CHI 2024

Exploring Context-Aware Mental Health Self-Tracking Using Multimodal Smart Speakers in Home Environments

Jieun Lim*, **Youngji Koh***, Auk Kim, Uichin Lee

KAIST KNU

* Equal contribution

Mental Health: A Rising Global Concern

1 in 8 people worldwide
live with a mental health problem

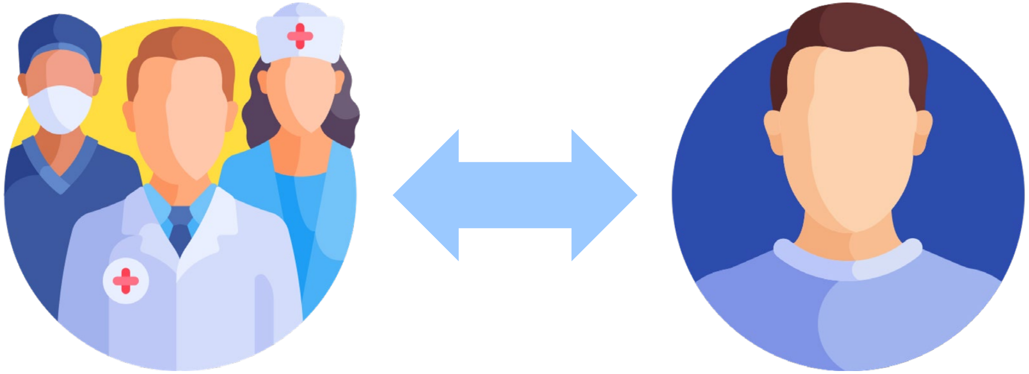
Institute of Health Metrics and Evaluation. 2019 Global Burden of Disease (GBD)

Self-Tracking: A Method for Mental Health Monitoring

Support Self-Reflection

Enhance self-awareness
of mental health

Help Clinical Decision-Making

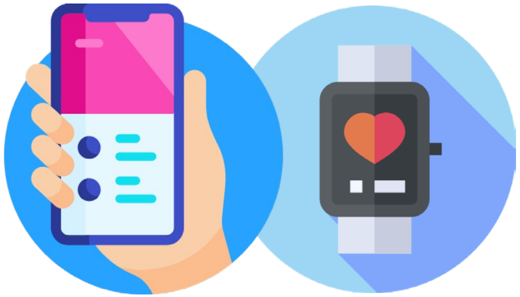


Bridge the information gap between
healthcare stakeholders and patients

ESM for Mental Health Self-Tracking

Diversity of the Experience Sampling Method (ESM) Technologies

Paper & Pen Method



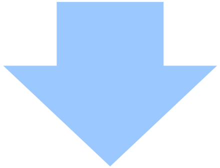
Mobile/Wearable Technology
(Wang et al., 2014)

Smart Speakers in Home Environments
(Wei et al., 2021)

Wang et al. "StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones." Ubicomp '14
Wei et al. "Understanding user perceptions of proactive smart speakers" IMWUT '21

Mental Health Self-Tracking in Home Environments

People with mental health issues often stay **indoors**



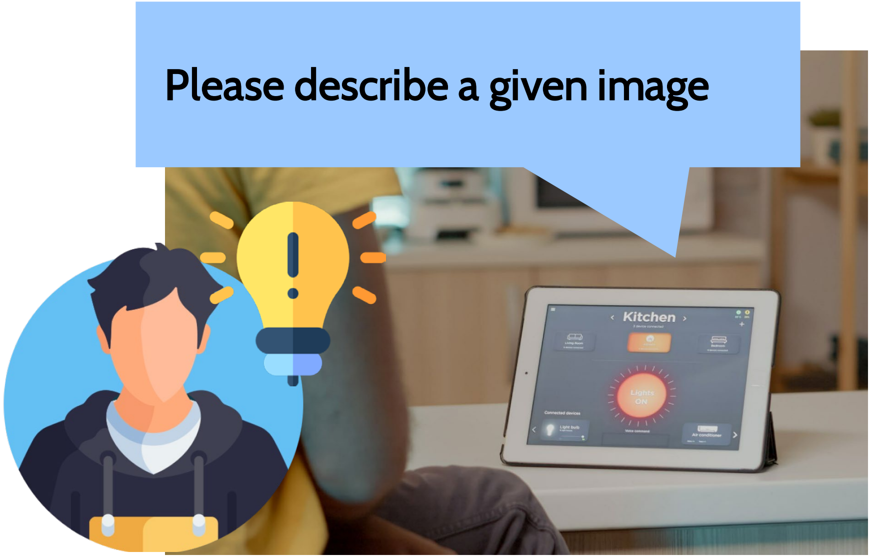
Need for **mental health self-tracking technology in homes** is increasing

Mental Health Self-Tracking with Multimodal Smart Speakers

Mental Health ESM often requires visual-verbal tasks

(e.g., Image description task for diagnosing depression or cognitive impairment)

Please describe a given image

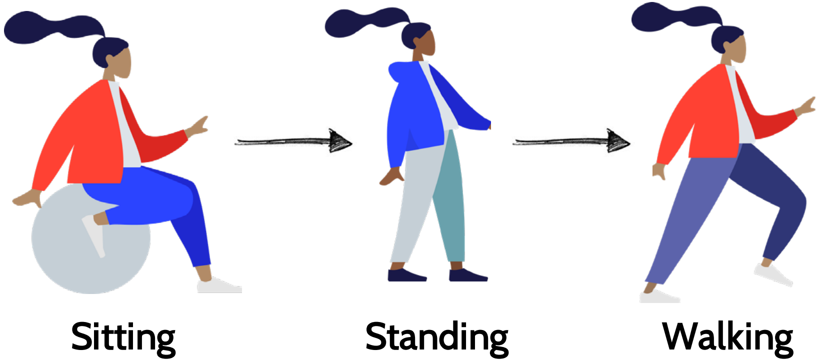


Please describe a given image

Opportune Timing for ESM Design in Home Environments

Identifying opportune moments in previous studies

Task Breakpoints
in Mobile/Desktop Environment
(Adamczyk et al., 2004)

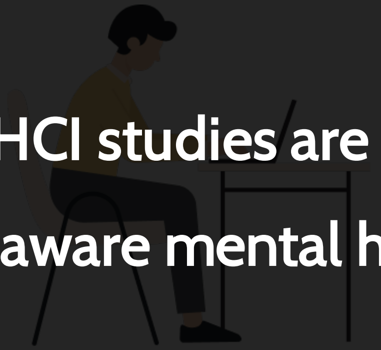


User Activity Transitions
in Mobile Environment
(Fischer et al., 2011)

Adamczyk et al., "If not now, when? The effects of interruption at different moments within task execution.", CHI ’04
Fischer et al., "Investigating episodes of mobile phone activity as indicators of opportune moments to deliver notifications.", MobileHCI ’11

Opportune Timing for ESM Design in Home Environments

Identifying opportune moments in previous studies



Task Breakpoints
in Desktop Environment
(Adamczyk et al., 2004)

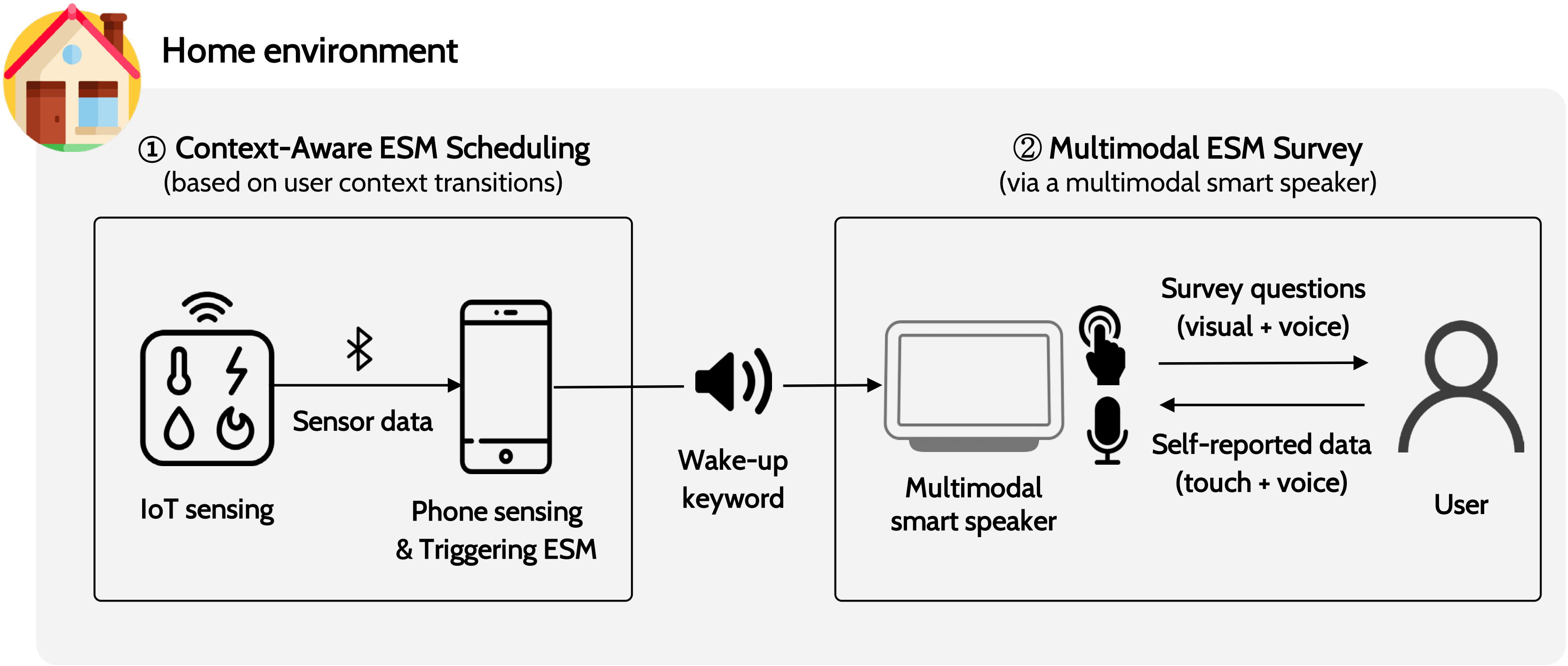
User Activity Transitions
in Mobile Environment
(Fischer et al., 2011)

HCI studies are still to investigate user experiences of context-aware mental health self-tracking using multimodal speakers

Adamczyk et al., "If not now, when? The effects of interruption at different moments within task execution.", CHI '04
Fischer et al., "Investigating episodes of mobile phone activity as indicators of opportune moments to deliver notifications.", MobileHCI '11

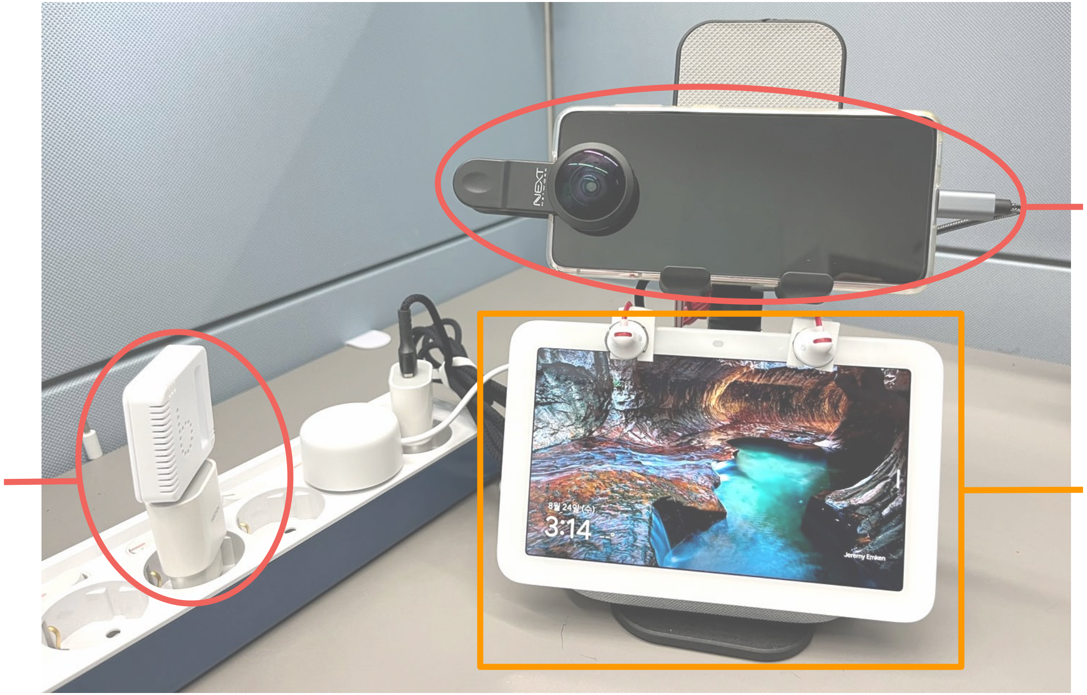
Context-Aware Self-Tracking System using Multimodal Speakers

Our System:



Context-Aware Self-Tracking System using Multimodal Speakers

Our System Prototype:



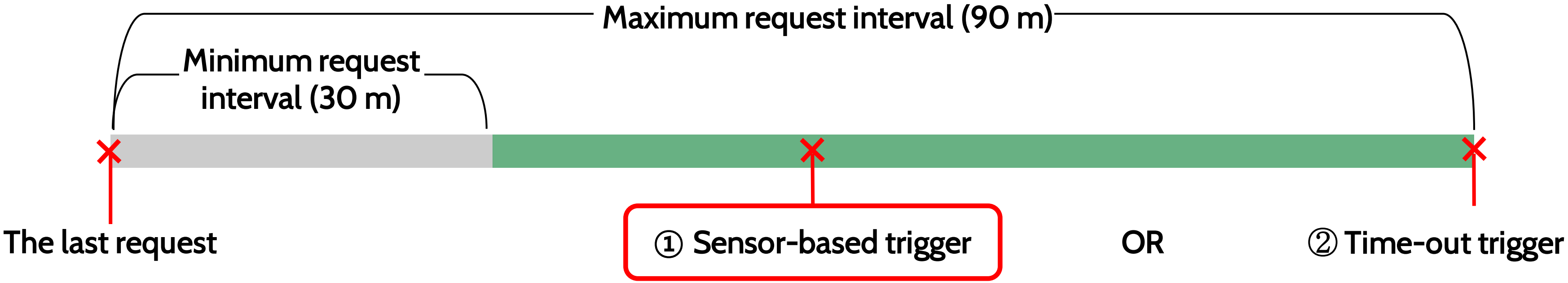
①-1 IoT Sensor
: Collect CO2 data

①-2 Speaker triggering app with wide-angle lens
: Collect noise, brightness, and # of people

② Multimodal speaker
: Provide voice and touch interactions

Context-Aware ESM Scheduling

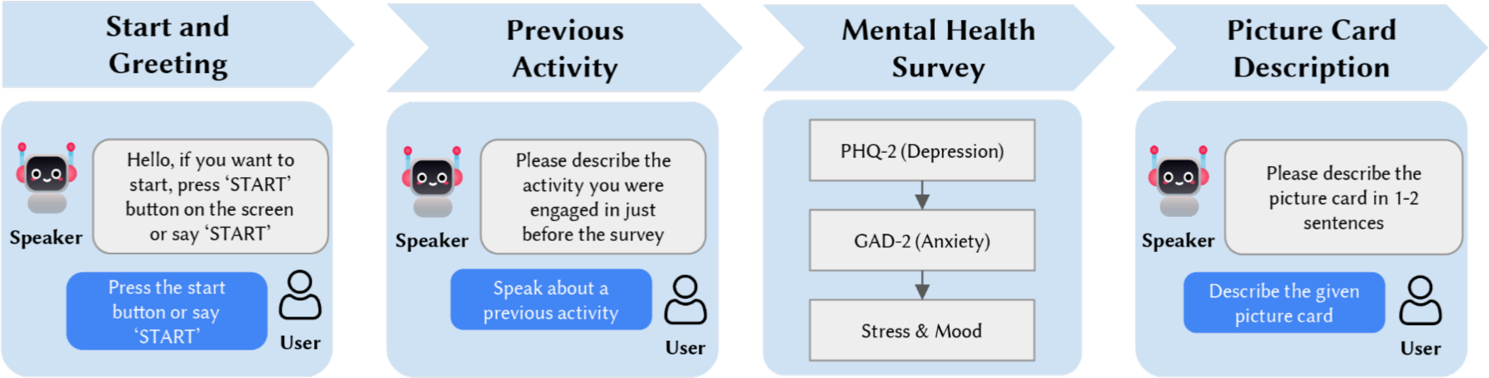
Determine **opportune moments for ESM** requests in **home environments**


Maximum request interval (90 m)
Minimum request interval (30 m)
The last request
① Sensor-based trigger
OR
② Time-out trigger

- Detect user context transitions using sensors:
  - **Auditory channel availability** using *Noise Sensor*
  - **Proximity to smart speakers** using *Light Sensor, CO2 Sensor, Camera*

Multimodal ESM Survey

ESM Task Steps and User Interface



Start and Greeting

Hello, if you want to start, press 'START' button on the screen or say 'START'

Speaker

Press the start button or say 'START'

User

Previous Activity

Please describe the activity you were engaged in just before the survey

Speaker

Speak about a previous activity

User

Mental Health Survey

PHQ-2 (Depression)

GAD-2 (Anxiety)

Stress & Mood

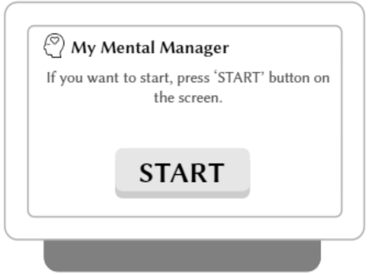
Picture Card Description

Please describe the picture card in 1-2 sentences

Speaker

Describe the given picture card

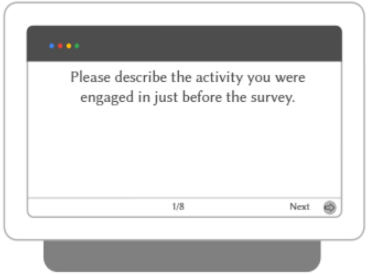
User



My Mental Manager

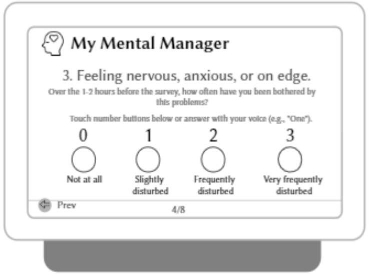
If you want to start, press 'START' button on the screen.

(1) Start and Greeting



Please describe the activity you were engaged in just before the survey.

(2) Previous Activity



My Mental Manager

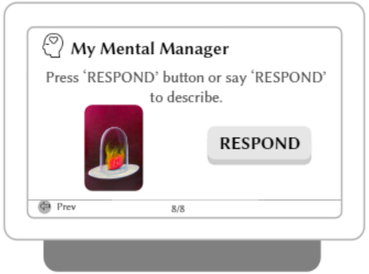
3. Feeling nervous, anxious, or on edge.

Over the 1-2 hours before the survey, how often have you been bothered by this problems?

Touch number buttons below or answer with your voice (e.g., "One").

0	1	2	3
○	○	○	○
Not at all	Slightly disturbed	Frequently disturbed	Very frequently disturbed

(3) Mental Health Survey



My Mental Manager

Press 'RESPOND' button or say 'RESPOND' to describe.

(4) Picture Card Description

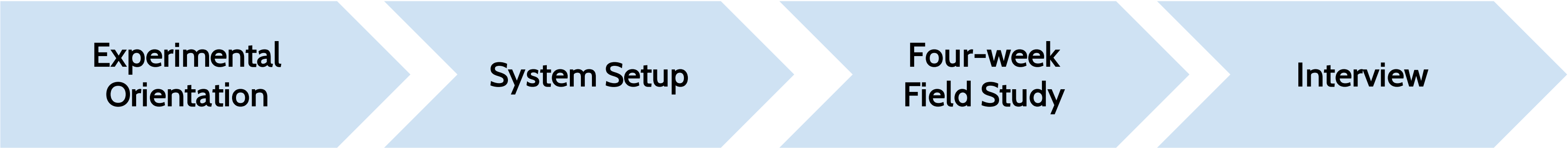
Research Questions

1. How do users perceive proactive mental health self-tracking using multimodal speakers?
2. How do ESM compliance rates change across different context transitions?
3. What are the preferred interaction modalities for responding to multimodal speakers?

Field Study Methods

Participants (N=20)

- Recruitment criteria
    - People who were diagnosed with at least mild depression (a PHQ 9 score of 5 or higher)
    - People who had private spaces at home or were single-person households
    - People who spent a minimum of 5 hours daily in their room, excluding sleep time



Experimental Orientation

System Setup

Four-week Field Study

Interview

Research Questions

1. **How do users perceive proactive mental health self-tracking using multimodal speakers?**
2. How do ESM compliance rates change across different context transitions?
3. What are the preferred interaction modalities for responding to ESM requests?

RQ1: Overall User Experience of Proactive Mental Health Self-Tracking using Multimodal Smart Speakers

Positive Aspects: **Proactive system** helped users to **gauge mental health status**

"I don't usually get a chance to ask myself these questions (related to mental health). But every hour or two, the system asks you how you're feeling or how stressed you are, and *<u>it gives me more opportunities to think about whether I've just gotten stressed</u>*." - P18

"Before, I had no idea about my moods. But when I got a chance to think about it (through the survey), I was like, *<u>'I see ... what was happening' and could relieve negative emotions</u>*." - P19

RQ1: Overall User Experience of Proactive Mental Health Self-Tracking using Multimodal Smart Speakers

Positive Aspects: **Human-like factors** made users **engaging to ESM requests**

"I was more focused on the question because the *speaker asked questions verbally*. Also, there's only one question on the screen. *It makes me concentrate on each question.*" - P12

"I felt like it's a person because the timing was not exactly regular. It's usually unpredictable when someone will contact you. So, *the timing of the speaker talking to me made me feel like a person.*" - P13

RQ1: Overall User Experience of Proactive Mental Health Self-Tracking using Multimodal Smart Speakers

Negative Aspects: **Machine-like interaction style** led to **boredom**

"The questions and pictures are repeated over and over again. *As the experiment progressed, I felt bored because the system became more habitual and predictable*." - P15

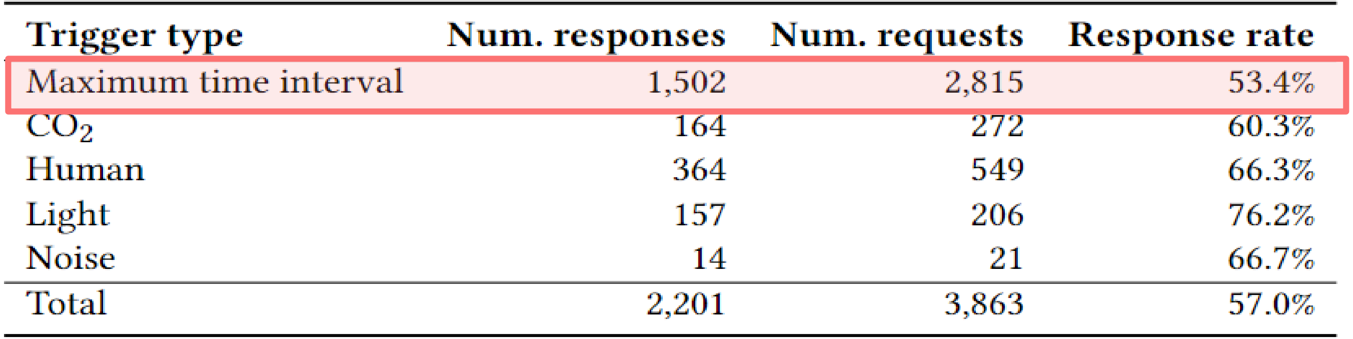
"I think it was annoying to keep asking the same questions over and over again. So, *there was a decrease in the sincerity of responses*." - P17

Research Questions

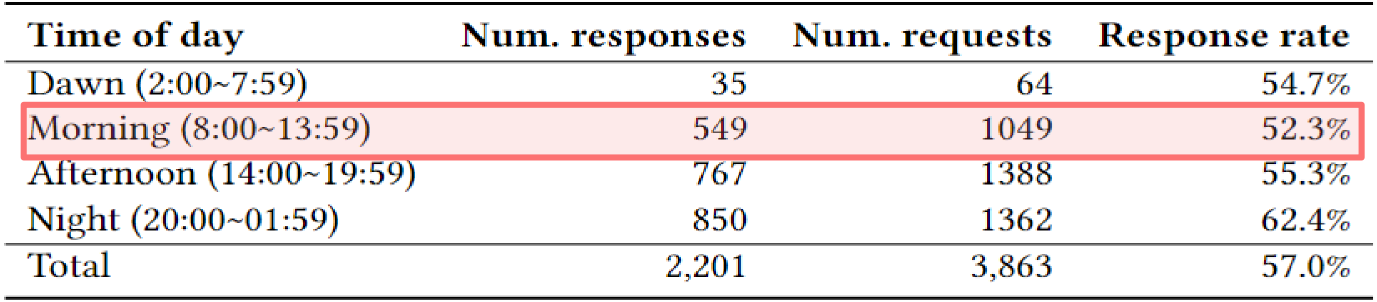
1. How do users perceive proactive mental health self-tracking using multimodal speakers?
2. **How do ESM compliance rates change across different context transitions?**
3. What are the preferred interaction modalities for responding to ESM requests?

RQ2: ESM Response Rates across Different Context Transition

ESM response rates were **lower** in the **time-out trigger condition** and **morning**



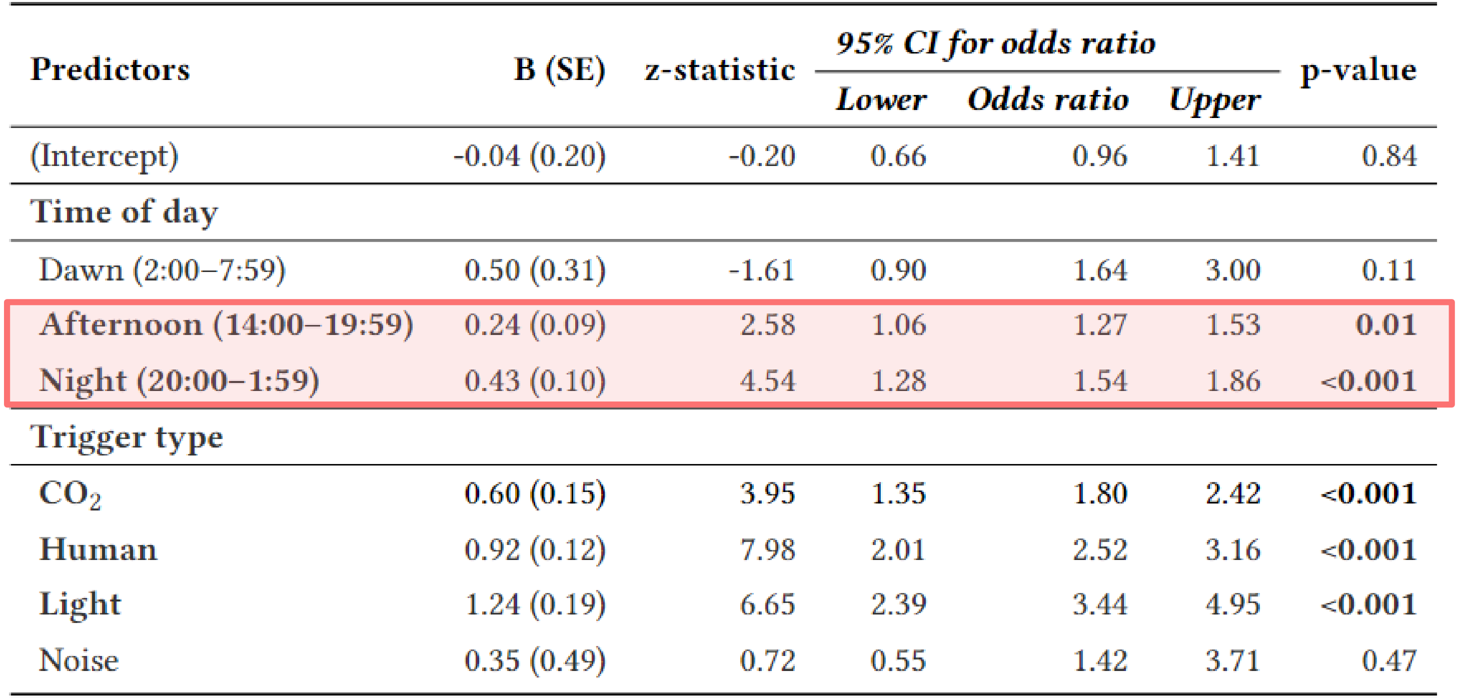
Trigger type	Num. responses	Num. requests	Response rate
Maximum time interval	1,502	2,815	53.4%
CO_2	164	272	60.3%
Human	364	549	66.3%
Light	157	206	76.2%
Noise	14	21	66.7%
Total	2,201	3,863	57.0%



Time of day	Num. responses	Num. requests	Response rate
Dawn (2:00~7:59)	35	64	54.7%
Morning (8:00~13:59)	549	1049	52.3%
Afternoon (14:00~19:59)	767	1388	55.3%
Night (20:00~01:59)	850	1362	62.4%
Total	2,201	3,863	57.0%

RQ2: ESM Response Rates across Different Context Transition

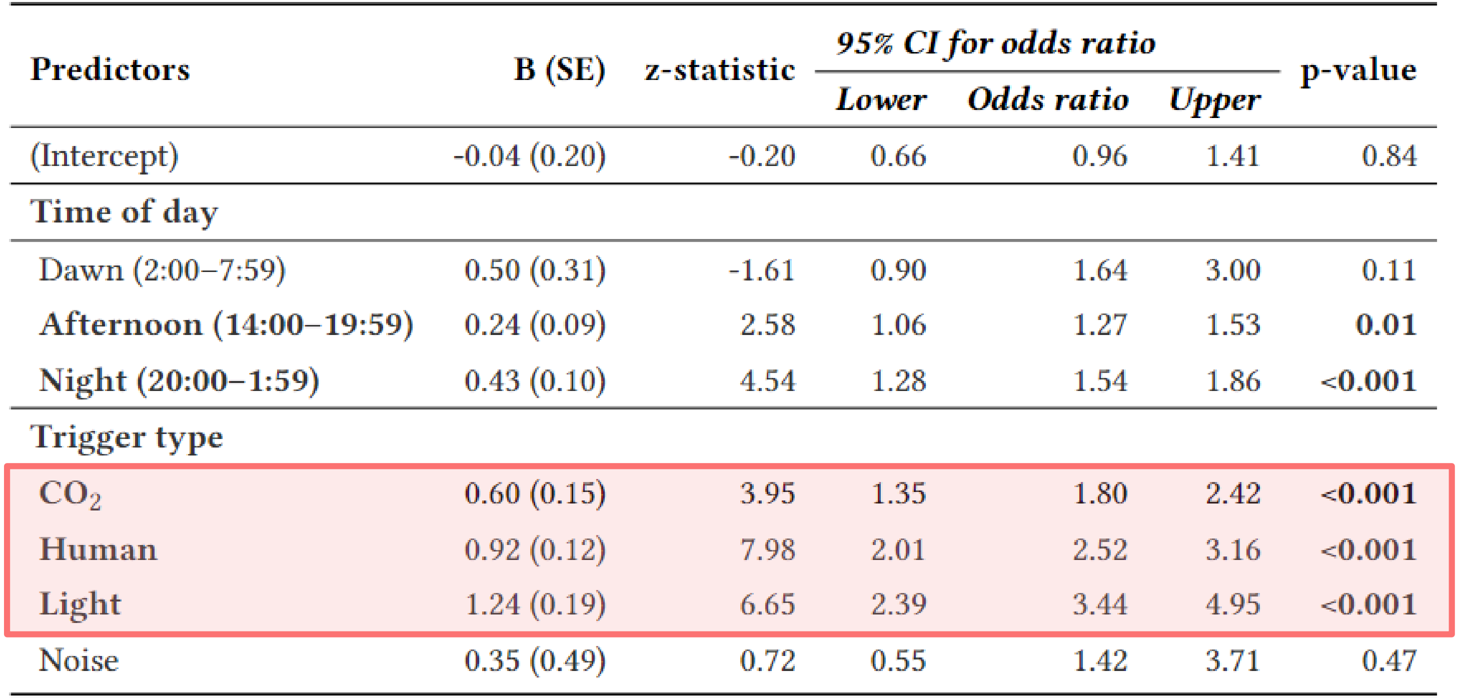
Responded more to ESM in the **afternoon** and **night** than in the morning



Predictors	B (SE)	z-statistic	*95% CI for odds ratio*			p-value
			Lower	*Odds ratio*	*Upper*	
(Intercept)	-0.04 (0.20)	-0.20	0.66	0.96	1.41	0.84
Time of day						
Dawn (2:00–7:59)	0.50 (0.31)	-1.61	0.90	1.64	3.00	0.11
Afternoon (14:00–19:59)	0.24 (0.09)	2.58	1.06	1.27	1.53	**0.01**
Night (20:00–1:59)	0.43 (0.10)	4.54	1.28	1.54	1.86	**<0.001**
Trigger type						
CO_2	0.60 (0.15)	3.95	1.35	1.80	2.42	**<0.001**
Human	0.92 (0.12)	7.98	2.01	2.52	3.16	**<0.001**
Light	1.24 (0.19)	6.65	2.39	3.44	4.95	**<0.001**
Noise	0.35 (0.49)	0.72	0.55	1.42	3.71	0.47

RQ2: ESM Response Rates across Different Context Transition

Responded more to ESM **when users were near the speakers** (CO2, Human, Light)



Predictors	**B (SE)**	**z-statistic**	***95% CI for odds ratio***			**p-value**
			Lower	***Odds ratio***	***Upper***	
(Intercept)	-0.04 (0.20)	-0.20	0.66	0.96	1.41	0.84
Time of day						
Dawn (2:00–7:59)	0.50 (0.31)	-1.61	0.90	1.64	3.00	0.11
Afternoon (14:00–19:59)	0.24 (0.09)	2.58	1.06	1.27	1.53	**0.01**
Night (20:00–1:59)	0.43 (0.10)	4.54	1.28	1.54	1.86	**<0.001**
Trigger type						
CO_2	0.60 (0.15)	3.95	1.35	1.80	2.42	**<0.001**
Human	0.92 (0.12)	7.98	2.01	2.52	3.16	**<0.001**
Light	1.24 (0.19)	6.65	2.39	3.44	4.95	**<0.001**
Noise	0.35 (0.49)	0.72	0.55	1.42	3.71	0.47

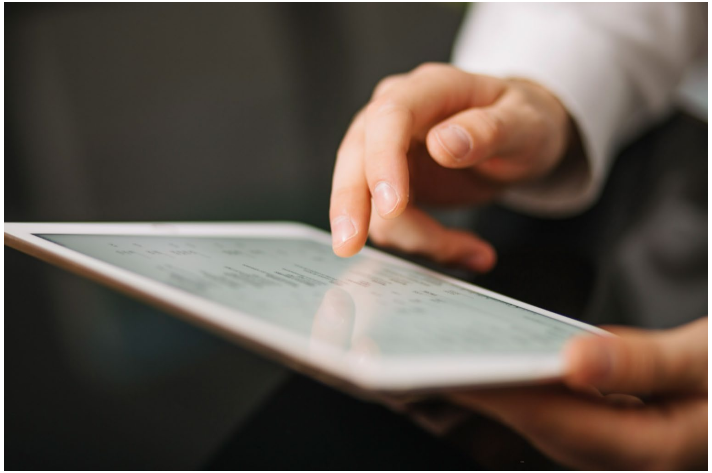
Research Questions

1. How do users perceive proactive mental health self-tracking using multimodal speakers?
2. How do ESM compliance rates change across different context transitions?
3. **What are the preferred interaction modalities for responding to ESM requests?**

RQ3: Interaction Modality Preferences based on User Context

Most users **preferred to respond with GUI** over VUI for multiple choice questions

Reasons for Using a **GUI**



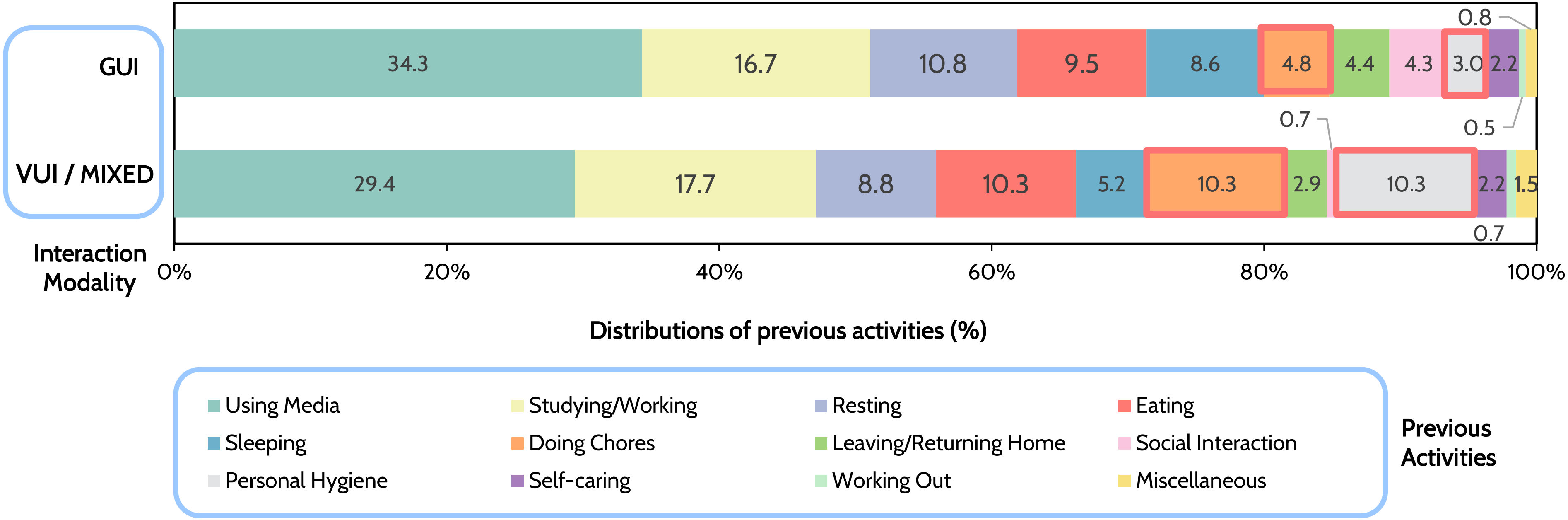
Limitation of VUI & Familiarity with GUI

Reasons for Using a **VUI**

Situations when hands are occupied

RQ3: Interaction Modality Preferences based on User Context

VUI/MIXED was more frequent than GUI in **doing chores and personal hygiene**



Summary of Key Findings

- RQ1: Overall User Experience
  - **Proactive self-tracking** can increase self-reflection regarding mental health
  - **Human-likeness** helped users engaging in answering mental health questions

- RQ2: ESM Response Rates across Different Context Transition
  - **User response rates** improved when ESM are requested in **context transitions**

- RQ3: Interaction Modality Preferences based on User Context
  - **Users' previous contexts** influenced their **interaction modality selection**

Discussion

Context Awareness for Modality Selection and Adaptation

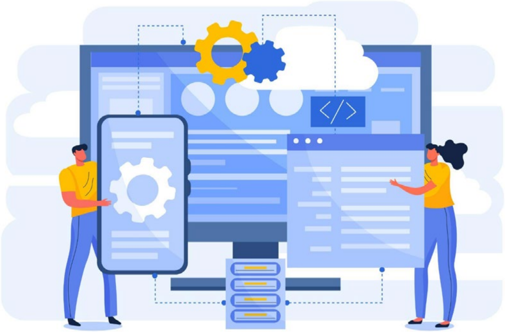
Detect user contexts and **adaptively select** an appropriate **interaction modality**

Sensor Selection in Home Environments

Use sensors that are capable of **detecting multiple users**

Design Implications

Consideration for ESM System Design



Consider **context-sensing in the multimodal ESM** interaction design

Consideration for Engaging ESM Interaction Design

Vary the tone and content as in context-tailored adaptations

CHI 2024

Exploring Context-Aware Mental Health Self-Tracking Using Multimodal Smart Speakers in Home Environments

Youngji Koh, KAIST
youngji@kaist.ac.kr

Takeaway Notes

- **Context-awareness** improves user compliance of ESM surveys and makes user feel it like human
- HCI studies continue to investigate **context-tailored adaptations** for making user engaging in ESM

*This research was supported by LGE-KAIST Digital Health Research Center (DHRC) and by the National Research Foundation (NRF) funded by the Korean government (MSIT)

KAIST KNU