

# 스마트 스피커의 선제적 상호작용 적절 시점에 대한 이해

## Opportune Moments for Proactive Interactions with Smart Speakers

차나래 Narae Cha 한국과학기술원 KAIST nr.cha@kaist.ac.kr	김아욱 Auk Kim 한국과학기술원 KAIST kimauk@kaist.ac.kr	박철영 Cheul Young Park 한국과학기술원 KAIST cheulyop@kaist.ac.kr
박민규 Mingyu Park 한국과학기술원 KAIST mininnov99@kaist.ac.kr	강수원 Soo Won Kang 한국과학기술원 KAIST sw.kang@kaist.ac.kr	이의진 Uichin Lee 한국과학기술원 KAIST uclee@kaist.ac.kr

### 요약문

현재 스마트 스피커는 사용자의 명령어에 반응하는 수동적인 서비스 제공에서 더 나아가, 먼저 말을 걸어 서비스를 제공하는 능동적인 선제적 서비스로 그 범위가 확장되고 있다. 선제적 서비스 제공 시에 스마트 스피커가 적절하지 않은 순간에 말을 걸면 사용자의 과업을 방해하고 부정적인 감정(짜증 등)을 초래하는 문제가 발생한다. 따라서, 말을 거는 타이밍이 선제적 서비스 제공에 있어서 매우 중요한 요소이다. 본 연구에서는 말을 걸기 적절한 순간을 탐험적으로 찾기 위해 스마트 스피커를 제작하여 일주일간 대학생 기숙사 방에 설치하여 데이터를 수집하였다. 스피커가 “지금 대화하기 좋은가요?”라는 질문을 하면 실험자는 “네/아니요”로 중단가능성(interruptibility)에 대해 응답하고, 그 때 하고 있던 행동을 서술하는 방식으로 데이터를 수집하였다. 40 명의 실험자의 자연스러운 생활 중에 총 3572 개의 데이터를 수집하였다. 집에서 가능한 행동을 분류하고, 행동과 관련하여 중단가능성에 영향을 주는 주요 요인들로 집중도/몰입도, 급함/바쁨, 신체/정신적 요인, 음성 채널 유효성이라는 네 가지 요인을 찾았다. 본 연구에서 찾은 요인들을 기반으로, 스마트 스피커가 대화하기 좋은 순간에 선제적 서비스를 제공할 수 있는 기능을 스마트 스피커에 추가할 수 있을 것이다.

### 주제어

스마트 스피커, 상호작용, 중단가능성(interruptibility)

### 1. 서론 및 연구 동기

스마트 스피커에 탑재된 인공지능 비서로서 음성 기반 상호작용을 통하여 정보 제공, 의사결정 등의 다양한 과업들을 지원한다. 현재의 스마트 스피커는 대부분 사용자의 명령에 반응하여 수동적으로 작업을 수행한다. 최근 들어 세계 스마트 스피커 시장을 선도하고 있는 구글 홈[9], 아마존 에코[3]에서 일정 알림이나 집의 안전 및 보안과 관련하여 선제적인 서비스를 제공하기 시작하였다.

선제적 서비스는 사용자가 잊고 있던 내용을 상기시키거나 상황에 맞는 서비스를 먼저 제공하므로 유용하다. 이전 연구들에 따르면 적절한 순간에 서비스를 제공하는 것이 매우 중요하다[2]. 적절하지 못한 순간에 선제적 서비스를 제공할 경우, 짜증을 유발하거나 원작업으로의 재개 지연이 일어날 수 있고[5], 운전 상황일 경우 안전 문제 또한 발생할 수도 있다[6, 7].

진행중인 작업에 대한 중단가능성(interruptibility)에 대한 기존 연구는 데스크 탑과 스마트폰부터 유비쿼터스 환경까지 다양한 분야에서 연구되었다. 예를 들어, 데스크탑에서 과업을 수행할 때는 과업

간의 전환이 일어날 때가 가장 적절한 순간이었다[1]. 스마트폰에서 알림을 보낼 때, 알림 내용 또는 알림을 보낸 사람과의 관계 등이 중단 가능성에 중요하게 영향을 주었다[8]. 자동차의 음성 서비스에서도 차량 센서 데이터와 도로 상황 등을 고려하여 대화에 적절한 순간을 찾았다[6, 7].

하지만 현재까지 스마트 스피커의 선제적 서비스에 대한 중단가능성과 적절한 순간은 연구되지 않았다. 스마트 스피커는 기존 연구한 데스크 탑, 스마트폰 등의 기기와는 달리 주로 편안한 상태로 머무르는 집 환경에서 사용한다는 점에서, 기존의 사무실 또는 자동차 환경과는 차이가 있다. 이러한 차이점에서, 스마트 스피커의 선제적 상호작용에 관한 중단가능성은 기존 연구와 다른 요인이 있을 것이라 예상하여, 이를 찾고자 하는 탐험적 연구를 수행하고자 한다.

## 2. 방법

집에서 자연스러운 생활을 하는 중에 스마트 스피커의 선제적 서비스가 제공됐을 때의 중단 가능성 및 적절한 순간을 이해하기 위하여 Experience Sampling Method (ESM) 방식으로 데이터를 수집하였다. 실험자의 거주 환경에 스마트 스피커를 설치한다. “지금 대화하기 좋은가요?” 라는 ESM 질문을 음성으로 물어보고, 사용자의 응답을 녹음하여 데이터로 수집하였다. 본 절에는 실험 환경, ESM 트리거 방식과 주기, 스마트 스피커 장치에 대해 설명한다.

### 2.1 실험 환경

집은 아파트, 주택, 기숙사 등 다양한 주거 형태 및 가족, 친구, 혼자 등 다양한 가구 구성원 형태가 있다. 본 연구에서는 2인 1실의 기숙사를 실험 환경으로 선택하였다. 기숙사의 구조(그림 1)는 2명이 한 방에서 생활하고, 화장실, 샤워실, 휴게실 등의 공용 공간을 다수의 학생들과 공유한다. 기숙사 환경은 룸메이트가 있기 때문에 함께 사는 사람이 주는 영향을 볼 수 있을 뿐만 아니라, 집 안에서 일어날 수 있는 다양한 활동이 한 공간에서 일어나기 때문에 풍부한 데이터를 수집할 수 있다는 장점이 있다.



그림 1 스마트 스피커가 설치된 상태의 기숙사 구조

### 2.2 ESM 작동 방식

ESM은 두가지 방식으로 트리거가 된다. 첫 번째는 랜덤 트리거이고, 두 번째는 실험자의 걸기 등 움직임이 있을 때 ESM 질문이 나오는, 움직임 감지 트리거이다. 기본 랜덤 방식에서 움직임 트리거를 추가한 이유는 활동 사이의 전환 중에 움직이는 경우가 중단가능성이 높다는 기존 연구가 있기 때문이다[4]. 움직임이라는 상황적 요소가 중단가능성에 주요하게 영향을 끼칠 수 있을 것이라는 예상 하에, 움직이는 상황의 데이터를 더 수집하기 위하여 해당 트리거를 추가하였다.

가장 최근의 ESM이 올린 시간으로부터 15분에서 25분 사이의 랜덤한 시간 후로 다음 ESM을 예약한다. 움직임 감지 트리거는 사용자의 움직임이 감지되면 바로 ESM 질문을 하도록 한다. 각 ESM 간의 간격이 평균 20분을 유지할 수 있도록 하였다.

ESM 작동 시간은 실험자가 깨어나는 아침 시간부터 잠들기 전인 밤까지 약 12시간으로, 실험자가 시간 구간을 직접 설정하도록 하였다.

### 2.3 움직임 감지

움직임을 감지하는 방식은 스마트폰 내장 카메라로 촬영한 실시간 사진들을 활용했다. 시간 연속적인 두

사진의 동위 픽셀 값을 비교하여 임계치 이상으로 차이가 있는지를 보고, 차이가 있는 픽셀 개수가 많으면(테스트를 통해 정한 한계점을 넘기면), 실험자의 움직임이 감지된 것으로 판단하였다. 실시간으로 촬영된 사진은 프라이버시 보장을 위해 비교 과정을 거치면 바로 지워지도록 하여 사진은 저장되지 않았다.

#### 2.4 스마트 스피커 장치

스마트 스피커 장치는 스마트폰(Galaxy S7), 블루투스 스피커(OA-SP600), 광각렌즈(NEXT-F30, 170° 화각)로 구성되어 있으며 종이 커버로 외관을 포장한 후 전원을 연결하여 높이 조절이 가능한 보면대를 활용해 설치하였다.

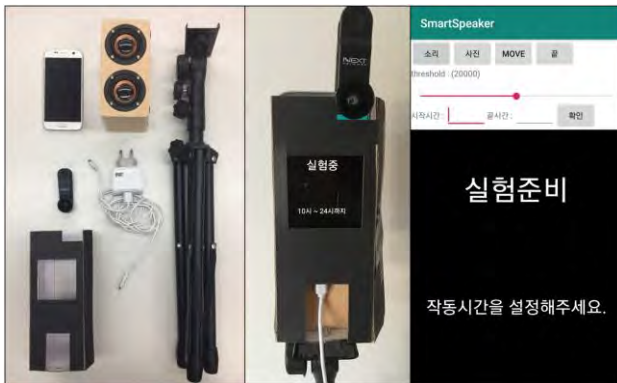


그림 3 (1) 스마트 스피커 부품, (2) 조립된 스마트 스피커, (3) ESM 앱 화면

스마트폰에는 안드로이드 스튜디오로 개발한 ESM 앱이 설치되었다. ESM 앱에는 질문을 듣고 소리 크기를 조절할 수 있는 기능, 방 사진을 찍을 수 있는 기능, 사용자의 움직임을 감지하는 기능, ESM 작동 시간을 설정하고, ESM 수집을 시작하는 버튼 등으로 이루어져 있다. 스마트폰은 블루투스 스피커와 연결하여, ESM 소리를 스피커를 통해 충분한 크기로 출력한다. 기숙사 환경 내부에서의 사용자 움직임 감지 범위를 넓히기 위해, 170° 광각렌즈를 장착하였다. 스피커 모양의 종이 커버에 넣어 조립을 하여 스피커의 외형을 갖추었다. 높이 조절 가능 보면대를 사용한 이유는, 각 기숙사의 가구 배치와 관계 없이, 사용자의 움직임을 감지하기 좋은 동일한 위치 및 높이에 설치하기 위함이다.

### 3. 데이터 수집

실험자는 2 인 1 실 기숙사에 살면서, 룸메이트와 함께 실험 참여가 가능한 사람을 모집하였다. 총 20 개 기숙사 환경에서 40 명의 데이터를 각 기숙사 당 1 주일 씩 수집하였다. 10 개의 스마트 스피커를 제작하여, 2 주간에 걸쳐 수집을 완료하였다.

먼저, OT 에서 실험의 목적과 방법에 대해 설명하고, 질문의 의미와 응답하는 방식에 대해 안내하였다. 카메라 사용과 응답 녹음에 관련하여 IRB 동의서를 받았다. 연구자가 직접 실험자의 기숙사에 방문하여 장비를 적절한 위치에 설치를 한 후, 실험자들은 일주일 간 본인이 지정한 시간에 작동하는 ESM 질의에 응답하였다.

스피커가 “대화하기 좋은가요?”라고 질문하면, 현재 대화하기 좋은지 아닌지 판단하여 “네” 또는 “아니요”라고 중단가능성에 대해 답변한다. 그리고 질문이 나오는 순간 무엇을 하고 있었는지 반드시 대답하여, 중단가능성에 영향을 끼친 상황이나 행동에 대한 데이터가 수집될 수 있도록 하였다.

일주일 간의 수집이 끝난 후, 1:1 인터뷰를 통해, 각 행동 별로 “네” 또는 “아니요”라고 응답한 이유를 물어보았다. 인터뷰를 통해 각 행동이 중단가능성에 영향을 끼치는 요인을 찾을 수 있었다.

### 4. 결과

총 3572 개의 응답이 수집되었다. 그 중 실험자가 기억이 나지 않아서 보충하지 못한 응답과 실험자가 실험을 잘못 이해하여 아예 다른 응답을 한 72 개의 응답을 제외하여, 총 3500 개의 응답을 분석하였다.

#### 4.1 집 행동 카테고리

10 개의 기숙사에서 각각 가장 응답이 많았던 날의 응답들을 뽑았다. 284 개의 샘플 데이터로 집에서 하는 행동을 분류하였다. 수정하는 과정을 여러 번 반복하여 다음과 같은 19 가지 카테고리가 포함된 최종 코드북을 만들었다. (1) 일/공부, (2) 게임, (3) 인터넷/스마트폰, (4) 비디오 보기(유튜브 등), (5) 휴식, (6) 먹기, (7) 잘 준비, (8) 잠, (9) 자고 일어남, (10) 머리 말리기/빋기, (11) 옷 갈아입기, (12)

화장하기/로션바르기, (13) 기숙사 내 공용 공간(화장실, 샤워실 등) 방문, (14) 집 밖 외출, (15) 공용 공간에서 돌아오기, (16) 집 밖에서 귀가, (17) 집안일(빨래 널기, 청소 등), (18) 대화하기, (19) 온라인 인터랙션(메신저, 통화 등).

#### 4.2 결과 전체 개요

중단 가능성이 높은 행동은 (1) 집 밖에서 귀가(98%가 “네”라고 응답), (2) 인터넷/스마트폰(92%), (3) 공용 공간에서 돌아오기(89%), (4) 휴식(85%), (5) 집안일 (85%), (6) 비디오 보기(72%) 등이다. 또한, 중단 가능성이 낮은 행동은 (1) 잠(4%), (2) 잘 준비(11%), (3)일/공부(21%), (4) 공용 공간 방문(26%), (5) 게임(36%), (6) 집 밖 외출(37%) 등이다.

#### 4.3 중단 가능성 주요 요인

인터뷰 응답에 따르면 중단 가능성에 영향을 끼친 주요 요인은 일반적으로 집중도/몰입도, 급함/바쁨, 신체/정신적 요인, 음성 채널 유효성으로 나눌 수 있다.

**집중도/몰입도** : 집중 혹은 몰입을 하고 있으면 “아니요”, 하고 있지 않으면 “네”라는 응답이 우세했다. 주로 일/공부, 게임, 인터넷/스마트폰, 비디오 보기, 소셜 인터랙션과 연관되었다.

**급함/바쁨** : 일/공부, 외출, 옷 갈아입기, 화장하기의 경우에서 주로 나타나며, 일의 제출 기한 혹은 외출 시간 등 정해진 기한이 가까워서 급한 경우에는 “아니요”의 응답이 우세했다.

**신체/정신적 요인** : 피곤함, 아픔, 우울함, 졸림 등의 상태일 때 “아니요” 라고 응답하는 경향을 보였다.

**음성 채널 유효성** : 소리를 듣거나, 말을 하는 것에 제약을 만드는 행동일 경우 “아니요” 라고 응답하는 경향을 보였다. 예를 들어, 헤어 드라이기로 머리를 말릴 때 소리가 잘 들리지 않아서 또는, 룸메이트와 대화 중일 때 대화하기가 힘들었다고 보고했다.

#### 5. 논의 및 결론

스마트 스피커가 선제적 서비스를 제공하려고 말을 거는 적절한 순간을 이해하기 위해 40명 실험자의 자연스러운 생활에서 총 3572개의 데이터를 수집하였다. 집에서의 행동과 관련하여, 중단가능성에 영향을 끼치는 주요 요인들을 찾았다. 사용자의 움직임 및 사회적 존재가 중단가능성에 어떻게 영향을 주는지 추가로 연구하여 선제적 대화 서비스 제공에 적절한 순간에 대해 더욱 연구해볼 수 있을 것이다.

#### 사사의 글

This research was supported by the KAIST-KU Joint Research Center, KAIST, South Korea (N11190249).

#### 참고 문헌

1. Piotr D. Adamczyk and Brian P. Bailey. 2004. If Not Now, when?: The Effects of Interruption at Different Moments Within Task Execution. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '04). ACM, New York, NY, USA, 271–278.
2. Ana Paula Chaves and Marco Aurélio Gerosa. 2019. How should my chatbot interact? A survey on human-chatbot interaction design. CoRR abs/1904.02743 (2019). arXiv:1904.02743
3. cnet. 2019. What Amazon’s Alexa will tell us in 2019. <https://www.cnet.com/news/what-amazon-alexa-will-tell-us-in-2019/>
4. Joyce Ho and Stephen S. Intille. 2005. Using Context-aware Computing to Reduce the Perceived Burden of Interruptions from Mobile Devices. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05). ACM, New York, NY, USA, 909–918.
5. Shamsi T. Iqbal and Brian P. Bailey. 2005. Investigating the Effectiveness of Mental Workload As a Predictor of Opportune Moments

- for Interruption. In CHI '05 Extended Abstracts on Human Factors in Computing Systems (CHI EA '05). ACM, New York, NY, USA, 1489–1492.
6. Auk Kim, Woohyeok Choi, Jungmi Park, Kyeyoon Kim, and Uichin Lee. 2018. Interrupting Drivers for Interactions: Predicting Opportune Moments for In-vehicle Proactive Auditory-verbal Tasks. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 4, Article 175 (Dec.2018), 28 pages.
  7. SeungJun Kim, Jaemin Chun, and Anind K. Dey. 2015. Sensors Know When to Interrupt You in the Car: Detecting Driver Interruptibility Through Monitoring of Peripheral Interactions. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 487–496.
  8. Abhinav Mehrotra, Mirco Musolesi, Robert Hendley, and Veljko Pejovic. 2015. Designing Content-driven Intelligent Notification Mechanisms for Mobile Applications. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. ACM, New York, NY, USA, 813–824.
  9. TechRepublic. 2019. Google Home: Cheat sheet. <https://www.techrepublic.com/article/google-home-the-smart-persons-guide/>