My Voice as a Daily Reminder: Self-Voice Alarm for Daily Goal Achievement

Jieun Kim
jk2345@cornell.edu
Cornell University
Ithaca, New York, USA

Hayeon Song
songhy@skku.edu
Sungkyunkwan University
Seoul, South Korea

ABSTRACT

Sticking to daily plans is essential for achieving life goals but challenging in reality. This study presents a self-voice alarm as a novel daily goal reminder. Based on the strong literature on the psychological effects of self-voice, we developed a voice alarm system that reminds users of daily tasks to support their consistent task completion. Over the course of 14 days, participants (N = 63) were asked to complete daily vocabulary tasks when reminded by an alarm (i.e., self-voice *vs.* other-voice *vs.* beep sound alarm). The self-voice alarm elicited higher alertness and uncomfortable feelings while fostering more days of task completion and repetition compared to the beep sound alarm. Both self-voice and other-voice alarms increased users' perceived usefulness of the alarm system. Leveraging both quantitative and qualitative approaches, we provide a practical guideline for designing voice alarm systems that will foster users' behavioral changes to achieve daily goals.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**.

KEYWORDS

Self-Voice, Daily Reminder, Mobile Alarm Design, Habit Formation, Language Learning

ACM Reference Format:
Jieun Kim and Hayeon Song. 2024. My Voice as a Daily Reminder: Self-Voice Alarm for Daily Goal Achievement. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA*. ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3613904.3641932

1 INTRODUCTION

A daily routine plays a significant role in changing one's lifestyle and achieving long-term goals. People make daily plans and consistent efforts to implement these plans. Adhering to a daily routine, however, is challenging in real life, as people are often distracted by miscellaneous tasks, chores, or other daily activities. These distractions take their attention away from their main goals or priorities and prevent them from remembering and completing their targeted tasks. Sticking to daily plans becomes even more demanding when the targeted task is new and unfamiliar [48, 62]. Consequently, even people who have strong behavioral intentions frequently fail to implement their plans [72, 74].

Recent advances in mobile technologies have introduced effective interventions that enable individuals to manage their daily behaviors. Leveraging mobile tracking technology, users can monitor their day-to-day task progress and adapt their behaviors toward meeting their goals [36]. Mobile alarms or notifications integrated into smartphones and wearable devices serve as daily reminders that increase users' awareness of their daily tasks [24, 82]. Many mobile alarm systems have been designed to assist individuals in achieving their long-term goals such as eating a healthy diet [87], regular exercise [63], medication adherence [58], improving sleeping patterns [57], and enhancing educational practice such as mathematics [25] and second language learning [94].

While mobile alarms have been researched for their effectiveness in behavior change, it remains unclear which alarm sound can effectively increase users' daily goal achievement. As a novel approach to nudge mobile users toward achieving their daily goals, this study introduces a *Self-Voice Alarm*, a mobile alarm that utilizes one's own voice as a daily goal reminder. Hearing one's own voice is not a common practice but serves as a unique experience. Humans' innate preference for self-relevant information allows them to process their own voices more effectively than other auditory information. A strong body of research highlights the potential of self-voice to automatically capture users' attention and facilitate their goal-oriented behavior. Increasing evidence from neuroscience studies indicates that listening to self-voice can promote brain activation, which elicits high arousal of emotions [2, 80] and increased attention [15, 23, 34]. Despite the prominent effects of self-voice in human emotional and cognitive processing, to our best knowledge, none of the existing studies in Human-Computer Interaction have attempted to employ self-voice in a self-regulatory alarm system. Furthermore, there has been limited research on the longitudinal effects of alarm sounds on users' attitudinal and behavioral changes.

The current study explores the psychological and behavioral impacts of self-voice alarms with the goal of enhancing users' automatic and consistent task achievement. By conducting an extensive review of the literature, we first examined how a self-voice alarm influences users' emotions and enjoyment of the task. Then, we observed if the self-voice can elicit their automatic responses to the alarm by considering both behavioral and self-report measures. As an effective alarm aims to encourage users' continuous and consistent task achievement, we examined how the self-voice alarm can increase task completion and repetition, while affecting the users' performance on the task. To assess our newly proposed alarm system, we also investigated users' evaluation of the alarm in terms

CHI '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0330-0/24/05
https://doi.org/10.1145/3613904.3641932

of its perceived ease of use, perceived usefulness, and intention to use it. We conducted a field experiment where university students (N = 63) were asked to use a mobile application that prompted a daily reminder using three different alarm sounds (i.e., *self-voice vs. other-voice vs. electronic beeping*). Then, we observed whether they decided to complete the daily tasks consistently depending on the alarm sounds over the two weeks of the experiment.

Our investigation sheds light on the potential for self-voice alarms to serve as a personalized intervention strategy and assist individuals in achieving their daily goals. While existing studies have acknowledged the distinctive impacts of the human voice, particularly self-voice, as an alarm sound on neurological, psychological, and behavioral changes, none have empirically examined these impacts within the context of goal achievement and mobile interventions. The effectiveness of voice alarms will be further enhanced with advances in voice synthesis technologies. For instance, voice cloning aims to generate a naturally synthesized voice that closely resembles an individual's voice by natural language processing techniques [3, 95]. The emerging technologies that can create realistic and personalized voices for user interfaces allow users to experience more engaging and persuasive interactions [13]. Accordingly, the design of user-specific alarm systems and their impact on users' perceptions and behavior deserves more attention.

2 LITERATURE REVIEW

2.1 Building a Daily Routine for Everyday Goals

A daily routine plays a vital role in achieving long-term behavioral goals in the pursuit of personal growth and self-improvement. Routines help individuals accomplish tasks easily, requiring minimal conscious effort for decision-making and enhancing task productivity [26, 27, 41]. Studies on habit formation have documented how to construct regular and consistent behavioral patterns. Habits are automatic actions that are learned through consistent behavioral discipline [41, 54]. Transforming a new behavior into a habit requires two stages: *habitual instigation* and *habitual execution* [27]. *Habitual instigation* is triggered by contextual cues associated with the goal behavior, followed by a sequence of actions that evolves as a higher-order target action. When a mental association between the contextual cues and a set of sub-actions is strengthened by repetition [54], *habitual execution* can be maintained as an automatic response to the trigger [85, 93].

In the initial stage of habit formation, increasing the frequency of the goal behavior in daily life is essential to gain automaticity and develop it as a long-term behavior [42, 43]. In habit formation, giving a contextual cue is important to trigger goal-oriented behavior. A contextual cue related to the target task can elicit the individual's sense of control over their own actions [75] and promote them to move toward the goal [60]. In this sense, our study aims to design a self-regulatory mobile reminder that encourages daily task completion and increases repetitions. As the entire process of habit formation takes more than 18 days [42], our study focuses on the first two weeks as the initial stage of habit formation and observes whether individuals conduct goal behaviors every day and maintain them for a given period of time. Adopting a mobile application, we developed a unique mobile alarm system that can effectively remind individuals of their daily goals and help them achieve the tasks consistently.

2.2 Human Voices as an Alarm Sound

As an alarm cue that can capture individuals' attention and motivate them to complete tasks, we employed the human voice as the alarm sound. In designing voice user interfaces, human voices have been a key design material that influences the production and perceptions of speech interactions [81]. Voice is a distinctive human trait that distinguishes one from others. Evolutionarily, humans have the advantageous ability to identify voices and discriminate them among multiple sounds [14, 52]. People automatically notice human voices and recognize them better than any other sounds. This is partly because paralinguistic cues of voice (e.g., pitch and variation) can provide the speaker's relevant physical and social characteristics, such as age, gender, personality, and emotional states [53, 96]. These acoustic cues form a voice signature that helps hearers identify the speaker and remember their voice in the future [6, 7].

Studies in the fields of ergonomics and cognitive engineering have shown the advantages of human voice in processing auditory information. Ulfvengren compared the cognitive capacity among various auditory stimuli, including simple tone sounds, environmental sounds (e.g., bird calls), speech, animal sounds, and traditional alarm sounds (e.g., beeps, bells, and buzzers). Speech alarms required the smallest effort to learn and remember the sound compared to the other sounds while the simple tone alarms required the longest time to learn [86]. Speech alarms can also deliver instructive messages compared to a simple, abstract ringtone, which benefits users in understanding the context of the alarm [46].

Highlighting the distinctiveness of voice, a voice alarm has been suggested as an emergency notification, such as for fire emergencies or aircraft warnings. Researchers have observed that a voice signal as an evacuation alarm makes the evacuation efficient and safe, delivering information faster and more accurately compared to a bell alarm [37, 46, 65, 66, 91]. For instance, voice alarm messages for an emergency evacuation prompted a shorter response time than an electronic sound alarm [68]. Studies particularly spotlight voice familiarity to facilitate processing voice information, as a familiar voice can increase speech recognition even in the presence of other competing sounds [33, 45]. An alarm with the sound of a parent's voice can successfully awaken a child from stage 4 deep sleep, outperforming conventional smoke alarm sounds [78]. A smoke alarm using a maternal voice awakened around 90% of sleeping children and prompted 85% of them to escape [77].

Voice familiarity influences how listeners perceive the speaker of a voice. When speakers use the same accent as the listener, it often leads to higher ratings of the speakers, aligning with the similarity-attraction effect [17, 18], although some studies show contrasting results [55]. Voice familiarity also matters in efficiently detecting and processing incoming information. Familiarity with the speaker helps speech comprehension and recognition [28], enabling more attentive and engaged listening [13]. Moreover, familiar voices can strengthen memory recall, as shown in the comparison between the voices of a familiar celebrity and an arbitrary speaker [22].

2.3 Effects of Self-Voice

People tend to process self-related information more importantly and positively. This self-biased trend is connected to the self-referencing effect, which is that any stimulus relevant to oneself can be remembered and recalled easily compared to other-relevant information [70]. The self-referencing effect has been adopted as an effective strategy in human-computer interaction to facilitate message processing and persuasion in digital environments. For instance, using a self-image in a digital advertisement elicited a positive brand attitude and higher purchase intention [1]. The use of self-referencing messages as an intervention to mitigate the risk of obesity led to an increase in engagement with physical activity and healthy diet [92].

Extensive research in neuroscience suggests that self-relevant stimuli elicit high arousal of emotions, leading to attentional bias. Given the limitation of human cognitive capacity, attentional resources tend to be allocated to emotional stimuli compared to neutral stimuli [12, 90]. Self-relevant information evokes electrodermal responses, indicating a high level of emotional reactivity [80], and activates the left inferior frontal and right anterior cingulate cortex, reflecting an attentional bias toward emotional signals [2].

This self-biased processing is also present when one's own voice is heard. People are good at recognizing their own voices compared to others' voices [30, 71, 73]. Moreover, listening to a pre-recorded self-voice engages higher emotional activation and greater attentional resources compared to others' voices [15, 34, 51]. In this regard, the emotional arousal triggered by a self-voice alarm can provoke affective experiences and promote cognitive processing that influences behavioral decisions [39].

2.3.1 Emotional Effects. Emotion is significant in determining how people perceive the environment and make behavioral decisions in everyday life. Listening to one's own voice, in fact, is not a pleasant experience. Holzman and Rousey [30] compared the ratings of voice recordings and found that listening to one's own voice induces negative reactions compared to listening to other's voices. If people do not know that their voice was used in the recording, however, the recording of their voice was evaluated as more attractive than others' voices [32]. These conflicting findings imply that the identification of a given voice critically influences the hearer's evaluation of the voice.

One possible explanation for why people do not like their recorded voice is that what people expect to hear differs from what they actually hear. Voice confrontation [29] is the cognitive dissonance between one's expectation and actual perception of their voice. The sound of a voice that we hear when we speak is different from the sound of our recorded voice due to the different routes of sound transmission. When we hear our own voice while speaking, the sound is transmitted directly through our bones to our ears. On the other hand, the sound of the voice recording is transmitted through the air to the ears, which makes the sound low-pitched and unfamiliar [49, 76]. This discrepancy creates distorted feelings from the listeners [29], which may lead them to feel awkward and display unpleasant and uncomfortable responses.

Aligned with the discrepancy, studies have suggested that increased voice familiarity (e.g., self and kinship) can lead to eerie feelings when the voice is recorded and mediated by voice user interfaces [13, 38]. Based on previous findings explaining how listening to self-voice elicits uncomfortable feelings, we hypothesize that a self-voice alarm will increase users' discomfort compared to other-voice alarms or beep sound alarms.

H1. Self-voice alarm will elicit *discomfort* compared to (a) other-voice and (b) beep sound alarm.

While previous studies have illustrated that self-voice might increase the feeling of discomfort in general, the emotions behind the discomfort have not been clarified yet. The discomfort produced by self-voice alarms may be accompanied by other emotions relevant to uncomfortable feelings. For instance, the discrepancy can lead to a sense of embarrassment and annoyance stemming from the unanticipated voices. Embarrassment emerges in a situation where individuals encounter socially awkward or humiliating events for themselves [50]. Therefore, when a self-voice alarm is sounded, one's concern about how other people might think about one's voice may induce embarrassment. Moreover, this will evoke the emotion of shame from being self-conscious about how other people might judge the voice and alarm [84].

In this study, we empirically test the effects of a self-voice alarm on a wide range of emotions in comparison with other auditory alarms. In particular, we investigate how the users will perceive the self-voice as an alarming sound compared to unfamiliar others' voices or electronic alarm sounds, proposing a research question:

RQ1. What other *emotions* does the self-voice alarm elicit from users compared to other types of alarm?

An uncomfortable feeling can be motivational in the context of behavior change. Pekrun [59] shed light on the importance of negative emotions in motivation and goal achievement, explaining that experiencing negative emotions in certain circumstances can activate motivational responses. While positive emotions elicit comfortable thinking that "all is good", negative emotions can evoke problem-solving thoughts and activate goal-directed behaviors. Rekimoto and Tsujita [69] proposed inconvenient interactions as a behavioral mechanism that encourages users to perform target tasks in order to relieve a sense of discomfort. In this view, a mild negative emotion can help users engage in new behaviors, leading to a long-term benefit for behavior change.

Contrary to the traditional design of usability-focused interaction that considers inconvenience or discomfort as a flaw, a short-term inconvenience can induce continuous interaction with an interface [8]. Using the benefits of inconvenient design, Oh et al. [56] designed a wake-up alarm that causes discomfort until the user completes a specific task, such as solving a math problem and taking a picture of it. Although the alarm may elicit an immediate feeling of unpleasantness, it may lead to users' satisfaction and fulfillment since it helps them successfully wake up [56]. From this perspective, the uncomfortable and awkward feelings elicited by self-voice as an alarm will nudge individuals to complete their daily tasks. Accordingly, we expect that a self-voice alarm will increase the daily task completion during the given period of time compared to the other types of alarms.

H2. Self-voice alarm will increase *daily task completion* compared to (a) other-voice and (b) beep sound alarm.

Since our study employs self-voice as the alarm sound for task reminders, users' perception of the alarm sound may affect their enjoyment of the task. Enjoyment is the key factor that determines the intrinsic motivation of the task, evoking the feeling that the task is inherently interesting and enjoyable [21]. Given that self-voice elicits negative and uncomfortable emotions [29, 30], a self-voice alarm may reduce task enjoyment. However, self-biased processing can enhance users' enjoyment of performing the task when they hear their own voice as the task reminder. Since our study is a pioneering work that presents the self-voice alarm as a daily goal reminder, we examine the effect of self-voice on users' task enjoyment with Research Question 2:

RQ2. How does hearing a self-voice as a daily alarm affect users' *enjoyment* of the task?

2.3.2 Cognitive and Behavioral Effects. Studies have shown the effectiveness of self-voice in attracting attention and motivating behavioral decisions. Bhatia and McCrickard [9] examined the effect of self-voice notifications on users' reactions while conducting a game and found that the self-voice notification resulted in higher interruption and faster reaction time than notifications using an unfamiliar other's voice. This result highlights the attention-capturing impact of self-voice over other auditory stimuli. A large body of literature shows that self-relevant stimuli are strongly related to both memory encoding and retrieval [16, 83]. In terms of voice delivery, a familiar voice elicits better recognition and memory [52]. Pucher et al. found that one's own voice increases performance in a memory task compared to unfamiliar voices [67].

If self-voice is used as a daily reminder, self-biased processing may affect users' cognitive reactions to the alarm by capturing their attention quickly and automatically. Reaction time assesses the automatic response to a given stimulus, with faster response times reflecting a higher degree of automaticity. Behavioral automaticity also serves as a measure of how automatically the targeted behavior is carried out across time and setting, enabling individuals to execute the behavior with minimal mental effort [5]. Therefore, we expect that a self-voice alarm will elicit a shorter reaction time to the alarm and a higher level of behavioral automaticity.

H3. Self-voice alarm will reduce *reaction time* to the alarm compared to (a) other-voice and (b) beep sound alarm.

H4. Self-voice alarm will increase *behavioral automaticity* compared to (a) other-voice and (b) beep sound alarm.

If the target behavior is automatically activated, the frequency of the behavior over a certain amount of time will also increase [43, 60]. Self-related cues associated with the target task may enhance users' recall of the task, so that they are likely to complete the task more frequently. This consistent and frequent task completion can also positively impact the users' task performance by gradually improving their task ability. If a self-voice alarm encourages people to conduct the goal task more often, their task performance might be improved as they have more chances to practice the task and produce a better result. Accordingly, we hypothesize that the self-voice alarm will make users repeat their tasks more and improve their task performance compared to the other types of alarm.

H5. Self-voice alarm will increase *task repetitions* compared to (a) other-voice and (b) beep sound alarm.

H6. Self-voice alarm will strengthen *task improvement* compared to (a) other-voice and (b) beep sound alarm.

Since the effects of self-voice have never been studied in daily mobile intervention so far, it is imperative to explore how mobile users perceive the given technology and intend to use it on their own. Mobile applications have been demonstrated as an effective means to facilitate users' continuous usage [11, 64]. Applying the Technology Acceptance Model (TAM) [19, 88], we observed how participants evaluated a mobile alarm using self-voice compared to other types of alarms. TAM introduces a theoretical model of technology use predicted by perceived ease of use and perceived usefulness. Perceived ease of use indicates the relative assessment of one's relief from a cognitive burden. Perceived usefulness is the degree to which individuals believe the system can enhance their performance. These two components, representing the motivating factors of human-computer interaction [19], influence the behavioral intention to use the system directly or indirectly. To test users' acceptance of our novel alarm system, we propose Research Question 3:

RQ3. How does using a self-voice alarm as a daily reminder affect users' evaluations of the alarm system in terms of *ease of use, usefulness,* and *intention to use*?

3 METHOD

3.1 Overview

The objective of this study is to explore the effects of three alarm sounds on users' daily goal achievement. Second-language vocabulary learning was selected as a daily goal task, as mobile interventions are increasingly used to encourage users' informal learning [10, 62]. The daily alarm served as a trigger for the learning task, with the sound of the alarm varying across the experimental groups: the self-voice group heard their own recorded voice; the other-voice group heard the recorded voice of another participant who has the same gender; and the beep sound group used a conventional electronic alarm tone (see details in *Materials*).

Throughout the experiment, the mobile application collected the participants' behavioral data, including response time to the alarm, the number of daily task completions, task repetitions and task performance over 14 days. Following the experiment, participants reported their emotional responses to the alarm sound, task motivation, perceived behavioral automaticity, and evaluations of the mobile application. Participants' overall experience of using the mobile alarm for two weeks was qualitatively analyzed using thematic analysis.

3.2 Participants

After this study was approved by the Institutional Review Board of the university, a total of 82 Korean university students were initially

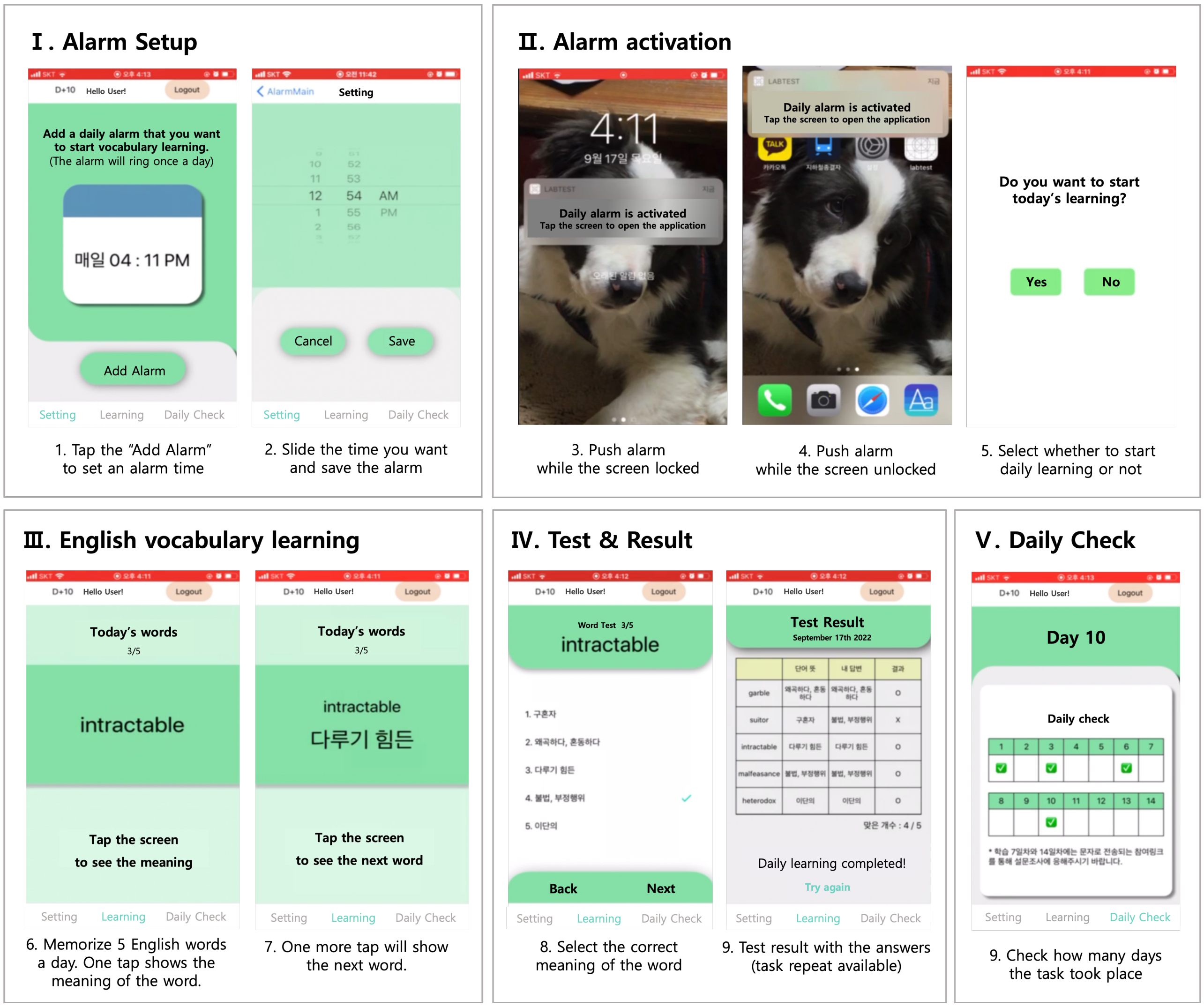

Figure 1: Mobile application instructions for English vocabulary learning. It displays the screens of the application, featuring five functionalities: alarm setup, alarm activation, English vocabulary learning, test and result, and a daily check.

recruited through the university's online and offline advertisement boards. Among them, 63 participants consented to install a mobile application, use their recorded voice, and permit data collection on their mobile usage patterns over a two-week period. The average age of the participants was 24 (SD = 2.57). Each alarm group consisted of 21 participants, controlling for gender (i.e., female = 14, male = 7). Most of the participants (n = 52) reported that learning English vocabulary is important to them. Nevertheless, only a few participants have conducted English vocabulary learning "often (n = 3)" or "always (n = 1)," while the majority stated they did it "rarely (n = 23)" or "never (n = 21)."

3.3 Apparatus

We developed a mobile application by which users can manage their daily English vocabulary learning goals. This application was designed to be used on mobile devices compatible with both Android and iOS systems. We allowed participants to select the alarm time based on their time availability. We also added a feature by which the participants can monitor their task completion progress. The application includes five key functionalities: alarm setup, alarm activation, English vocabulary learning, test and result, and a daily check (Fig 1).

3.3.1 Alarm Setup. All verified participants were granted access to the mobile application using assigned identification numbers provided by the researcher. Once they logged in to the application, participants could set a daily alarm by selecting a time when the alarm would sound. They could also adjust the alarm time whenever they wanted.

3.3.2 Alarm Activation. The alarm was activated only once a day, ensuring that participants were exposed to the alarm sound on a controlled basis. When the alarm was activated at the scheduled time, participants heard one of the alarm sounds assigned by the researchers. A pop-up screen suggested the mobile learning tasks, asking if participants wished to initiate their daily learning session. To quiet the alarm, participants were required to select the "yes" or "no" options displayed on the screen. Tapping the "yes" button initiated the learning task while selecting the "no" button terminated the application. Participants who did not start the task could complete it later in the day.

3.3.3 English Vocabulary Learning. The primary learning task included memorizing five advanced English words each day and testing the acquired knowledge with a quiz. During the memorization phase, each English word was presented along with its corresponding meaning and reviewed by tapping the screen. This process could be repeated until participants decided to start the daily quiz by pressing the "Quiz" button.

3.3.4 Test and Result. We included a quiz as part of the learning task. The quiz asked participants to select the correct meaning of the English word among five different choices. After taking the quiz, participants could view their test scores. They also had the option to revisit the daily vocabulary and retake the quiz if they wanted.

3.3.5 Daily Check. The application provided an independent section where participants could track their daily progress in both learning and quizzes over the two weeks. This section showed a 14-day calendar that highlighted the dates when tasks were successfully completed. This visual representation served as feedback for the participants' learning engagement, encouraging them to use the mobile application.

3.4 Materials

3.4.1 Alarm Recording and Selection. One week ahead of the experiment, participants were instructed to record their voices while reading a specific statement out loud: *"One of the most effective ways to increase vocabulary is to consistently memorize words every day. During the experiment period, a learning alarm will ring once a day. The meaning of the alarm is as follows: It is time to begin your daily learning task."* From these recordings, the last sentence was extracted to be the alarm sound. After creating the set of alarm sounds, we asked two research confederates to listen to the recordings and rate the extent to which each recording sounded clear and comprehensible from 1 to 5. Based on the ranking of scores, the highest-rated 21 recordings were selected as the alarm sound for the experiment while counterbalancing the gender ratio across the groups (i.e., 14 females and 7 males).

3.4.2 Alarm Assignment. Participants whose voice recordings were selected as alarm sounds were assigned to the self-voice alarm group. Their recordings were integrated into the application and employed as task reminders. The same recordings were used for participants in the other-voice group, matching the gender between the recorded voice's speaker and its listener. Since the other-voice alarm condition was manipulated as unfamiliar voices of others, we checked how participants in the other-voice alarm group would perceive the alarm voice by asking them to rate the unfamiliarity of the given alarm voice out of 5 (M = 4.84, SD = 1.83). The remaining participants in the beep sound group were exposed to an electronic bell sound sourced from a free licensed alarm website.

3.4.3 Vocabulary for Daily Learning Task. As the learning material, we built a vocabulary list based on the Graduate Record Examination (GRE), which includes advanced-level English vocabulary (e.g., "fulsome," "galvanize," "encomiastic," "penitent," "stolid"). In order to measure participants' knowledge of GRE vocabulary, a pretest was conducted asking participants to match ten GRE words with correct meanings among five options for each word. The average rate of correct answers was 26%. The GRE words used in the pretest were excluded from the actual learning task during the experiment.

3.5 Procedure

The potential participants were informed about the experimental procedure (Fig 2). If they consented to participate, we asked them to record their voice following our instructions. After a week, we delivered a mobile application download link and instructed them to install the application for the assigned 14 days. To provide more concrete and comprehensive guidelines on the use of the application, instruction videos explaining the overall procedures of application download, alarm setting, and mobile learning tasks were presented.

One day before the experiment started, all participants were asked to download and log in to the application so they could set the alarm for when they wanted to receive the reminder. As the participants' behavioral data were tracked on the mobile server, the researchers could check if they properly followed the instructions or not. While participants were required to hear the alarm every day, they were able to decide themselves whether to carry out the task or not. To encourage their voluntary participation, we informed them that there was no penalty for not completing the daily task.

After two weeks, the participants were invited to complete a post-experiment survey asking about their experience of using the mobile application. An open-ended question was included at the end of the survey to ask about their overall experience of using the mobile alarm for two weeks and check if they encountered any difficulties in using the application during the experiment. All participants received monetary compensation for their participation.

3.6 Measures

In order to investigate the psychological and behavioral responses to the alarm system, both self-report and behavioral tracking methods were used. During the experiment, participants' behavioral responses were tracked through the mobile application. For the manipulation check, we provided five descriptions of the alarm sound (i.e., "The alarm sound used in the experiment was [music/ animal sound/ electronic beep/ my voice/ other's voice]") and asked participants to respond to each statement using a 5-point Likert scale (1 = Strongly Disagree, 5 = Strongly Agree). Then, an open-ended question allowed participants to describe their experience of using the daily alarm for two weeks.

3.6.1 Behavioral Responses. Participants' behavioral responses during their use of the mobile application were measured regarding *reaction time, daily task completion* and *task repetition.*

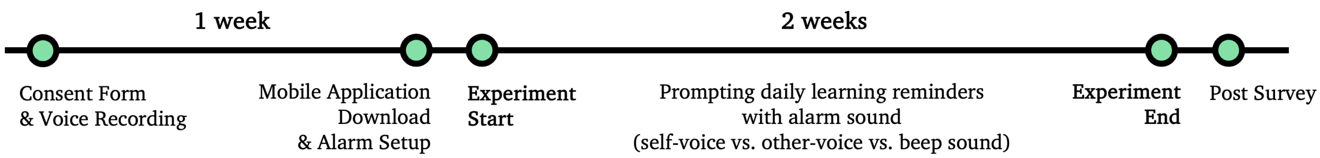

Figure 2: Experimental procedure. It illustrates the sequence of experimental tasks that participants underwent before, during, and after the experiment.

- *Reaction Time* measured the time (in seconds) participants needed to notice and react to the alarm sound. We collected the timestamps when the alarm sounded and when the participants clicked either "yes" or "no" to cease the alarm. The time difference between the timestamps was calculated as the reaction time.
- *Daily Task Completion* indicates the number of days that participants completed the daily task. If participants completed both the learning and test sessions on a given day, their daily achievement was confirmed. Daily task completion does not count the number of tasks repeated on the same day. For example, even if a person performed the target task twice a day, the daily goal achievement was the same as for another who completed the task once a day. This measure presents how consistently the mobile task was completed during the experiment.
- *Task Repetition* is the total number of tasks repeated by participants during the 14-day experiment. Participants who responded to the alarm by clicking either "yes" or "no" could still carry out the learning task later at their preferred time in the day. Therefore, task repetition tells whether the participant repeated the task regardless of the alarm prompt. For example, if a person performed the mobile task twice a day, the total task repetition would be two times.

3.6.2 Emotions. We measured 14 types of emotions, asking participants to report their emotional responses to the alarm sound. Individual emotions were measured by a single item using a 5-point scale ranging from 1 (low) to 5 (high). The basic human emotions were derived from the PANAS scale [35] measuring five positive emotions (i.e., *excited, enthusiastic, inspired, determined, alert)* and five negative emotions (i.e., *nervous, upset, scared, afraid, distressed*). In addition to these ten basic emotions, we measured four more emotions that self-voice can cause: *discomfort, shame, annoyance,* and *embarrassment.* The items for the measures were consistent with the 5-point PANAS scale.

3.6.3 Task Enjoyment. We measured enjoyment of the task as the participants' inherent satisfaction with the daily task. Enjoyment (α = .93) was measured with a 7-point scale to rate 7 adjectives for describing the learning task: boring, enjoyable, entertaining, exciting, fun, interesting, and pleasant [61].

3.6.4 Behavioral Automaticity. Behavioral automaticity (α = .78) was measured by the Self-Report Behavioral Automaticity Index (SRBAI) [26]. Studies of habit formation employ the SRBAI as an independent scale measuring the extent to which the target behavior is performed automatically, confirming its high reliability [26, 42, 89]. The questions were: "When the alarm was activated, [I started the learning task automatically; I started the task without having to consciously remember; I started the task without thinking; I started the task before I realized I was doing it.]"

3.6.5 Task Improvement. Participants' performance on the learning task was indicated by their scores on the vocabulary test. To measure the participants' improvement in test scores, their *pre-score* and *post-score* were collected. *Pre-score* reflects the score on the pretest that participants took before starting the experiment. *Post-score* was devised as the average of the scores that participants had received for two weeks. Since they had been allowed to take the test as many times as they wanted, their latest attempt score was used for the average. Both pre-score and post-score were standardized as the score out of 10. Lastly, *task improvement* was determined as the gap between pre-score and post-score.

3.6.6 Alarm System Evaluations. Participants' acceptance of the newly developed alarm system was assessed based on the Technology Acceptance Model [88] in terms of perceived ease of use, perceived usefulness, and intention to use.

- *Perceived Ease of Use* (α = .84) was evaluated using four questions: "My interaction with the alarm system is clear and understandable;" "Interacting with the alarm system does not require a lot of my mental effort;" "I find the alarm system to be easy to use;" "I find it easy to use the alarm system to do what I want it to do."
- *Perceived Usefulness* (α = .87) was measured by the following items: "Using the alarm system improves my performance in my learning;" "Using the alarm system in my learning increases productivity;" "Using the alarm system enhances my effectiveness in my learning;" "I find the alarm system to be useful in my learning."
- *Intention to Use* (α = .95) was measured by two questions: "If possible, I intend to use the alarm system;" and "Given that I have access to the alarm system, I predict that I would use it."

3.6.7 Open-ended Question. To understand participants' overall experience with the daily alarm over two weeks, our survey included an open-ended question at the end: "Please feel free to share any thoughts or feelings about your experience during the experiment." With the collected answers, we conducted a thematic analysis to extract meaningful insights from the textual data. We first thoroughly read all the responses to familiarize ourselves with them. Then, we labeled and clustered the excerpts to generate and develop the overarching themes. To ensure the participants' anonymity, we allocated unique numerical identifiers corresponding to the experimental conditions: S1-S21 for the self-voice group; O1-O21 for the other-voice group; and B1-B21 for the beep sound group. The average length of responses was 70 words.

4 RESULTS

Our dataset collected through a mobile tracker revealed that 54 out of 63 participants (86%) completed the experiment by using the mobile application for two weeks following our study instructions. Every participant in the self-voice group used the application for two weeks without any dropouts, while two participants in the other-voice group and seven participants in the beep sound group stopped using the application before conducting the post-survey. As dropping out indicated participants' intentional refusal to use the application, their incomplete behavioral data were treated as an informative dropout and included in the analysis [47]. As a result, the behavioral data collected from 63 participants were used in the final analysis, which includes the 54 participants who completed the survey and 9 participants who stopped using the application before the experiment ended.

To find significant differences between the three experimental groups (i.e., self-voice, other-voice, and beep sound), a multivariate analysis of variance (MANOVA) was performed. Given that the self-report results had an unequal sample size for each group, the Bonferroni post hoc test was used for pairwise comparisons across groups. All statistical analysis was conducted using IBM SPSS software [79].

4.1 Manipulation Check

The alarm sound was successfully manipulated for each group. The self-voice group ($F(2, 51) = 27.39$, $p < .001$), the other-voice group ($F(2, 51) = 19.66$, $p < .001$), and the beep sound group ($F(2, 51) = 13.74$, $p < .001$) were able to identify the alarm sound as we manipulated.

4.2 Emotions

To answer **Research Question 1**, we asked the participants how they felt when they heard the daily alarm in terms of 14 emotions. Figure 3 visualizes the significant differences among the ten positive and negative emotions, in addition to four emotions that are relevant to uncomfortable feelings.

4.2.1 Basic Emotions. First, we measured participants' responses to positive emotions (i.e., excited, enthusiastic, inspired, determined, and alert). All three alarm groups showed consistently low ratings of positive emotions: excited ($M = 1.48$, $SD = .46$), enthusiastic ($M = 1.90$, $SD = .82$), inspired ($M = 2.42$, $SD = .41$) and determined ($M = 2.84$, $SD = .74$). On the other hand, the self-voice alarm significantly evoked alertness ($M = 3.95$, $SD = .79$) compared to the other-voice alarm ($M = 3.5$, $SD = .43$), $SE = .18$, $p = .04$, and beep sound alarm ($M = 3.37$, $SD = .38$), $SE = .2$, $p = .01$. This result indicates that the self-voice alarm had no significant impact on general positive emotions while eliciting higher alertness from the participants.

For negative emotions, we found significant differences between the other-voice group and the beep sound group regarding the emotions of being nervous, upset, and scared. First, the self-voice group ($M = 3.31$, $SD = .71$) and other-voice group ($M = 3.32$, $SD = .58$) showed higher nervousness than the beep sound group ($M = 2.32$, $SD = 1.27$), $F(2, 51) = 7.06$, $p = .002$. On the other hand, participants using the other-voice alarm ($M = 1.99$, $SD = .86$) were less upset than those who heard the self-voice alarm ($M = 2.82$, $SD = .77$), $SE = .22$, $p = .002$, and the beep sound alarm ($M = 2.73$, $SD = .33$), $SE = .24$, $p = .013$. The other-voice group reported a lower level of being scared ($M = 1.46$, $SD = .5$) compared to the beep sound group ($M = 2.22$, $SD = 1.24$), $SE = .26$, $p = .01$. Significant differences were not found among the three groups for being afraid, $F(2, 51) = 1.67$, $p = .21$, and distressed, $F(2, 51) = 2.68$, $p = .07$.

4.2.2 Emotions Related to Uncomfortable Feelings. As for the emotions that were related to uncomfortable feelings, we measured discomfort, shame, annoyance, and embarrassment. The results show that discomfort was dominant in the self-voice alarm group ($M = 4.28$, $SD = .7$) compared to the other-voice group ($M = 3.59$, $SD = .34$), $SE = .16$, $p < .001$, and the beep sound group ($M = 3.37$, $SD = .38$), $SE = .17$, $p < .001$, supporting **H1-a** and **H1-b**. In addition, the self-voice group ($M = 3.94$, $SD = 1.01$) and other-voice group ($M = 3.54$, $SD = .41$) experienced a higher level of shame when they heard the alarm than the beep sound group ($M = 2.25$, $SD = .26$), $F(2, 51) = 26.13$, $p < .001$. The other-voice alarm ($M = 2.8$, $SD = .8$) caused a lower level of annoyance than the beep sound alarm ($M = 3.69$, $SD = 1.19$), $SE = .3$, $p = .01$, while having no significant difference from the self-voice alarm group, $SE = .27$, $p = .3$. In terms of embarrassment, the self-voice group, other-voice group, and beep sound group had no significant difference, $F(2, 51) = 1.94$, $p = .15$.

4.3 Behavioral Responses

We collected the participants' behavioral responses regarding *daily task completion, reaction time,* and *task repetition* across the alarm groups (Table 2). Figure 4 displays the trajectory of daily task completions and repetitions by the three alarm groups during the 14 days of the experiment.

4.3.1 Daily Task Completion. Participants more consistently used the mobile application when they were reminded of the learning task by the self-voice alarm than the beep sound alarm. A significant difference in daily task completion was not found between the self-voice group ($M = 13.14$, $SD = .91$) and the other-voice group ($M = 10.8$, $SD = 4.44$), rejecting **H2-a**, $SE = 1.23$, $p = .15$. However, the self-voice group completed the daily tasks more consistently than the beep sound group ($M = 9.14$, $SD = 5.24$), confirming **H2-b**, $SE = 1.23$, $p = .006$.

4.3.2 Reaction Time. Since the reaction time was highly skewed (*Skewness* = 3.89, $SE = .3$), log transformation was used to normalize the distribution (*Skewness* = .06, $SE = .3$). There was no significant difference between the beep sound group ($M = 4.01$, $SD = .1.51$), the self-voice group ($M = 3.96$, $SD = .93$), and the other-voice group ($M = 3.84$, $SD = .1.65$), $F(2, 60) = .07$, $p = .92$. Therefore, **H3-a** and **H3-b** were rejected (Table 2).

4.3.3 Task Repetition. A significant difference in task repetition was found between the self-voice group ($M = 14$, $SD = 2.6$) and the beep sound group ($M = 9.71$, $SD = 5.56$), indicating that the self-voice alarm elicited more task repetition than the beep sound alarm, $SE = 1.5$, $p = .01$. The difference between the self-voice group and the other-voice group ($M = 12.71$, $SD = 5.78$) was not significant, $SE = 1.5$, $p = .67$. Thus, we rejected **H5-a** and accepted **H5-b**.

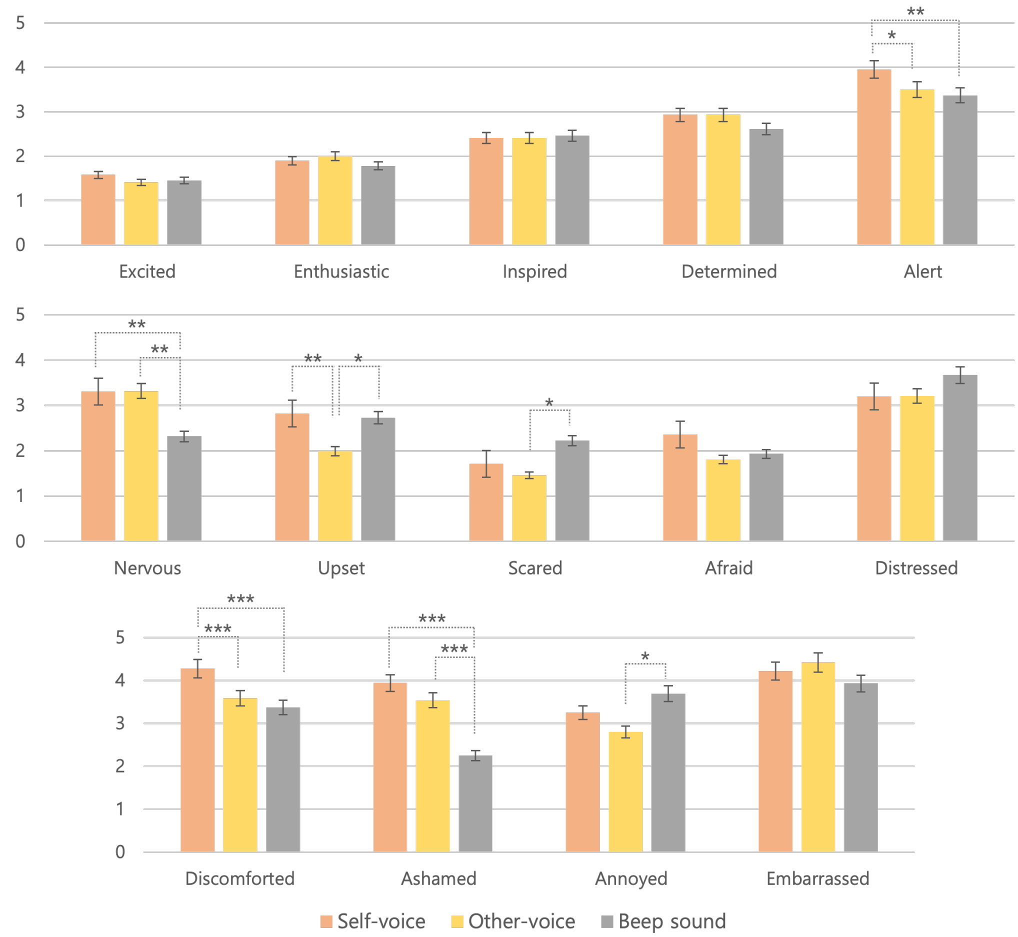

Figure 3: Comparison of Emotions between Alarm Groups
(*: p < .05, **: p < .01, ***: p < .001)

Table 1: MANOVA: Alarm Effects on Self-Report Measures

	Self-voice alarm		Other-voice alarm		Beep sound alarm			
	M	*SD*	*M*	*SD*	*M*	*SD*	$F(2, 51)$	p
Task Enjoyment	4.45	.85	5.45	.6	4.58	.29	12.7	<.001***
Behavioral Automaticity	5.54	.32	5.55	.29	5.16	.15	9.49	<.001***
Ease of Use	6.01	.44	6.29	.7	5.69	.21	5.6	.006**
Usefulness	4.67	.42	4.48	.44	3.89	.38	15.43	<.001***
Intention to Use	3.11	.89	3.49	.93	3.03	.65	1.54	.22

Note. p < .05*, < .01**, < .001***

4.4 Task Enjoyment

Research Question 2 explored the level of enjoyment of the learning task across the three alarm groups (Table 1). Interestingly, the other-voice alarm group experienced the greatest enjoyment of the task compared to the other two groups, $F(2, 51) = 12.7$, $p < .001$. Participants exposed to the other-voice alarm ($M = 5.45$, $SD = .6$) showed greater enjoyment than those exposed to the self-voice alarm ($M = 4.44$, $SD = .85$), $SE = .21$, $p < .001$, and those exposed to the beep sound alarm ($M = 4.58$, $SD = .29$), $SE = .23$, $p = .002$. A significant difference between the self-voice alarm group and the beep sound group was not found, $SE = .22$, $p = .85$.

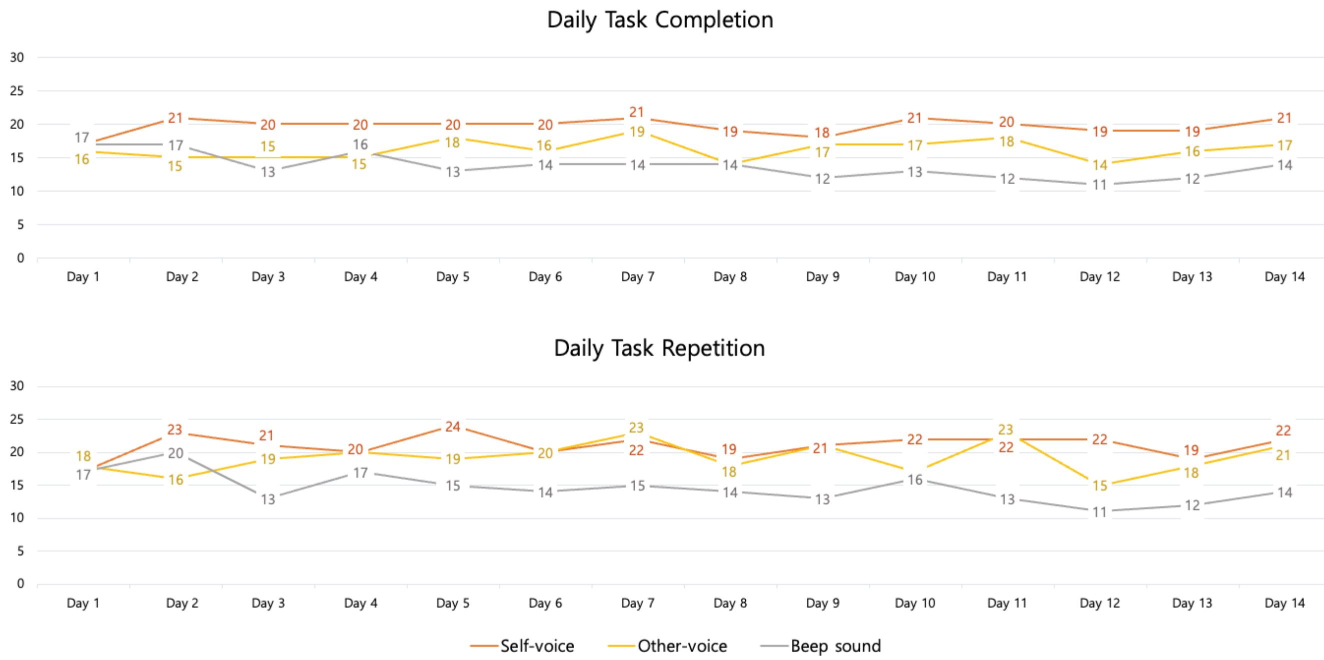

Figure 4: Daily Task Completion and Repetition by Alarm Group over 14 days. The upper graph illustrates the number of participants who completed the daily task by each group. The lower graph demonstrates the number of task repetitions by each group (*n* = 21) each day.

Table 2: MANOVA: Alarm Effects on Behavioral Responses

	Self-voice alarm		Other-voice alarm		Beep sound alarm			
	M	*SD*	*M*	*SD*	*M*	*SD*	*F*(2, 60)	*p*
Daily Task Completion	13.14	.91	10.8	4.44	9.14	5.24	5.25	.008**
Reaction Time	3.96	.93	3.84	1.65	4.01	1.51	.07	.92
Task Repetition	14	2.6	12.71	5.78	9.71	5.56	4.28	.01*

Note. p < .05*, < .01**, < .001***

4.5 Behavioral Automaticity

Our self-report results show that the self-voice and other-voice elicited higher behavioral automaticity than the beep sound alarm (Table 1). The SRBAI scores of the self-voice group (M = 5.54, SD = .32) and the other-voice group (M = 5.55, SD = .29) were significantly higher than those of the beep sound group (M = 5.16, SD = .15), $F(2, 51)$ = 9.49, p = .001, but the difference between the self-voice group and the other-voice group was not significant, SE = .08, p = .98. Thus, **H4-a** was rejected and **H4-b** was accepted.

4.6 Task Performance

On average, participants in the self-voice alarm group did not display any significant difference in their vocabulary test scores from those in the other-voice and beep sound alarm groups (Fig 5). The differences in pre-test score between the self-voice alarm group (M = 4.38, SD = .1.74), other-voice group (M = 5.05, SD = 2.48), and beep sound group (M = 5.21, SD = 1.25) were not statistically significant, $F(2, 51)$ = .96, p = .38. The overall test score improved across the conditions, showing no significant differences between self-voice alarm group (M = 8.81, SD = 1.75), other-voice group (M = 8.82, SD = 2.94), and beep sound group (M = 7.88, SD = 2.44), $F(2, 51)$ = 1.47, p = .23. Although the average task improvement of the self-voice group (M = 4.43, SD = 2.47) was slightly higher than that of the other-voice (M = 3.76, SD = 2.66) and beep sound groups (M = 2.67, SD = 2.65), the differences were statistically insignificant, $F(2, 51)$ = 1.93, p = .15. Thus, **H6-a** and **H6-b** were rejected.

4.7 Alarm System Evaluations

To answer **Research Question 3**, participants evaluated the mobile alarm system in terms of its ease of use, usefulness, and intention to

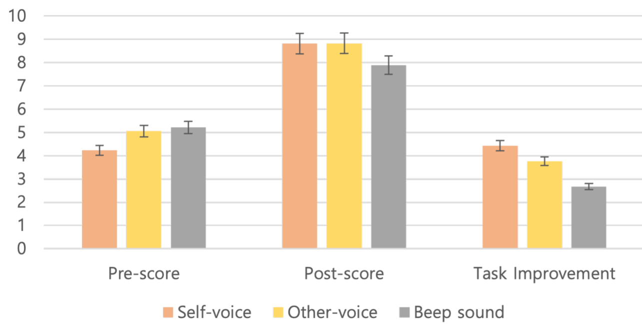

Figure 5: Comparison of Task Performances and Improvement between Alarm Groups
(*: $p < .05$, **: $p < .01$, ***: $p < .001$)

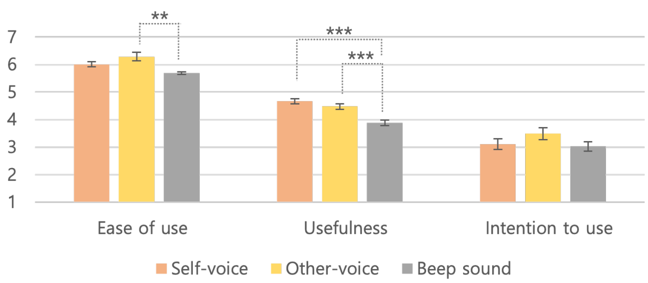

Figure 6: Comparison of Alarm System Evaluations between Alarm Groups
(*: $p < .05$, **: $p < .01$, ***: $p < .001$)

use (Table 1). Fig 6 shows the comparisons between the conditions for the three variables.

4.7.1 Ease of Use. A significant difference in ease of use was revealed between the other-voice (M = 6.29, SD = .7) and the beep sound group (M = 5.68, SD = .2), SE = .17, p = .004. However, the self-voice group (M = 6.01, SD = .44) showed no significant differences from the other-voice group, SE = .16, p = .23, and the beep sound group, SE = .17, p = .14.

4.7.2 Usefulness. The ratings for usefulness of the self-voice (M = 4.67, SD = .42) and other-voice groups (M = 4.48, SD = .44) were significantly higher than of the beep sound group (M = 3.89, SD = .38), $F(2, 51)$ = 15.43, $p < .001$. Compared to the beep sound group, participants evaluated the alarm system as more useful when it provided the self-voice alarm, SE = .14, $p < .001$, or other-voice alarm, SE = .14, $p < .001$. The difference between the self-voice and other-voice groups was not significant, SE = .13, p = .35.

4.7.3 Intention to Use. The alarm type did not make any significant differences in the intention to use between the self-voice group (M = 3.11, SD = .89), the other-voice group (M = 3.49, SD = .93), and the beep sound group (M = 3.03, SD = .65), $F(2, 51)$ = 1.54, p = .22.

4.8 Qualitative Responses

In general, the qualitative responses showed similar trends to the quantitative results. However, they provided more intuitive thoughts about participants' experience of using the alarm and explained the reasons behind their statistical results.

4.8.1 Evaluations of the Alarm Sound. Consistent with our quantitative results, participants in the self-voice alarm group mentioned that listening to their own voice as an alarm was an unpleasant experience (S19, 20), making them feel embarrassed (S3, S9, S17). Some felt ashamed and uncomfortable when they were with other people and the alarm sounded (S13). Several participants pointed out that when the alarm was ringing in a public space, it surprised people around them (S15, S19), which made them feel nervous whenever the alarm was about to ring (S8, S12).

Similarly, participants in the other-voice alarm group also mentioned that they felt shame (O9) and embarrassment (O5, O11) when they heard the alarm sound. They also worried that the alarm sound would surprise people in the public space (O11, O15), which made them feel nervous (O18, O20). Some participants who heard the other-voice alarm reported that the alarm sound scared them, especially when they were alone (O15).

While participants expressed negative responses to using the voice alarms, some shared positive feedback. They found that the human voice encouraged them to continue using the app, which positively contributed to their task completion (O7, O8, O11). By contrast, participants in the beep sound alarm group described their experience as boring (B6) and suggested adding more enjoyable and engaging features to facilitate learning (B8).

4.8.2 Evaluations of the Application. Participants exhibited positive responses towards the application, illustrating its usefulness for learning. Those in the self-voice group expressed enjoyment in learning with the app and appreciated the high-quality vocabulary the app offered (S9, S15). The voice alarm groups acknowledged that the application helped them engage in the task regularly (S4, S6, S20, O6, O16). However, participants in the beep sound alarm group mentioned that while the app's prompting for task continuation was helpful (B13), they didn't find it particularly useful for learning (B4).

4.8.3 Suggestions for Improvement. Participants provided suggestions for enhancing the application. Many participants across the groups mentioned that the alarm sound was loud (S14, S18, O7, B2, B10), suggesting a volume adjustment feature (S16, O3, O5, B14, B16). In addition, they proposed the option to mute the alarm or switch it to vibration mode in public spaces to avoid disturbance (S2, S21, O9, O15). Regarding language learning features, they recommended adding an option that allows users to listen to pronunciations for vocabulary (S2, S3, S6, S11, O3, O10, O12, O14, B1, B7) and a feature that enables users to review previously learned vocabulary (S5, O3, O16, O21).

5 DISCUSSION

It is never easy to perform a targeted behavior consistently for daily goal achievement. The literature on habit formation emphasizes the significance of automatic behavior triggered by cues that remind individuals of their desired behavior, especially in the initial stages of habit formation [5, 42, 43]. Therefore, we aimed to observe whether the human voice alarms not only capture users' attention but also motivate them to regularly complete their daily tasks.

Our results highlight the effectiveness of a self-voice alarm in helping users pursue their daily goals for two weeks. While the self-voice alarm elicited a higher level of alertness and discomfort, it increased the actual number of daily task completions and task repetitions compared to the beep sound alarm. We found that human voice alarms evoke a greater sense of behavioral automaticity and increase users' perceived usefulness of the alarm system. Our qualitative data analysis supported our quantitative findings by demonstrating that while voice alarms triggered unpleasant emotions, they assisted participants in sustaining their learning. Unexpected results were found for the other-voice alarm, revealing a higher level of enjoyment of the task and perceived ease of use compared to the self-voice and beep sound alarm. The other-voice alarm also induced the lowest level of upset and annoyed feelings.

Our study tested the use of a self-voice alarm as a novel method to enhance users' continuous task completion, investigating its psychological and behavioral effects over two weeks. While prior researchers have explored the immediate effects of self-voice on user perceptions, to our best knowledge, none examined its long-term implications in the context of habit formation. In a two-week-long field experiment, we observed and analyzed participants' behaviors and responses to the alarm in real-life settings, significantly enhancing the ecological validity of our experiment. Furthermore, our study adopted both quantitative and qualitative approaches to comprehensively measure various facets of user responses. In the following sections, we highlight several key discussion points to address the findings regarding the study's hypotheses and research questions.

5.1 Self-Voice as a Daily Reminder: Leveraging Discomfort for Alertness, Enhancing Daily Goal Achievement

As implied by the inconvenient interaction design concept [69], uncomfortable feelings can play a pivotal role in promoting behavioral changes. Our two-week experiment demonstrated that a self-voice alarm that causes discomfort can effectively help people initiate a task every day, increasing their daily task completion rate. In our qualitative responses, participants described the self-voice alarms as disturbing and uncomfortable while acknowledging those features as helpful to continue their learning. Thus, we conjecture that non-positive feelings elicited by self-voice could have a positive influence on daily behavioral change.

Previous research exploring the emotional impact of self-voice revealed that listening to one's own voice can cause a discrepancy between what they expect and what they actually hear [29, 30] which creates eerie feelings [13]. Aligning with this, we found that self-voice induced a higher level of discomfort when it was used as an alarm compared to the other's voice and a beeping sound. Beyond identifying the discomfort, however, our study clarified the specific emotions that self-voice may induce, considering 14 types of emotions. Notably, the self-voice alarm did not affect participants' positive emotions including being excited, enthusiastic, inspired, and determined. Given the nature of the alarm which aims to interrupt users and increase their attention and arousal, it may not inherently induce positive emotions regardless of the alarm type. On the other hand, hearing one's own voice as an alarm elicited higher alertness than other alarms. This heightened state of alertness is a significant result for alarm design, as the major role of an alarm is to attract users' attention and encourage them to take appropriate actions. This alerted state can enhance attention [12, 90], which is crucial to maintaining focus on target goals in a daily setting.

In addition to emotional arousal, self-voice alarms played a significant role in promoting behavioral changes. During our 14-day experiment, participants in the self-voice group completed daily tasks for more days and completed a greater number of tasks than those in the beep sound group. This result indicates that simply changing the alarm sound can help people achieve their daily goals more successfully. Sticking to a plan in the long term is critically important to achieving long-term goals like health management, learning, and diet change. As the habit instigation stage can lead to habit execution [27], placing an instigating cue in daily contexts is helpful to initiate the target behaviors. Nevertheless, people experience difficulties in implementing their behaviors even when they have a strong intention to form a habit [72, 74]. As a way to execute users' behavioral intentions, our study demonstrates that using a self-voice alarm can effectively help users achieve the daily goal task regularly.

Moreover, the self-voice alarm encouraged participants to perform the task multiple times on the same day. Researchers have highlighted the repetition of goal tasks in the process of habit formation, particularly when the goal behavior was not stably formed. Initial repetitions lead to a high level of automaticity, which can gradually reduce the required number of repetitions until reaching a fully automated behavior [26, 43]. As the self-voice alarm motivated users to perform the target task more often, it enabled them to become familiar with the task and gain automaticity in the initial stage of habit formation. Although frequent task completion did not lead to immediate task improvement within our two-week experiment time frame, day-to-day goal achievement will become the foundation for establishing a routine that yields long-term benefits.

5.2 Other-Voice as a Task Motivator: Increasing Task Enjoyment While Reducing Feelings of Being Upset and Annoyed

We also discovered unexpected effects of the other-voice alarm on reducing the levels of negative emotions while increasing the enjoyment of the task compared with the beep sound alarm. This interesting difference between the other-voice and beep sound alarms highlights the comparison between human voice and machine-generated sounds. Since the human voice possesses interpersonal features displayed through its paralinguistic cues [96], a naturalistic human voice might be preferred to a machine's electronic sound. Previous studies have primarily focused on the voices of familiar others, such as relatives and friends [13], teachers [67], and well-known celebrities [20], to be understood easily and recognized faster. However, limited studies have discovered the motivating impacts of an unfamiliar other's voice.

Our study revealed that an unfamiliar other's voice can significantly lower feelings of being upset and annoyed compared to the beep sound alarm. Moreover, the other-voice group reported the highest level of enjoyment toward the learning task compared to the self-voice and beep sound alarm groups. As enjoyment is closely associated with intrinsic motivation, these findings emphasize the role of others' voices as a motivating resource to encourage behavior change. In a learning context, intrinsic motivation helps learners continue their learning and achieve their learning goals [40]. Although other-voice alarms did not increase task completion nor improve task performance, enhancing task enjoyment will benefit users by providing a more positive learning experience.

This motivational effect of others' voices can be connected to the benefit of peer learning. For instance, social learning theory suggests that involving others in one's learning process can facilitate their learning experience, increasing enjoyment and motivation [4]. Although our study did not include other users in one's task performance, daily exposure to other users' voices can evoke a sense of social presence, which is the feeling of being together with other social actors [44]. However, the sound of an unfamiliar voice could also scare the users when they are alone, as reported in the qualitative data. Thus, more attention should be paid to the other-voice alarm as a motivating resource to encourage behavior change.

5.3 Human Voice as a Learning Nudge: Improving Behavioral Automaticity and Perceived Usefulness for Learning

Previous studies have shown the advantages of the human voice in processing auditory information. Speech alarms have been shown to be effective in recalling the alarm messages compared to other types of sounds, while simple tone alarms took the longest time to remember [86]. Speech alarms also allow users to understand the context of the alarm [57]. In an emergency, speech alarms delivered information faster and more accurately and made people evacuate more efficiently compared to bell alarms [37, 46, 65, 66, 91]. What previous studies did not reveal, however, is the differential emotional response toward human voice and machine sound such as simple beep sound alarm.

Our quantitative and qualitative results consistently indicated that both self-voice and other-voice alarms evoked higher levels of shame and nervousness than the beep sound alarm. Participants using self-voice and other-voice alarms reported that they felt ashamed when the alarm sounded in the public space and nervous preceding the alarm. These emotional influences could also impact the participants' reaction time. Previous research suggested that using one's voice as a notification could reduce response time [9]. Contrary to the earlier findings, however, we did not find a significant impact of the human voice alarm on reaction time compared to the beep sound alarm. One plausible explanation for this insignificance is that participants in public spaces who felt shame and nervousness might have stopped the alarm by pressing any button regardless of the sound. Since they had to choose between "yes" or "no" buttons, their deliberation on whether to start the task or not might have caused similar reaction times across the groups.

Despite similar reaction times across the groups, participants who heard the human voice alarms still reported a greater level of task automaticity compared to those who heard the beep sound alarm. This finding sheds light on the potential of using human voices (including self-voice and other-voice) in mobile interventions. Behavioral automaticity is a key feature that facilitates habit formation [42, 87] by reducing the mental burden required to initiate behaviors [5]. As the participants perceived that they performed the daily task automatically in response to the voice alarm, a voice alarm system may effectively promote their task behaviors, requiring lower cognitive load.

Furthermore, those using the self-voice and other-voice alarms evaluated the application as more useful compared to those using the beep sound alarm. Our qualitative data, which aligned with our quantitative findings, suggested that participants using self-voice and other-voice alarms perceived the alarm of the application to be useful for their learning. By contrast, participants in the beep sound group did not find the alarm to be useful for their learning even though they acknowledged its continuous nudging as beneficial. Participants from each group evaluated the application's usefulness differently despite its identical features (except for the alarm sound) based on their subjective perceptions. Perceived usefulness is the key component to accepting new technology. When evaluating a mobile application, users' perception of usefulness is strongly associated with continuous use [11, 64] and satisfaction with its use [31]. In this respect, human voice alarm systems can build a positive user experience when assisting users' daily goal behavior.

5.4 Design Guidelines for Goal Task Alarm

Based on the analysis using both quantitative and qualitative approaches, we found significant trade-offs for each type of alarm, regarding the effectiveness of behavior change and user experience. First, self-voice alarms elicited heightened alertness leading to increased task completion and repetition. In this sense, the self-voice alarm can be considered the most practically effective goal reminder among others by increasing users' behavioral changes within two weeks. However, the discomfort and uneasy feelings associated with hearing one's own voice may impair the user experience. Conversely, the other-voice alarm has advantages in reducing negative emotions such as being upset and annoyance while increasing task enjoyment. However, it does not provide the same level of alertness or personal connection as the self-voice alarm, resulting in a less effective impact on users' goal behaviors. While both self-voice and other-voice evoked shame and nervousness in public settings, they were perceived to be effective in eliciting automatic responses to the alarm sound and useful for pursuing the daily learning task. By contrast, the beep sound alarm may elicit neutral emotional states compared to the voice alarms–neither positive, uncomfortable, nor negative,–but their mundane interactions failed to enhance the users' attention and motivation toward the goal tasks.

These trade-offs suggest that while each alarm type has its advantages, there are accompanying drawbacks to consider. Considering the pros and cons of each alarm type, the alarm can be framed with diverse design approaches. According to the voice design strategies proposed by Sutton et al. [81], voice can be utilized in the user interfaces adopting *individualization* and *context awareness*.

Individualization enables users to select the voice to be used on their devices, accommodating their preferences. While some people prefer to use self-voice alarms, expecting higher achievements of daily goal tasks, others would prefer beep sound alarms to avoid emotional discomfort. Adjusting the voice parameters such as tone and pitch can affect users' perceptions of voice familiarity or preferences reflecting different personas or emotional states of the voice speakers. As mentioned in our qualitative data, an option to adjust the alarm volume may be crucial to meet their preference for alarm use.

Another strategy, context awareness, involves employing multiple voices tailored to various contexts. This allows users to choose from multiple voice alarms based on the situation. As the self-voice and other-voice alarms have different influences on motivation and emotional activation, their contextual implications may vary. For instance, the initial stage of behavior change can benefit from the self-voice alarm's intense emotional activation. Users can self-discipline by performing an unfamiliar behavior and establishing a daily routine with self-voice. As the behavior becomes familiar, however, they can employ other-voice alarms to increase task enjoyment and alleviate the negative emotions triggered by the self-voice alarm. While the self-voice and other-voice alarms are effective in prompting automatic responses to the task, they cause embarrassment in public spaces. To avoid their arousal of shame and nervousness, users will need an option to shift to a beep sound or vibration mode when they are in a shared space. As voice becomes the key material in designing user interfaces, choosing the right source of voice aligned with users' preferences and context is essential to the effective design of voice user interfaces.

6 LIMITATIONS AND FUTURE WORK

Our quantitative results demonstrated the advantages of self-voice and other-voice alarms over a beep sound alarm. Concurrently, our qualitative data analysis partially explained the participants' emotional responses in relation to their dispositional and situational factors. Yet, it is still unclear what critical factors affect users' behavioral intentions and task completion, impacting their long-term behavior change. As our experiment took place over 14 days, which is a shorter period than needed for habit formation, our findings have limited implications for habit execution. Given that habit formation requires 18 to 256 days [42], future researchers can investigate the effect of voice alarms in the longer term, identifying users' behavioral patterns and task improvement during this time period, which will have stronger implications for habit formation.

As a follow-up study, further investigation is required to clarify the impacts of voice features including voice familiarity. While our study included unfamiliar others' voices, a comparative analysis can extend to the voices of familiar others, such as friends, teachers, and family members. In addition, future researchers can consider the verbal contents and nonverbal vocal signals delivered through voice alarms. Given that speech alarms can convey meaningful messages and present diverse social categories through vocal cues, the contents and vocal signals of the alarm itself could significantly influence users' engagement in the target behavior.

7 CONCLUSION

With the help of today's technology, it has never been easier to take one's own photo and record one's own voice. While there is a strong body of literature showing the powerful effects of self-voice and self-image, they have not been studied or applied, particularly in the context of daily goal attainment. We hope more studies will be conducted to reveal the potential of self-relevant information in motivation and long-term behavior change and their processes. The findings of these studies will help people achieve their long-term goals more easily and effectively.

REFERENCES

[1] Sun Joo Ahn and Jeremy N Bailenson. 2011. Self-endorsing versus other-endorsing in virtual environments. *Journal of Advertising* 40, 2 (2011), 93–106.
[2] Paul P Allen, Edson Amaro, Cynthia HY Fu, Steven CR Williams, Michael Brammer, Louise C Johns, and Philip K McGuire. 2005. Neural correlates of the misattribution of self-generated speech. *Human brain mapping* 26, 1 (2005), 44–53.
[3] Sercan Arik, Jitong Chen, Kainan Peng, Wei Ping, and Yanqi Zhou. 2018. Neural voice cloning with a few samples. *Advances in neural information processing systems* 31 (2018).
[4] Albert Bandura and Richard H Walters. 1977. *Social learning theory*. Vol. 1. Englewood cliffs Prentice Hall.
[5] John A Bargh. 1994. The four horsemen of automaticity: Intention, awareness, efficiency, and control as separate issues. (1994).
[6] Oliver Baumann and Pascal Belin. 2010. Perceptual scaling of voice identity: common dimensions for different vowels and speakers. *Psychological Research PRPF* 74, 1 (2010), 110–120.
[7] Pascal Belin, Patricia EG Bestelmeyer, Marianne Latinus, and Rebecca Watson. 2011. Understanding voice perception. *British Journal of Psychology* 102, 4 (2011), 711–725.
[8] Steve Benford, Chris Greenhalgh, Gabriella Giannachi, Brendan Walker, Joe Marshall, and Tom Rodden. 2013. Uncomfortable user experience. *Commun. ACM* 56, 9 (2013), 66–73.
[9] Saurabh Bhatia and Scott McCrickard. 2006. Listening to your inner voices: investigating means for voice notifications. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1173–1176.
[10] Stephan Böhm and Georges Philip Constantine. 2016. Impact of contextuality on mobile learning acceptance: An empirical study based on a language learning app. *Interactive Technology and Smart Education* (2016).
[11] Charles Buabeng-Andoh. 2018. Predicting students' intention to adopt mobile learning: A combination of theory of reasoned action and technology acceptance model. *Journal of Research in Innovative Teaching & Learning* (2018).
[12] Luis Carretié, Manuel Martín-Loeches, José Antonio Hinojosa, and Francisco Mercado. 2001. Emotion and attention interaction studied through event-related potentials. *Journal of cognitive neuroscience* 13, 8 (2001), 1109–1128.
[13] Sam WT Chan, Tamil Selvan Gunasekaran, Yun Suen Pai, Haimo Zhang, and Suranga Nanayakkara. 2021. KinVoices: Using voices of friends and family in voice interfaces. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–25.
[14] Colin Cherry. 1953. Cocktail party problem. *Journal of the Acoustical Society of America* 25 (1953), 975–979.
[15] Tatiana Conde, Óscar F Gonçalves, and Ana P Pinheiro. 2015. Paying attention to my voice or yours: An ERP study with words. *Biological psychology* 111 (2015), 40–52.
[16] Sheila J Cunningham, David J Turk, Lynda M Macdonald, and C Neil Macrae. 2008. Yours or mine? Ownership and memory. *Consciousness and cognition* 17, 1 (2008), 312–318.
[17] Nils Dahlbäck, Seema Swamy, Clifford Nass, Fredrik Arvidsson, and Jörgen Skågeby. 2001. Spoken interaction with computers in a native or non-native language-same or different. In *Proceedings of INTERACT*. Citeseer, 294–301.
[18] Nils Dahlbäck, QianYing Wang, Clifford Nass, and Jenny Alwin. 2007. Similarity is more important than expertise: Accent effects in speech interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 1553–1556.
[19] Fred D Davis. 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly* (1989), 319–340.
[20] Ricardo Molina de Figueiredo. 1999. Perceptual recognition of celebrity voices using random spliced speech. (1999).
[21] Edward L Deci and Richard M Ryan. 2013. *Intrinsic motivation and self-determination in human behavior*. Springer Science & Business Media.
[22] Marina England et al. 2021. The Effects of Voice Familiarity on Attention and Memory During the Dichotic Listening Test. (2021).

[23] Charles Fernyhough and James Russell. 1997. Distinguishing one's own voice from those of others: A function for private speech? *International Journal of Behavioral Development* 20, 4 (1997), 651–665.

[24] Brianna S Fjeldsoe, Alison L Marshall, and Yvette D Miller. 2009. Behavior change interventions delivered by mobile telephone short-message service. *American journal of preventive medicine* 36, 2 (2009), 165–173.

[25] Teresa Franklin and Li-Wei Peng. 2008. Mobile math: Math educators and students engage in mobile learning. *Journal of computing in higher education* 20, 2 (2008), 69–80.

[26] Benjamin Gardner, Charles Abraham, Phillippa Lally, and Gert-Jan de Bruijn. 2012. Towards parsimony in habit measurement: Testing the convergent and predictive validity of an automaticity subscale of the Self-Report Habit Index. *International Journal of Behavioral Nutrition and Physical Activity* 9, 1 (2012), 1–12.

[27] Benjamin Gardner, L Alison Phillips, and Gaby Judah. 2016. Habitual instigation and habitual execution: Definition, measurement, and effects on behaviour frequency. *British journal of health psychology* 21, 3 (2016), 613–630.

[28] Emma Holmes, Ysabel Domingo, and Ingrid S Johnsrude. 2018. Familiar voices are more intelligible, even if they are not recognized as familiar. *Psychological science* 29, 10 (2018), 1575–1583.

[29] Philip S Holzman, Andrew Berger, and Clyde Rousey. 1967. Voice confrontation: A bilingual study. *Journal of Personality and Social Psychology* 7, 4p1 (1967), 423.

[30] Philip S Holzman and Clyde Rousey. 1966. The voice as a percept. *Journal of Personality and Social Psychology* 4, 1 (1966), 79.

[31] Yueh-Min Huang, Yen-Ting Lin, and Shu-Chen Cheng. 2010. Effectiveness of a mobile plant learning system in a science curriculum in Taiwanese elementary education. *Computers & Education* 54, 1 (2010), 47–58.

[32] Susan M Hughes and Marissa A Harrison. 2013. I like my voice better: Self-enhancement bias in perceptions of voice attractiveness. *Perception* 42, 9 (2013), 941–949.

[33] Ingrid S Johnsrude, Allison Mackey, Hélène Hakyemez, Elizabeth Alexander, Heather P Trang, and Robert P Carlyon. 2013. Swinging at a cocktail party: Voice familiarity aids speech perception in the presence of a competing voice. *Psychological science* 24, 10 (2013), 1995–2004.

[34] Jonas T Kaplan, Lisa Aziz-Zadeh, Lucina Q Uddin, and Marco Iacoboni. 2008. The self across the senses: an fMRI study of self-face and self-voice recognition. *Social cognitive and affective neuroscience* 3, 3 (2008), 218–223.

[35] Kyle Kercher. 1992. Assessing subjective well-being in the old-old: The PANAS as a measure of orthogonal dimensions of positive and negative affect. *Research on Aging* 14, 2 (1992), 131–168.

[36] Hadi Kharrazi, Robin Chisholm, Dean VanNasdale, and Benjamin Thompson. 2012. Mobile personal health records: an evaluation of features and functionality. *International journal of medical informatics* 81, 9 (2012), 579–593.

[37] Michiharu Kimura and Jonathan D Sime. 1989. Exit choice behaviour during the evacuation of two lecture theatres. *Fire Safety Science* 2 (1989), 541–550.

[38] Marino Kimura and Yuko Yotsumoto. 2018. Auditory traits of" own voice". *PloS one* 13, 6 (2018), e0199443.

[39] Paul R Kleinginna and Anne M Kleinginna. 1981. A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation and emotion* 5, 4 (1981), 345–379.

[40] Lauren Koff and Rebecca Mullis. 2011. Nutrition education and technology: can delivering messages via new media technology effectively modify nutrition behaviors of preschoolers and their families? *Journal of Nutrition Education and Behavior* 43, 4 (2011), S40.

[41] Phillippa Lally and Benjamin Gardner. 2013. Promoting habit formation. *Health psychology review* 7, sup1 (2013), S137–S158.

[42] Phillippa Lally, Cornelia HM Van Jaarsveld, Henry WW Potts, and Jane Wardle. 2010. How are habits formed: Modelling habit formation in the real world. *European journal of social psychology* 40, 6 (2010), 998–1009.

[43] Phillippa Lally, Jane Wardle, and Benjamin Gardner. 2011. Experiences of habit formation: a qualitative study. *Psychology, health & medicine* 16, 4 (2011), 484–489.

[44] Kwan Min Lee. 2004. Presence, explicated. *Communication theory* 14, 1 (2004), 27–50.

[45] Susannah V Levi, Stephen J Winters, and David B Pisoni. 2011. Effects of cross-language voice training on speech perception: Whose familiar voices are more intelligible? *The Journal of the Acoustical Society of America* 130, 6 (2011), 4053–4062.

[46] Huiyang Li, Xianghong Sun, and Kan Zhang. 2007. Voice alarm system in emergency evacuation. In *International Conference on Engineering Psychology and Cognitive Ergonomics*. Springer, 723–730.

[47] Roderick JA Little and Donald B Rubin. 1989. The analysis of social science data with missing values. *Sociological methods & research* 18, 2-3 (1989), 292–326.

[48] Justina Yat-Wa Liu, Patrick Pui-Kin Kor, Claire Pik-Ying Chan, Rick Yiu-Cho Kwan, and Daphne Sze-Ki Cheung. 2020. The effectiveness of a wearable activity tracker (WAT)-based intervention to improve physical activity levels in sedentary older adults: A systematic review and meta-analysis. *Archives of Gerontology and Geriatrics* 91 (2020), 104211.

[49] Dieter Maurer and Theodor Landis. 1990. Role of bone conduction in the self-perception of speech. *Folia phoniatrica* 42, 5 (1990), 226–229.

[50] Rowland S Miller and June Price Tangney. 1994. Differentiating embarrassment and shame. *Journal of Social and Clinical Psychology* 13, 3 (1994), 273–287.

[51] Katsuki Nakamura, Ryuta Kawashima, Motoaki Sugiura, Takashi Kato, Akinori Nakamura, Kentaro Hatano, Sumiharu Nagumo, Kisou Kubota, Hiroshi Fukuda, Kengo Ito, et al. 2001. Neural substrates for recognition of familiar voices: a PET study. *Neuropsychologia* 39, 10 (2001), 1047–1054.

[52] Clifford Nass and Li Gong. 2000. Speech interfaces from an evolutionary perspective. *Commun. ACM* 43, 9 (2000), 36–43.

[53] Clifford Nass and Kwan Min Lee. 2001. Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. *Journal of experimental psychology: applied* 7, 3 (2001), 171.

[54] David T Neal and Wendy Wood. 2009. Automaticity in situ and in the lab: The nature of habit in daily life. *Oxford handbook of human action* (2009), 442–457.

[55] Andreea Niculescu, George M White, See Swee Lan, Ratna Utari Waloejo, and Yoko Kawaguchi. 2008. Impact of English regional accents on user acceptance of voice user interfaces. In *Proceedings of the 5th Nordic conference on Human-computer interaction: building bridges*. 523–526.

[56] Kyue Taek Oh, Jisu Ko, Jaemyung Shin, and Minsam Ko. 2022. Using Wake-Up Tasks for Morning Behavior Change: Development and Usability Study. *JMIR Formative Research* 6, 9 (2022), e39497.

[57] Oladapo Oyebode, Mona Alhasani, Dinesh Mulchandani, Tolulope Olagunju, and Rita Orji. 2021. SleepFit: a persuasive mobile app for improving sleep habits in young adults. In *2021 IEEE 9th International Conference on Serious Games and Applications for Health (SeGAH)*. IEEE, 1–8.

[58] Samir Patel, Laura Jacobs-Kantor, Lorraine Marshall, Clark Ritchie, Michelle Kaplinski, Parvinder S Khurana, and Richard J Katz. 2013. Mobilizing your medications: an automated medication reminder application for mobile phones and hypertension medication adherence in a high-risk urban population.

[59] Reinhard Pekrun. 1992. The impact of emotions on learning and achievement: Towards a theory of cognitive/motivational mediators. *Applied psychology* 41, 4 (1992), 359–376.

[60] Wei Peng, Lin Li, Anastasia Kononova, Shelia Cotten, Kendra Kamp, Marie Bowen, et al. 2021. Habit formation in wearable activity tracker use among older adults: qualitative study. *JMIR mHealth and uHealth* 9, 1 (2021), e22488.

[61] Wei Peng, Hayeon Song, Jinyoung Kim, and Tom Day. 2016. The influence of task demand and social categorization diversity on performance and enjoyment in a language learning game. *Computers & Education* 95 (2016), 285–295.

[62] Xuan-Lam Pham, Thi-Huyen Nguyen, Wu-Yuin Hwang, and Gwo-Dong Chen. 2016. Effects of push notifications on learner engagement in a mobile learning app. In *2016 IEEE 16th International Conference on Advanced Learning Technologies (ICALT)*. IEEE, 90–94.

[63] Meihua Piao, Hyeongju Ryu, Hyeongsuk Lee, Jeongeun Kim, et al. 2020. Use of the healthy lifestyle coaching chatbot app to promote stair-climbing habits among office workers: exploratory randomized controlled trial. *JMIR mHealth and uHealth* 8, 5 (2020), e15085.

[64] Naincie Pindeh, Norbayah Mohd Suki, and Norazah Mohd Suki. 2016. User acceptance on mobile apps as an effective medium to learn Kadazandusun language. *Procedia Economics and Finance* 37 (2016), 372–378.

[65] Guylène Proulx. 2007. Response to fire alarms. *Fire Protection Engineering* 33 (2007), 8.

[66] Guylene Proulx and Jonathan D Sime. 2006. To prevent 'panic'in an underground emergency: why not tell people the truth?. In *Fire Safety Science—Proceedings of the Third International Symposium*. Routledge, 843–852.

[67] Michael Pucher, Bettina Zillinger, Markus Toman, Dietmar Schabus, Cassia Valentini-Botinhao, Junichi Yamagishi, Erich Schmid, and Thomas Woltron. 2017. Influence of speaker familiarity on blind and visually impaired children's and young adults' perception of synthetic voices. *Computer Speech & Language* 46 (2017), 179–195.

[68] David Purser. 2010. Comparisons of evacuation efficiency and pre-travel activity times in response to a sounder and two different voice alarm messages. In *Pedestrian and evacuation dynamics 2008*. Springer, 121–134.

[69] Jun Rekimoto and Hitomi Tsujita. 2014. Inconvenient interactions: an alternative interaction design approach to enrich our daily activities. In *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces*. 225–228.

[70] Timothy B Rogers, Nicholas A Kuiper, and William S Kirker. 1977. Self-reference and the encoding of personal information. *Journal of personality and social psychology* 35, 9 (1977), 677.

[71] Christine Rosa, Maryse Lassonde, Claudine Pinard, Julian Paul Keenan, and Pascal Belin. 2008. Investigations of hemispheric specialization of self-voice recognition. *Brain and cognition* 68, 2 (2008), 204–214.

[72] Alexander J Rothman, Austin S Baldwin, Andrew W Hertel, and Paul T Fuglestad. 2011. Self-regulation and behavior change: Disentangling behavioral initiation and behavioral maintenance. (2011).

[73] Clyde Rousey and Philip S Holzman. 1967. Recognition of one's own voice. *Journal of Personality and Social Psychology* 6, 4p1 (1967), 464.

[74] Paschal Sheeran. 2002. Intention—behavior relations: a conceptual and empirical review. *European review of social psychology* 12, 1 (2002), 1–36.

[75] Richard M Shiffrin and Walter Schneider. 1977. Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological review* 84, 2 (1977), 127.

[76] Linda I Shuster and John D Durrant. 2003. Toward a better understanding of the perception of self-produced speech. *Journal of Communication Disorders* 36, 1 (2003), 1–11.

[77] Gary A Smith, Thitphalak Chounthirath, and Mark Splaingard. 2019. Effectiveness of a voice smoke alarm using the child's name for sleeping children: a randomized trial. *The Journal of pediatrics* 205 (2019), 250–256.

[78] Gary A Smith, Mark Splaingard, John R Hayes, and Huiyun Xiang. 2006. Comparison of a personalized parent voice smoke alarm with a conventional residential tone smoke alarm for awakening children. *Pediatrics* 118, 4 (2006), 1623–1632.

[79] IS Statistics. 2013. IBM Corp. Released 2013. IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp. *Google Search* (2013).

[80] Motoaki Sugiura, Ryuta Kawashima, Katsuki Nakamura, Ken Okada, Takashi Kato, Akinori Nakamura, Kentaro Hatano, Kengo Itoh, Shozo Kojima, and Hiroshi Fukuda. 2000. Passive and active recognition of one's own face. *Neuroimage* 11, 1 (2000), 36–48.

[81] Selina Jeanne Sutton, Paul Foulkes, David Kirk, and Shaun Lawson. 2019. Voice as a design material: Sociophonetic inspired design strategies in human-computer interaction. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–14.

[82] Dallas Swendeman, Nithya Ramanathan, Laura Baetscher, Melissa Medich, Aaron Scheffler, W Scott Comulada, and Deborah Estrin. 2015. Smartphone self-monitoring to support self-management among people living with HIV: Perceived benefits and theory of change from a mixed-methods, randomized pilot study. *Journal of acquired immune deficiency syndromes (1999)* 69, 0 1 (2015), S80.

[83] Cynthia S Symons and Blair T Johnson. 1997. The self-reference effect in memory: a meta-analysis. *Psychological bulletin* 121, 3 (1997), 371.

[84] June Price Tangney, Rowland S Miller, Laura Flicker, and Deborah Hill Barlow. 1996. Are shame, guilt, and embarrassment distinct emotions? *Journal of personality and social psychology* 70, 6 (1996), 1256.

[85] Robert Tobias. 2009. Changing behavior by memory aids: a social psychological model of prospective memory and habit development tested with dynamic field data. *Psychological review* 116, 2 (2009), 408.

[86] Pernilla Ulfvengren. 2003. *Design of natural warning sounds in human-machine systems*. Ph. D. Dissertation. KTH.

[87] Janne Vainio, Kirsikka Kaipainen, and Ilkka Korhonen. 2014. Habit change as a learning process: Design framework for mobile interventions. In *IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*. IEEE, 801–804.

[88] Viswanath Venkatesh and Fred D Davis. 2000. A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management science* 46, 2 (2000), 186–204.

[89] Bas Verplanken and Sheina Orbell. 2003. Reflections on past behavior: a self-report index of habit strength 1. *Journal of applied social psychology* 33, 6 (2003), 1313–1330.

[90] Patrik Vuilleumier. 2005. How brains beware: neural mechanisms of emotional attention. *Trends in cognitive sciences* 9, 12 (2005), 585–594.

[91] John Wheale. 1982. Decrements in performance on a primary task associated with the reaction to voice warning messages. *British Journal of Audiology* 16, 4 (1982), 265–272.

[92] John G Wirtz. 2020. Does it matter who's at risk?: Testing message framing and narrative as moderators of the effects of self-referencing in obesity prevention messages targeting Hispanic adults living in the US. *Health marketing quarterly* 37, 1 (2020), 22–40.

[93] Wendy Wood, Jeffrey M Quinn, and Deborah A Kashy. 2002. Habits in everyday life: thought, emotion, and action. *Journal of personality and social psychology* 83, 6 (2002), 1281.

[94] Qun Wu. 2015. Designing a smartphone app to teach English (L2) vocabulary. *Computers & Education* 85 (2015), 170–179.

[95] Qicong Xie, Xiaohai Tian, Guanghou Liu, Kun Song, Lei Xie, Zhiyong Wu, Hai Li, Song Shi, Haizhou Li, Fen Hong, et al. 2021. The multi-speaker multi-style voice cloning challenge 2021. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 8613–8617.

[96] Galit Yovel and Pascal Belin. 2013. A unified coding strategy for processing faces and voices. *Trends in cognitive sciences* 17, 6 (2013), 263–271.