TandemTrack: Shaping Consistent Exercise Experience by Complementing a Mobile App with a Smart Speaker

Yuhan Luo
University of Maryland
College Park, MD, USA
yuhanluo@umd.edu

Bongshin Lee
Microsoft Research
Redmond, WA, USA
bongshin@microsoft.com

Eun Kyoung Choe
University of Maryland
College Park, MD, USA
choe@umd.edu

ABSTRACT
Smart speakers such as Amazon Echo present promising opportunities for exploring voice interaction in the domain of in-home exercise tracking. In this work, we examine if and how voice interaction complements and augments a mobile app in promoting consistent exercise. We designed and developed TandemTrack, which combines a mobile app and an Alexa skill to support exercise regimen, data capture, feedback, and reminder. We then conducted a four-week between-subjects study deploying TandemTrack to 22 participants who were instructed to follow a short daily exercise regimen: one group used only the mobile app and the other group used both the app and the skill. We collected rich data on individuals' exercise adherence and performance, and their use of voice and visual interactions, while examining how TandemTrack as a whole influenced their exercise experience. Reflecting on these data, we discuss the benefits and challenges of incorporating voice interaction to assist daily exercise, and implications for designing effective multimodal systems to support self-tracking and promote consistent exercise.

Author Keywords
Self-tracking; multimodal interaction; smart speaker; mobile app; exercise; field deployment study.

CCS Concepts
•Human-centered computing → Human computer interaction (HCI); User centered design; Sound-based input / output; Auditory feedback; User studies;

INTRODUCTION
Technologies such as wearable devices and mobile apps have been developed to assist exercise, supporting workout planning and tracking, as well as providing progress feedback. While these technologies can help people fulfill their needs related to exercise, a sizable percentage of the population that could benefit from these technologies still does not engage with them [37], failing to meet the recommended level of exercise guidelines [7].

CHI'20, April 25–30, 2020, Honolulu, HI, USA

ISBN 978-1-4503-6708-0/20/04...$15.00
DOI: https://doi.org/10.1145/3313831.3376616

In recent years, voice interaction has been rapidly integrated into people's daily lives, with the introduction of home-based smart speakers such as Amazon Echo [5] and Google Home [11]. In addition to playing music and responding to simple queries, smart speakers provide third-party applications (e.g., Alexa skills, Google Actions), which support a variety of activities including in-home workouts. In particular, many of these voice applications support people's workout (e.g., 7-minute Workout [3], 30-day Push-up Challenge [1]) through an exercise program. As of December 2019, more than 2,000 Alexa skills are listed under the "Health & Fitness" category.

Voice interaction on smart speakers holds promises to support consistent exercise in three aspects. First, with the hands-free interaction capability, voice interaction can lower the data capture burden during exercise. Second, in the home environment where people do not necessarily carry their mobile phone all the time, the voice prompt from a smart speaker can increase the chance for them to receive the exercise reminder. Third, smart speakers' lack of mobility can be helpful, forcing a consistent exercise location, which aids in creating habits and routines [42]. However, smart speakers' main drawback as an exercise supporting device is in its lack of a visual component, a valuable medium for providing feedback and promoting self-reflection. Popular mobile apps for exercise, on the other hand, provide a visual interface while neglecting the convenience that the smart speaker may provide.

In this light, we examine how a smart speaker's voice interaction complements and augments a mobile app in supporting consistent exercise. We designed and developed TandemTrack, a multimodal system coupling a mobile app and an Alexa skill, which offers exercise regimen on sit-ups and push-ups, captures workout repetitions, provides performance feedback, and sends daily reminders. We then conducted a four-week between-subjects study deploying TandemTrack with 22 participants, followed by semi-structured interviews. All the participants had not have the habit of performing sit-ups and push-ups regularly. During the study, 11 participants used only the mobile app to perform exercise (which we refer to as the "M" group), while the other 11 interacted with both the mobile app and Alexa skill (the "MA" group).

To the best of our knowledge, this work is the first study that uncovers the factors influencing how people choose between a mobile app and smart speaker's voice interaction to perform daily exercise. We found that, even though the two groups did not differ in their exercise adherence and performance, they interacted with the mobile app differently: compared with the M group, the MA group spent less time on reviewing their exercise feedback, but edited their exercise data more often. When it comes to choosing between the app and the skill to perform and capture their exercise, MA participants' decisions were influenced by their personal preferences, proximity to Amazon Echo, exercise environment, the social context around them, and the technical issues they encountered using Amazon Echo. We also provide an empirical understanding of benefits and challenges of using smart speakers' voice interaction in supporting exercise regimen, data capture, feedback, and reminders. Furthermore, reflecting on the data from the deployment study and follow-up interviews, we discuss implications for designing an integrated multimodal experience for self-tracking and consistent exercise.

RELATED WORK

In this section, we review existing work on mobile health and voice assistants that support health and fitness activities. We particularly focus on the behavior change techniques employed in these work, and identify opportunities for multimodal systems to support self-tracking and promote consistent exercise.

Mobile Apps for Health & Fitness

The mobile health (mHealth) app market has been expanding over the past decades, reducing the costs of patient care and improving consumer health [30, 48]. Mobile phones' rich sensing capabilities make it possible to collect personal data and send behavior change cues in situ [54]. In 2017, 325,000 mobile health apps were available on Apple and Google Play Store, and the number is continuously growing [43]. According to the national survey on mHealth, 58% of mobile phone users have downloaded and used health apps, with an increasing interest in fitness training [34]. Researchers have found that current fitness app users are more physically active and have lower body mass index (BMI), compared with non-users and past-users [23, 37]. However, this relationship was moderated by exercise barriers such as inconvenience, lack of time, and environmental constraints. In other words, people who reported more exercise barriers were less likely to engage with the fitness apps, and therefore do not benefit from the fitness support provided by these apps [37].

To support people in overcoming exercise barriers, researchers and developers have created technologies incorporating behavior change techniques [26, 27, 29, 40, 46, 62]. Of particular interest is the following four techniques, which are commonly employed in many existing health and fitness apps: exercise guidance [26, 40], self-tracking [26], performance feedback [26, 27, 40], and reminders [46]. Popular fitness training apps such as Nike Training Club [18], SWORKIT [21], and Keep Trainer [14], provide over hundreds of exercise guidance in various length and intensity, ranging from basic stretch to full-body workout. To help people understand the proper postures, the guidance usually take them through a series of exercises with a video demonstration; some leveraged 3D techniques to explain the details of body movement [55, 56]. While guiding people to exercise, many apps also capture their exercise data and provide feedback on their performance. For example, Nike Training Club [18] has an activity panel, which summarizes individuals' total workouts, highlights their achieved milestones, and provides a list of their workout history. To encourage consistent exercise routine, Keep Trainer [14] allows people to pick an exercise time, and sends them daily reminders. Despite of being equipped with important behavior change techniques, the capabilities of these apps are limited in one modality—the mobile phone. In this work, we complement mobile apps with a smart speakers as a way to enhance existing mHealth infrastructure [60].

Although not part of health and fitness apps, conversational agents (CAs) have been explored in supporting people to gain awareness and reflect on their health and activity data [32, 36]. Commercial apps such as Lark [16] and HealthyBot [12] serve as a "Chatbot," actively initiating conversations with people by asking about their daily activities and well-being. By processing the natural language input and generating responses as feedback, the CAs present opportunities to engage people with their personal health data in new ways [32]. However, such communication requires people to manually input their activity information, which can add extra interaction burden [32].

Voice Assistants for Health & Fitness

With the release of home-based smart speakers such as Amazon Echo and Google Home, voice-activated assistants have gained a great popularity in recent years: it has been predicted that over 75% U.S. household will own at least one smart speaker by 2020 [9]. In the HCI community, we see a growing interest in learning about how people use smart speakers in their daily life [28, 38, 44, 45, 47]. Researchers have found several benefits that smart speakers offer, including improved accessibility for elders and disabled people [45], playfulness for kids [28], and perceived companionship [38, 47, 59].

While smart speakers are primarily used for playing music and querying information [58], numerous voice applications (e.g., Alexa skills, Google Actions)—the equivalent of apps on the mobile phone—have been developed by third parties. In 2017, researchers conducted a content analysis on Amazon Alexa skill and Google Assistant Store and found 309 voice applications in the "Health & Fitness" category, among which 72 (23.3%) supported fitness training and 56 (18.1%) could be paired with a mobile app, website, or other devices [25]. With the exponential growth, over 2,000 Alexa skills are listed under the "Health & Fitness" category as of December 2019.

Taking Alexa skills as an example, many support fitness training by providing audio-based exercise regimens (e.g., 7-Minute Workout [3], 5-Minute Plank [2]). For instance, 7-Minute Workout [3] guides people through 14 sets of workouts by informing them of the upcoming workout before the next set and starting the timer upon receiving the invoke word "*Ready*". These skills make it easy to follow the exercise regimen with the hands-free interaction, but most of them lack support for data capture and exercise feedback—a valuable means to engage people with their exercise data. On the other hand, a few health & fitness mobile apps have a skill version (e.g., Fitbit skill [8], MyFitnessPal skill [17]), which allows people to ask questions about their data captured by other devices (e.g., mobile phone, wearable device), but does not support data capture using voice.

Although not using smart speakers, researchers have explored the opportunities to use voice-enabled interface to support healthy living [59, 64]. Turunen and colleagues created a multimodal health and fitness companion consisting of a mobile app and a physical intelligent agent to provide health advice [59]. While their work focused on examining the dialogue details and speech recognition accuracy, we are interested in exploring how people use the mobile app and the smart speaker in the context of performing and tracking daily exercise.

TANDEMTRACK

Our goal of this work is to examine if and how voice interaction complements and augments a mobile app in promoting consistent exercise. Therefore, TandemTrack should be considered as a research prototype that situates people to interact with a smart speaker and a mobile phone in a simple exercise context, not as the most powerful exercise intervention. In addition, we aimed to maximize the advantages of the voice interaction and the mobile app, instead of competing one against the other. To effectively answer our research question, we revised the initial design of TandemTrack, which was proposed in our earlier work [52]. Here, we describe our design goals, TandemTrack design components, and implementation details.

Design Goals

DG1. Lower Exercise Barriers

Given that lack of time, inconvenience, and environment constraints are common barriers to exercise [33, 50, 51], we aim to lower these barriers through a simple and basic exercise regimen that allows people to complete the exercise quickly at any location. To this end, we promote healthy exercise in the simplest way possible by choosing a common indoor exercise routine such as sit-ups and push-ups that are beneficial and can be completed with minimal efforts.

DG2. Leverage Smart Speakers' Unique Characteristics

The Amazon Echo has a set of features and capabilities: while some of them overlap with those of mobile devices, others are unique. Because we ultimately want to leverage the potential synergy that would come from effectively combining these two interaction modalities, TandemTrack was designed to explore what modality would better suit people's needs and preferences in delivering exercise regimen, supporting data capture, providing feedback, and sending reminders.

DG3. Promote Consistency in Exercise Time and Location

Performing a task every day, especially at the same time and place, leads to habit formation or strengthening of existing habits, where habits are behaviors that become automatic after they have been performed consistently and repeatedly [35]. Therefore, we aim to encourage exercise on a daily basis, ideally taking place at the same time and the same place.

TandemTrack Components

In the following, we describe how we achieved the above design goals with TandemTrack's four key components.

Providing a Simple yet Beneficial Exercise Regimen

Unlike many fitness apps that provide diverse and complicated workout guidance, TandemTrack minimizes the complexity of the exercise process to allow people to start exercising easily (DG1). TandemTrack delivers an exercise regimen that alternates between sit-ups and push-ups by day, focusing on the timing and repetitions (DG3). Each exercise session consists of three timed sets and two breaks in-between; each set lasts 30 seconds and each break lasts up to 60 seconds. Figure 1 shows an example process of using the TandemTrack skill to do push-ups. To allow people to control their exercise pace, TandemTrack provides options for resetting the current set and skipping the current break.

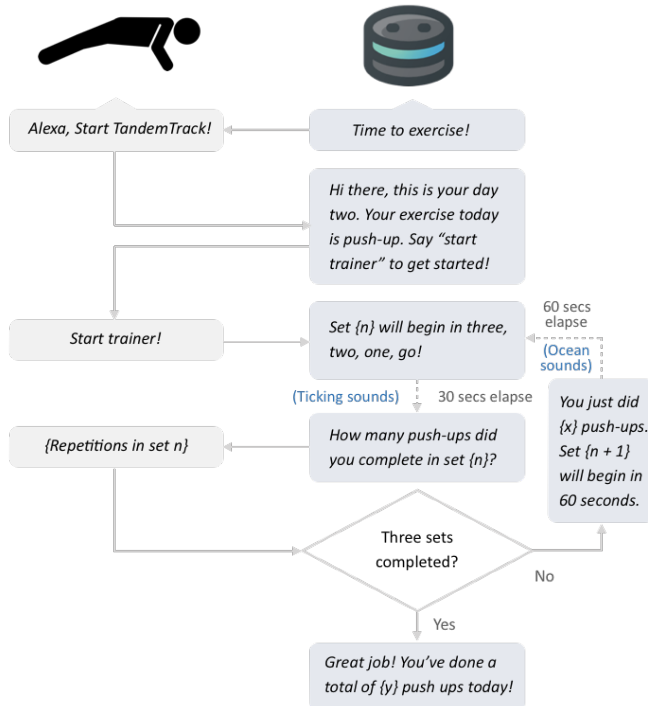

Figure 1. An example of using the TandemTrack skill to do push-ups.

We chose to alternate between sit-ups and push-ups because (1) these two exercises can be performed at most locations within a short amount of time (less than five minutes); (2) they are well-known and commonly used in general fitness tests, and even fitness novices can quickly get started without specific instructions; (3) despite the short exercise duration, the intensity of sit-ups and push-ups helps build up muscle strength and enhance glucose metabolism [53, 63]; (4) sit-ups and push-ups use two different muscle groups, and thus alternating between the two allows each muscle group to rest on the day of the other exercise [39, 49].

Enabling Voice-Activated Data Capture

Exercise data can be captured at different dimensions, ranging from simple entries to detailed metrics. For example, RunKeeper [20] measures runners' performance by capturing their running distance, average pace, time spent, path, heart rate, calorie burned, etc. JEFIT [13], on the other hand, enables people to manually capture exercise-related information in various formats such as number of repetitions, text notes, body metrics (e.g., height, weight), and photos.

In contrast, TandemTrack lowers the data capture burden (DG1): it only captures the number of repetitions per set for sit-up/push-up, by prompting people right after the completion of each set when the number is still fresh in their mind. The mobile app pops up a text field for people to manually type the numbers; the Alexa skill asks *"how many sit-ups/push-ups did you complete?"* and then takes people's voice responses as input (DG2). Unlike existing fitness training skills which do not support data capture or only provide feedback on the data captured by other devices, the TandemTrack skill supports data capture independently of other devices. In case of typos on the mobile app, or inaccurate voice recognition of Alexa, TandemTrack allows people to edit their repetition data after exercise using the mobile app.

Enriching Feedback Through Multimodal Interaction

With the exercise data collected, the TandemTrack app provides visual feedback in three parts (Figure 2): (1) the daily exercise feedback on the top (A) shows the sit-up/push-up repetitions; people can swipe left and right to review the repetitions of sit-ups or push-ups they completed on each day; (2) the middle part (B) summarizes the longest streak (i.e., the consecutive days of doing exercise) and the total number of completed exercise sessions; (3) the aggregated feedback on the bottom shows a series of time-based visualizations, including exercise streak view (C1), sit-up performance (C2), and push-up performance (C3).

Complementing the visual feedback, TandemTrack skill provides voice feedback in a Q&A manner (DG2). People can ask about their "*exercise summary*" to receive a summary of their exercise progress, including total exercise sessions, current streak, longest streak, and average sit-up/push-up repetitions per session. They can also ask specific questions on their exercise records, such as "*how many push-ups did I complete yesterday*" and "*what is my best sit-up performance*."

Facilitating Exercise Routine with Daily Reminders

Because forgetfulness is one of the most common reasons that prevent people from building a habit [54], TandemTrack asks people to set a daily reminder by picking a time when they want to exercise (DG3). Considering that people may not be closed to the Amazon Echo when the reminder is sent, TandemTrack reminds people using both modalities: a mobile notification from the app and an Alexa reminder from the Amazon Echo. To force a consistent exercise routine, TandemTrack does not allow people to change their reminder time once it is set.

Implementation

We implemented the TandemTrack mobile app using Kotlin [15]. For the TandemTrack Alexa skill, we used Alexa Skills Kit (ASK) [4] to build the front-end voice interface, and Node.js [19] for the back-end service, which is hosted on the Amazon Web Servic Lambda [6]. The TandemTrack skill is distributed through Amazon Alexa skill's beta test that is only available to people who are invited to the test.

To explore what modality would better suit people's needs and preferences in delivering exercise regimen, supporting data capture, providing feedback, and sending reminders, we strove to provide equivalent features for both the mobile app and Alexa skill. In this way, people could use the app and the skill without being influenced by extra or missing features on either modality, and we could examine the specific advantages and drawbacks of voice interaction (on smart speakers) regarding each feature. However, the mobile phone and the Amazon Echo has its own characteristics, making some of the interaction experience inherently different. For example, the mobile notification stays on the screen as long as the user does not remove it; but the voice reminder sent from the Amazon Echo does not have the same visibility due to the ephemeral nature of voice.

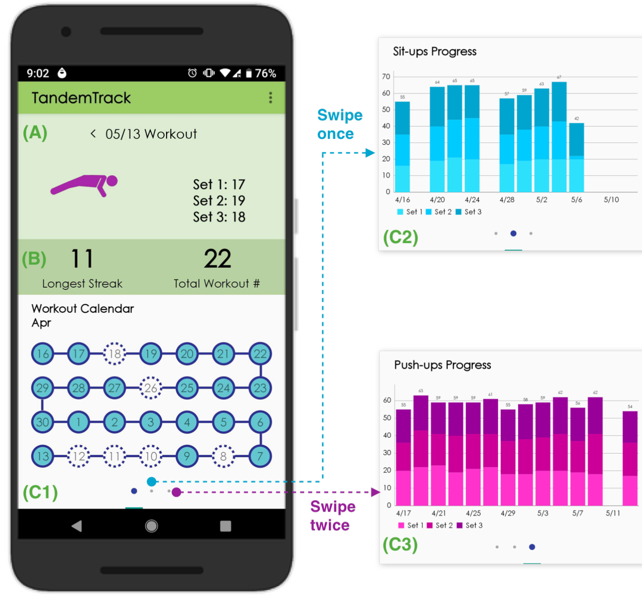

Figure 2. The home screen of the TandemTrack mobile app: the daily exercise feedback (A); a summary of the longest streak and complete exercise sessions (B); a series of aggregated feedback—exercise streak view (C1), sit-up progress (C2), and push-up progress (C3).

The TandemTrack app and the skill shared access to the same database on Google Firebase [10], which synchronizes people's exercise data on both devices. TandemTrack allows people to hear their exercise feedback from the skill while reviewing the visual feedback on their phone, but it does not support people to perform exercise with the app and the skill at the same time due to the technical difficulties.

METHOD

We conducted a between-subjects study with 22 participants for four weeks. The study was approved by the university's Institutional Review Board.

Participants

We advertised the study through the university mailing list. Among the 44 people who responded, 25 met our inclusion criteria: individuals who (1) are over 18; (2) own an Android phone that runs an operating system 6.0.0 or above; (3) have stable Internet access at home; (4) are motivated to do short strength exercise (i.e., sit-ups & push-ups) daily and collect their exercise data; and (5) currently do not do sit-ups or push-ups regularly (i.e., two times or more per week).

With an aim to create a balanced group assignment regarding participants' gender, age, living environment, and motivation levels[1], we assigned 12 participants to use the TandemTrack mobile app only (M) and the other 13 to use both the mobile app and the Alexa skill (MA). To ensure that all the MA participants were first-time smart speaker users, we intentionally assigned the three participants who already owned a smart speaker (e.g., Amazon Echo, Google Home) to the M group (M-4, M-8, M-9)[2].

One MA participant withdrew from the study because his phone was broken. We also excluded the data of one M participant and one MA participant who were found to be housemates and overlapped for two weeks participating in the study. We thus analyzed the data of the remaining 22 participants (11 in M, 11 in MA; 11 female, 11 male). They were 16 graduate students, 3 undergraduate students, and 3 full-time university staff, and their ages ranged from 20 to 61 (M = 26.0, SD = 8.3). Five M participants and five MA participants were non-native English speakers. All participants lived off-campus with housemates or family members during the study period, except for M-11 who lived independently in an apartment.

Study Procedure

The study consisted of three stages: (1) study tutorial, (2) four-week deployment with TandemTrack, and (3) debriefing interview. After completing the study, each participant received a $60 Amazon gift card as compensation. We loaned each MA participant an Amazon Echo Dot (3rd generation), which they returned back at the end of the study.

Study Tutorial

At the beginning of the study, we scheduled an in-person tutorial with each participant in our lab. We introduced the study procedures, instructed participants to install the TandemTrack mobile app, and helped them set a reminder. We explained that the reminder time could not be changed throughout the study, and asked participants to carefully pick a time when they were most likely to exercise. To demonstrate how TandemTrack works, we asked participants to perform a full session of situps by following the TandemTrack app's exercise regimen, on the Yoga mat we prepared. To make sure that participants understand how to interpret the visual feedback within the app, we explained each part of the visualization in detail.

With MA participants, we additionally created a new Amazon account for each of them, and demonstrated how to connect the Echo Dot to Wifi. Given that MA participants were new to smart speakers, we also showed them how to communicate with Alexa using basic voice queries (e.g., asking weather, playing music). We then showed them how to perform exercise and ask feedback with the TandemTrack skill. To ensure that MA participants would not receive duplicated exercise reminders from their Alexa app (i.e., the companion app of Amazon Echo), we asked them to turn off the notification option for the Alexa app (they would still receive the exercise reminder from the TandemTrack app and the Echo Dot). In addition, we asked MA participants to put the Echo Dot at a place where they could do exercise nearby, and clarified that TandemTrack is a multimodal system—their exercise would be synchronized on both devices. As such, MA participants were encouraged to explore the app and the skill as they liked.

The tutorial lasted about 30 minutes for M participants, and 60 minutes for MA participants. At the end of the tutorial, each participant received a study manual (including a physical copy and a digital copy) that described common usages of the TandemTrack app. The manual that MA participants received also included usages of the Echo Dot and the TandemTrack skill. At the end of that day, we confirmed with MA participants that they have connected the Echo Dot to Wifi.

Field Deployment

The day after the study tutorial, participants started using TandemTrack to exercise for the following four weeks. At the end of each week, we sent each participant a weekly diary using Google form, asking them to briefly respond to three short questions: (1) "*What was your experience like with TandemTrack over the past week*?", (2) "*Did you experience any technical difficulties with TandemTrack*?", and (3) "*Is there anything you want to share with us*?" The purpose of sending a weekly diary was to collect participants' feedback on TandemTrack as the study progressed, and to check if they encountered any technical issues. We did not send a weekly diary at the end of the forth week, because we were able to talk with participants during the debriefing interviews.

Debriefing Interview

We conducted a semi-structured interview with each participants at the end of the study. To help participants better recall their exercise experience, we asked them to refer their exercise feedback on the TandemTrack app while having them describe their exercise locations, environments, and reasons for missing the exercise on certain days. We also asked how they used the four key features (i.e., exercise regimen, data capture, feedback, and reminder) provided by TandemTrack. For MA participants, we asked additional questions regarding their preferences and use of the mobile app and the Alexa skill, and their experience in interacting with the Echo Dot. Each interview lasted 20 to 45 minutes.

Data Analysis

We first compared the two groups regarding their exercise adherence (i.e., complete sessions, longest and average streak) and interaction with the TandemTrack app (e.g., opening the app, reviewing feedback) using an independent t-test. We then examined how participants' exercise performance (i.e., average repetitions per session) changed as the study progressed using a linear regression analysis. We also examined the factors affecting their exercise adherence using a multiple linear regression, as well as the factors affecting their performance change using a multiple logistic regression.

We audio-recorded all the interviews, took verbatim notes, and digitized all the weekly diaries. We analyzed the qualitative data from interviews and diaries together to answer the following questions: (1) why participants missed exercise;

[1]We asked participants "*How motivated are you to form a strength exercise habit (i.e., push-ups, sit-ups)?*" in the screening questionnaire to capture their motivation levels using a five-point scale (from 1: "*Not motivated at all*" to 5: "*Very motivated*").

[2]We use M-# and MA-# to denote a participant ID in two groups.

M participants	M-1	M-2	M-3	M-4	M-5	M-6	M-7	M-8	M-9	M-10	M-11	Average	Longest streak	Average streak
App sessions	28	20	28	17	24	12	27	13	15	15	17	19.63	10.90	8.27
MA participants	**MA-1**	**MA-2**	**MA-3**	**MA-4**	**MA-5**	**MA-6**	**MA-7**	**MA-8**	**MA-9**	**MA-10**	**MA-11**	**Average**	**Longest streak**	**Average streak**
App sessions	19	3	15	8	4	13	10	8	12	7	22	11 (57%)		
Skill sessions	0	17	2	5	20	0	14	11	6	15	1	8.27 (43%)	8.44	6.02
Total sessions	19	20	17	13	24	13	24	19	18	22	23	19.27		

Table 1. The number of exercise sessions each participant completed with the TandemTrack app & the skill.

Engagement	M	MA	
	app	app	skill
# Opening TandemTrack	26.36	20.72	12.90
# Exercise sessions	19.63	11	8.27
Average time spent on reviewing feedback (seconds)	12.70	6.45	NA
# Swiping the daily feedback	37.09	31.72	NA
# Swiping the aggregated feedback	44.63	33.45	NA
# Tapping the mobile notification	13.09	7.81	NA
Average time elapse between reminder and exercise (minutes)	296	337	252
# Editing exercise data	0.09	1	NA

Table 2. Descriptive summary of participants' engagement with TandemTrack during the entire study period.

(2) what they liked or disliked about the four key features provided by TandemTrack (exercise regimen, data capture, feedback, reminder); (3) what factors influenced MA participants' choice of the modality in interacting with TandemTrack; and (4) what challenges participants encountered when interacting with TandemTrack. The first author used an inductive approach to first identify codes as labels from participants' elaborations pertaining to the above questions. The process was complemented with a top-down (deductive) approach in that we specifically organized the codes (findings) according to how participants reacted to each of the four features. This first phase resulted in a list of codes with associated quotes categorized under the four features, which were later re-grouped and organized into potential themes. This qualitative analysis complemented the quantitative results drawn from adherence, performance, and usage logs.

RESULTS

Throughout the four-week deployment study, TandemTrack collected 428 exercise sessions (217 sit-up & 211 push-up sessions), 2,892 interaction logs with the mobile app, and 445 utterances with the skill. In addition, we collected 66 weekly diary entries, in which participants described their experience (including usability issues) with TandemTrack and reasons for missed exercise.

Exercise Behavior: Adherence and Performance

The two groups—M and MA—did not differ in their exercise completion rate, change in performance, and exercise adherence (i.e., complete sessions, longest and average streak length; see Table 1). Therefore, we combined the two groups for the remaining analyses on examining the factors affecting their exercise adherence and change in performance.

We found that, as the study progressed, the performance for sit-ups has increased for 11 participants, decreased for two, and stayed the same for nine. Participants who completed more sit-up sessions were more likely to increase their sit-up performance ($OR = .52$, $p = .04$). On the other hand, the performance for push-ups has increased for eight participants, decreased for one, and stayed the same for 13. Participants who completed more push-up sessions did not necessarily improve their push-up performance.

From the weekly diaries and interviews, we found the following reasons why participants missed the exercise: busyness and prior commitments ($n = 13$); physical difficulties ($n = 6$); forgetfulness ($n = 6$); procrastination ($n = 6$); and missing a reminder ($n = 4$).

Usage of TandemTrack

In this section, we report how participants used TandemTrack based on the interaction logs and interview data. Table 2 shows how participants in both groups engaged with different features of TandemTrack. In comparison to M participants ($M = 26.36$, $SD = 6.82$), MA participants ($M = 20.72$, $SD = 6.59$) opened the TandemTrack app less often, $t(20) = 2.20$, $p = .04$. This result is expected because MA participants could also use the skill, which provided the same features. In what follows, we focus on what MA participants preferred between the app and the skill to conduct different actions and reasoning behind their choice.

Exercise Regimen & Data Capture

In total, MA participants completed 212 exercise sessions, 57% of which were completed and captured using the TandemTrack app (see Table 1). We observed a large differences among the MA participants in what they used to exercise and capture data. Out of 11 MA participants, five used the TandemTrack skill more than the app for exercise regimen and data capture. These five MA participants had better exercise adherence (completed session: 21.8, longest streak length: 11, average streak length: 8.9) than the other six MA participants who used the TandemTrack app more (completed session: 17.2, longest streak length: 6, average streak length: 3.6), though we did not run a statistics test due to the small sample size. Notably, two MA participants (MA-1 and MA-6) did not use the skill at all, which we explain later.

MA participants shared their rationale for choosing between the app and the skill to exercise during the interview: their decisions were influenced by their personal preference, proximity to the Echo Dot, their exercise environment, the social context around them, and the technical issues they encountered using the Echo Dot.

Personal Preferences: MA-2, MA-5, MA-7, and MA-10 showed a strong preference for the TandemTrack skill over the app for performing and capturing exercise, as they discovered the benefits that voice interaction offers. First, the hands-free interaction made the data capture more convenient. MA-10, who preferred using TandemTrack skill especially for push-ups, explained that "*When doing push-ups, you're putting a lot [of] strength to your hands, so it's a bit difficult to type [on the phone], but Alexa makes it easier*." In addition, MA-2 and MA-7 called out that the voice-based exercise regimen helped reduce distraction, so that they could focus on their performance. For example, MA-2 remarked, "*I don't have to look at my phone, but just listen to it [the Echo Dot] and do my exercise. It was less distracting, and I got to focus on my body and my performance.*"

On the contrary, MA-1 and MA-6 resisted interacting with the Echo Dot, because they did not feel comfortable talking to Alexa. MA-1 explained that "*This is not about privacy, I just did not feel comfortable talking to a personified machine. If it doesn't have name, maybe I'll try to talk to it.*" Similarly, MA-6 noted that "*It's something like almost human but not. The way to interact with it is not intuitive, it's awkward.*"

Proximity: Participants usually exercised in their home, but occasionally, they also reported exercising at the gym, friends' home, private rooms in the library, and in their offices. At home, some participants always exercised at the same place (e.g., bedroom) while others switched places (e.g., living room and basement) depending on their convenience. It was common for participants to exercise with the TandemTrack skill if they happened to be close to the Echo Dot when they received the exercise reminder. For example, MA-9 said "*I have my phone with me all the time, but if I was there when Echo reminded me, I would use it because it's convenient.*"

Exercise Environment: To effectively exercise with the TandemTrack skill, participants needed enough space to do sit-ups and push-ups while being proximate enough to the Echo Dot. For example, MA-10 rearranged her room by moving a desk next to her bed, and found it was no longer easy for her to exercise using the skill: "*This narrowed the space between the desk, where I put Alexa, and my bed. So I had to do exercise near the door of my room, but I couldn't use Alexa anymore—it's not close enough*."

Social Context: When the Echo Dot was placed at a location where other house members could access, participants worried that exercising with the TandemTrack skill might interrupt others. For example, MA-8 and MA-9 tended to exercise using the TandemTrack skill when other house members were not around, because they did not want to bother others by speaking to Alexa in the early morning or late night.

Although most MA participants did not bring up privacy concerns while exercising with the TandemTrack skill, MA-4 emphasized that exercise is "*a personal, and private activity*," thus she did not want to exercise with the skill when other people were around: "*I'm a shy person, and I don't want to speak out loud to let others know that I'm doing this exercise*."

On the other hand, how other house members interacted with the Echo Dot also affected participants' tendency to use the skill. For example, MA-3 herself was not concerned about exercising using the skill, but the way her son interacted with Alexa annoyed her husband; as a result, she used the TandemTrack skill less often: "*ever since my four-year old son found he could ask Alexa 'Knock Knock Jokes,' he kept shouting at it every day. My husband got really annoyed so he unplugged the Echo a couple of times*."

Voice Recognition Error: Six MA participants (3 international, 3 native English speakers) faced voice recognition error, which discouraged some of them from using the skill. Sometimes Alexa could not recognize the command "*Start TandemTrack*" accurately; in those cases, Alexa either responded with "*I'm not sure*" or started playing a random music. The latter was more frustrating than the former because participants had to first stop the music before trying to start the TandemTrack skill again. Voice recognition error also occurred when entering data. In such cases, participants had an option to use the mobile app to edit the data entries. In fact, we found that, in comparison to M participants (M = 0.09, SD = .30), MA participants (M = 1, SD = 1.34) edited their exercise data more frequently using the app: $t(20) = -2.19$, $p = .04$.

Because of these errors, some MA participants had an impression that the voice interaction was fragile—once an error occurs, there is no way to retract. For example, it took over 10 times for MA-11 to successfully invoke the TandemTrack skill during his first tryout. When the skill failed to recognize his repetitions accurately, he corrected the number in the TandemTrack app later, and decided to use the mobile app without trying the skill again: "*I think it just doesn't pick up my accent.*"

Exercise Feedback

We found that participants who spent more time reviewing exercise feedback on the mobile app completed more exercise sessions (b = .41, p = .04), achieved a higher longest streak length (b = .70, p = .02) and a higher average streak length (b = .07, p = .03). While looking at how participants used TandemTrack's feedback features, we found that in comparison to M participants (M = 12.70, SD = 6.41), MA participants (M = 6.54, SD = 4.09) spent less time on the app reviewing exercise feedback in seconds, $t(20) = 2.73$, $p = .01$. Although MA participants had an option to check the exercise feedback through the skill's voice interaction in a Q&A manner, not many participants used it: over the course of 28 days, only three MA participants asked five questions in total.

Figure 3 shows *when* participants checked the feedback in relation to the exercise time. A majority of M participants and the four MA participants who heavily used the app for exercise (MA-1, MA-3, MA-6, MA-11) frequently reviewed their exercise feedback on the TandemTrack app around their exercise time, as M-1 noted, "*I just naturally looked at the chart after exercise*." Because they were already interacting with the TandemTrack app—either to initiate the exercise or to finish logging the exercise, checking the feedback was easy to tag on during this opportune moment. However, we noticed an exception—M-4, who did not necessarily review her feedback around her exercise time. She explained, "*Usually I checked*

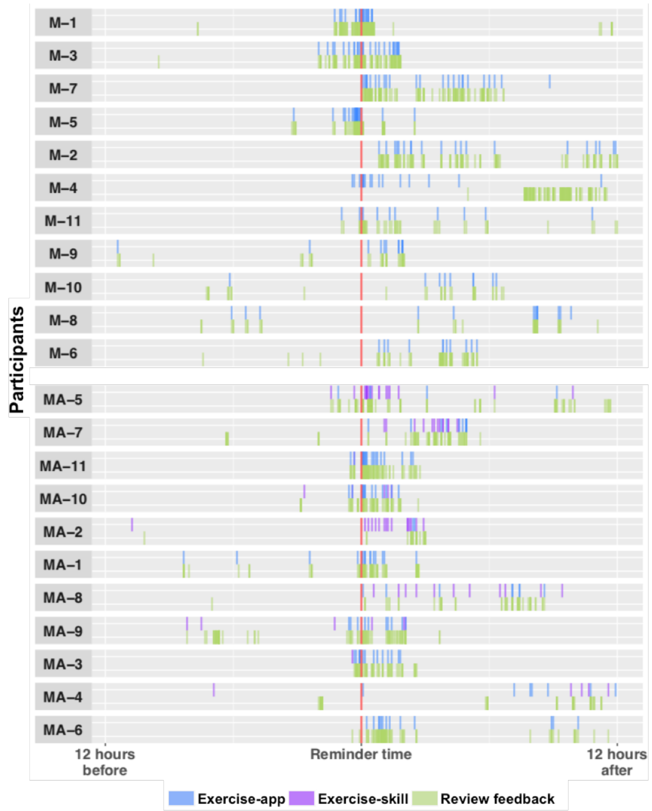

Figure 3. The time distribution of participants' interaction log with the TandemTrack app and their exercise entries (on the TandemTrack app & the skill) over the study period. Each row represents each participant's data, in an descending order of their complete exercise sessions.

the visualization when I was thinking about exercise, or when I knew I had missed exercise. It just popped up randomly."

On the other hand, exercising with the skill did not necessarily prompt MA participants to check the feedback neither from the app nor from the skill (with the exception of MA-5 who liked the feedback on the app, and MA-7 who wanted to check whether the data was stored correctly). This behavior was partly due to the voice-only, ephemeral nature of the interaction that does not provide visual cue, which was exemplified by MA-2: "*it* [checking the feedback] *just didn't pop up into my mind.*"

Participants in both groups acknowledged that the feedback on the TandemTrack app was informative to know how exercise progressed and encouraged them to stay motivated in completing the four-week challenge. For example, M-2 mentioned, "*I felt accountable when I saw this type of visualization. When I saw that I missed a day, I kind of want to keep doing it more so I don't miss anything in the future.*" But sometimes, the number of repetitions did not necessarily reflect participants' actual performance. For example, M-7 mentioned: "*to challenge myself, I intentionally did the sit-ups with a harder posture with my feet in the air. That's why the number doesn't look great.*" In addition, the performance-focused feedback (see Figure 2 C2 and C3) sometimes made participants stressful, as M-4 noted that "*I don't like seeing the chart showing how you're doing, because I know that I was not that good.*"

MA participants rarely asked the TandemTrack skill questions about their exercise data. During the interview, MA participants shared various reasons for not using the voice interaction for feedback. Seven MA participants explained that they did not feel the need to ask Alexa questions given that the feedback on the app already had sufficient information. Two participants forgot what voice commands to use for asking questions (MA-3, MA-4), and one found the responses from Alexa not insightful (MA-10).

Reminder

With a multiple linear regression model, we found that participants who exercised closer to their reminder time consistently completed more exercise sessions (b = -.01, p = .03), and achieved a higher average streak length (b = -.01, p = .04), but not necessarily a higher longest streak length. Participants in both groups found that the reminders on both devices were helpful, especially when the reminder time was a part of their daily routines. For example, M-1 used the exercise reminder as his wake-up alarm, which turned out to be effective: "*It (the reminder) is critical at the beginning, especially for the first week, but once I get into the habit—getting up at 8:00 AM every day, I don't really need that.*" It is noteworthy that M-1 completed all of the exercise sessions (28 days). However, the initial setting of reminder time did not always work for everyone: participants often found themselves not ready to exercise when receiving the reminder. For those who received many notifications on their phone, the exercise reminder from the app was easy to be ignored.

Although there was no difference between the two groups regarding the consistency of their exercise time, we found that for MA participants, the exercise sessions completed on the TandemTrack skill (M = 252, SD = 229.76) were closer to their reminder time than the sessions completed on the mobile app (M = 337, SD = 325.29), $t(210)$ = 3.08, p < .01. During the interview, MA participants acknowledged that the voice reminder was more noticeable and accountable than the mobile notification. For example, MA-5 noted, "*it's easy to swipe the notification (on the mobile phone) unintentionally. The voice (reminder) is more accountable in making you be aware of, it's time to exercise.*" Even MA-1 and MA-6, who resisted using the Echo Dot, acknowledged the effectiveness of the voice reminder: "*I do feel weird about it, but it did get me to exercise because I'm not always on my phone*" (MA-1). In addition, to get ready for exercise, MA-5 and MA-10 hoped that the voice reminder could specify which type of exercise (sit-ups or push-ups) should be performed, so that they do not need to look it up on the app.

DISCUSSION

In this section, we discuss the lessons learned from the study and implications for effectively combining a smart speaker and a mobile app. We also discuss the limitations of the study and opportunities for future work.

Reflecting on the Exercise Adherence and Performance

We found no difference between M and MA group regarding their exercise adherence and performance; we suspect that this result is due to two reasons. First, there were large individual differences in MA participants' choice between the skill and the app, which were affected by many factors we described earlier (e.g., personal preferences, proximity, social context). Four MA participants (MA-1, MA-3, MA-6, MA-11) heavily used TandemTrack's mobile app, and thus their usage was similar to that of M group. On the other hand, five MA participants (MA-2, MA-5, MA-7, MA-8, MA-10) used both the skill and the app, with a clear preference toward the skill. Given that the goal of this work is to examine ways in which we can complement the two modalities in designing technologies for exercise support, it is much more important to look at *how* participants interacted with each of the modalities and *why*, rather than simply looking at the adherence and performance. We believe this work contributes to such an understanding.

Second, participants' behaviors might have been affected by other factors that we have not captured in this study. As shown in Table 1, three M participants (M-1, M-3, M-7) achieved the adherence rate of 93% or higher whereas none of MA participants achieved such high adherence rates, which leads us to question if we neglected an important factor to consider in assigning participants to groups (e.g., habit strength [61], self-regulation [22]), or whether the result is due to the small sample size. Reflecting on our group assignment, although M participants' app use data served as an important baseline, the findings around MA participants' app and skill use revealed more interesting insights. Therefore, in future studies that compare a novel system with a conventional one, assigning more participants to use the novel system can be an effective strategy to help researchers know more about the novel system.

Given the results, we learned that having people commit and stick to a simple exercise regimen—which takes 90 seconds per day, with the total interaction being less than 5 minutes—is very hard. Among the reasons why people missed the exercise (page 6), not many can be addressed by technologies, especially when people are reluctant to do the exercise. Simply providing a multimodal experience therefore would not motivate people to create a daily exercise routine. However, we believe that many interesting opportunities exist in *enriching* people's exercise experience by integrating multimodal interaction, if people are motivated to engage.

How Can Smart Speakers Complement the Mobile?

Optimizing Performance By Voice-Enabled Exercise Regimen

For fitness experts who are already familiar with the forms and routines, the visual interface can be distracting. We noticed that some of our participants paid a close attention to their forms and postures, and were keen to achieve high performance (M-2, M-7, MA-2, MA-7). While exercising using the TandemTrack skill, MA-2 and MA-7 reported that the voice-based exercise regimen helped reduce distraction, allowing them to focus on their performance. Therefore, a multimodal exercise regimen can be tailored by individuals' exercise knowledge and skill level [41]. For those who already know correct postures, voice interaction can serve as the primary interface, to which the visual interaction can be added when it is necessary.

Facilitating Data Capture With Hands-Free Interaction

We found that hands-free voice interaction complemented the mobile app's capability in data capture, increasing the convenience for participants to record their exercise data while performing exercise, especially when they were intensively using their hands (i.e., doing push-ups). Although our study only focused on sit-ups and push-ups for the purpose of reducing the exercise complexity, voice interaction shows promising opportunities to support other exercises that require using hands (e.g., plank, bodypump, dumbbell).

On the other hand, voice interaction did not always work smoothly due to voice recognition errors; participants reported situations when Alexa could not accurately recognize their voice input. In our study, the option of editing one's exercise data using the mobile app helped participants correct their repetitions incorrectly recognized by Alexa, which suggested the benefits of using the mobile app to complement voice interaction's limitations.

Enriching Feedback Combining Voice and Visual

Our results showed that the more time participants spent on reviewing their exercise feedback using the TandemTrack app, the better adherence they were likely to achieve. This finding corresponds to previous research on using technologies to encourage physical activities, which suggested that feedback is important in promoting behavior change through increasing self-awareness and self-reflection [27, 31, 57].

When it comes to interacting with their exercise feedback, MA participants showed a strong preference for the mobile app over the skill. We suspect that the low usage of voice feedback (from smart speakers) was due to the difficulties in discovering and remembering the voice commands as well as the lack of (1) coordination between the app and the skill in delivering feedback; (2) additional interesting information given the rich information provided by the mobile app; and (3) support for data exploration. To improve and enrich individuals' experience with multimodal feedback, it is important to leverage the advantages of the two modalities in a synergistic manner for different scenarios. For data overview, we can use the mobile app as the primary interface, while enabling people to ask specific questions through speech leveraging the voice input from the phone [24]. In addition, instead of waiting for people to initiate a query, the skill can learn about individuals' interaction pattern with the app, and proactively prompt them with the type of data they are interested in, especially before or after exercise—"the critical reflection moments."

Delivering Effective Multimodal Reminders

Reminders are powerful nudges, as we found that participants who exercised consistently closer to their reminder time completed more exercise sessions and achieved better average streak length. Although the two groups did not differ regarding the consistency of their exercise time, we found that for MA participants, the exercise sessions completed on the TandemTrack skill were closer to their reminder time than the sessions completed on the mobile app. During the interviews, MA participants also acknowledged that in comparison to the mobile notification, the voice reminder is more noticeable and reliable. The biggest challenge in designing effective reminders is to reach people at the right time in the right place.

Often times, participants ignored the mobile notification either because they accidentally swiped it or they had many other notifications on their phone; for MA participants, they missed the voice reminder when they were not close to the Echo Dot. In the context of supporting consistent exercise, we suggest designing "proximity-based reminders"—determining the best modality to send the reminder based on people's proximity to the smart speaker and their mobile phone, within a certain range of the reminder time.

Technical Limitations of the Existing Platform

Several inherent limitations within the development platform (Alexa Skills Kit, or ASK) prevented us from designing an ideally "integrated" multimodal system. First, the current version of Alexa skills cannot send their own voice reminders; the only way to deliver notifications from third-party skills is through a visual cue (i.e., a spinning yellow light) on the Echo device, which made it hard to discern the source of the notifications. Therefore, we used Alexa's native reminder, which speaks out "time to exercise." In this way, participants were aware that the reminder was for TandemTrack, but we could not capture whether participants actually responded to the reminder because Alexa's reminder is one-way communication and is not tied with the TandemTrack skill. As a result, participants would still receive the reminder even if they had already completed exercise before the reminder time.

Second, Alexa's voice interface is vulnerable to voice recognition errors. For example, the time-out window of Alexa's voice interface is eight seconds. If the system fails to receive voice input within this time-out window, the conversation session will end. In such a case, people need to initiate the session from the beginning to finish a task. In our study, exercising with the TandemTrack skill requires more than one round of voice interaction: participants needed to first invoke the skill, start training, and then report their repetitions three times. Within the interaction flow, any failure of voice recognition could impede a successful completion of the task.

Third, due to Amazon's limited support for third-party skills, we could not provide a truly integrated multimodal system, where people can use both the phone and smart speaker to control TandemTrack at the same time—for example, controlling the skill through the mobile app. Although some music apps (e.g., Spotify) allow people to control music streaming on the Echo device through both voice commands (e.g., "Pause") and the mobile app (e.g., "pause" button), this real-time synchronization is not well supported for other third-party skills.

For these limitations and yet having to rely on the ASK for the development of TandemTrack, we could not leverage the synergy between the two modalities to their best. However, these lessons are valuable in helping us understand what makes a technology an integrated multimodal system: one modality complements the other (and vice versa) when each of them lacking in a specific ability (e.g., editing exercise data on the mobile app). If the ASK allows, it would be better to show the captured data in real-time through the app when a person captures data through voice, as a way to give assurance.

Study Limitations and Future Work

To capture sit-up and push-up repetitions, we relied on self-reported data, which could have affected our data accuracy (e.g., participants could have entered inaccurate repetitions). Our participants were predominantly students (86.36%), which may limit the generalizability of the results. We also acknowledge the small sample size of 22 participants.

As the first step to explore the opportunities of supporting self-tracking and promoting consistent exercise leveraging multimodal interaction, we identified rich insights regarding how voice interaction on smart speakers influenced participants' exercise experience in different environmental and social contexts. Going forward, we aim to extend TandemTrack to provide an integrated, multimodal self-tracking experience. Our next step is to enhance the synergy between the mobile app and the Alexa skill in terms of supporting exercise regimen, data capture, feedback and reminders. Furthermore, with the advances of multimodal technologies, exploring other interaction modalities beyond mobile apps and smart speakers (e.g., wearable devices, virtual reality) holds great promise for supporting in-home exercise training and tracking.

CONCLUSION

We presented the design and evaluation of TandemTrack, a multimodal system comprised of a mobile app and an Alexa skill to support daily exercise. To examine how voice interaction complements and augments the mobile app, we conducted a four-week between-subjects study with 22 participants, followed by debriefing interviews. We identified the factors affecting participants' exercise adherence and performance change, and their preferences and use of TandemTrack's exercise regimen, data capture, feedback, and reminders with the two modalities. We uncovered the benefits and challenges of using voice interaction on smart speaker as a daily exercise assistant. Our results generated interesting insights regarding how participants' personal preferences, proximity to the device, living environment, social context, and voice recognition errors influence their use of the TandemTrack app and the skill. With the lessons learned, we discuss design implications for effectively combining the mobile app and voice interaction in promoting consistent exercise.

ACKNOWLEDGMENT

We thank Daniel Smolyak, who helped implement the TandemTrack app. We also thank Amanda Lazar, Niklas Elmqvist, Pramod Chundury, and Young-Ho Kim for their valuable feedback on an earlier version of the paper. The work was in part supported by National Science Foundation #1753452.

REFERENCES

[1] 30 Day Pushup Challenge. https://www.amazon.com/dp/B078Z5QMH7. Accessed: 2020-01-08.

[2] 5-Minute Plank Workout. https://www.amazon.com/cubic-ai-5-Minute-Plank-Workout/dp/B06XHTCB3Z. Accessed: 2020-01-08.

[3] 7 Min Fitness Challenge. http://7minworkoutapp.net/. Accessed: 2020-01-08.

[4] Alexa Skill Kit. https://developer.amazon.com/en-US/alexa/alexa-skills-kit. Accessed: 2020-01-08.

[5] Amazon Echo. https://www.amazon.com/echo. Accessed: 2020-01-08.

[6] Amazon Web Service (AWS) Lambda. https://aws.amazon.com/lambda/. Accessed: 2020-01-08.

[7] Centers for Disease Control and Prevention: Physical Activity Basics. https://www.cdc.gov/physicalactivity/basics/index.htm. Accessed: 2020-01-08.

[8] Fitbit Skill. https://www.amazon.com/dp/B01CH4BP28. Accessed: 2020-01-08.

[9] Gartner Predicts 75% of US Households will Have Smart Speakers by 2020. https://tinyurl.com/yfzt33jm. Accessed: 2020-01-08.

[10] Google Firebase. https://firebase.google.com/. Accessed: 2020-01-08.

[11] Google Home. https://store.google.com/gb/product/google_home. Accessed: 2020-01-08.

[12] HealthBot. https://healthbot.in/. Accessed: 2020-01-08.

[13] JEFIT. https://www.jefit.com/. Accessed: 2020-01-08.

[14] Keep Workout and Fitness Trainer. https://www.keepkeep.com/. Accessed: 2020-01-08.

[15] Kotlin. https://kotlinlang.org/. Accessed: 2020-01-08.

[16] Lark. https://www.lark.com/. Accessed: 2020-01-08.

[17] My fitnesspal Amazon Alexa Skill. https://www.amazon.com/gp/product/B07QN179C5. Accessed: 2020-01-08.

[18] Nike Training Club App. https://www.nike.com/us/en_us/c/nike-plus/training-app. Accessed: 2020-01-08.

[19] Node.js. https://nodejs.org/en/. Accessed: 2020-01-08.

[20] RunKeepper. https://runkeeper.com/running-app. Accessed: 2020-01-08.

[21] SWORKIT. https://sworkit.com/. Accessed: 2020-01-08.

[22] Janice M Brown, William R Miller, and Lauren A Lawendowski. 1999. The self-regulation questionnaire. (1999).

[23] Jennifer K Carroll, Anne Moorhead, Raymond Bond, William G LeBlanc, Robert J Petrella, and Kevin Fiscella. 2017. Who uses mobile phone health apps and does use matter? A secondary data analytics approach. *Journal of medical Internet research* 19, 4 (2017), e125. http://doi.org/10.2196/jmir.5604

[24] Eun Kyoung Choe, Bongshin Lee, and Seung-Won Hwang. 2018. Personal Data Exploration with Speech on Mobile Devices. In *AVI 2018 Workshop on Multimodal Interactoin for Data Visualization*.

[25] Arlene E Chung, Ashley C Griffin, Dasha Selezneva, and David Gotz. 2018. Health and Fitness Apps for Hands-Free Voice-Activated Assistants: Content Analysis. *JMIR mHealth and uHealth* 6, 9 (2018), e174. http://doi.org/10.2196/mhealth.9705

[26] David E Conroy, Chih-Hsiang Yang, and Jaclyn P Maher. 2014. Behavior change techniques in top-ranked mobile apps for physical activity. *American journal of preventive medicine* 46, 6 (2014), 649–652. http://doi.org/10.1016/j.amepre.2014.01.010

[27] Sunny Consolvo, Katherine Everitt, Ian Smith, and James A Landay. 2006. Design requirements for technologies that encourage physical activity. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM, 457–466. http://doi.org/10.1145/1124772.1124840

[28] Stefania Druga, Randi Williams, Cynthia Breazeal, and Mitchel Resnick. 2017. Hey Google is it OK if I eat you?: Initial Explorations in Child-Agent Interaction. In *Proceedings of the 2017 Conference on Interaction Design and Children*. ACM, 595–600. http://doi.org/10.1145/3078072.3084330

[29] Brian J Fogg. 2009. A behavior model for persuasive design. In *Proceedings of the 4th international Conference on Persuasive Technology*. ACM, 40. http://doi.org/10.1145/1541948.1541999

[30] Macro Gutierrez, Ramon Moreno, and Marina Rebelo. 2017. Information and communication technologies and global health challenges. In *Global Health Informatics*. Elsevier, 50–93. http://doi.org/10.1016/B978-0-12-804591-6.00004-5

[31] Florence-Emilie Kinnafick, Cecilie Thøgersen-Ntoumani, and Joan L Duda. 2014. Physical activity adoption to adherence, lapse, and dropout: a self-determination theory perspective. *Qualitative health research* 24, 5 (2014), 706–718. http://doi.org/10.1177/1049732314528811

[32] Rafal Kocielnik, Lillian Xiao, Daniel Avrahami, and Gary Hsieh. 2018. Reflection Companion: A Conversational System for Engaging Users in Reflection on Physical Activity. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 2 (2018), 70. http://doi.org/10.1145/3214273

[33] Eveliina E Korkiakangas, Maija A Alahuhta, and Jaana H Laitinen. 2009. Barriers to regular exercise among adults at high risk or diagnosed with type 2 diabetes: a systematic review. *Health promotion international* 24, 4 (2009), 416–427. http://doi.org/10.1093/heapro/dap031

[34] Paul Krebs and Dustin T Duncan. 2015. Health app use among US mobile phone owners: a national survey. *JMIR mHealth and uHealth* 3, 4 (2015), e101. http://doi.org/10.2196/mhealth.4924

[35] Phillippa Lally and Benjamin Gardner. 2013. Promoting habit formation. *Health Psychology Review* 7, sup1 (2013), S137–S158. http://doi.org/10.1080/17437199.2011.603640

[36] Minha Lee, Sander Ackermans, Nena van As, Hanwen Chang, Enzo Lucas, and Wijnand IJsselsteijn. 2019. Caring for Vincent: A Chatbot for Self-Compassion. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. http://doi.org/10.1145/3290605.3300932

[37] Leib Litman, Zohn Rosen, David Spierer, Sarah Weinberger-Litman, Akiva Goldschein, and Jonathan Robinson. 2015. Mobile exercise apps and increased leisure time exercise activity: a moderated mediation analysis of the role of self-efficacy and barriers. *Journal of medical Internet research* 17, 8 (2015), e195. http://doi.org/10.2196/jmir.4142

[38] Irene Lopatovska and Harriet Williams. 2018. Personification of the Amazon Alexa: BFF or a mindless companion. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*. ACM, 265–268. http://doi.org/10.1145/3176349.3176868

[39] Paula M Ludewig, Molly S Hoff, Erin E Osowski, Shane A Meschke, and Peter J Rundquist. 2004. Relative balance of serratus anterior and upper trapezius muscle activity during push-up exercises. *The American journal of sports medicine* 32, 2 (2004), 484–493. https://doi.org/10.1177/0363546503258911

[40] Elizabeth Jane Lyons and Claire Hatkevich. 2013. Prevalence of behavior changing strategies in fitness video games: theory-based content analysis. *Journal of medical Internet research* 15, 5 (2013), e81. http://doi.org/10.2196/jmir.2403

[41] Harm op den Akker, Valerie M Jones, and Hermie J Hermens. 2014. Tailoring real-time physical activity coaching systems: a literature survey and model. *User modeling and user-adapted interaction* 24, 5 (2014), 351–392. https://doi.org/10.1007/s11257-014-9146-y

[42] Rosemary Pimm, Corneel Vandelanotte, Ryan E Rhodes, Camille Short, Mitch J Duncan, and Amanda L Rebar. 2016. Cue consistency associated with physical activity automaticity and behavior. *Behavioral Medicine* 42, 4 (2016), 248–253. http://doi.org/10.1080/08964289.2015.1017549

[43] Markus Pohl. 325,000 mobile health apps available in 2017. https://research2guidance.com/325000-mobile-health-apps-available-in-2017/. Accessed: 2020-01-08.

[44] Martin Porcheron, Joel E Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice interfaces in everyday life. In *proceedings of the 2018 CHI conference on human factors in computing systems*. ACM, 640. http://doi.org/10.1145/3173574.3174214

[45] Alisha Pradhan, Kanika Mehta, and Leah Findlater. 2018. Accessibility came by accident: use of voice-controlled intelligent personal assistants by people with disabilities. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 459. http://doi.org/10.1145/3173574.3174033

[46] Andrew Prestwich, Macro Perugini, and Robert Hurling. 2009. Can the effects of implementation intentions on exercise be enhanced using text messages? *Psychology and Health* 24, 6 (2009), 677–687. http://doi.org/10.1080/08870440802040715

[47] Amanda Purington, Jessie G Taft, Shruti Sannon, Natalya N Bazarova, and Samuel Hardman Taylor. 2017. Alexa is my new BFF: social roles, user satisfaction, and personification of the amazon echo. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 2853–2859. http://doi.org/10.1145/3027063.3053246

[48] IQVIA Institute Report. The Growing Value of Digital Health. https://www.iqvia.com/institute/reports/the-growing-value-of-digital-health. Accessed: 2020-01-08.

[49] B Ricci, M Marchetti, and F Figura. 1981. Biomechanics of sit-up exercises. *Medicine and science in sports and exercise* 13, 1 (1981), 54–59. http://www.ncbi.nlm.nih.gov/pubmed/6452564

[50] Ron Shor and Anat Shalev. 2014. Barriers to involvement in physical activities of persons with mental illness. *Health promotion international* 31, 1 (2014), 116–123. http://doi.org/10.1093/heapro/dau078

[51] Cindy HP Sit, John H Kerr, and Irene TF Wong. 2008. Motives for and barriers to physical activity participation in middle-aged Chinese women. *Psychology of Sport and exercise* 9, 3 (2008), 266–283. http://doi.org/10.1016/j.psychsport.2007.04.006

[52] Daniel Smolyak, Bongshin Lee, and Eun Kyoung Choe. 2018. TandemTrack: Promoting Consistent Exercise Leveraging Multimodal Training and Tracking. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, LBW543. http://doi.org/10.1145/3170427.3188536

[53] Emmanuel Stamatakis, I-Min Lee, Jason Bennie, Jonathan Freeston, Mark Hamer, Gary O'Donovan, Ding Ding, Adrian Bauman, and Yorgi Mavros. 2017. Does strength-promoting exercise confer unique health benefits? A pooled analysis of data on 11 population cohorts with all-cause, cancer, and cardiovascular mortality endpoints. *American journal of epidemiology* 187, 5 (2017), 1102–1112. http://doi.org/10.1093/aje/kwx345

[54] Katarzyna Stawarz, Anna L Cox, and Ann Blandford. 2014. Don't forget your pill!: designing effective medication reminder apps that support users' daily routines. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. ACM, 2269–2278. http://doi.org/10.1145/2556288.2557079

[55] Thomas Stütz, Michael Domhardt, Gerlinde Emsenhuber, Daniela Huber, Martin Tiefengrabner, Nicholas Matis, and Simon Ginzinger. 2017a. An interactive 3D health app with multimodal information representation for frozen shoulder. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, 3. https://doi.org10.2196/10.1145/3098279.3098562

[56] Thomas Stütz, Gerlinde Emsenhuber, Daniela Huber, Michael Domhardt, Martin Tiefengrabner, Gertie Janneke Oostingh, Ulrike Fötschl, Nicholas Matis, and Simon Ginzinger. 2017b. Mobile Phone–Supported Physiotherapy for Frozen Shoulder: Feasibility Assessment Based on a Usability Study. *JMIR rehabilitation and assistive technologies* 4, 2 (2017), e6. https://doi.org10.2196/rehab.7085

[57] Lie Ming Tang and Judy Kay. 2017. Harnessing Long Term Physical Activity DataâĂŤHow Long-term Trackers Use Data and How an Adherence-based Interface Supports New Insights. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 2 (2017), 26. https://doi.org/10.1145/3090091

[58] Georgina Torbet. Smart speakers used primarily for music and information - and shopping. https://www.digitaltrends.com/home/smart-speakers-survey/. Accessed: 2020-01-08.

[59] Markku Turunen, Jaakko Hakulinen, Olov Ståhl, Björn Gambäck, Preben Hansen, Mari C Rodríguez Gancedo, Raúl Santos de La Cámara, Cameron Smith, Daniel Charlton, and Marc Cavazza. 2011. Multimodal and mobile conversational health and fitness companions. *Computer Speech & Language* 25, 2 (2011), 192–209. https://doi.org/10.1016/j.csl.2010.04.004

[60] Upkar Varshney. 2014. Mobile health: Four emerging themes of research. *Decision Support Systems* 66 (2014), 20–35. https://doi.org/10.1016/j.dss.2014.06.001

[61] Bas Verplanken and Sheina Orbell. 2003. Reflections on past behavior: a self-report index of habit strength. *Journal of Applied Social Psychology* 33, 6 (2003), 1313–1330. http://doi.org/10.1111/j.1559-1816.2003.tb01951.x

[62] Joshua H West, P Cougar Hall, Carl L Hanson, Michael D Barnes, Christophe Giraud-Carrier, and James Barrett. 2012. There's an app for that: content analysis of paid health and fitness apps. *Journal of medical Internet research* 14, 3 (2012), e72. http://doi.org/10.2196/jmir.1977

[63] Justin Yang, Costas A Christophi, Andrea Farioli, Dorothee M Baur, Steven Moffatt, Terrell W Zollinger, and Stefanos N Kales. 2019. Association Between Push-up Exercise Capacity and Future Cardiovascular Events Among Active Adult Men. *JAMA network open* 2, 2 (2019), e188341–e188341. http://doi.org/10.1001/jamanetworkopen.2018.8341

[64] Ugan Yasavur, Christine Lisetti, and Naphtali Rishe. 2014. Let's talk! speaking virtual counselor offers you a brief intervention. *Journal on Multimodal User Interfaces* 8, 4 (2014), 381–398. https://doi.org/10.1007/s12193-014-0169-9