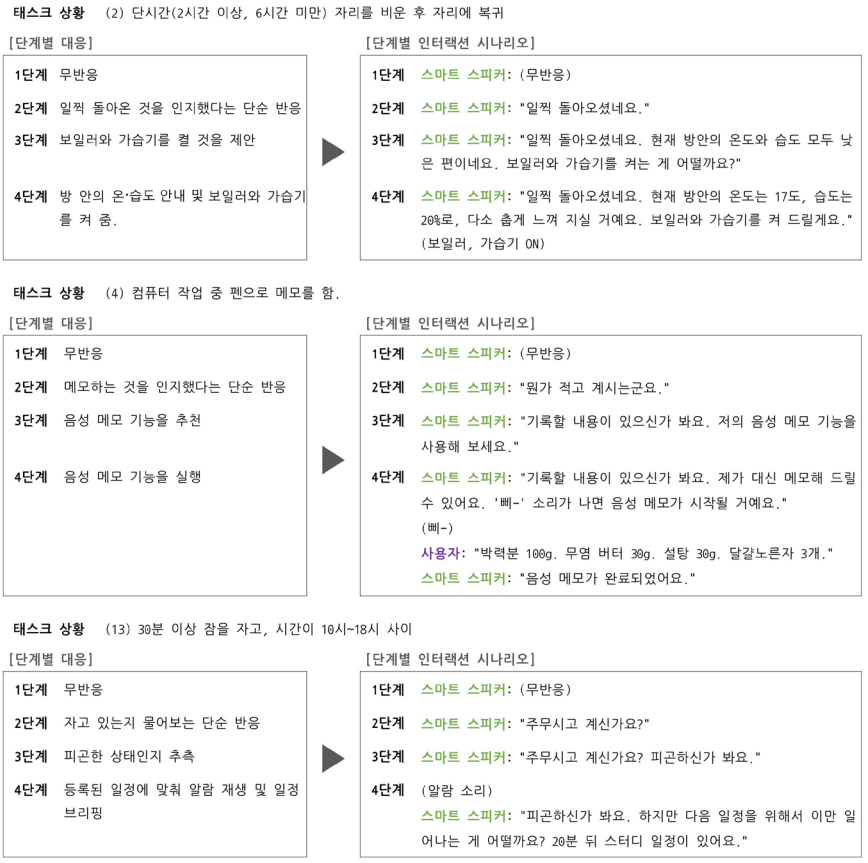

Figure 6 Example scenarios for task situation 2, 4, and 13

4. 능동적 인터랙션에 대한 선호도 평가

4. 1. 실험 목적

작성된 시나리오를 바탕으로 태스크 상황별로 어떤 수준의 능동적 인터랙션이 가장 선호되는지 알아보고자 사용자 평가를 수행하였다. 또한, 개인의 외향성(Extraversion)에 따라 선호하는 능동적 대응에 차이가 있는지도 검증하고자 하였다. 타푸스, 샤푸슈, 마타리치(Tapus, Ţăpuş & Matarić, 2008)와 나스, 이관민(Nass & Lee, 2000)에 따르면 사람들은 자신의 외·내향적 성향과 유사한 특성을 보이는 로봇 또는 인터페이스를 더 선호한다고 한다. 이를 바탕으로 능동적 반응 수준이 증가할수록 더 적극적으로 기능을 어필하거나 감성 대화를 시도하며, 수다스럽게 대응하는 시나리오에 대해서도 유사한 결과가 나타나는지 알아보고자 하였다.

4. 2. 실험참여자

실험에는 스마트 스피커의 주요 사용층인 20~30대 남녀 32명(남: 16명, 여: 16명)이 참여하였으며, 이들의 평균나이는 24.2세였다. 참여자들은 실험에 앞서 외향성을 측정할 수 있는 10개의 질문 문항에 답변하였다. 이 문항은 심리학계에서 인정받는 성격 검사지 중 하나인 HEXACO-PI-R의 60문항 설문지 중 외향성 항목에 해당하는 질문으로 구성하였다(Table 3). 모든 문항은 '매우 그렇다(5점)', '그런 편이다(4점)', '중간 정도(3점)', '그렇지 않은 편이다(2점)', '전혀 그렇지 않다(1점)'로 평가하였다. 이 설문지에 따르면 10개 질문 문항에 대한 점수의 평균치가 기준치인 3.51보다 크면 외향적인 성향을 나타내며, 기준치보다 낮으면 내향적인 성향을 나타낸다. 참여자의 응답에 따라 평균값을 계산하여 기준치와 비교한 결과, 외향적인 성향을 나타낸 참여자는 15명, 내향적인 성향을 나타낸 참여자는 17명이었다.

Table 3 Questionnaire for measuring extraversion in HEXACO-PI-R 60 items version

번호	질문
1	전반적으로 나 자신에 대해 만족하는 편이다.
2	단체 모임에서 나의 의견을 잘 나타내지 않는 편이다.
3	주로 혼자 하는 일보다는 다른 사람들과 적극적으로 상호작용하는 일을 더 좋아한다.
4	나는 거의 매일 명랑하고 낙천적인 편이다.
5	나는 별로 인기가 없는 편이라고 느낀다.
6	단체에서 나는 다른 사람의 눈치를 보지 않고 내 의견을 적극적으로 말한다.
7	새로운 환경에서 내가 제일 먼저 하는 일은 친구를 사귀는 것이다.
8	나는 다른 사람에 비해 별로 생기가 없고 활동적이지 않다.
9	나는 가끔 하찮은 인간이라고 생각할 때가 있다.
10	나는 종종 내가 속한 집단의 대변인 역할을 한다.

4. 3. 실험 환경 및 절차

실험은 WoZ(Wizard of Oz) 방식으로 진행하였다. 이를 위해 태스크 상황별로 데스크 위에서 일어나는 능동 인터랙션의 상황 연출을 위한 제품들(예. 공기청정기, 열풍기, 가습기, 스탠드 등)을 Figure 7에서 보여주듯이 책상 위에 두었다. 그리고 3.4절에서 설명한 VUI 시나리오는 TTS(Text to Speech)로 미리 제작하여 상황에 따라 컴퓨터로 재생할 수 있도록 준비하였다.

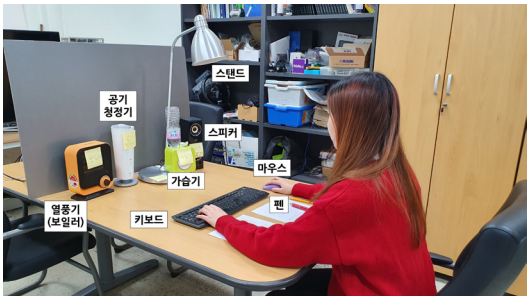

Figure 7 Environment of user test

실험 절차는 다음과 같았다. 먼저 참여자들에게 자리 복귀, 컴퓨터 작업, 공부, 미팅 또는 회의, 잠, 간식 섭취, 기타의 7종류 데스크 액티비티에 대해 충분히 설명한 후 순서의 영향을 받지 않도록 이를 임의의 순서로 제시하였다. 그 후 각 태스크 상황별로 인터랙션이 재현될 수 있는 태스크를 부여하여 이를 인터랙션 시나리오대로 수행하도록 요구하였다. 예로, 태스크 상황 7번의 공부 중에 조도가 낮은 상황을 재연하기 위해 실험실의 불을 끈 채, 참여자에게 산수 문제를 풀며 공부를 하는 중이라고 상상할 것을 요청하였다. 이렇듯 참여자가 부여된 태스크를 수행하는 중에 네 수준 중 하나의 능동적 인터랙션 유형에 해당하는 TTS를 재생하였다. 특히 태스크 상황 1, 3, 4, 8의 인터랙션 시나리오는 사용자의 답변이 요구되는 시나리오였으므로, 이 경우 참여자들에게 정해진 시나리오대로 발화 혹은 행동하도록 요구하였다. 네 수준의 인터랙션이 있으므로 참여자들은 동일한 태스크를 네 번씩 반복하였다. 참여자들은 네 종류의 인터랙션 유형을 모두 경험한 후에 능동적 반응 수준에 대한 선호도를 1~4위로 평가하였으며, 순위 결정의 이유를 기술하였다.

Table 4 The ANOVA result of preference for each task situation

		1 장시간 자리를 비운 후 자리에 복귀		2 단시간 자리를 비운 후 자리에 복귀		3 1시간 동안 거의 움직이지 않고 컴퓨터 작업을 함		4 컴퓨터 작업 중 펜으로 메모를 함		5 컴퓨터 작업 중 자정이 넘음		6 1시간 째 공부를 하고 있음		7 어두운 방안에서 공부를 함	
인자	자유도	F-값	p-값	F-값	p-값	F-값	p-값	F-값	p-값	F-값	p-값	F-값	p-값	F-값	p-값
능동수준	3	47.60	0.000**	44.24	0.000**	25.76	0.000**	8.47	0.000**	16.92	0.000**	10.02	0.000**	5.75	0.001**
외향성	1	0.00	1.000	0.00	1.000	0.00	1.000	0.00	1.000	0.00	1.000	0.00	1.000	0.00	1.000
성별	1	0.00	1.000	0.00	1.000	0.00	1.000	0.00	1.000	0.00	1.000	0.00	1.000	0.00	1.000
능동수준 x 외향성	3	1.01	0.390	0.74	0.531	6.20	0.001**	2.61	0.054	2.24	0.087	1.03	0.383	0.70	0.552
능동수준 x 성별	3	0.56	0.642	1.52	0.213	1.50	0.218	0.23	0.876	1.54	0.207	0.41	0.744	0.95	0.421
오차	116														
총계	127														

		8 공부 중 자정이 넘음		9 공부 중 집중이 되지 않아 움직임이 많음		10 미팅/회의 중 대화 소리가 큼		11 미팅/회의 중 박수 소리가 남		12 미팅/회의 중 웃음 소리가 남		13 낮에 30분 이상 책상에 엎드려 잠		14 밤에 30분 이상 책상에 엎드려 잠	
인자	자유도	F-값	p-값	F-값	p-값	F-값	p-값	F-값	p-값	F-값	p-값	F-값	p-값	F-값	p-값
능동수준	3	0.48	0.698	1.76	0.158	2.25	0.086	2.95	0.036*	18.00	0.000**	10.44	0.000**	3.66	0.014*
외향성	1	0.00	1.000	0.00	1.000	0.00	1.000	0.00	1.000	0.00	1.000	0.00	1.000	0.00	1.000
성별	1	0.00	1.000	0.00	1.000	0.00	1.000	0.00	1.000	0.00	1.000	0.00	1.000	0.00	1.000
능동수준 x 외향성	3	1.10	0.352	0.30	0.828	0.55	0.651	2.46	0.066	1.24	0.298	0.73	0.536	0.46	0.711
능동수준 x 성별	3	0.90	0.446	0.61	0.607	1.92	0.131	1.31	0.275	2.02	0.115	2.93	0.037*	0.51	0.677
오차	116														
총계	127														

		15 간식 섭취 후 트림을 함		16 재채기를 함		17 기침을 함		18 움직임이 많음		19 한숨을 쉼	
인자	자유도	F-값	p-값	F-값	p-값	F-값	p-값	F-값	p-값	F-값	p-값
능동수준	3	15.22	0.000**	19.43	0.000**	15.65	0.000**	16.17	0.000**	9.28	0.000**
외향성	1	0.00	1.000	0.00	1.000	0.00	1.000	0.00	1.000	0.00	1.000
성별	1	0.00	1.000	0.00	1.000	0.00	1.000	0.00	1.000	0.00	1.000
능동수준 x 외향성	3	0.86	0.462	1.38	0.253	1.22	0.306	1.09	0.358	0.17	0.917
능동수준 x 성별	3	0.51	0.677	1.86	0.140	0.98	0.403	0.12	0.947	0.65	0.583
오차	116										
총계	127										

5. 실험 결과

실험 데이터는 유의수준 α=5%로 ANOVA(Analysis of Variance)와 Tukey 사후분석을 수행하였다. Table 4는 각 태스크 상황별 선호도에 대한 통계분석 결과이다. 태스크 상황 8, 9, 10을 제외한 모든 상황에서 네 가지 능동적 반응 수준에 따른 선호도에 유의한 차이를 보였다. 그러나 사용자의 외향성과 성별에 따라서는 선호도에 유의한 차이가 없었다.

	1** 장시간 자리를 비운 후 자리에 복귀	2** 단시간 자리를 비운 후 자리에 복귀	3** 1시간 동안 거의 움직이지 않고 컴퓨터 작업을 함	4** 컴퓨터 작업 중 펜으로 메모를 함	5** 컴퓨터 작업 중 자정이 넘음	6** 1시간 째 공부를 하고 있음	7** 어두운 방안에서 공부를 함	8 공부 중 자정이 넘음	9 공부 중 집중이 되지 않아 움직임이 많음	10* 미팅/회의 중 대화 소리가 큼
무반응	3.72	3.66	3.28	2.88	3.44	1.66	3.00	2.66	2.66	2.44
단순 반응	2.66	2.84	3.03	3.03	2.59	2.56	2.78	2.34	2.63	2.53
예측 및 제안	2.16	1.69	1.69	1.89	2.00	2.88	2.13	2.50	2.09	2.16
능동적 대응	1.47	1.81	2.00	2.22	1.97	2.91	2.09	2.50	2.63	2.88
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.698	0.158	0.086

	11** 미팅/회의 중 박수 소리가 남	12** 미팅/회의 중 웃음 소리가 남	13** 낮에 30분 이상 책상에 엎드려 잠	14** 밤에 30분 이상 책상에 엎드려 잠	15 ** 간식 섭취 후 트림을 함	16 ** 재채기를 함	17 ** 기침을 함	18** 움직임이 많음	19** 한숨을 쉼
무반응	2.31	1.72	2.53	2.72	1.59	2.97	3.00	2.00	2.78
단순 반응	2.97	2.53	2.84	2.69	2.75	3.28	3.09	2.25	3.16
예측 및 제안	2.25	2.31	2.94	2.66	2.44	1.91	2.28	2.22	1.97
능동적 대응	2.47	3.44	1.69	1.94	3.22	1.84	1.63	3.53	2.09
p-value	0.036	0.000	0.000	0.014	0.000	0.000	0.000	0.000	0.000

Figure 8 Overall results of user test
(The use of the same alphabetic characters indicates that there was no significant difference when α=0.05 according to Tukey test)

5. 1. 태스크 상황별 능동적 인터랙션의 선호도

19개의 태스크 상황별로 각 인터랙션 유형에 대한 선호 순위는 Figure 8과 같았다. 대부분의 태스크 상황에서 스마트 스피커가 3단계 또는 4단계의 능동적 인터랙션을 시도하는 것이 무반응보다 높은 선호도 순위를 보였다. 2단계의 단순 반응 인터랙션에 대한 선호도의 경우 무반응 인터랙션과 평균 순위는 유사했으나 정성적 평가에서는 더 부정적인 것으로 나타났다. 대부분의 태스크 상황에서 참여자들은 단순 반응 인터랙션이 아무런 도움이 되지 않으며, 기분이 나쁘고 거부감이 든다는 의견을 남겼다.

(1) 정서적 교류에 대한 기대

자리 복귀에 해당하는 1과 2번 태스크 상황에서는 무반응과 3·4단계 인터랙션 간의 선호도 격차가 가장 컸다. 3·4단계의 능동적 인터랙션에 대해 사용자의 안부를 물어보기 때문에 정서적으로 안정감이 들었다는 긍정적 평가가 많았다. 15번 태스크 상황에서 트림과 같은 생리현상에 반응하는 것은 민망하다는 참여자들의 평가에 따라 무반응이 가장 순위가 높게 나왔다. 하지만 일부 참여자들의 경우 속이 더부룩하진 않은지 물어보는 3단계 인터랙션 시나리오에 대해 적절한 정도의 관심이며, 걱정해 주는 느낌이라고 평가하였다. 이러한 평가로

미루어 보아 사용자들은 스마트 스피커와 정서적 교류를 기대한다고 볼 수 있는데, 이는 김준한, 김유정, 김병준, 윤수경, 김민준, 이중식(Kim, Kim, Kim, Yun, Kim & Lee, 2018)이 청소년을 대상으로 한 연구에서 스마트 스피커 에이전트에 대해 언제든 참을성 있게 대화를 들어주며, 단순히 경청하는 것뿐만 아니라 격려를 해줄 것을 기대했다는 결과에서도 나타난다.

(2) 콘텍스트 중단에 대한 우려

공부와 미팅/회의에 해당하는 6~12번 태스크 상황에서는 네 수준의 인터랙션 유형 간 선호도 순위의 격차가 적거나 무반응 인터랙션이 더 높은 순위로 나타나기도 하였다. 공부에 해당하는 태스크 상황의 경우 참여자들은 공부에 몰입하고 있기에 스마트 스피커가 말을 거는 것은 오히려 방해라고 평가하였다. 또한, 미팅/회의와 같이 여러 명이 함께 있는 상황에서도 태스크 상황에 방해되거나 대화의 흐름이 끊기는 것을 우려하는 평가가 많았다. 따라서 사용자가 특정 태스크에 몰입하고 있는 경우이거나 여럿이 함께 일을 하는 상황에서는 스마트 스피커의 개입은 최소화되는 것이 좋다.

(3) 추천 사항의 제안 vs. 실행

VUI 시스템이 보일러 켜기나 스트레칭과 같은 추천 사항을 단순히 제안만 하는 것과 실행까지 직접 해주는 인터랙션에 대해서는 참여자들의 의견이 분분하였다. 참여자 대부분은 추천 사항을 실행해 주는 4단계의 능동적 대응 인터랙션에 대해 편리하기는 하지만 실행을 원치 않아 취소해야 하는 경우가 있기에 제안 정도가 적절하며, 사용 여부에 대한 의사를 사용자 스스로 결정하는 것이 좋다는 의견을 제시하였다. 그러나 태스크 상황 16, 17의 공기 질, 기침과 같이 건강과 관련 있는 경우에는 추천 사항을 직접 실행해 주는 것을 더 선호하였다. 태스크 상황 3의 능동적 인터랙션에 대해서도 확실한 스트레칭 유도가 좋았다는 의견과 더불어 건강(피로 누적)과 관련된 것이기에 적극적으로 개입해도 기분이 나쁘지 않았다는 의견이 있었다.

5. 2. 성향에 따른 인터랙션 선호도의 차이

Figure 9와 같이 태스크 상황 3, 4, 5, 11에서는 참여자들의 성향과 인터랙션 유형 간 교호작용에 유의한 차이가 존재하였다. 교호작용도를 살펴보면 공통적으로 내향적 성향의 참여자들은 4단계의 능동적 대응 인터랙션보다 3단계의 예측 및 제안 인터랙션을 더 선호하였고, 외향적 성향의 참여자들은 반대의 모습을 보였다. 익숙지 않은 공간에서 실험 진행자와 실험이 진행되는 실험 환경상 내향적 성향의 참여자들은 의자에서 일어나 스트레칭하거나 소리 내어 음성 메모를 하는 실험 태스크에 부담을 느꼈을 수 있다. 이러한 실험 환경상의 요인과 더불어 이는 문요한, 김기준, 신동희(Moon, Kim & Shin, 2016)가 소개한 유사성 매력(similarity attraction)에 의한 결과라 할 수 있다. 이 이론에 따르면 민족성, 정치적 성향 그리고 성격 등은 다른 사람들이 자신과 유사하다고 느끼게 만드는 요소이며, 사람들은 자신과 유사한 상대방에게 더 호감을 느끼고 소통하고 싶어 한다. 즉, 외향적 성향의 사용자들은 더 외향적이면서 능동적인 스마트 스피커에 더 호감을 느꼈을 것이다.

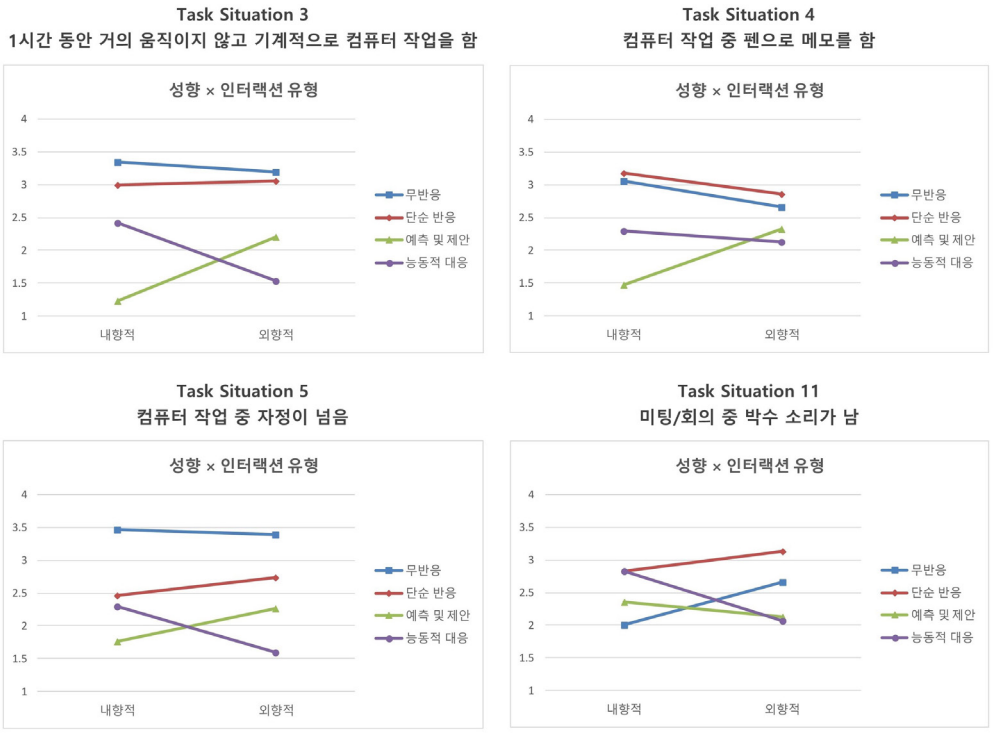

Figure 9 Interaction effects between extraversion and proactive interaction type for task situation 3, 4, 5, 11

6. 결론 및 고찰

본 연구에서는 사용자의 데스크 액티비티를 자동으로 인지하고, 인지된 태스크 상황에 따라 능동적 인터랙션을 시도하는 스마트 스피커의 유용성에 대한 연구를 수행하였다. 데스크 액티비티를 19가지의 태스크 상황으로 구체화하였고, 인터랙션을 능동적 수준에 따라 무반응, 단순 반응, 예측 및 제안, 능동적 대응의 4단계로 구분하여 단계별로 유용할 것으로 예상되는 VUI 시나리오 스크립트를 도출하였다. 작성된 스크립트는 TTS로 제작하여 WoZ 환경에서 태스크 상황별 능동적 인터랙션의 선호도를 평가하는 실험을 진행하였다. 그 결과, 대부분의 태스크 상황에서 예측 및 제안(3단계) 또는 능동적 대응(4단계) 인터랙션이 무반응(1단계)보다 선호도 순위가 높은 것으로 나타났다. 이를 보면, 사용자의 태스크 맥락과는 상관없이 어떤 반응도 시도하지 않는 현재의 스마트 스피커보다 먼저 말을 걸어주고 대화를 시도하는 능동적 성격의 스마트 스피커를 사용자들은 더 긍정적으로 생각한다고 볼 수 있다. 한편, 단순 반응(2단계) 인터랙션은 대부분의 태스크 상황에서 부정적으로 평가되었다. 태스크 상황 3, 4, 5, 11에서는 참여자들의 성향과 인터랙션 유형 간 교호작용에 유의한 차이가 존재하였다. 유사성 매력에 따라 외향적 성향의 참여자들은 능동적 대응 인터랙션을 예측 및 제안 인터랙션보다 더 선호하였고, 내향적 성향의 참여자들은 이와 반대되는 결과를 보였다.

본 연구의 결과는 능동적인 스마트 스피커 개발 시 반응 시나리오와 시나리오별 적절한 반응의 수준을 결정하는 데 활용될 수 있다. 본 연구는 대화 시작 후 1~2회 내의 짧은 음성 인터랙션에 대한 시나리오만을 대상으로 하였다. 여러 단계에 걸친 긴 음성 인터랙션에 관한 사용자 반응에 대해서는 향후 추가적인 연구가 필요할 것으로 보인다. 더불어 본 연구는 능동적 인터랙션의 장기적 사용에 대한 사용자 반응은 고려하지 못한 한계가 있다. 초기에는 긍정적으로 평가한 능동적 인터랙션이라 하더라도 장기적 관점에서의 유용성에 대한 논의가 필요하고, 능동적 인터랙션의 콘텐츠 또한 지속적으로 개발되어 다양화될 필요성이 있다.

References

1. Agarwal, A., Jain, M., Kumar P., & Patel, S. (2018). Opportunistic Sensing with MIC Arrays on Smart Speakers for Distal Interaction and Exercise Tracking. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Calgary, AB, pp. 6403-6407
2. Araujo, T. (2018). Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Computers in Human Behavior, 85*, 183-189.
3. Ashton, M. C., & Lee, K. (2009). The HEXACO-60: A short measure of the major dimensions of personality. *Journal of Personality Assessment, 91*, 340-345.
4. Dey, A. K., & Abowd, G. D. (2000). Towards a better understanding of context and context-awareness. *Proceedings of the What, Who, Where, When, and How of Context-Awareness Workshop*, CHI 2000 Conference on Human Factors in Computer Systems. New York: ACM.
5. Gemmeke, J. F., Ellis, D. P. W., Freedman, D., Jansen, A., Lawrence, W., Moore, R. C., Plakal, M., & Ritter, M. (2017). Audio set: An ontology and human-labeled dataset for audio events. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 776-780). IEEE.
6. Goetz, J., Kiesler, S., & Powers, A. (2003). Matching robot appearance and behavior to tasks to improve human-robot cooperation. In *Robot and Human Interactive Communication 2003*, 55-60.
7. Hong, E., Cho, K., & Choi, J. (2017). 스마트홈 대화형 인터페이스의 의인화 효과: 음성-채팅 인터랙션 유형에 따른 실험 연구[Effects of Anthropomorphic Conversational Interface for Smart Home : An Experimental Study on the Voice and Chatting Interactions]. *한국 HCI 학회 논문지[Journal of the HCI Society of Korea], 12*(1), 15-23.
8. Kim, H., Ko, J., Kim, S., & Hwang, H. (2019). 스마트스피커의 보이스와 외형 개성이 사용자 만족도에 미치는 영향[Effects of voice and appearance personality of smart speaker on user satisfaction]. *Journal of the Ergonomics Society of Korea, 38*(6), 527-541.
9. Kim, H., Ko, J., Kim, S., & Hwang, H. (2020). 청자의 지루함: 스마트스피커에 적합한 음성정보의 길이는?[Listeners' boredom: how long can the voice information of a smart speark be?]. *Archives of Design Research, 33*(1), 151-163.
10. Kim, J., Kim, Y., Kim, B., Yun, S., Kim, M., & Lee, J. (2018). Can a Machine Tend to Teenagers' Emotional Needs?: A Study with Conversational Agents. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (p. LBW018). ACM.
11. Kinsella, B., & Mutchler, A. (2019). Smart Speaker Consumer Adoption Report 2019. https://voicebot.ai/wp-content/uploads/2019/03/smart_speaker_consumer_adoption_report_2019.pdf (accessed 13/11/19).
12. Kong, Q., Xu, Y., Wang, W., & Plumbley, M. D. (2018). Audio set classification with attention model: A probabilistic perspective. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 316-320). IEEE.
13. Krishnan, N. C., & Cook, D. J. (2014). Activity recognition on streaming sensor data. *Pervasive and Mobile Computing, 10*, 138-154.
14. Lee, C., & Helal, S. (2003). Context Attributes: An Approach to Enable Context-awareness for Service Discovery. In *Proceedings of the Symposium on Applications and the Internet (SAINT)*, 22-30.
15. Lee, J., Kim, K., Lee, S., & Shin, D. (2015). Can autonomous vehicles be safe and trustworthy? Effects of appearance and autonomy of unmanned driving systems. *International Journal of Human-Computer Interaction, 31*, 682-691.
16. Liao, Q. V., Davis, M., Geyer, W., Muller, M., & Shami, N. S. (2016). What can you do?: Studying social-agent orientation and agent proactive interactions with an agent for employees. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems* (pp. 264-275). ACM.
17. Lu, C., & Fu, L. (2009). Robust location-aware activity recognition using wireless sensor network in an attentive home. *IEEE Transactions on Automation Science and Engineering, 6*(4), pp. 598-609.

18. Moon, Y., Kim, K. J., & Shin, D. H. (2016). Voices of the internet of things: An exploration of multiple voice effects in smart homes. In *International Conference on Distributed, Ambient, and Pervasive Interactions* (pp. 270–278). Springer, Cham.
19. Nass, C., & Lee, K. (2000) Does computer-generated speech manifest personality? an experimental test of similarity-attraction. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 329–336). ACM.
20. Nass, C., Steuer, J., & Tauber, E. (1994). Computers are social actors. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 72–78). ACM.
21. Nguyen, Q. N., Ta, A., & Prybutok, V. (2019). An integrated model of voice-user interface continuance intention: The gender effect. *International Journal of Human–Computer Interaction, 35*(15), 1362–1377.
22. Reeves, B., & Nass, C. I. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge university press.
23. Shi, Y., Yan, X., Ma, X., Lou, Y., & Cao, N. (2018). Designing Emotional Expressions of Conversational States for Voice Assistants. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems* (pp. 1–6). Montréal, Canada.
24. Tapus, A., Ţăpuş, C., & Matarić, M. J. (2008). User–robot personality matching and assistive robot behavior adaptation for post-stroke rehabilitation therapy. *Intelligent Service Robotics, 1*(2), 169–183.
25. Yu, C., Barsim, K. S., Kong, Q., & Yang, B. (2018). Multi-level attention model for weakly supervised audio classification. *arXiv preprint arXiv:1803.02353*.

사용자의 데스크 액티비티에 따른 스마트 스피커의 능동적 인터랙션 연구

김헌[1], 이소향[2*]

[1]서울과학기술대학교 기계시스템디자인공학과, 교수, 서울, 대한민국

[2]서울과학기술대학교 나노IT디자인융합대학원 디자인기술융합전공, 학생, 서울, 대한민국

초록

연구배경 사용자들은 높은 의인화 속성을 가진 지능형 에이전트를 더 긍정적으로 평가하는 경향을 가진다. 하지만 현재 스마트 스피커와의 대화는 사용자의 일방적 호출어 발화에 의해 대화가 시작되며, 대화 시 사용자의 상황에 대한 정보나 취향 등은 고려되지 않는다. 이는 사람-사람 간의 일반적인 의사소통 특성과는 다르다. 이에 본 연구에서는 사용자의 데스크 액티비티에 초점을 두고, 이를 인지할 수 있는 시스템과 인지된 정보를 이용하여 능동적인 대화를 수행하는 스마트 스피커에 대해 능동적 반응의 수준별 사용자 선호도를 평가하였다.

연구방법 먼저, 마이크와 사운드 센서, 인체 감응 센서, 조도 센서를 이용한 딥러닝 기반 모델과 규칙 기반 모델을 거쳐 사용자의 데스크 액티비티를 판별하는 시스템 콘셉트를 정의하였다. 그리고 이 시스템으로 판단 가능한 태스크 상황을 19가지로 구분하였고, 인터랙션의 능동적 수준을 무반응, 단순 반응, 상황 예측 및 그에 따른 제안, 그리고 제안사항을 실행하는 능동적 대응의 네 단계로 구분하여 구체적인 시나리오를 도출하였다. 그 후, 태스크 상황별 각 태스크 시나리오에 대한 사용자 평가를 통해 어떤 수준의 능동적 인터랙션이 더 선호되는지를 평가하였다.

연구결과 대부분의 태스크 상황에서 상황 예측 및 제안과 능동적 대응 인터랙션이 무반응보다 더 선호되는 것으로 나타났으며, 단순 반응은 부정적으로 평가되었다. 특히 실험참여자들은 스마트 스피커와의 정서적 교류를 기대하며, 특정 태스크에 몰입하고 있거나 여럿이 함께 있는 상황에서는 콘텍스트 중단에 대해 우려하는 것으로 나타났다.

결론 사용자의 콘텍스트에 따른 스마트 스피커의 능동적 대화 시도는 일정 수준 이상의 인지 정확도가 뒷받침된다면 매우 유용할 것이며, 이는 확장된 사용자 경험을 제공할 것으로 기대된다.

주제어 스마트 스피커, 음성 사용자 인터페이스, 데스크 액티비티, 콘텍스트 인지, 능동적 인터랙션