

# Interacting with Smart Speakers based on Finger Touch Gestures

Huhn Kim<sup>1</sup>, Hojeong Im<sup>2\*</sup>

<sup>1</sup>Department of Mechanical System Design Engineering, Professor, Seoul National University of Science and Technology, Seoul, Korea

<sup>2</sup>Department of Design and Engineering, Student, Seoul National University of Science and Technology, Seoul, Korea

---

## Abstract

**Background** The smart speaker market is growing significantly every year, but its usability still does not meet the expectations of users. Voice interaction with existing smart speakers requires a pre-process consisting of a user's call and the smart speaker's response on the call, which has a limitation in that it is difficult to control errors immediately. Therefore, a new interaction is required that allows for more immediate error control while shortening the pre-call process of the existing smart speaker.

**Methods** First, twelve unit operations essential for using the main functions of the smart speaker were derived, and various touch gestures posed by the user's one hand were devised. Second, a survey was conducted to decide a suitable gesture for each unit operation, and a smart speaker interaction based on the finger touch gestures was designed. Third, in order to evaluate the ease of learning and usability of the designed gesture interaction, usability evaluation was performed using the Wizard of Oz technique.

**Results** As a result of the usability evaluation, the participants rated the use of finger touch gestures as useful, and almost all gestures were evaluated for fast learning and high usability. However, the one-tap gesture and the rubbing gesture on the top were found to have learning and usability problems, which needed improvement.

**Conclusions** Although there is a limitation that the touch gesture can be used only when the smart speaker is in a position close to the user, such as at a bedside or on a desk, it can be operated very quickly without a pre-call process. Thus, the touch gesture is expected to be applied to smart speakers in the future.

**Keywords** Smart Speaker, Finger Touch, Gesture Design, Multimodal Interaction

---

This work was supported by the Research Program funded by Seoul National University of Science and Technology and published based on Master's thesis of the corresponding author in Seoultech 2021.

\*Corresponding author: Hojeong Im (424555@naver.com)

*Citation:* Kim, H., & Im, H. (2021). Interacting with Smart Speakers based on Finger Touch Gestures. *Archives of Design Research*, 34(4), 157-171.

<http://dx.doi.org/10.15187/adr.2021.11.34.4.157>

**Received :** Jun. 25. 2021 ; **Reviewed :** Aug. 22. 2021 ; **Accepted :** Aug. 22. 2021

**pISSN** 1226-8046 **eISSN** 2288-2987

**Copyright :** This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits unrestricted educational and non-commercial use, provided the original work is properly cited.

## 1. 서론

스마트 스피커란 음성인식과 인공지능 기술을 결합하여 다양한 서비스를 제공하는 인공지능 비서이다. 과학정보통신부(2019)에 따르면 2019년 591만대에서 2020년 861만대로 스마트 스피커의 판매량은 45.7% 증가했으며 사용자 수가 증가하는 추세이다. 현재 상용화된 스마트 스피커의 인터랙션은 음성인식 및 발화를 중심으로 하는 음성 인터랙션을 기본으로, 터치 디스플레이, 음장 센서, 적외선 센서, 온습도 센서, 그리고 카메라 등을 이용한 다양한 인터랙션 방식을 함께 제공하고 있다. 보이스봇(2020)의 조사에 따르면 스마트 스피커의 주된 설치 위치는 거실(43.2%), 침실(45.5%), 주방(41.5%), 그리고 홈 오피스(11.5%)라 한다. 2019년 조사 결과와 비교하였을 때 거실에 설치한 비율은 44.4%에서 43.2%로 감소하였지만, 침실은 37.5%에서 45.5%로, 주방은 32.7%에서 41.5%로 증가하였다. 이러한 공간별 설치 위치의 변화는 스마트 스피커가 침실, 주방과 같은 독립된 공간에서 더욱더 많이 사용되고 있음을 의미한다. 침실의 침대 머리맡, 주방의 싱크대나 식탁 위, 그리고 오피스의 책상 위 등에 놓여 있는 스마트 스피커는 사용자가 음성 명령과 더불어 손을 뻗어서 직접 인터랙션하기도 용이할 수 있다.

하지만, 컨슈머인사이트(2019)에 따르면 스마트 스피커 사용자들의 사용 만족도는 45%로 절반에도 미치지 못하는 것으로 조사되었다. 사용자들은 스마트 스피커와 자연스러운 대화가 가능하지 않다는 점을 낮게 평가하였다. 한국소비자원(2017)에 따르면, 스마트 스피커의 주된 불편사항으로는 일상사용 환경에서의 음성인식 미흡(56.7%), 기기와 사용자 간의 자연스러운 연속 대화 곤란(45.7%), 그리고 소음을 음성 명령으로 인식하는 작동오류 발생(37.0%) 등이 꼽혔다. 영어가 모국어인 사용자와 그렇지 않은 사용자 간 구글 홈 사용경험 차이를 조사한 파예, 시플릿(Pyae & Scifleet, 2019)에 따르면, 언어의 유창함이 스마트 스피커와의 긍정적 사용자 경험에 영향을 미친다고 한다. 이러한 문제들은 포처론, 어민, 리브, 샤프스(Porcheron, Fischer, Reeves & Sharples, 2018)와 파이, 조엘슨(Pyae & Joelsson, 2018)의 아마존 에코와 구글 홈 사용자에게 대한 실사용 행태 조사에서도 비슷하게 나타났다. 웨이, 랜데이(Wei & Landay, 2018), 박과 이(Park & Lee, 2020)는 물건 구입이나 음식 배달과 같은 문장형 질문이나 대화가 2~3번 연속적으로 이어져야 하는 경우 사용자가 스마트 스피커를 활용하는 데 한계를 느낀다고 하였다. 또한, 신종규(Shin, 2020)에 따르면 사용자가 발화를 잘못했을 때 이를 즉각적으로 수정할 수 없는 점도 문제라고 하였다. 이와 같은 불편사항은 발화를 기반으로 한 음성 인터랙션의 특성과 기술적 한계로 인해 발생하는 문제점으로 스마트 스피커의 사용성을 저하시켜 사용자가 사용을 중단하게 만드는 동기가 될 수 있다.

일반적인 스마트 스피커의 명령 처리 과정은 사용자의 음성 호출과 그에 대한 스마트 스피커의 응답으로 이루어진 ‘사전 과정’을 반드시 필요로 한다. 예를 들어, SKT의 스마트 스피커인 NUGU는 사용자가 “아리아~”라는 호출어를 부르면 NUGU는 “땡~”하는 호출 응답을 하며 다음 명령을 대기한다. 이는 주변 소음이나 사람들 간의 대화 소리와 스마트 스피커를 향한 음성 명령을 구분하기 위해 필요한 과정이다. 하지만 이러한 ‘사전 과정’은 반복적이고 단순한 명령을 하거나 수행한 명령을 반복하고자 할 때에도 항상 거쳐야 해서 스마트 스피커의 사용 과정을 매우 번거롭게 만든다. 장순규(Jang, 2018)에 따르면, 매 명령마다 스마트 스피커를 반복하여 호출하는 대화 방식에 대해 70%의 사용자들이 부정적으로 응답하였다고 한다. 특히, 스마트 스피커는 사람들 간의 대화소리, TV나 라디오 등 근처에 있는 다른 음성 디바이스를 통해서 의도치 않게 호출되는 문제도 있다. 심지어 TV에서 나온 유사한 호출 음성에 의해 아마존 에코가 의도치 않은 상품 주문을 시도했다는 보고도 있었다(버지 Verge, 2017). 알란와르, 발라지, 티안, 양, 스리바스타바(Alanwar, Balaji, Tian, Yang & Srivastava, 2017)는 이러한 문제를 해결하고자 소나(Sonar) 기술을 이용하여 스마트 스피커 근처에 사용자가 있을 경우에만 발화되는 시스템을 제안하였다. 맥밀란, 브라운, 카와구치, 자베르, 벨렝게르, 쿠프스오카(McMillan, Brown, Kawaguchi, Jaber, Belenguer & Kuzuoka, 2019)는 호출어 대신 스마트 스피커를 바라보는 사용자의 시선을 이용하여 호출어 반복 발화의 번거로움을 해소하고 자연스러운 연속 대화를 가능하게 하는 Tama란 스마트 스피커 플랫폼을 개발하였다. 포미칼스키, 워니악, 워니악,

그루지야, 자오, 로마노프스키(Pomykalski, Woźniak, Woźniak, Grudzień, Zhao & Romanowski, 2020)는 사용자 조사를 통해 스마트 스피커 호출어 대신에 사용가능한 다섯 가지 제스처로 박수 한번, 박수 두 번, 스냅(Snap), 오른쪽 스와이프(Swipe), 손 흔들기 제스처를 제안하였고 스냅과 손 흔들기 제스처가 가장 선호됨을 보였다.

한편, 일반적인 제어장치의 조작 어려움이나 한계를 극복하고자 음성과 제스처를 활용한 멀티모달 인터랙션에 대한 연구들도 많이 수행되었다. 라이즈, 알소스(Rise & Alsos, 2020)에 따르면 제스처 기반 인터랙션은 TV, 모바일폰, 드론 등의 조작에 폭넓게 활용되고 있다. 안바라산, 이(Anbarasan & Lee, 2018)는 음성과 비접촉 제스처를 통한 스마트홈 제어가 고령자에게 매우 유용함을 보였다. 반면, 호프만, 타이롤러, 웬드, 헨제(Hoffmann, Tyroller, Wende & Henze, 2019)는 스마트홈의 컨트롤러로 음성 명령이나 터치 디스플레이를 활용하는 것이 비접촉 제스처보다 더 선호됨을 보였다. 반 덴 호벤, 마 잘렉(van den Hoven & Mazalek, 2011)은 디지털 기기의 제스처 조작은 다른 일반적인 조작방식에 비해 덜 제약적이고 더 직관적인 인터랙션을 가능하게 해주기 때문에 디지털 기기와의 커뮤니케이션을 더 수월하게 해준다고 하였다. 스마트 스피커를 사용하는 사용자의 경우에도 음성 명령과 더불어 스마트 스피커가 가까이 있을 때는 그냥 툭 치거나 쓰다듬는 등의 인터랙션으로 특정 명령을 수행하고 싶을 수 있다.

이에 본 연구에서는 스마트 스피커의 사전 호출 과정을 단축시켜 주는 새로운 핑거터치(Finger touch) 제스처 기반의 스마트 스피커 인터랙션을 제안하였다. 핑거터치 제스처란 사용자가 스마트 스피커를 손가락으로 접촉한 상태에서 특정 제스처를 취하면 해당 제스처에 매핑되어 있는 기능을 스마트 스피커가 수행해주는 인터랙션이다. 사용자는 음성으로 명령을 수행하거나 핑거터치 제스처를 이용하여 스마트 스피커와 인터랙션을 수행할 수 있다. 이 때, 사용자의 핑거터치 제스처는 터치 센서나 카메라와 같은 하드웨어를 이용하여 인식하거나 스마트 스피커에 제스처를 취할 때 발생하는 소리를 딥러닝으로 분류하는 방법 등으로 인식이 가능하다. 스마트 스피커가 침실 머리맡, 주방의 식탁 위, 사무실 책상 위 등 사용자의 손에 쉽게 닿는 위치에 있을 경우에는 Figure 1과 같이 핑거터치 제스처가 호출어를 대신함으로써 보다 신속하게 스마트 스피커를 호출할 수 있어 사전 호출 응답의 과정을 생략하고 바로 음성 명령을 수행할 수 있다. 또한, 특정 핑거터치 제스처에 특정 단위조작(예. 일시정지, 다음 등)을 매핑한 경우에는 해당 제스처를 바로 취함으로써 스마트 스피커에게 음성으로 명령(예. 음악 일시정지 해줘~, 다음 곡 재생해줘~ 등)하는 과정조차 생략이 가능해진다.

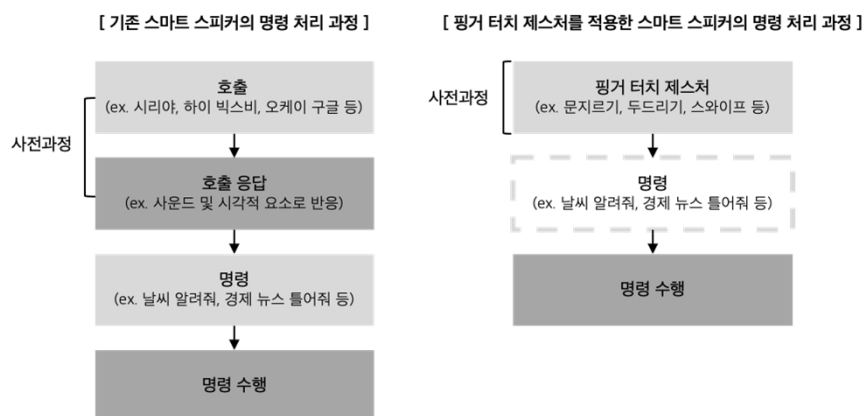


Figure 1 Interacting with a smart speaker based on finger touch gestures

## 2. 스마트 스피커 인터랙션의 단위조작 선정

적절한 핑거터치 제스처를 디자인하기 위해서는 먼저 스마트 스피커와의 사용자 인터랙션에 필수적인 단위조작들에 대한 정의가 필요하다. 이에 본 연구에서는 사용자들이 주로 사용하는 스마트 스피커의 기능들을 먼저 조사하였다. 컨슈머인사이트(2018)의 국내 스마트 스피커 사용자들에 대한 설문조사 결과에 따르면, 사용자들은 주로 음악재생, 날씨정보, 블루투스 스피커, TV 조작, 대화, 검색, 알람, 뉴스 브리핑, 리모컨 찾기, 무드 등 기능을 주로 사용하고 있다고 한다. 한편, 박미나(2018)에 따르면, 영국 사용자들은 음악재생, 질의/응답, 알람설정, 뉴스/날씨, 다른 스마트 기기와의 연동, 메모 기록, 달력/스케줄 동기화, 온라인 제품 구매, 그리고 게임과 같은 기능을 주로 이용하고 있다고 한다. 이러한 기능들의 사용자 조작은 스피커 호출, 일시정지, 다시재생, 취소와 같은 기본적인 단위조작들로 구성되어 있다. 따라서 스마트 스피커와의 인터랙션을 위한 핑거터치 제스처는 이러한 단위조작들에 적용되는 것이 합리적이다. 이에 본 연구에서는 스토셋, 블레싱(Stöbel & Blessing, 2010), 김, 송(Kim & Song, 2011), 송, 김(Song & Kim, 2012)이 분류한 단위 조작들을 기반으로 스마트 스피커에서 자주 사용되는 기능들의 조작에 필수적인 12개의 단위조작들을 Table 1과 같이 도출하였다. 전체 단위조작의 수는 12개지만 Pause(일시정지)와 Replay(다시재생), Next(다음)와 Previous(이전), Confirm.yes(승인)와 Confirm.no(거절), Volume up(소리 키우기)과 Volume down(소리 줄이기)은 서로 대응되는 조작으로서 하나로 묶을 수 있으므로, 선정된 단위조작은 총 8개의 군집으로 볼 수 있다.

Table 1 List of the 12 selected unit operations

|    | 단위조작               | 설명                           |
|----|--------------------|------------------------------|
| 1  | Pause (일시정지)       | 수행 중인 기능을 일시 멈춤              |
| 2  | Replay (다시재생)      | 일시정지로 멈추었던 기능을 다시 재생         |
| 3  | Cancel (취소/종료)     | 수행 중인 기능을 취소하거나 종료           |
| 4  | Repeat (반복)        | 수행했던 것 또는 수행 중인 것을 다시 반복     |
| 5  | Next (다음)          | 순차적 정보를 제공 중에 다음 정보로 이동      |
| 6  | Previous (이전)      | 순차적 정보를 제공 중에 이전 정보로 이동      |
| 7  | Confirm.yes (승인)   | 스마트 스피커의 제안을 승인              |
| 8  | Confirm.no (거절)    | 스마트 스피커의 제안을 거절              |
| 9  | Volume up (소리증가)   | 스마트 스피커의 소리를 키움              |
| 10 | Volume down (소리감소) | 스마트 스피커의 소리를 줄임              |
| 11 | Wake up (발화문구)     | 스마트 스피커를 호출                  |
| 12 | Reinput (발화재입력)    | 마지막 음성 명령을 즉각 취소하고 새 명령을 재입력 |

## 3. 스마트 스피커에 적합한 핑거터치 제스처 디자인

김현(Kim, 2012)은 제스처 조작은 직관적이어야 하며, 기억하기 용이하고, 자연스러워서 팔이나 손 등에 무리가 가지 않아야 하고, 각 제스처들이 명확하게 구분 가능해야 한다고 하였다. 본 연구에서는 이러한 제스처의 디자인 요건과 스마트 스피커의 형태를 고려하여 사용자들이 한 손으로 취할 수 있는 핑거터치 제스처들을 고안하였다. 시중에서 판매되고 있는 스마트 스피커의 주된 형태는 원 혹은 사각 기둥 형태이다. 이를 고려하여 스마트 스피커의 윗면과 옆면을 제스처 인식이 가능한 부위로 정하고, Figure 2와 같이 총 20가지의 제스처를 디자인하였다. 이들은 스마트폰이나 블루투스 이어폰과 같은 터치 제스처를 사용하는 타 기기에도 자주 사용되는 제스처들로서, 직관적이고 단순하여 사용자들이 쉽게 학습이 가능한 것들로

구성하였다. 탭은 스마트 스피커의 표면을 손가락으로 가볍게 터치하는 것이며, 두드리기는 손가락 끝을 이용하여 표면을 순차적으로 두드리는 제스처이다. 문지르기란 손바닥으로 표면을 2~3회 문지르는 것이며, 동그라미는 손가락 끝을 이용하여 스마트 스피커의 윗면에 동그라미를 그리는 제스처이다. 스와이프는 손가락 끝으로 상하 혹은 좌우 방향으로 직선을 그리는 제스처이다. 방향성을 가진 스와이프 제스처의 경우 사용자간 동작을 쉽게 취할 수 있도록 좌우 스와이프는 윗면에, 상하 제스처는 옆면에 취하는 것으로 한정하였다.



Figure 2 List of the 20 finger touch gestures

## 4. 단위조작별 적합한 핑거터치 제스처의 결정

### 4. 1. 설문 참여자

2장에서 선정한 12개의 단위조작별 적합한 핑거터치 제스처가 무엇인지를 결정하기 위해 설문조사를 수행하였다. 설문에는 여성 19명, 남성 39명으로 총 56명(평균연령: 21.4세)이 참여하였다. 그들 중 스마트 스피커 사용 경험이 있는 응답자는 14명, 경험이 없는 응답자는 42명이었다. 사용 경험이 있다고 응답한 참가자들의 최장 사용기간은 2년 7개월이었으며, 스마트 스피커의 사용 빈도는 네이버 클로바, 카카오 미니, 애플 홈팟, 그리고 구글 홈 순이었다.

### 4. 2. 설문 방법

설문조사는 사전에 섭외된 참여자들을 대상으로 온라인으로 진행하였는데, 사전 설문을 통해 설문 참여자의 나이, 성별, 그리고 스마트 스피커 사용경험을 먼저 조사하였다. 사전설문 후, Figure 3과 같이 핑거터치

제스처에 대한 설명과 각 제스처의 시연 영상을 보여주며 제스처를 학습시켰다. 그 후, Table 2와 같은 상황별 스마트 스피커 조작 시나리오를 주고 해당 상황과 조작 기능에 대해 가장 적합하다고 생각되는 제스처를 세 개씩 골라 순위를 매기도록 하였다. 모든 설문이 종료된 후에는 핑거터치 제스처의 종합적인 사용 만족도에 대해 5점 척도로 평가하도록 하였다.

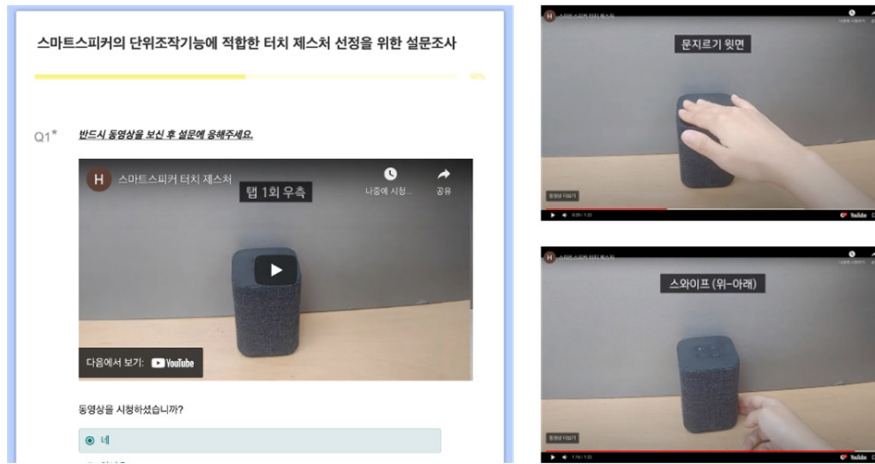


Figure 3 Online survey

Table 2 Task scenario for each unit operation

| 단위조작 |                   | 시나리오   |
|------|-------------------|--|
| 1    | Pause<br>(일시정지)   | 스마트 스피커가 일정을 읽어주는 중에 거래처로부터 중요한 전화가 왔습니다. 전화를 받기 위해 당신은 일정 리마인드를 일시정지하려고 합니다.<br>• 스마트 스피커 : “오늘 일정은 오후 3시부터 오후 7시까지...”<br>• (전화벨)<br><br>스마트 스피커에 제스처를 취하여 일정 리마인드를 일시정지시키려 할 때, 어떤 제스처가 적절한지 1순위/2순위/3순위를 선택해 주십시오. |
| 2    | Replay<br>(다시재생)  | 당신은 스마트 스피커로 음악을 듣던 중에 택배가 와서 음악을 일시정지시켰습니다. 당신은 택배를 받고 난 후 아까 들던 음악을 다시 재생시키려고 합니다.<br><br>스마트 스피커에 제스처를 취하여 음악을 다시 재생시키려 할 때, 어떤 제스처가 적절한지 1순위/2순위/3순위를 선택해 주십시오.  |
| 3    | Cancel<br>(취소/종료) | 당신은 스마트 스피커와 스무고개 게임을 하고 있습니다. 게임을 하던 중 지루해져 게임을 취소하려고 합니다.<br><br>스마트 스피커에 제스처를 취하여 진행 중인 게임을 취소한다고 할 때, 어떤 제스처가 적절할지 1순위/2순위/3순위를 선택해 주십시오.  |
| 4    | Repeat<br>(반복)    | 당신은 외출 준비를 위해 드라이기로 머리를 말리며 오늘의 날씨를 들으려고 합니다. 하지만 드라이기 소리로 인해 날씨 정보를 제대로 듣지 못해 오늘의 날씨를 다시 들으려 합니다.<br><br>스마트 스피커에 제스처를 취하여 날씨를 다시 듣는다고 할 때, 어떤 제스처가 적절할지 1순위/2순위/3순위를 선택해 주십시오.                                       |
| 5    | Next<br>(다음)      | 당신은 스마트 스피커에게 문자 메시지를 읽어달라고 했습니다. 첫 문자의 내용이 스펀문자여서 그만 듣고 다음 문자로 넘어가고자 합니다.<br><br>스마트 스피커에 제스처를 취하여 다음 문자를 듣도록 조작한다고 할 때, 어떤 제스처가 적절할지 1순위/2순위/3순위를 선택해 주십시오.  |
| 6    | Previous<br>(이전)  | 당신은 스마트 스피커로 라디오를 듣다가 다음 채널로 넘어갔습니다. 당신은 이전의 채널을 듣고자 합니다.<br><br>스마트 스피커에 제스처를 취하여 이전 채널로 조작한다고 할 때, 어떤 제스처가 적절할지 1순위/2순위/3순위를 선택해 주십시오.   |

|    |                       |  |
|----|-----------------------|--|
| 7  | Confirm.yes<br>(승인)   | 당신은 스마트 스피커로 내일 오전 7시에 알람을 설정하고자 합니다.<br>• 사용자 : “아침 7시에 알람 맞춰줘”<br>• 스마트 스피커 : 내일 오전 7시에 알람을 맞춰드릴까요?<br><br>이 상황에서 스마트 스피커에게 알람을 맞추도록 승인을 하기 위해 제스처를 취하여 조작한다고 할 때, 어떤 제스처가 적절할지 1순위/2순위/3순위를 선택해 주십시오.   |
| 8  | Confirm.no<br>(거절)    | 스마트 스피커와 휴대폰을 연결시켜 이용 중에 전화가 걸려왔습니다. 스마트 스피커는 전화를 받을지 물어보았고, 거절을 하고 싶은 상황입니다.<br><br>이때 스마트 스피커에 거절을 위한 제스처를 취하여 조작한다고 할 때, 어떤 제스처가 적절한지 1순위/2순위/3순위를 선택해 주십시오.  |
| 9  | Volume up<br>(소리증가)   | 당신은 날씨가 궁금해 스마트 스피커에게 오늘의 날씨를 물어보았습니다. 스마트 스피커가 답변을 해주는 상황에서 소리가 작아, 듣는 도중 소리를 키우고 싶어졌습니다.<br><br>이 상황에서 소리를 키우기 위해 스마트 스피커에 제스처를 취하여 조작한다고 할 때, 어떤 제스처가 적절할지 1순위/2순위/3순위를 선택해 주십시오.   |
| 10 | Volume down<br>(소리감소) | 당신은 방에서 혼자 음악을 듣고 있습니다. 그러던 중 집 밖에 누가 찾아온 듯한 소리가 들려, 음악 소리를 줄이려고 합니다.<br><br>이 상황에서 소리를 줄이기 위해 스마트 스피커에 제스처를 취하여 조작한다고 할 때, 어떤 제스처가 적절할지 1순위/2순위/3순위를 선택해 주십시오.  |
| 11 | Wake up<br>(발화문구)     | 당신은 스마트 스피커에 명령을 하려고 합니다.<br><br>기가지니, 오케이 구글 등과 같은 발화문구 대신 어떤 제스처가 적절할지 1순위/2순위/3순위를 선택해 주십시오.  |
| 12 | Reinput<br>(발화재입력)    | 3명의 친구들이 집에 놀러와 스마트 스피커를 이용해 집에서 가까운 카페에서 커피를 배달시키려고 합니다. 아이스 아메리카노 3잔과 아이스 카페라떼 1잔을 주문하기 위해, 스마트 스피커를 호출하였습니다.<br>• 사용자 : “카페에서 배달시켜줘”<br>• 스마트 스피커 : 어떤 제품을 주문할까요?<br>• 사용자 : “아이스 아메리카노 2잔이랑 아이스 카페라,,,”<br><br>당신은 주문을 하던 중 아이스 아메리카노 3잔을 시켜야 하는 것을 깨닫고 방금 전의 발화문을 취소하고 다시 주문하고자 합니다. 기존 발화를 취소하고 재발화를 하기 위해 스마트 스피커에 제스처를 취하여 조작한다고 할 때 어떤 제스처가 적절할지 1순위/2순위/3순위를 선택해 주십시오. |

#### 4. 3. 설문 결과

Table 3은 설문 결과를 보여준다. 순위별 가중치를 부여하여 각 단위조작에 대해 1순위로 선정한 제스처에는 9점, 2순위로 선정한 제스처에는 3점, 3순위로 선정한 제스처에는 1점씩 부여하여 점수를 합산하였다. Table 3은 최종 합산점수를 기준으로 단위조작별 3등까지의 제스처를 보여주고 있다.

**Table 3** Experiment results and final mapping between unit operations and finger touch gestures (The darker the color, the higher the ranking score; the underlined cells mark the final selection)

| 단위조작 |                       | 1등                              | 2등                             | 3등                   |
|------|-----------------------|---------------------------------|--------------------------------|----------------------|
| 1    | Pause<br>(일시정지)       | 260점<br><u>탭 1회 윗면</u>          | 172점<br>탭 2회 윗면                | 55점<br>두드리기<br>1회 윗면 |
| 2    | Replay<br>(다시재생)      | 282점<br><u>탭 1회 윗면</u>          | 165점<br>탭 2회 윗면                | 76점<br>탭 1회 우측       |
| 3    | Cancel<br>(취소/종료)     | 110점<br>탭 2회 윗면                 | 73점<br><u>두드리기<br/>1회 윗면</u>   | 70점<br>문지르기 윗면       |
| 4    | Repeat<br>(반복)        | 202점<br><u>동그라미</u>             | 79점<br>스вай프<br>(우→좌)          | 66점<br>두드리기<br>2회 윗면 |
| 5    | Next<br>(다음)          | 174점<br><u>스вай프<br/>(좌→우)</u>  | 107점<br>탭 2회 우측                | 81점<br>탭 1회 우측       |
| 6    | Previous<br>(이전)      | 171점<br>탭 2회 좌측                 | 158점<br><u>스вай프<br/>(우→좌)</u> | 93점<br>탭 1회 좌측       |
| 7    | Confirm.yes<br>(승인)   | 181점<br><u>탭 1회 윗면</u>          | 169점<br>동그라미                   | 103점<br>탭 2회 윗면      |
| 8    | Confirm.no<br>(거절)    | 97점<br>스вай프<br>(우→좌)           | 89점<br><u>탭 2회 윗면</u>          | 85점<br>문지르기 윗면       |
| 9    | Volume up<br>(소리증가)   | 412점<br><u>스вай프<br/>(아래→위)</u> | 84점<br>스вай프<br>(좌→우)          | 54점<br>문지르기 우측       |
| 10   | Volume down<br>(소리감소) | 395점<br><u>스вай프<br/>(위→아래)</u> | 87점<br>스вай프<br>(우→좌)          | 45점<br>문지르기 좌측       |
| 11   | Wake up<br>(발화문구)     | 217점<br><u>탭 2회 윗면</u>          | 105점<br>문지르기 윗면                | 81점<br>두드리기<br>1회 윗면 |
| 12   | Reinput<br>(발화재입력)    | 130점<br>동그라미                    | 124점<br><u>문지르기 윗면</u>         | 85점<br>탭 2회 윗면       |

각 단위조작별로 설문 결과에서 1등 점수를 얻은 제스처들을 우선적으로 매핑하다 보면, 같은 제스처가 서로 다른 조작에 중복되어 1순위로 평가된 경우들이 있었다. 이에 순위점수와 더불어 단위조작의 전후 관계, 기능적 연관성도 함께 고려하여 다음과 같이 단위조작별 적합한 핑거터치 제스처를 최종 결정하였다.

- Pause(일시정지)와 Replay(다시재생): ‘탭 1회 윗면’의 동일한 제스처가 높은 점수로 1순위였다. 이 둘은 명확한 전후 관계를 가진 단위조작으로서 일시정지 후에 다시 재생되기 때문에 혼동의 여지가 없으므로 같은 제스처를 매핑하였다.

- Cancel(취소/종료)과 Wake up(발화문구): ‘탭 2회 윗면’이라는 같은 제스처가 1순위로 평가되었는데, 이 둘은 조작 상황이 겹칠 수 있어 동일한 제스처를 매핑할 수는 없었다. 이에 보다 높은 점수를 받은 Wake up(발화문구) 기능에 ‘탭 2회 윗면’ 제스처를 매핑하였고, Cancel(취소/종료)에는 2순위로 평가된 ‘두드리기 1회 윗면’ 제스처를 매핑하였다.

- Repeat(반복)과 Reinput(발화재입력): 둘 다 ‘동그라미’ 제스처가 1순위로 평가되었다. 이에 보다 높은 점수를 받은 Repeat(반복)에 ‘동그라미’ 제스처를 매핑하였고, Reinput(발화재입력)에는 2순위인 ‘문지르기 윗면’ 제스처를 매핑하였다.



- Next(다음)와 Previous(이전): 이 두 단위조작은 유사하나 서로 반대의 방향성을 가진 기능이다. 이에 같은 유형에 방향성만 다른 제스처인 ‘스와이프(좌→우)’, ‘스와이프(우→좌)’를 각각 매핑하였다.

- Confirm.yes(승인)와 Confirm.no(거절): 이 두 단위조작도 서로 방향이 반대인 조작으로서, 순위점수를 고려하여 동일한 제스처 유형에 탭 횟수만 다른 제스처인 ‘탭 1회 윗면’, ‘탭 2회 윗면’을 각각 매핑하였다. Confirm.yes(승인)에 매핑된 ‘탭 1회 윗면’ 제스처의 경우, Pause(일시정지)와 Replay(다시재생) 기능과 동일한 제스처이나 스마트 스피커의 제안에 대해 응답할 경우에만 사용되는 조작이기 때문에 서로 조작이 충돌될 여지는 없다.

- Volume up(소리증가)과 Volume down(소리감소): 소리를 조절하는 동일 유형에 방향만 반대인 조작이다. 순위점수가 가장 높게 평가된 방향성만 다른 스와이프 제스처인 ‘스와이프(아래→위)’, ‘스와이프(위→아래)’를 각각 매핑하였다.

정리하자면, 12개의 단위조작에 매핑된 최종 핑거터치 제스처의 종류는 상하좌우 방향의 스와이프를 하나로 간주한다면 탭 1회, 탭 2회, 두드리기, 동그라미, 문지르기, 그리고 스와이프 제스처로 총 6개여서 사용자가 학습하여 사용하기에 부담 없는 개수였다. 본 설문 후 진행된 핑거터치 제스처의 종합적 사용만족도에 대해서는 평균 3.79점으로 핑거터치 제스처가 스마트 스피커의 사용성에 도움을 줄 것이라고 응답한 참여자들이 대체로 많았다.

## 5. 핑거터치 제스처에 대한 사용성 평가

### 5. 1. 실험참여자

사용성 평가 실험에는 총 20명(남성 13명, 여성 8명)이 참여하였으며, 이들 모두는 스마트 스피커나 스마트폰의 빅스비 및 쉬리와 같은 스마트 에이전트에 대한 사용 경험을 가지고 있었다.



Figure 4 Experimental environment

### 5. 2. 실험 목적 및 환경

본 실험은 4장에서 매핑된 단위조작별 핑거터치 제스처들의 학습용이성과 사용성을 평가하는 것이 목적이었다. 실험시스템은 핑거터치 제스처로 조작되는 스마트 스피커를 Figure 4와 같이 오즈의 마법사

기법을 이용하여 진행자가 실험참여자의 핑거터치 제스처를 눈으로 보고 그에 매핑된 명령수행 결과를 노트북에 연결된 스마트 스피커 음성으로 제공해주었다. 마법사 역할을 맡은 진행자 2와 실험참여자 사이에는 Figure 4와 같이 가벽을 세워 진행자 2(마법사)의 기능 조작을 가림으로써 실험참여자가 스마트 스피커를 실제 사용 환경처럼 느끼고 실험에 좀 더 집중할 수 있도록 하였다.

### 5. 3. 실험 절차

실험은 크게 세 단계로 제스처 학습, 제스처 학습 평가시험, 그리고 사용성 평가 순으로 진행하였다. 먼저, 실험참여자에게 실험에 대한 전반적인 설명을 한 후 제스처-기능 매핑도와 제스처별 시연 영상을 통해 제스처를 학습하도록 하였다. 그 후, 제스처에 대한 충분한 학습이 이뤄졌음을 확인하기 위해 진행자는 참여자의 단위조작별 제스처 학습 수준을 평가하였다. 평가시험 결과에서 실험참여자가 12개 단위조작 중 11개 이상을 맞춰야만 사용성 평가 단계로 넘어가도록 하였다.

사용성 평가는 Table 4와 같은 총 12개의 사용 태스크 시나리오에 대해, 실험참여자가 스마트 스피커에 직접 해당 태스크에 적합한 제스처를 취하는 방식으로 진행하였다. 각 시나리오에 적합한 제스처를 취한 후, 실험참여자는 해당 태스크를 얼마나 쉽게 수행하였는지를 7점 척도로 평가하였다. 참여자가 올바른 제스처를 취하지 못한 경우에는 해당 태스크에 대한 평가를 진행하지 않고 다음 태스크로 넘어갔다. 모든 태스크가 완료된 후 실험참여자는 음성만으로 조작하는 것보다 핑거터치 제스처를 함께 사용하는 것이 얼마나 유용한지를 7점 척도로 답하였으며, 추가했으면 하는 제스처나 개선 방향에 대한 의견을 제시하도록 하였다.

Table 4 Task scenarios for each unit operation

| 단위조작 |                       | 시나리오   |
|------|-----------------------|--|
| 1    | Pause<br>(일시정지)       | 현재 재생 중인 교통정보를 일시정지 하십시오.                                    |
| 2    | Replay<br>(다시재생)      | 일시정지한 교통정보를 다시 재생하십시오.                                       |
| 3    | Cancel<br>(취소/종료)     | 현재 울리는 알람을 종료하십시오.   |
| 4    | Repeat<br>(반복)        | 오늘의 일정을 알려드리겠습니다. 오늘의 일정을 다시 듣기 하십시오.                        |
| 5    | Next<br>(다음)          | 오늘의 일정을 알려드리겠습니다. '다음' 기능을 이용하여 내일의 일정을 들으십시오.               |
| 6    | Previous<br>(이전)      | 오늘의 일정을 알려드리겠습니다. '이전' 기능을 이용하여 어제의 일정을 들으십시오.               |
| 7    | Confirm.yes<br>(승인)   | 스마트 스피커의 제안에 대해 승인하십시오.                                      |
| 8    | Confirm.no<br>(거절)    | 스마트 스피커의 제안에 대해 거절하십시오.                                      |
| 9    | Volume up<br>(소리증가)   | 현재 재생 중인 뉴스의 소리를 키우십시오.                                      |
| 10   | Volume down<br>(소리감소) | 현재 재생 중인 뉴스의 소리를 줄이십시오.                                      |
| 11   | Wake up<br>(발화문구)     | 스마트 스피커를 호출하십시오.   |
| 12   | Reinput<br>(발화재입력)    | 스마트 스피커를 호출하여 '네이버 주가'를 묻다가 발화재입력 기능을 이용하여 '카카오 주가'를 물어보십시오. |

### 5. 4. 실험 결과

실험참여자들이 제스처를 학습하는 데 걸린 시간은 평균 410초(약 7분)였으며, 학습 평가시험을 통과하기까지의 평균 응시횟수는 2.35회였다. 학습 평가시험에서 단위조작 제스처별 오답 횟수는 Figure 5와 같았다. 대응되는 방향성을 가진 5, 6번(Next-Previous)과 9, 10번(Volume up-down) 조작 제스처는 학습이 매우 빨랐다. 반면, 3번(Cancel: 취소/종료 두드리기 1회 윗면) 조작의 오답

횟수는 26번으로 다른 문항들보다 상대적으로 매우 높았다. 분산분석 결과, 남녀 간에는 오답 횟수에 유의한 차이가 없었지만( $F(1,239)=0.04, p=0.834$ ), 단위조작 제스처 간에는 오답 횟수에 유의한 차이가 존재하였다( $F(11,239)=10.38, p=0.0000$ ).

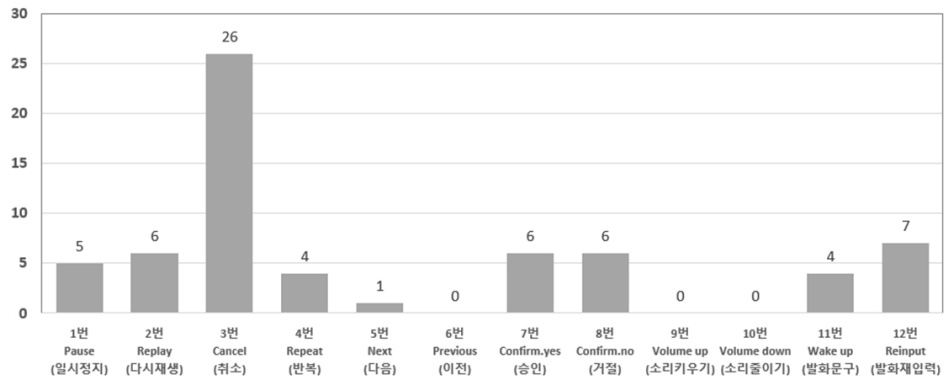


Figure 5 Number of incorrect answers in learning assessment

Figure 6은 Table 4의 태스크 시나리오를 핑거터치 인터랙션으로 수행하는 동안의 단위조작별 오답횟수를 보여준다. 3번 (Cancel:취소/종료 두드리기 1회 윗면)과 12번(Reinput:발화재입력- 문지르기 윗면) 제스처 조작에서 오답횟수가 다른 조작들에 비해 많음을 알 수 있다.

Figure 7은 각 태스크를 완료하는데 까지 걸린 시간과 태스크 완료 후 평가한 제스처의 사용용이성 점수 평균을 보여준다. 실험참여자들은 평균적으로 2.57초만에 각 태스크를 완료했으며, 평균 6.56점(7점 척도)으로 핑거터치 제스처 조작이 용이했다고 평가하였다. 통계분석 결과, 다른 조작 제스처들 간에는 큰 차이가 없었으나 12번(Reinput:발화재입력- 문지르기 윗면) 제스처의 경우, 태스크 완료 시간도 유의하게 더 오래 걸렸으며( $F(11,200)=27.07, p=0.000$ ), 용이성 점수도 상대적으로 더 낮았다( $F(11,202)=5.23, p=0.000$ ). 이는 발화 재입력이란 조작이 다른 단위조작들과는 달리 실험참여자들에게 덜 익숙했고 문지르기 제스처도 생소했기 때문으로 보인다. 한편, 핑거터치 제스처 조작의 유용성에 대한 종합적인 평가 점수는 5.2점으로 대체로 유용하다는 평가를 받았다.

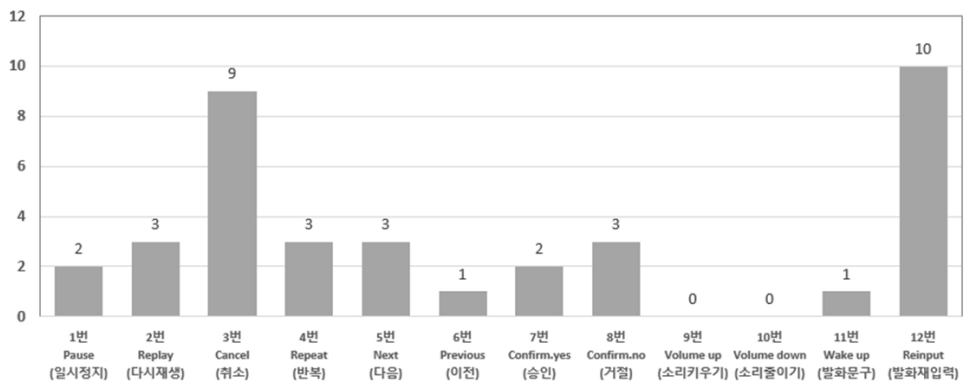


Figure 6 Number of incorrectly answered person in usability test

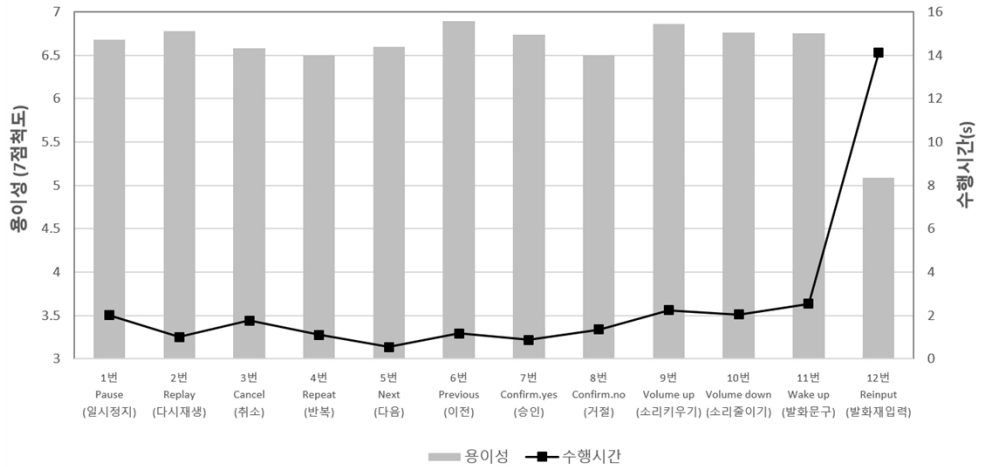


Figure 7 Ease of use and task completion time in usability test

## 6. 결론 및 고찰

본 연구는 스마트 스피커를 사용할 때 항상 호출어(예. 아리아~, OK구글~ 등)로 먼저 스피커를 호출한 후에 특정 음성 명령을 해야 하는 번거로움을 해결해줄 수 있는 핑거터치 제스처 기반의 멀티모달 인터랙션을 제안하고 그것의 학습용이성과 사용성을 평가하였다.

본 연구에서는 스마트 스피커의 주요 기능 조작에 필요한 총 12개의 단위조작을 도출하였고, 원 혹은 사각기둥 모양의 스마트 스피커 조작에 사용 가능한 핑거터치 제스처 20종을 고안하였다. 그 후, 단위조작과 제스처 간 최적의 매핑을 결정하기 위해 각 단위조작에 적합한 제스처를 선택하는 설문조사를 수행하였다. 설문 결과와 단위조작의 성격을 고려하여 최종적으로 단위조작별 핑거터치 제스처를 다음과 같이 매핑하였다: 1) 일시정지-탭 1회 윗면, 2) 다시재생-탭 1회 윗면, 3) 취소/종료-두드리기 1회 윗면, 4) 반복-동그라미, 5) 다음-스вай프(좌→우), 6) 이전-스вай프(우→좌), 7) 승인-탭 1회 윗면, 8) 거절-탭 2회 윗면, 9) 소리 증가-스вай프(아래→위), 10) 소리감소-스вай프(위→아래), 11) 발화문구-탭 2회 윗면, 12) 발화제일력-문지르기 윗면.

12개의 단위조작에 매핑된 제스처의 종류는 상하좌우 스와이프를 하나로 간주하면 탭 1회, 탭 2회, 두드리기, 동그라미, 문지르기, 그리고 스와이프로 총 6개였다. 본 연구에서는 최종 결정된 핑거터치 제스처의 학습용이성과 사용성을 오즈의 마법사 기법을 이용하여 평가하였다. 실험 결과, 방향성을 가진 스와이프 제스처는 학습이 매우 빨랐으나, 취소/종료 조작에 매핑된 ‘두드리기 1회 윗면’ 제스처의 경우 학습이 매우 느림을 알 수 있었다. 또한, 참여자들은 스마트 스피커와의 핑거터치 인터랙션이 유용하다고 평가하였으나, 발화제일력 -문지르기 윗면 제스처의 경우 태스크 수행 시간도 오래 걸렸고 사용용이성도 상대적으로 낮게 평가되었다. 따라서 이 두 단위조작의 경우 조작 자체를 제외하거나 제스처에 개선이 필요함을 알 수 있었다.

본 연구에서 제안한 핑거터치 제스처는 음성 명령을 보조하는 수단으로서, 스마트 스피커가 침대 머리맡이나 책상 위와 같이 사용자가 근접한 위치에 있을 때만 사용 가능하다는 한계가 있다. 하지만 특정 호출어를 불러야 하는 사전 호출의 과정 없이 신속하게 스마트 스피커에 명령을 내릴 수 있다는 점에서 충분히 그 유용성을 가지며 이는 실험을 통해서도 입증되었다. 또한, 음악 재생 중 “다음 곡 재생”을 스와이프(좌→우) 제스처로

바로 명령하는 예와 같이 아예 음성명령 없이 빠르게 조작할 수 있는 장점이 있다. 특히 다른 사람들과 대화 중에 스피커에서 울리는 알람을 꺼야 하거나 스피커로 음악을 듣는 중에 전화를 받기 위해 음악을 멈춰야 하는 등 사용자가 음성을 다른 목적으로 사용해야 해서 스마트 스피커에게 음성으로 명령하기가 어렵거나 부자연스러운 상황에서도 유용하게 쓰일 수 있다. 또한 핑거터치 제스처는 특정 영역을 정확하게 터치하거나 눌러야 하는 기존 스마트 스피커들에서의 터치 입력이나 버튼 조작과는 달리 제품의 표면을 손가락으로 탭하거나 문지르는 등의 제스처를 취하면 되므로 더 빠르게 다양한 조작을 수행하는 것을 가능하게 해준다. 실험 결과를 보면 두드리기나 문지르기와 같은 제스처는 사용자들이 일반적으로 접하는 제스처가 아니므로 학습성이나 사용성에는 문제가 있었다. 하지만 탭, 동그라미, 스와이프 제스처의 경우 기존 스마트 스피커에 적용되어 발화문구 호출이나 반복 등의 음성 명령 조작을 대신한다면 음성 보다는 손으로 직접 조작하는 것이 더 나은 다양한 상황에 유용하게 쓰일 수 있을 것이다.

본 연구에는 다음과 같은 한계점이 있다. 첫째, 3장에서 디자인된 핑거터치 제스처는 원기둥 또는 사각기둥 형태의 스마트 스피커를 기준으로 하였기에 구 모양이나 조약돌 형태의 스마트 스피커에는 적합하지 않을 수 있다. 둘째, 4장에서 결정된 단위조작별 핑거터치 제스처는 평균 연령 21.4세인 젊은 설문 참여자들의 스테레오타입 (Stereotype)이었다. 따라서 다른 연령대의 사용자나 첨단 IT 기기들에 대한 숙련도 등에 따른 핑거터치 인터랙션의 다양성을 반영하지는 못했다는 한계가 있다. 셋째, 5장에서 핑거터치 제스처 각각의 사용성을 주관적으로 평가하였으나 이는 핑거터치 제스처를 음성 명령과 함께 사용하는 멀티모달 환경에서의 평가는 아니었다. 따라서 음성 명령과 핑거터치 둘 다를 사용하여 스마트 스피커를 조작 가능한 상황에서 핑거터치 인터랙션의 유용성과 사용성에 대한 평가가 필요하다. 이에 대해서는 임호정(Im, 2021)이 후속 연구를 진행하였는데, 숨은 그림 찾기와 단어 기억하기와 같은 주 태스크를 수행하면서 음성이나 핑거터치로 스마트 스피커를 조작하는 부 태스크를 수행시켰더니 90% 이상의 빈도로 음성 명령보다 핑거터치 제스처를 더 많이 사용하였다. 이는 스마트 스피커와의 인터랙션에 핑거터치 제스처가 유용함을 보여준 것이다.

## References

1. Alanwar, A., Balaji, B., Tian, Y., Yang, S., & Srivastava, M. (2017, November). Echosafe: Sonar-based verifiable interaction with intelligent digital agents. In *Proceedings of the 1st ACM Workshop on the Internet of Safe Things* (pp. 38-43).
2. Lee, J. S. (2018, October). Speech and Gestures for Smart-Home Control and Interaction for Older Adults. In *Proceedings of the 3rd International Workshop on Multimedia for Personal Health and Health Care* (pp. 49-57).
3. Consumer Insight. (2019). *AI스피커, 스마트폰 AI 넘어서나[AI speaker beyond smartphone AI]*. Retrieved from [https://www.consumerinsight.co.kr/voc\\_view.aspx?no=2940&id=ins02\\_list&PageNo=1&schFlag=0](https://www.consumerinsight.co.kr/voc_view.aspx?no=2940&id=ins02_list&PageNo=1&schFlag=0)
4. Ha, J., & Kim, S. (2017). 인공지능 스피커의 과거와 현재, 그리고 미래[The past, present, and future of artificial intelligence speakers]. *한국컴퓨터정보학회지 [Journal of Korean Society of Computer Information]*, 25(2), 1-8.
5. Hoffmann, F., Tyroller, M. I., Wende, F., & Henze, N. (2019, November). User-defined interaction for smart homes: voice, touch, or mid-air gestures?. In *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia* (pp. 1-7).
6. Im, H. (2021). *스마트 스피커 터치 제스처 디자인과 사운드 딥러닝의 적용 가능성 [Design of smart speaker touch gesture and applicability of sound deep learning]*. (Master's Thesis). Seoul National University of Science and Technology, Seoul, Korea.
7. Jang, S., & Yun, J. (2018). User experience study on use of repetitive wake word in conversational speech interaction using AI voice assistant. *KSDS Conference Proceeding*, 222-223.
8. Kim, H. (2012). Interacting with gestures: taxonomy and requirements. *Journal of the Ergonomics Society of Korea*, 31(4), 475-481.

9. Kim, H., & Song, H.W. (2011). 보다 직관적인 비접촉 조작 제스처를 위한 디자인 방향성 제시 [Towards designing more intuitive touchless operations based on hand gestures]. *Journal of Korean Society of Design Science*, 25(1), 269–277.
10. Korea Consumer Agency. (2017). 인공지능(AI) 가전제품 문제점 및 개선방안 – 음성인식 스피커를 중심으로[Problems and Improvement Plans for Artificial Intelligence Home Appliances – Focusing on Voice Recognition Speakers]. *한국소비자원 조사보고서[Korea Consumer Agency Report]*, 1–33.
11. McMillan, D., Brown, B., Kawaguchi, I., Jaber, R., Solsona Belenguer, J., & Kuzuoka, H. (2019). Designing with Gaze: Tama—a Gaze Activated Smart-Speaker. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW), 1–26.
12. Park, S., & Lee, Y. (2020). User experience of smart speaker visual feedback type: the moderating effect of need for cognition and multitasking. *Archives of Design Research*, 33(2), 181–199.
13. Pomykalski, P., Woźniak, M. P., Woźniak, P. W., Grudzień, K., Zhao, S., & Romanowski, A. (2020). Considering Wake Gestures for Smart Assistant Use. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–8.
14. Porcheron, M., Fischer, J. E., Reeves, S., & Sharples, S. (2018). Voice interfaces in everyday life. In *proceedings of the 2018 CHI conference on human factors in computing systems*, 1–12.
15. Pyae, A., & Joelsson, T. N. (2018). Investigating the usability and user experiences of voice user interface: a case of Google home smart speaker. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, 127–131.
16. Pyae, A., & Scifleet, P. (2019). Investigating the role of user's English language proficiency in using a voice user interface: A case of Google Home smart speaker. In *2019 CHI Conference on Human Factors in Computing Systems*, 1–6.
17. Rise, K., & Alsos, O. A. (2020, July). The Potential of Gesture-Based Interaction. In *International Conference on Human-Computer Interaction* (pp. 125–136). Springer, Cham.
18. Shin, J., Jo, I., Lim, W., & Kim, S. (2020). 음성기반 지능형 시스템의 사용자 만족도에 영향을 미치는 인터랙션 설계변수들의 정의 및 분류방안 [A Few Critical Design Parameters Affecting User's Satisfaction in Interaction with Voice User Interface of AI-Infused Systems]. *Journal of the Ergonomics Society of Korea*, 39(1), 73–86.
19. Song, H. W., & Kim, H. (2012). Towards establishing a touchless gesture dictionary based on user participatory design. *Journal of the Ergonomics Society of Korea*, 31(4), 515–523.
20. Stöβel, C., Wandke, H., & Blessing, L. (2009, February). Gestural interfaces for elderly users: help or hindrance?. In *International Gesture Workshop* (p. 269–280). Springer, Berlin, Heidelberg.
21. The Verge. (2017). *Amazon's Alexa started ordering people dollhouses after hearing its name on TV*. Retrieved July 27, 2017 from <https://www.theverge.com/2017/1/7/14200210/amazon-alexa-tech-news-anchor-order-dollhouse>
22. van den Hoven, E., & Mazalek, A. (2011). Grasping gestures: Gesturing with physical artifacts. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 25(3), 255.
23. Voicebot (2020). Smart speaker consumer adoption report 2020. Retrieved April 8, 2021 from <https://voicebot.ai/2020/04/30/yes-the-bedroom-is-now-the-most-popular-location-for-smart-speakers-heres-why-and-what-it-means/>
24. Wei, Z., & Landay, J. A. (2018). Evaluating speech-based smart devices using new usability heuristics. *IEEE Pervasive Computing*, 17(2), 84–96.

# 핑거터치 제스처를 이용한 스마트 스피커 인터랙션

김현<sup>1</sup>, 임호정<sup>2\*</sup>

<sup>1</sup>서울과학기술대학교 기계시스템디자인공학과, 교수, 서울, 대한민국

<sup>2</sup>서울과학기술대학교 나노IT디자인융합대학원 디자인기술융합전공, 학생, 서울, 대한민국

---

## 초록

**연구배경** 스마트 스피커 시장은 매년 크게 성장하고 있지만 그것의 사용성은 여전히 사용자의 기대 수준에 미치지 못하고 있다. 기존 스마트 스피커와의 음성 인터랙션은 사용자의 호출과 스마트 스피커의 호출 응답으로 이루어진 사전 호출 과정이 반드시 필요하며, 즉각적인 오류 제어가 어렵다는 한계를 가지고 있다. 이에 기존 스마트 스피커의 사전 호출 과정을 단축시키면서 보다 즉각적인 오류 제어를 할 수 있는 새로운 인터랙션이 요구된다.

**연구방법** 먼저, 스마트 스피커의 주요 기능사용에 필수적인 단위조작들을 12개로 도출하였으며, 사용자의 한 손으로 취할 수 있는 다양한 터치 제스처들을 구상하였다. 그 후, 단위조작별 적합한 제스처를 선정하는 설문조사를 수행하여 핑거터치 제스처 기반의 스마트 스피커 인터랙션을 디자인하였다. 디자인된 제스처 인터랙션의 학습용이성과 사용용이성을 평가하기 위해 오즈의 마법사 기법을 이용한 사용성 평가를 수행하였다.

**연구결과** 사용성 평가 결과, 참여자들은 핑거터치 제스처를 이용한 스마트 스피커와의 인터랙션에 대해 유용하다고 평가하였으며, 거의 모든 제스처들의 학습이 빠르고 사용용이성도 높게 평가되었다. 하지만 두드리기 1회 윗면이나 문지르기 윗면 제스처의 경우 학습성이나 사용성에 문제가 있는 것으로 드러나 개선이 필요하였다.

**결론** 핑거터치 제스처는 침대 머리맡이나 책상 위와 같이 스마트 스피커가 사용자에게 근접한 위치에 있을 때만 사용 가능하다는 한계가 있으나 사전 호출의 과정 없이 매우 신속하게 조작할 수 있어 향후 스마트 스피커에 적용될 수 있을 것으로 기대된다.

**주제어** 스마트 스피커, 핑거터치, 제스처 디자인, 멀티모달 인터랙션

---