Understanding Task Differences to Leverage the Usability and Adoption of Voice Assistants (VAs)

Isabela Motta[(✉)] and Manuela Quaresma

LEUI | Laboratory of Ergodesign and Usability of Interfaces, PUC-Rio, Rio de Janeiro, Brazil
`isabelamotta@aluno.puc-rio.br, mquaresma@puc-rio.br`

Abstract. Voice Assistants (VAs) are becoming increasingly popular, but evidence shows that users' utilization of features is limited to few tasks. Although the literature has shown that usability impact VA adoption, little is known about how usability varies across VA tasks and its relation to task adoption by users. To address this gap, we conducted usability tests followed by debriefing sessions with Siri and Google Assistant users, assessing usability measures of six features and uncovering reasons for task usage. The results showed that usability varied across tasks regarding task completeness, error number, error types, and user satisfaction. Checking the weather and making phone calls had the best usability measures, followed by playing songs and sending messages, whereas adding appointments to a calendar and searching for information were the most incomplete and frustrating interactions. Furthermore, usability-related factors such as perceived ease of use and the interaction's hands/eyes-free nature influenced task adoption. Nevertheless, we also identified other task-independent factors that affect VA usage, such as use context (i.e., place, task content), VAs' personality, and preferences for settings. Our main contributions are recommendations for VA design, highlighting that attending to tasks separately is paramount to understanding specific usability issues, task requirements, users' perceptions of features, and developing design solutions that leverage VAs' usability and adoption.

Keywords: Voice Assistants · Usability · Technology adoption

1 Introduction

Voice Assistants (VAs), such as Apple Siri and Google Assistant, are artificial intelligence-powered virtual agents that can perform a range of tasks in a system, which users interact through a voice interface that may be supported by a visual display [40]. They run on several devices, such as earphones, smart speakers, and smartphones, and were estimated to be in use in over four billion devices by 2020 [26]. The projections for VAs indicate that interfaces for human-computer interaction (HCI) are in the midst of a paradigm shift from visual interfaces to hands-free, voice-based interactions [40].

Despite the growth in VA adoption, evidence shows that users do not utilize all tasks available in these systems. By 2020, Amazon's Alexa was able to perform over 70.000 skills in the USA [37]. Nevertheless, studies have shown that users' utilization of these

devices is limited to tasks such as checking the weather, adding reminders, listening to music, and controlling home automation [1, 5, 10, 11, 18, 30, 33, 34, 38, 39, 42]. The discrepancy in task availability and feature adoption may suggest that user experience across tasks is heterogeneous, leading to the underutilization of VAs.

The literature indicates varied causes for VA usage and abandonment that may account for such incongruity. Among these motivators, we hypothesize that usability-related factors may be essential for task adoption. While users' attitudes and data privacy concerns are influential [4, 6, 17, 24, 27, 28], these aspects are usually determinant for VA usage as a whole interface rather than for specific features. Conversely, users perceive tasks to have varying levels of difficulty [22] and satisfaction with task types is affected by different factors [14], suggesting inconsistent usability across features.

Although usability issues have been extensively covered in the field of voice interaction, most publications tend to study the VA as a single entity instead of attending to system features separately. Whereas such an approach is useful for understanding users' general impressions and highlighting VAs' major strengths and flaws, it is necessary to regard task specificities to comprehend why users abandon some features. To the extent of our knowledge, no study has examined differences in the usability of VA tasks and their relationship to task adoption.

Thus, this study aimed to assess usability variations in six VA tasks and its relations to VA adoption. Two research questions were developed:

- **RQ1:** How does usability vary across VA tasks?
- **RQ2:** How is usability related to task adoption in VAs?

To answer these questions, we conducted usability tests followed by debriefing sessions with users of Siri and Google Assistant (GA). Participants performed six tasks in both Siri and GA on smartphones: check the weather, make a phone call, search for information, play a song, send a message, and add an appointment to a calendar. In the debriefing sessions, users talked about their perceptions of the tasks and stated reasons for adopting – or not – such VA features in their routines. Our findings showed that task completeness, number and types of errors, and satisfaction varied across tasks, and usability-related factors were mentioned as motivators for feature adoption and abandonment. Moreover, other factors such as customization, VAs' personality, and use context impacted usage. Our main contributions are recommendations for VA design.

2   Related Work

2.1   Task Adoption by VA Users

Although VA adoption has been steadily increasing over the years, several studies have shown that users' utilization of features is limited to a small set of tasks. Commonly used features reported in the literature are playing music, checking the weather, and setting timers, alarms, and reminders [1, 5, 10, 11, 18, 30, 33, 34, 38, 39, 42]. Users also utilize VAs for looking up information [1, 10, 11, 33, 34] such as recommendations on places to eat or visit [20, 42], recipes [21], information about sports and culture [21, 42], and for learning-related activities [12, 21, 33]. Another frequently mentioned task in the

literature is controlling Internet of Things (IoT) devices such as lights and thermostats [1, 5, 10, 11, 33, 34, 39, 42], although these features are more commonly performed through smart speakers. Moreover, VAs are used for entertainment purposes: telling jokes, playing games, and exploring the VAs' personality [5, 10, 12, 34, 38, 39]. Other tasks such as creating lists, sending messages, checking the news, and managing calendar appointments are relatively underused, as shown by industry reports on commercially available VAs [25, 29, 41].

2.2   Usability and VA Adoption

As mentioned above, several factors impact VA adoption. Particularly, perceived ease of use has been demonstrated to cause a significant effect on VA usage [27] and may be related to task completeness and effort. On the one hand, due to the use of speech, voice interaction is considered easy and intuitive [31], and the possibility of a hands/eyes-free interaction is valued by users [20] and motivates them to adopt a VA [22, 28, 30]. On the other hand, errors throughout interactions may lead to frustration and underutilization. Purington et al. [35] analyzed Amazon Echos' online reviews and identified that technical issues with the VAs' functioning were associated with decreased user satisfaction levels. As argued by Lopatovska et al. [19], unsatisfactory interactions cause users to lower their VA usage over time.

A central theme around errors in user-VA interaction is VAs' conversational capabilities. While some users expected to have human-like conversations with their VAs, actual system capabilities lead to disappointment and eventual abandonment of the device [5, 22, 30]. Speech recognition problems also impose entry barriers to new users. Motta and Quaresma [28] observed through an online questionnaire that one of the reasons for smartphone users not to adopt a VA was poor query recognition. Likewise, Cowan et al. [6] conducted focus groups with infrequent users and showed that speech recognition problems are a core barrier to VA usage.

Specifically, studies indicated that users consider VAs inefficient to recognize accents and are limited in terms of supported languages [6, 11, 15, 20, 23]. Furthermore, VAs fail to bear contextual references, such as users' physical locations and information provided in past interactions [1, 10, 11, 20–22, 36]. Given these conversational limitations, users often need to adapt their speech to match the VA's capacity, which is an obstacle to VA adoption [28]. Speech adaptations include pronouncing words more accurately [8, 30], removing words, using specific terms, speaking more clearly, changing accents [22], and removing contextual references [1, 22, 30]. Additionally, users have reported needing to check if their commands were understood in a visual interface or provide manual confirmations, making interactions slower [6, 39].

2.3   Differences in Usability Across VA Tasks

The literature provides indications that VA tasks may have different usability. Firstly, Luger and Sellen [22] identified that users judge tasks to be simple (e.g., setting reminders, checking the weather) or complex (e.g., launching a call, writing an email). The authors note that failures such as query misrecognition in complex tasks were more frequent, leading users to feel that the VA could not be trusted for certain activities and

limit their usage to simple features. These results echo Oh et al.'s [30] findings, who conducted a 14-day study in which users interacted with a VA, Clova, in a realistic setting (i.e., their homes). The study's participants reported that interaction failures caused them to stop performing complicated and difficult tasks and focus on simple features with reliable results (e.g., weather reports, listen to music).

Further evidence on task differences was provided by Kiseleva et al. [14], who conducted a user study to evaluate variables impacting satisfaction in the use of Cortana for three task types: "control device" (e.g., play a song, set a reminder), "web search," and "structured search dialogue" (tasks that required multiple steps). The authors found that, while user satisfaction was negatively correlated with effort for all task types, the "device control" and "structured search dialogue" tasks were the only ones in which satisfaction was positively correlated with query recognition quality and task completion. That is, for the "web search" activity exclusively, achieving the desired result with good speech recognition did not necessarily guarantee satisfaction. A similar tendency was observed by Lopatovska et al. [19], who collected users' experiences with Alexa through online diaries. The study showed that, even though most participants reported interactions' success to be positively related to satisfaction, there were occasions in which successful interactions were rated as unsatisfactory and vice-versa.

The literature described above provides indications that usability variations across tasks may contribute to discrepancies in task adoption by VA users. As little is known around such a topic, it is necessary to investigate how usability may vary in different VA tasks and how it is related to task adoption in VAs.

3 Method

3.1 Participants and Test's Format

The study's participants were Brazilian smartphone users who used at least one VA on their smartphones – Siri and/or Google Assistant (GA) – at least once a month. The users were recruited by social media and chat apps.

The usability tests had a within-subject design (2 × 6) in which participants had to perform six tasks using Siri and GA on a smartphone. To increase participants' immersion in the test, all tasks revolved around the scenario of a musical concert that was to occur, in which participants were hypothetically interested in attending.

We selected the VAs and tasks based on the literature and a previously conducted survey with Brazilian smartphone users. We chose Siri and GA since these were the most commonly used VAs among our survey's respondents. Tasks with different adoption rates were selected to assess whether usability would vary across tasks and cause discrepancies in task adoption. The tasks selected for this study were: searching for information online, checking the weather, making a phone call, playing a song, adding an appointment to a calendar, and sending a message.

3.2 Apparatus

Two smartphones were used in the usability tests: a Motorola G4 Play running Android 7.1.1. OS for GA and an iPhone XR running iOS 12.4.1. for Siri. We recorded users

in video and audio using a webcam placed above their hands, a notebook's camera positioned towards their faces, and apps capturing the smartphones' screens (Fig. 1).
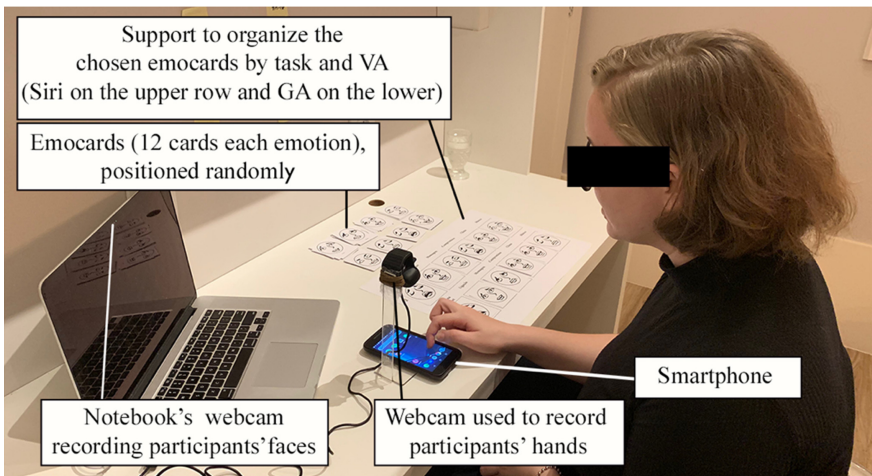

Fig. 1. Usability test set up.

Moreover, to represent user satisfaction with tasks, we employed the emocards tool [7]. According to Desmet et al. [7], having users describe their emotions towards a product may be challenging since they are difficult to verbalize and users' answers may be affected by cognitive involvement. To tackle this problem, the authors developed the emocards [7]: a set of 16 cards that picture cartoon faces with eight different emotional expressions (eight male faces and eight female faces). The expressions represent emotions that combine levels of two emotional dimensions: Pleasantness and Arousal.

We chose the emocards for this study to help users illustrate their emotional responses and understand how – in terms of pleasantness and arousal – different tasks impact users' satisfaction. We highlight that the emocards were not considered an objective metric for measuring exact levels of satisfaction. Instead, our primary goal in employing this tool was to start a conversation between participants and the moderator [7] during the debriefing sessions to clarify the reasons for users' choice of emotional responses. Each emocard was printed 12 times (for two rounds of six tasks), and a support was developed to help participants registering their preferred emocards.

3.3   Procedure

The usability tests were arranged in three parts: 1) introduction, 2) two rounds of task performance, and 3) debriefing session. In the introduction, participants read and signed a term of consent and filled a digital form to gather profile data. The moderator provided an oral explanation concerning the experiment's goal and procedure.

Users performed tasks through Siri and GA. Independently of participants' previous experience with the VAs, each round started with guidance on how to activate the VA,

followed by a training session. Thereafter, all participants performed a trial task (ask for a joke). The initial preparation and training served the purpose of getting participants familiarized with the system and its functioning so that learnability issues would not affect the results. To mitigate order presentation bias, tasks were presented in random order. For the same reason, half of the users started with Siri and half with GA.

Participants were instructed to complete the tasks using the VAs. The moderator orally presented the tasks, one at a time, through a previously scripted instruction. Cards with the information necessary to complete the tasks were provided to participants. Thus, the moderator would give the oral instruction and point to the necessary information, as in the following example: *"You want to attend Sandy and Junior's concert, but you don't want to go alone. So, you decide to call this friend's cellphone* [points to the contact's name on the card] *to ask if he wants to go with you."* These procedures were validated through two pilot studies. We observed that entirely written instructions not only led to confusion but also influenced participants' queries, as they would read the instruction instead of creating their own phrases.

Following each task, participants were instructed to choose one among the eight randomly positioned emocards to illustrate how they felt during the interaction. After the tests' two rounds, debriefing interviews were conducted, in which participants were asked about why they chose each of the emocards. The moderator also asked users if and why they utilized a VA to perform the six tasks in their routines. The tests were conducted from September to November 2019, and sessions lasted around 45 min.

3.4 Data Analysis

For the data analysis, we reviewed the video and audio recordings and analyzed task usability by measuring task completeness, number and types of errors per task, and user satisfaction. We attributed four levels of completeness to measure mean task completeness: *completed* - user completes the task in the first attempt; *completed with effort* - user completes the task but has to try two or more times to do so; *partially completed* - user only completes a part of the task successfully (e.g., schedules an event to the calendar at the correct day but misses the place); *incomplete* - user gives up or achieves a failed result (e.g., plays a song different from what was asked). As for the error types, we described and recorded errors throughout interactions for each task. We considered errors to be any VA output that did not directly answer a request or moved the interaction forwards (e.g., asking the appointment's time). An affinity diagram was made in a bottom-up approach [2] to identify similarities and create categories of errors.

To assess the effects of varied tasks on user satisfaction, the number of emocards chosen by participants for each task was accounted. This analysis was graphically represented to identify patterns in users' preferred emotional responses (regarding pleasantness and arousal) towards the tasks. Moreover, to understand the causes of satisfaction variations, we related the emocards to arguments stated by participants in the debriefing interviews for choosing them. For this cross-checking, we categorized users' claims through an affinity diagram [2].

We employed a similar procedure to analyze users' reasons for adopting tasks in their daily VA usage. Firstly, we transcribed participants' answers and then created affinity diagrams [2] to find categories of reasons to use or not the VA for each task.

4 Results

4.1 Task Usability

In this section, we address RQ1: "*How does usability vary across VA tasks?*"

Task Completeness. Figure 2 shows that the tasks had large variations concerning their completeness. Checking the weather and making a phone call were completed by all participants, and only 8% of the interactions required effort to be successful. Moreover, most participants were able to play a song, but 18% of the trials were ineffective. Sending a message was considered complete for only 30% of the interactions, and most participants had to issue their commands more than once to complete the task. The search task had the greatest number of failed interactions (43%), whereas adding an appointment to the calendar had the highest mean for partially complete outcomes (63%) and the lowest task completeness mean (3%), as only one participant was able to complete the task in his first try.
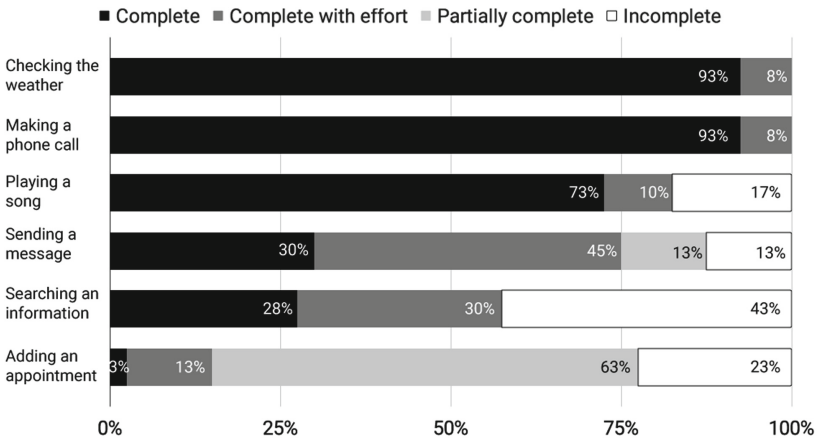

Fig. 2. Task completeness per task (n = 40 interactions per task, i.e., 20 on Siri and 20 on GA).

Number of Errors Per Task. Variations in the number of errors followed similar patterns to task completeness scores (Fig. 3). Few errors happened during weather reports and phone calls, whereas playing a song led to a higher number of errors. Akin to completeness results, in which the calendar task had the lowest completeness rate, "adding an appointment to the calendar" had the largest number of errors. Nevertheless, although "sending a message" had a smaller count of incomplete results when compared to "searching information" (Fig. 2), fewer errors happened during the web search. This difference may mean that participants either gave up the search task sooner or recovered more quickly from errors throughout "sending a message" failures.

Error Types. As completeness and error number, error types varied across tasks. We identified nine error categories: 1) *Query misrecognition*; 2) *Unrequired task;* 3) *System*
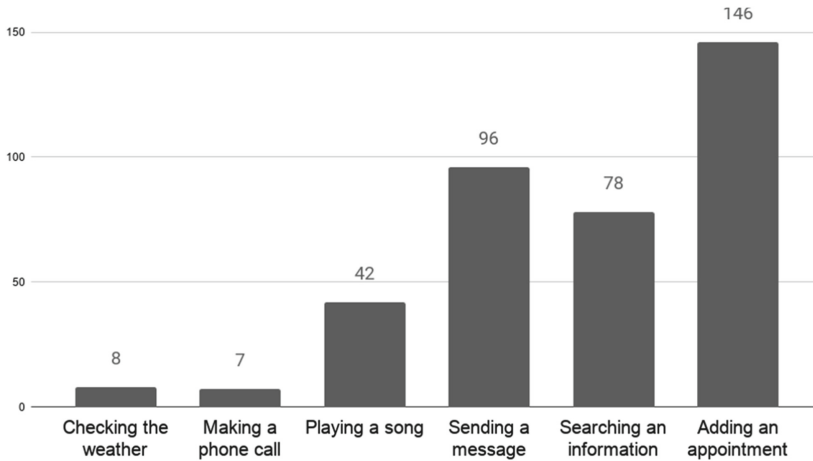
Fig. 3. Number of errors per task (nTotal = 377).

error; 4) *Interruption*; 5) *No capture;* 6) *Editing error;* 7) *Cancellation/confirmation error;* 8) *Request for visual/manual interaction;* 9) *Wrong application.*

Query Misrecognition errors were the most frequent error for all features and happened when VAs failed to recognize users' speech correctly. These issues made users repeat themselves and led to subsequent errors such as Unrequired Task (i.e., performing the wrong task), which was also observed for all task types. In particular, VAs frequently misrecognized appointments' information (i.e., the name, date, time, or place), forcing participants to either start over or edit a scheduled event. As for "playing a song," Query Misrecognition errors were less recurrent than we expected. Since we intentionally chose a song with a title in English, compelling participants to mix two languages (Portuguese and English), we supposed that VAs would fail to understand the song's title. Nevertheless, such an issue only happened six times.

Interruptions occurred when VAs stopped capturing users' inputs midway. No Capture errors were failures from VAs to capture *any* user input (i.e., the assistant did not "hear" participants from the beginning). We observed both issues for all task types, but they were more recurrent for "sending a message" and "adding an appointment."

This tendency might be attributed to the lengthiness of commands issued for these tasks. The message comprised all of the concert's information, resulting in commands with several words. For the same reason, scheduling an event led to long queries when users tried to say all information at once (e.g., "Add to my calendar 'Show Sandy & Junior' on November 9th at 9:30 pm at Parque Olímpico"). No Capture errors were also identified for "step-by-step" interactions (e.g., "What is the appointment's date and time?"), since participants had trouble matching VAs' timing to start input capture.

Contrarily to the issues mentioned above, System Errors were observed mostly during search and sending a message. Systems Errors were bugs or the VAs' inability to fulfill an inquiry. On several occasions, Siri answered search requests by saying, "Sorry, I don't have an answer for that," causing users to give up, especially when the assistant kept repeating the same output after participants adjusted their commands (e.g., changed

wording). Likewise, outputs as "Ops, try again" in the message task were frequent and required participants to restate their messages, an effortful work.

Cancellation/Confirmation Errors were problems for confirming or canceling actions and occurred in the calendar and message tasks. Editing Errors were also exclusive for these tasks and happened when users tried to change a message's content or an event's information. We believe that such problems arose exclusively for these features due to their characteristics. Confirmations and cancellations were unnecessary for most tasks as VAs presented outputs without requiring users' permission. Similarly, results for other tasks were not editable and iterations on previous outcomes were unfeasible.

Furthermore, we considered Requests for Manual/Visual interactions as errors since VAs are supposedly a hands/eyes-free interface, and the literature has shown that users deem such requests as failures [22]. On occasions, participants had to read a search result on the screen, complete the appointments' details manually, or press a button to play the song. Finally, Wrong Application errors happened when users explicitly asked the VA to send a message through a specific app (WhatsApp, as requested by the moderator), but it executed the task using a different application.

User Satisfaction. Figure 4 illustrates the number of each emocard chosen by participants for the test's tasks.
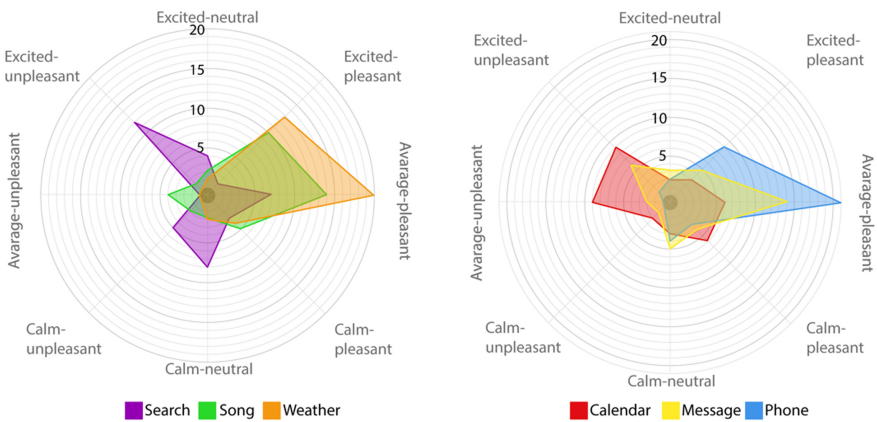

Fig. 4. Distribution of selected emocards for each task.

Except for one participant who chose an unpleasant emotion for the phone task, no other unpleasant emotions were evoked towards making a phone call and checking the weather. A mild preference for neutral emotions was also observed for both tasks. Similarly, playing a song and sending a message elicited mostly pleasant emotions on participants. Nevertheless, users preferred unpleasant emotions on seven occasions for the song task, and we observed two additional points of concentration in participants' emotional preferences for the message task: the "calm-neutral" and "excited-unpleasant" emocards. Participants selected unpleasant emotions towards searching for information and adding an event to the calendar in most cases. We identified a substantial convergence

of emotional preferences for the "excited-unpleasant" emocard for both tasks, but the same does not occur for the "excited-pleasant" emocard. A preference for the "calm-neutral" emotion can also be observed for the search task.

Causes for Variations in Satisfaction. The leading causes for variations in user satisfaction were identified in the arguments stated by participants to reason their choices in the debriefing sessions and organized into the following categories: 1) *task completeness,* 2) *perceived ease of use*, 3) *expectations for VAs capabilities and limitations*, 4) *hands/eyes-free aspects of the interaction*, and 5) *individual preferences for settings.*

Firstly, *task completeness* yielded variations in user satisfaction for all tasks. On the one hand, participants who selected pleasant emocards stated that they felt happy and satisfied with completing the tasks. On the other hand, unpleasant emotions were related to unsatisfactory results. As for the neutral emotions, some participants stated that they faced unexpected outcomes but ultimately managed to complete the task.

"Excited-pleasant" emocards were chosen when VAs' answers were perfect or surprising: "*It entered the site and gave me the temperature without any mistakes*" (P10; Weather). Some participants stated that their preference for the "average-pleasant" emocard was due to interactions that were good enough to complete the task: "*It picked a random* [song's] *live version that I didn't really choose, but it handled what I wanted*" (P9; Song). Similarly, "calm-pleasant" emotions were related to incomplete or deficient results: "*It wasn't exactly perfect. Siri put* [the appointment] *as if it was from 9 pm to 9:30 pm, not starting at 9:30 pm*" (P7; Calendar). Contrarily, answers that were insufficient to fulfill requests or too different from users' expectations caused them to choose the "excited-unpleasant" emocard: "*I was upset. Siri didn't even try giving me a result similar to what I asked. It just said 'I didn't understand.'*" (P9; Search). Thus, task completeness may strongly affect users' preferences for pleasantness, and the quality of VAs actions might impact emotions' levels of arousal.

Secondly, the debriefings' results show that interactions' *perceived ease of use* may affect the emotion's levels of pleasantness and arousal for all tasks. Overall, the "excited-pleasant" and "average-pleasant" emocards were related to the quick, easy, objective, and automatic interactions. Unpleasant emocards were selected due to hardships in interactions, illustrating frustration, sadness, anger, annoyance, and disappointment. The perceived ease of use category showed how error types affected satisfaction.

Concerning VAs' communication capacity, the easiness in communicating with the VA was commented by participants in the multi-step calendar task and for playing a song, which required participants to speak in English. Two participants stated that they chose an "excited-pleasant" emotion due to cues given by the VAs that helped them adding the appointment: "*I was happy with Siri because it asked me everything: 'What's the date?' and I answered. (…) It was really easy to set up everything*" (P5; Calendar). Similarly, participants who preferred the "excited-neutral" emotions expressed surprise for playing a song more easily than they expected: "*I was surprised it understood my request in English because I generally try, and it doesn't understand*" (P5; Song). Differently, not being able to communicate with the system led to confusion and frustration: "*I couldn't find the logic that Siri understands. (…) I felt incompetent. Awful. As if I don't know how to communicate*" (P20; Calendar).

Particularly, query misrecognition was related to the "excited-unpleasant" emocard for sending messages, adding events to the calendar, and searching for information. "*Google was kind of dumb. I said everything. All the information was there, and it simply put everything as the* [appointment's] *title*." (P15; Calendar). As mentioned before, such obstacles made users repeat themselves, eliciting unpleasant emotions especially when they could not complete the task even after various interactions: "*Siri kept repeating the same question, and it made me upset because I had already said it several times and it didn't recognize it*" (P12; Calendar). Additionally, interruptions led to the selection of the "excited-unpleasant" emocard: "*It made me anxious. It did not wait for me to end the message. (…). I didn't know how to explain to Siri not to interrupt me after the 'period'*". (P18; Message).

Furthermore, Editing Errors impacted users' preferred emotions for the message and calendar tasks. A user who chose a "calm-neutral" emocard argued that: "*Although I needed some interactions, I could send the message I wanted*" (P17; Message). Nonetheless, a participant selected the emotion "excited-unpleasant" because: "*Google asked me for the* [appointment's] *end-time. I don't know why, but it set it up for* [the following year]. *I tried to edit it but I couldn't, so I just had to save it as it was.*" (P7; Calendar). Therefore, being able to edit outcomes and recover from errors may affect user satisfaction since it determines whether interactions will be successful.

We also identified that users might have had varied *expectations for VAs' capabilities and limitations* for different tasks. Adding an event to the calendar was considered complex due to the number of steps needed to achieve it, as was the recognition of slangs, question intonation, queries that mixed languages, and editing. Contrarily, checking the weather, making a phone call, and searching the internet were considered basic for a VA: "*Checking the weather is a simple request. So, there's a smaller chance of a communication fault.*" (P14; Weather).

Participants who selected pleasant emotions despite facing hardships argued that they did not blame the VA. Rather, they believed to have expressed their queries inadequately or considered the task too complicated: "*It got it* [song's name] *wrong, but it was my fault because I said 'fells' instead of 'fails'*" (P6; Song); "*I put a slang, and it didn't recognize the slang. It wasn't perfect, but I was expecting that. It's too hard*." (P8; Message). Differently, errors that happened throughout simple activities evoked unpleasant emotions: "*I couldn't complete the task that is supposed to be banal. It gave me what I wanted, but not in the right way*" (P3; Search). Hence, expectations of VAs' capabilities for each task may affect the interactions' perceived ease of use and perceived quality of the outcomes, impacting satisfaction.

Moreover, the *hands-free and eyes-free aspects of the VAs* impacted users' preferred emotional responses, but such issues affected participants differently. Users who chose "calm-pleasant" and "calm-neutral" emotions argued that they could complete the task despite having to finish it visually or manually: "*If it had read the article's text out loud, I would have been more satisfied, but it showed the results, so I was satisfied*." (P13, Search). However, unpleasant emotions were selected by some users, who argued that this compromises the voice interaction's advantages. "*If I say, 'create an event' and it answers, 'touch the screen,' it's almost an insult. I'm using the assistant because I don't want to touch the screen or because I can't touch the screen*" (P13, Calendar).

Finally, we observed that participants had different *preferences for settings* in checking the weather, making phone calls, and playing songs. For the weather task, participants had different opinions on the amount of information displayed on the interface. While some preferred more detailed weather prevision, others liked an objective answer. "*The interface was interesting because Siri showed me the temperatures for every hour of the day*" (P21, Weather). For making phone calls, Siri's option to call the house's or cellphone's number also divided opinions. Some users positively evaluated this feature, but others considered it an extra step to complete the task and pledged that VAs should know their preferred number and directly call it. "*Siri even gave me the options: 'Cellphone or house?', and I was like: 'Wow! Cellphone!'*" (P15, Phone). For both tasks, participants' individual preferences influenced the levels of arousal of their emotional responses, ranging from the "excited-pleasant" to the "calm-neutral" emocards, indicating a mild influence on user satisfaction.

However, users were more sensitive to their preferred apps to listen to music. The VAs executed users' commands by launching different music apps, evoking variations on both levels of pleasantness and arousal of users' emotional responses. Two participants who chose "excited-pleasant" emocard stated that they were happy because they completed the task, and the VA launched the app in which they usually listen to music. "*This one* [emocard] *is 'really happy' because I personally listen to music through YouTube much more often.*" (P8, Song). On the other hand, three participants who preferred unpleasant emotions stated that despite being able to finish the task, the apps launched by the VAs were not the ones they commonly use: "*iPhone insists on taking me to iTunes or Apple Music*" (P4, Song).

4.2   Task Adoption

During the debriefing sessions, we asked users to state whether they used the test's tasks in their daily VA usage and for which reasons. Figure 5 illustrated users' task adoption. Interestingly, while adoption means for all other tasks are in line with the usability scores, "search for information" was the most frequently employed feature by users. Despite having the largest number of incomplete outcomes (Fig. 2), placing third in the number of errors (Fig. 3), and yielding negative emotions on users for almost half of the interactions (Fig. 4), 17 out of 20 participants reported using a VA to search for information in their routines. Below, we present reasons stated by users to adopt tasks, addressing RQ2 (*"How is usability related to task adoption in VAs?"*).

We identified that usability is influential to task adoption, but other factors also affect tasks differently and impact VA usage. The motivators for feature usage were similar to the causes for variations in satisfaction: 1) *Hands/eyes-free interaction;* 2) *Ease of use;* 3) *Expectations and trust in VAs;* 4) *Preferences for settings and knowledge about VA features;* 5) *Usage context and task content;* and 6) *VAs' personality.*

In the first place, the possibility to interact with their smartphone without requiring their hands or eyes was the most cited reason for using a VA (16 participants). This motivator was mentioned for all tasks. Users argued that voice interaction is beneficial when manual interactions are not safe, practical, or possible. "*I use the assistant a lot when I'm driving, it's safer for me both having it read the message out loud and send the message by voice. So, I can be focused on the road and won't put my life in danger*"
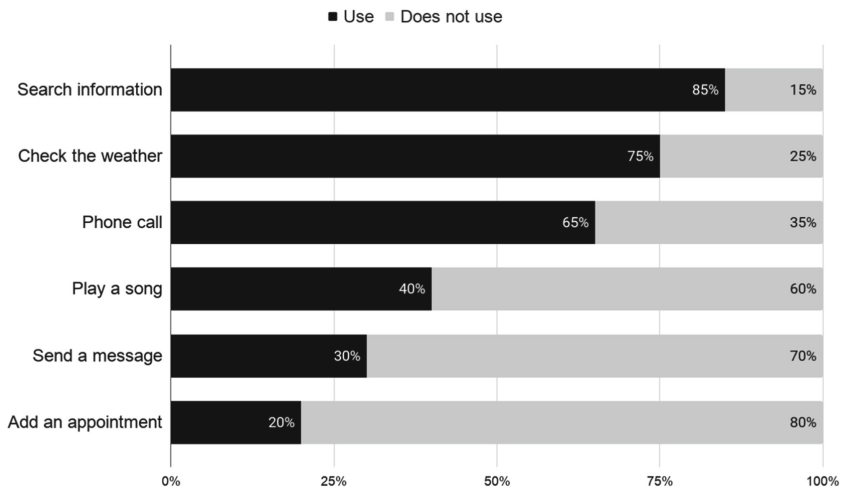
■ Use ■ Does not use

Search information	85%	15%
Check the weather	75%	25%
Phone call	65%	35%
Play a song	40%	60%
Send a message	30%	70%
Add an appointment	20%	80%

0%   25%   50%   75%   100%

Fig. 5. Adoption rate by task (n = 20 participants).

(P11; Message). Use cases mentioned were cooking, driving, riding a bike, when their hands are dirty or injured, when they are multitasking, or when their smartphone is far away: "*Usually, I use it a lot while cooking. (…) I use it to search lots of things, like, 'I need a recipe for this.'*" (P13, Search).

Notwithstanding, the "eyes-free" characteristic of voice interaction also imposed barriers for some users in specific tasks. Ten users mentioned that activities such as handling a calendar, sending a message, and playing songs naturally require visual attention. Six participants preferred to add appointments manually because they needed to organize their appointments by colors and tags or visualize other scheduled events before adding a new one. "*I do it* [adding events] *by hands because it gives me more options to edit the event, and I like things to be organized*" (P16; Calendar). As for sending a message, five participants said that they preferred to look at unread messages before choosing which ones to answer and that VAs did not support features such as stickers, emojis, and audio messages. "*I don't know if Siri can do this, but I put little hearts and flowers emojis in my messages*" (P21; Message). Besides, five participants mentioned that they usually do not have a specific song in mind, so they need to look at the screen to decide what to play. "*Actually, sometimes I don't know what I want to listen to. I look at the album's list to choose.*" (P19; Song).

The second motivator observed in our analysis was the perceived ease of use of performing tasks in VAs. Eleven participants reported that their VA usage was driven by user-VA interaction's easiness and quickness when compared to typing. As with the previous category, easy and quick interactions were a reason to perform all tasks. Specifically, for getting weather reports, three participants mentioned being used to voice interaction. "*I use Google* [Assistant] *a lot to check the weather. I think it's faster. (…) I just got used to it.*" (P16; Weather)*; "I don't even know where to click* [to check the weather on the smartphone]*, so I feel it's easier by voice.*" (P8; Weather).

However, some users perceived VAs to be harder or slower than manual interactions or claimed they are used to graphical interfaces. These perceptions did not affect a particular activity specifically. Likewise, the belief that VAs have issues recognizing users' commands was a task-independent concern for 15 out of 20 participants. Users believed that they should speak with VAs in a specific way, which led to a feeling of limitation that prevents them from using a VA. *"I guess that whenever I talk to her* [Siri]*, I set up a logic in my head to communicate with her. It's not natural for me."* (P20, Search). Contrary to our results, in which VAs misrecognized few queries mixing two languages, some users reported having trouble with bilingual commands in their daily usage. *"Words in English, first names, names in German, French. I have to say them* [to the VA]*, but I'm not sure how to pronounce it, so the odds of getting it right are low. So, I avoid these* [queries]*"* (P9, Message).

Past experiences seem to affect task adoption, as participants showed different trust levels for specific features. *"I say 'call* [wife's name]*' or something like this, and* [the assistant] *does it in just one step. (…) I have a history of positive results with it"* (P11; Phone). The trust in the VA determined their expectations for the outcomes of future interactions. Users mentioned that VAs are reliable for making phone calls and checking the weather. *"Checking the weather is the easiest, the most basic* [task]*. (…) I never had a problem with it"* (P3; Weather). Conversely, negative experiences with sending messages and playing songs led users to lose their trust in VAs. *"When I choose to type, the message is bigger or more elaborated, and I am sure Google* [Assistant] *will not recognize it."* (P10; Message); *"I don't even try* [asking it to play the user's playlist]*. I tried once or twice, but it got it all wrong, so I don't even try it."* (P13; Song).

Individual preferences for settings and the lack of knowledge of the VAs' capabilities for some features also caused users to adopt or not specific tasks. Six participants did not know that VAs could send messages, add appointments to a calendar, or play a song. Moreover, five participants said that they did not send messages through a VA because they used WhatsApp, and they were unaware of the VAs integration with this app. Similarly, six users claimed that their preferred applications to play songs and manage a calendar were not supported by their VAs. *"[Siri] goes straight to Apple Music, and I never use it (…)* [Nor Apple Music] *nor Siri"* (P4; Music).

Our results showed that the usage context and task content might impact users' decisions regarding performing activities through a VA. For example, four users explained that they only listen to music through a VA if they have already chosen which song to play. Moreover, three participants mentioned the complexity of the task's content to search for information (P2) and add appointments to the calendar (P3, P9). *"It depends on the* [appointment's] *complexity. If it's just 'I have an exam on X day,' then I can do it by voice. But if it's more complex, with a longer description, I'd rather type"* (P3; Calendar). Also, independent of the feature, using the VA was considered inappropriate or embarrassing by six users for situations in which one must be silent (e.g., during class), when there is much background noise, or when other people are listening to interactions. *"If I just got out of the doctor's office and I need to schedule a follow-up appointment (…), and I'm in front of the clerk, I won't open the assistant and speak in front of them. So, I just type"* (P15, Calendar).

Finally, six participants stated that they use a VA because they like its personality. Users mentioned liking voice interaction, enjoying testing and training the VA, and feeling futuristic or "high-tech." *"I feel very smart when I use it. Like, 'Man, I'm a genius! I was able to call by asking the assistant. And I didn't need to use my hands'. It feels magic when these things happen. I feel very clever, technological, and modern".* (P16; Phone). This factor was also not related to a specific feature.

5 Discussion

Our study aimed to assess variations of usability in six VA tasks and their relations to VA adoption. Below, we provide recommendations for VA design by discussing our results in light of existing literature.

Firstly, **designers should examine VA features separately to assess discrepancies in task usability and motivators for task adoption.** Our findings highlight that VA tasks vary in task completeness, number and types of errors, and user satisfaction. Moreover, users adopt or neglect tasks based on usability-related factors, such as the hands/eyes-free nature of interactions, ease of use, and expectations and trust in VAs. Thus, analyzing tasks individually is essential to uncover a feature's particular usability issues that may otherwise be overlooked in general analysis and understand how it is related to VA adoption. Nonetheless, in line with previous research [6, 11, 15, 20, 22, 23, 39], our findings showed that query misrecognition problems and unrequired task performance were common across all analyzed features. The results also echoed literature indicating that other, more task-independent factors unrelated to usability impacted VA adoption, such as use context [6, 12, 17–19] and the VA's personality [9, 12, 22]. Hence, a comprehensive understanding of user-VA interaction is also vital to designing adequate interactions and leveraging VA adoption.

Secondly, **designers should consider users' perception of task easiness and provide support for complex tasks**. As indicated [22, 30], we observed that participants had varied expectations for how easy and satisfactory interactions with different tasks should be. Likewise, a single activity may vary in perceived complexity depending on its content (e.g., songs with foreign names may be considered more difficult for the VA). Although users are more forgiving of mistakes made in complex tasks [22], expectations and trust impact feature usage, causing users to engage in simple tasks more frequently (e.g., search) than complex activities (e.g., message).

The effects of such perceptions on task adoption may be substantial. As shown in previous studies and this study's results, voice search has a high adoption rate [1, 10, 11, 33, 34] despite the inadequate usability measures observed in this study. We hypothesize that such discrepancy may be due to users considering it an easy activity, among other factors. Notably, we believe users employ voice search in their routines to find other, more straightforward information than the one required on the test (i.e., the name of a band's next tour). Instead of abandoning the search feature when faced with errors, users might adjust their behavior to perform simpler searches that match the VAs' capacity. This possibility is reinforced by Kiseleva et al. [14], who indicated that task completeness and speech recognition are not linked to satisfaction with web searches in VAs, pointing that other advantages may outweigh failures.

Considering users' perceptions of task complexity, designers should understand which tasks users believe to be more challenging and unfold the reasons leading to these beliefs. Thereafter, specific solutions should be employed. As an example, we observed that multi-step activities (i.e., sending messages and adding events to a calendar) are more prone to interruptions or failures in capturing users' inputs. Thus, VAs should provide clear indicators to users about when to speak.

Despite designers' efforts to facilitate interactions, errors are deemed to occur. Therefore, **VAs should facilitate error recovery for users, and such support may vary for different tasks.** Our results converge with previous studies suggesting that task completeness and perceived ease of use strongly affect satisfaction [14, 35] and may shape users' trust and expectations for future interactions with VAs [3, 5, 22, 38]. Therefore, the possibility to recover from mistakes may restore both users' trust and satisfaction with the system. For example, despite having the second-largest number of obstacles, the message task had relatively positive satisfaction scores, as successful editing led to a great number of completed interactions.

To facilitate error recovery, VA responses are essential resources for users to understand trouble sources and handle errors [13, 32] However, as discussed, task characteristics may lead to different types of errors and require specific solutions. On the one hand, tasks in which VAs directly show outcomes are not editable (e.g., search), so it is essential that VAs explicitly present error sources. Responses such as "Sorry, I don't understand" do not display any useful information for error recovery and should be avoided. On the other hand, features that require confirmations or several pieces of information may be challenging for users, and therefore explicit or implicit (i.e., questions) instructions may be presented. Notwithstanding, as subsequential failures may further escalate frustration for editable activities, VAs should avoid repeatedly requesting the same information and allow users to follow other paths.

The hands and eyes-free nature of VA interaction is paramount for VA adoption overall. This interactional characteristic allows users to multitask and cause participants to perceived voice interaction as easier and more manageable than typing on some occasions. We believe that the convenience of a hands/eyes-free interaction may be another factor leveraging the "search" task, as such benefit may counterbalance eventual errors. Conversely, requests for manual interaction (e.g., confirmations) or visual attention (e.g., displaying a search's results on screen) were negatively evaluated by this study's participants, reinforcing previous literature findings [20, 22, 28, 30]. Hence, **VAs should always present interaction outputs auditorily**: the minimum amount of information necessary to fulfill users' commands should be presented out loud, and other complementary elements may be visually displayed if a screen is available (e.g., smartphone, tablet).

Nonetheless, the results indicate that users consider some activities to naturally require visual attention (e.g., scrolling to choose a song, organizing a calendar, or sending emojis on a message). Such statements highlight that, although users value voice-based commands, this characteristic alone may not be enough for users to transition from visual to voice interactions for specific activities. **Designers should understand differences in tasks' requirements to adapt visual tasks to voice interfaces adequately.** Comprehension of how users perform an activity indicates relevant factors and ways to

translate features from graphical interfaces to voice interaction. Designers may employ methods such as task analysis to gather such awareness. For example, knowing that users eventually need to visualize a set of options to choose a song, VAs could offer music recommendations features based on user data.

Moreover, **VAs' settings should be customizable.** It was observed that participants' preferences for settings yielded variations in user satisfaction and affected task adoption. As indicated in previous studies [5, 20, 30], users desire to customize their VAs for a personalized experience. However, this study showed that participants were more susceptible to express frustration for undesired settings for music apps than for checking the weather or making phone calls, indicating that users attribute varied values for specific settings on different tasks.

Finally, **VAs' capabilities should be presented to users.** Our findings show that users do not engage in some features because they are unaware of the possibility of performing these activities. Likewise, several studies have indicated that VAs have issues of information transparency, including privacy-related data [1, 6, 16, 39], supported features [30], and system functioning [20]. Such evidence supports the hypothesis that the way information is presented to users is inappropriate, leaving out essential system descriptions, instructions, feedback, and privacy clarifications. Thus, designers must attend to information presentation in voice interaction.

6   Conclusion

Voice Assistant (VA) usage has been shown to be heterogeneous across features. Although the literature indicates that usability affects VA adoption, little was known about how usability varies across VA tasks and its relationship to task adoption. To address this gap, we conducted usability tests followed by debriefing sessions with VA users, assessing usability measures of six features and uncovering reasons for task usage. Our findings showed that task completeness, number and types of errors, and satisfaction varied across tasks, and usability-related factors were mentioned as motivators for feature adoption and abandonment. Moreover, other factors such as customization, VAs' personality, and usage context also impacted VA adoption. Our main contributions are recommendations for the design of VAs. We highlight that, while comprehensive approaches are still vital, attending to tasks separately is paramount to understand specific usability issues, task requirements, and users' perceptions of features, and consequently develop design solutions that leverage VAs' usability and adoption.

Nevertheless, our study had some limitations. Firstly, data privacy, which has been demonstrated to affect VA usage [4, 6, 24, 34], was not mentioned by our participants. This absence may be due to the usability tests' setting - as users were not sharing their own data -, and the debriefing sessions' dynamics, since participants were instructed to talk about each task separately instead of evaluating VAs generally. Furthermore, this study only evaluated six tasks, and therefore further examinations are needed to understand how other tasks' characteristics may impact VA usage. Finally, additional studies are needed to increase our recommendations' significance by indicating specific design solutions (e.g., how to present information about the VA to users appropriately).

References

1. Ammari, T., Kaye, J., Tsai, J.Y., Bentley, F.: Music, search, and IoT. ACM Trans. Comput. Hum. Interact. **26**, 1–28 (2019). https://doi.org/10.1145/3311956
2. Barnum, C.: Usability Testing Essentials. Ready, Set… Elsevier, Burlington (2011)
3. Beneteau, E., Guan, Y., Richards, O.K., et al.: Assumptions Checked. Proc. ACM Interact. Mob. Wearable Ubiquit. Technol. **4**, 1–23 (2020). https://doi.org/10.1145/3380993
4. Burbach, L., Halbach, P., Plettenberg, N., et al.: "Hey, Siri", "Ok, Google", "Alexa". Acceptance-relevant factors of virtual voice-assistants. In: 2019 IEEE International Professional Communication Conference (ProComm), pp. 101–111. IEEE (2019)
5. Cho, M., Lee, S., Lee, K.-P.: Once a kind friend is now a thing: understanding how conversational agents at home are forgotten. In: Proceedings of the 2019 on Designing Interactive Systems Conference, pp. 1557–1569. Association for Computing Machinery, New York (2019)
6. Cowan, B.R., Pantidi, N., Coyle, D., et al.: What can i help you with? In: Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services, pp. 1–12. ACM, New York (2017)
7. Desmet, P., Overbeeke, K., Tax, S.: Designing products with added emotional value: development and application of an approach for research through design. Des. J. **4**, 32–47 (2001). https://doi.org/10.2752/146069201789378496
8. Doyle, P.R., Edwards, J., Dumbleton, O., et al.: Mapping perceptions of humanness in intelligent personal assistant interaction. In: Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services. Association for Computing Machinery, New York (2019)
9. Festerling, J., Siraj, I.: Alexa, what are you? Exploring primary school children's ontological perceptions of digital voice assistants in open interactions. Hum. Dev. **64**, 26–43 (2020). https://doi.org/10.1159/000508499
10. Garg, R., Sengupta, S.: He is just like me: a study of the long-term use of smart speakers by parents and children. Proc. ACM Interact. Mob. Wearable Ubiquit. Technol. **4** (2020). https://doi.org/10.1145/3381002
11. Huxohl, T., Pohling, M., Carlmeyer, B., et al.: Interaction guidelines for personal voice assistants in smart homes. In: 2019 International Conference on Speech Technology and Human-Computer Dialogue (SpeD), pp. 1–10. IEEE (2019)
12. Kendall, L., Chaudhuri, B., Bhalla, A.: Understanding technology as situated practice: everyday use of voice user interfaces among diverse groups of users in urban India. Inf. Syst. Front. **22**, 585–605 (2020). https://doi.org/10.1007/s10796-020-10015-6
13. Kim, J., Jeong, M., Lee, S.C.: "Why did this voice agent not understand me?": error recovery strategy for in-vehicle voice user interface. In: Adjunct Proceedings - 11th International ACM Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI 2019, pp. 146–150 (2019)
14. Kiseleva, J., Williams, K., Jiang, J., et al.: Understanding user satisfaction with intelligent assistants. In: Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval, pp. 121–130. ACM, New York (2016)
15. Larsen, H.H., Scheel, A.N., Bogers, T., Larsen, B.: Hands-free but not eyes-free. In: Proceedings of the 2020 Conference on Human Information Interaction and Retrieval, pp. 63–72. ACM, New York (2020)

16. Lau, J., Zimmerman, B., Schaub, F.: Alexa, are you listening? Privacy perceptions, concerns and privacy-seeking behaviors with smart speakers. Proc. ACM Hum. Comput. Interact. **2** (2018). https://doi.org/10.1145/3274371

17. Li, Z., Rau, P.-L.P., Huang, D.: Self-disclosure to an IoT conversational agent: effects of space and user context on users' willingness to self-disclose personal information. Appl. Sci. **9**, 1887 (2019). https://doi.org/10.3390/app9091887

18. Lopatovska, I., Oropeza, H.: User interactions with "Alexa" in public academic space. Proc. Assoc. Inf. Sci. Technol. **55**, 309–318 (2018). https://doi.org/10.1002/pra2.2018.145 05501034

19. Lopatovska, I., Rink, K., Knight, I., et al.: Talk to me: exploring user interactions with the Amazon Alexa. J. Librariansh Inf. Sci. **51**, 984–997 (2019). https://doi.org/10.1177/096100 0618759414

20. Lopatovska, I., Griffin, A.L., Gallagher, K., et al.: User recommendations for intelligent personal assistants. J. Librariansh Inf. Sci. **52**, 577–591 (2020). https://doi.org/10.1177/096 1000619841107

21. Lovato, S.B., Piper, A.M., Wartella, E.A.: Hey Google, do unicorns exist? Conversational agents as a path to answers to children's questions. In: Proceedings of the 18th ACM International Conference on Interaction Design and Children, pp. 301–313. Association for Computing Machinery, New York (2019)

22. Luger, E., Sellen, A.: "Like having a really bad PA": the gulf between user expectation and experience of conversational agents. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 5286–5297. Association for Computing Machinery, New York (2016)

23. Maués, M.P.: Marcela Pedroso Maués Um olhar sobre os assistentes virtuais personificados e a voz como interface. Dissertation, Pontifical Catholic University of Rio de Janeiro (2019)

24. McLean, G., Osei-Frimpong, K.: Hey Alexa … examine the variables influencing the use of artificial intelligent in-home voice assistants. Comput. Hum. Behav. **99**, 28–37 (2019). https://doi.org/10.1016/j.chb.2019.05.009

25. Meeker, M.: Internet trends 2016. In: Kleiner Perkins (2016). https://www.kleinerperkins.com/perspectives/2016-internet-trends-report/. Accessed 11 Feb 2021

26. Moar, J., Escherich, M.: Hey Siri, how will you make money? In: Juniper Research (2020). https://www.juniperresearch.com/document-library/white-papers/hey-siri-how-will-you-make-money. Accessed 11 Feb 2021

27. Moriuchi, E.: Okay, Google!: an empirical study on voice assistants on consumer engagement and loyalty. Psychol. Mark. **36**, 489–501 (2019). https://doi.org/10.1002/mar.21192

28. Motta, I., Quaresma, M.: Opportunities and issues in the adoption of voice assistants by Brazilian smartphone users. Rev. Ergodes. HCI **7**, 138 (2019). https://doi.org/10.22570/erg odesignhci.v7iEspecial.1312

29. Newman, N.: The future of voice and the implications for news. In: Reuters Institute (2018). https://reutersinstitute.politics.ox.ac.uk/our-research/future-voice-and-implications-news. Accessed 11 Feb 2021

30. Oh, Y.H., Chung, K., Ju, D.Y.: Differences in interactions with a conversational agent. Int. J. Environ. Res. Public Health **17**, 3189 (2020). https://doi.org/10.3390/ijerph17093189

31. Pearl, C.: Designing Voice User Interfaces. O'Reilly (2016)

32. Porcheron, M., Fischer, J.E., Reeves, S., Sharples, S.: Voice interfaces in everyday life. In: Proceedings of the Conference on Human Factors in Computing Systems (2018)

33. Pradhan, A., Mehta, K., Findlater, L.: "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, pp. 1–13. Association for Computing Machinery, New York (2018)

34. Pridmore, J., Zimmer, M., Vitak, J., et al.: Intelligent personal assistants and the intercultural negotiations of dataveillance in platformed households. Surveill. Soc. **17**, 125–131 (2019). https://doi.org/10.24908/ss.v17i1/2.12936

35. Purington, A., Taft, J.G., Sannon, S., et al.: "Alexa is my new BFF": social roles, user satisfaction, and personification of the Amazon Echo. In: Proceedings of the Conference on Human Factors in Computing Systems, pp. 2853–2859 (2017)

36. Rong, X., Fourney, A., Brewer, R.N., et al.: Managing uncertainty in time expressions for virtual assistants. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 568–579. Association for Computing Machinery, New York (2017)

37. Statista: Total number of Amazon Alexa skills in selected countries as of January 2020 (2020). https://www.statista.com/statistics/917900/selected-countries-amazon-alexa-skill-count/. Accessed 11 Feb 2021

38. Trajkova, M., Martin-Hammond, A.: "Alexa is a toy": exploring older adults' reasons for using, limiting, and abandoning echo. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–13. Association for Computing Machinery, New York (2020)

39. Weber, P., Ludwig, T.: (Non-)Interacting with conversational agents: perceptions and motivations of using chatbots and voice assistants. In: Proceedings of the Conference on Mensch Und Computer, pp. 321–331. Association for Computing Machinery, New York (2020)

40. West, M., Kraut, R., Ei Chew, H.: I'd blush if I could. In: UNESCO Digital Library (2019). https://unesdoc.unesco.org/ark:/48223/pf0000367416.page=1. Accessed 11 Feb 2021

41. White-Smith, H., Cunha, S., Koray, E., Keating, P.: Technology tracker - Q1 2019. In: Ipsos MORI Tech Tracker (2019). https://www.ipsos.com/sites/default/files/ct/publication/documents/2019-04/techtracker_report_q12019_final.pdf. Accessed 11 Feb 2021

42. Yang, X., Aurisicchio, M., Baxter, W.: Understanding affective experiences with conversational agents. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1–12. Association for Computing Machinery, New York (2019)