Differences Between Smart Speakers and Graphical User Interfaces for Music Search Considering Gender Effects

Florian Habler
University of Regensburg
Regensburg, Germany
Florian.Habler@student.ur.de

Marco Peisker
University of Regensburg
Regensburg, Germany
Marco.Peisker@student.ur.de

Niels Henze
University of Regensburg
Regensburg, Germany
Niels.Henze@ur.de

ABSTRACT

The ubiquitous availability of smart speakers allows hands- and eyes-free interaction through Voice User Interfaces (VUIs). Controlling music playback is the most commonly used feature of VUIs. Previous work investigated how users naturally interact with smart speakers and suggested that users' gender could affect the devices' usability. The usability of commercial devices compared to other interactive systems and the effects of users' gender is, however, unclear. Therefore, we conducted a study with 20 participants using an Amazon Echo Dot and a laptop device. Participants searched for artists and titles using a Graphical User Interface (GUI) and a VUI. In addition, they performed different tasks such as saving a song in a playlist or adding songs into a queue. The analysis revealed that the VUI provides significantly lower usability because it lacks features, requires higher mental effort, and provides confusing answers. In contrast to previous concerns, the analysis did not reveal significant device×gender effects.

CCS CONCEPTS

• **Human-centered computing → Natural language interfaces**; **Ubiquitous and mobile computing**; **Graphical user interfaces**.

KEYWORDS

Smart speaker, graphical user interface, voice user interface, gender, music search

ACM Reference Format:
Florian Habler, Marco Peisker, and Niels Henze. 2019. Differences Between Smart Speakers and Graphical User Interfaces for Music Search Considering Gender Effects. In *MUM 2019: 18th International Conference on Mobile and Ubiquitous Multimedia (MUM 2019), November 26–29, 2019, Pisa, Italy.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3365610.3365627

1 INTRODUCTION & BACKGROUND

Voice interaction has become increasingly popular in recent years. All major smartphone platforms, including Apples' iOS and Google's Android, come with an integrated conversational agent that allows users to interact through speech input and output. Microsoft included Cortana and Apple included Siri into their most recent operating systems. In addition, dedicated smart speakers that feature conversational agents by Amazon, Google, Apple, and Alibaba, Baidu, Microsoft, and Samsung are available. In 2019, Amazon announced that smart speakers with their conversational agent Alexa have been sold over 100 million times[1] and media reports[2] suggests that there will be over 200 million installed smart speakers at the end of 2019. Recent work aims to bring conversational agents to an increasing number of domains. An example is the work by Pearson et al. who compared smart speakers with human responses to guide the development of conversational agents for supporting people living in slums [27]. Despite clearly providing less relevant answers than human responders, the conversational agent received over double the number of questions. The authors attribute this difference to the much shorter response time of the computational system. The work highlights the potential of conversational agents. Even if their answers have a lower quality than answers by a human counterpart, the availability and the instantaneous responses can cancel out the limitations.

Previous work showed that smart speakers are mainly used to control music streaming services. Maita not only showed that 73% of the persons not using a smart speaker envision to use it for music playback but that over 90% of the persons owning a smart speaker use it for music playback [23]. Observing how Google Home is used in 88 homes over 110 days, Bentley et al. showed that with 40% of all requests, controlling music playback is the most commonly used feature [2]. Similarly, Sciuto et al. showed for users' of Amazon's Alexa that voice commands to control music playback are more often issued than any other type of command [32].

Research investigating the interaction with smart speakers focused on how users naturally interact with the device by analyzing, for example, log files [2, 32] and product reviews [10, 28]. It has been investigated how users react to errors [18, 20, 24], make search requests [11], and ask for recommendations [19] using VUIs. Recent work even determined the effects of a smart speaker's gender and language and showed that the VUIs' gender is less important for the provided user experience than the used language [12]. The usability that smart speakers provide for commonly used features, such as music playback, compared to other interactive systems is, however, unclear. Enabling hands- and eyes-free interaction, smart speakers have clear benefits compared to other interactive systems. Comparing the usability of smart speakers and other interactive systems is, however, important for learning how to identify their limitations and to improve their usability.

MUM 2019, November 26–29, 2019, Pisa, Italy

ACM ISBN 978-1-4503-7624-2/19/11…$15.00
https://doi.org/10.1145/3365610.3365627

[1] https://www.theverge.com/2019/1/4/18168565/amazon-alexa-devices-how-many-sold-number-100-million-dave-limp

[2] https://techcrunch.com/2019/04/15/smart-speakers-installed-base-to-top-200-million-by-year-end/

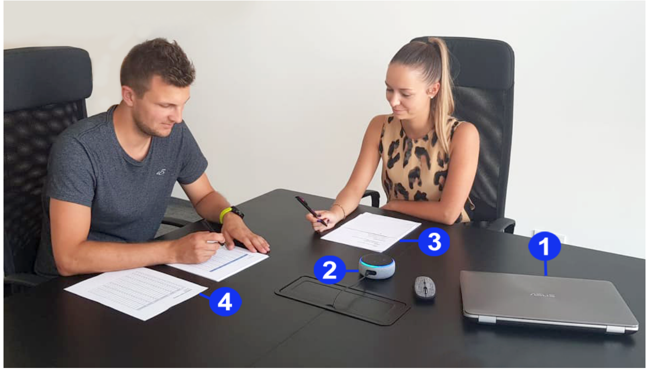

Figure 1: Study setup of participant (right) searching music with the VUI (2). List of interaction tasks (3), sheets with questionnaires and a stop watch (4). The computer with the GUI (1) was closed during these tasks.

Only little is known about the effects of demographic characteristics on the interaction with conversation agents. From other domains, we know that users' gender can have an effect on, for example, interface preferences [31] how artifacts are designed [30]. Based on data from six participants Myers et al. found no correlation between age, gender, and obstacles encountered with a VUI [24]. Unsurprisingly, Myers et al. subsequently showed that experience with voice user interfaces and technical confidence have an effect on interacting with unfamiliar voice user interfaces [25]. Nonetheless, previous work repeatedly warned that current commercial devices can have effects on gender bias [13, 21, 26]. Besides other concerns, it has been argued that women's voices are harder to recognize because of biased training data [22] and even that voice recognition is naturally sexist [29]. It is, however, unclear if the usability of commercial smart speakers is indeed affected by users' gender.

Overall, with smart speakers, a new category of interactive systems currently becomes widely adopted. Today, smart speakers' most often used feature is music playback [2, 23, 32]. While previous work focused on analyzing how users naturally interact with commercial smart speakers (e.g. [2, 32]), it is unclear how their usability compares to other interactive systems. This is not only important to improve smart speakers' usability but also to determine if users' demographic characteristics affect the devices' usability.

In this paper, we present the results from a controlled experiment with 10 female and 10 male participants to determine the effect of the device and the user's gender on usability. Participants interacted with Spotify, one of the most widely used streaming services for music, either using a graphical user interface or using an Amazon Echo Dot. While smart speakers enable eyes- and hands-free interaction, we found that their usability is lower compared to a GUI. Reasons for the lower usability include missing features, a higher required mental effort, and confusing answers. In contrast to concerns expressed by previous work, the analysis did not reveal significant device×gender effects on the objective and subjective measures.

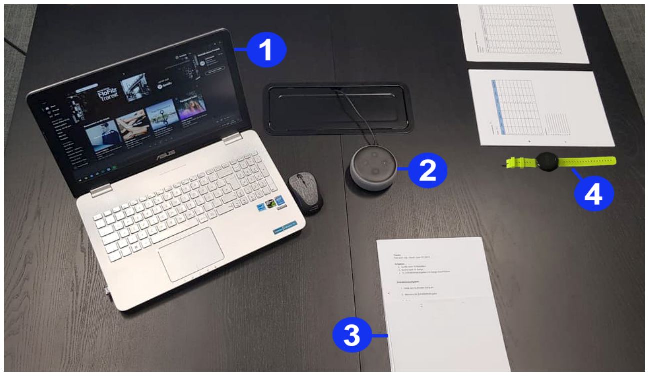

Figure 2: Study setup of participants searching music with GUI (1), VUI (2), list of artists and songs (3) and sheets with questionnaires (4).

2 METHOD

We conducted a controlled experiment to determine if the used interface (VUI or GUI) or the users' gender (female or male) have significant effects on usability when controlling music playback. 10 female and 10 male participants used the music streaming service Spotify. They used an Amazon Echo Dot that comes with Amazon's virtual assistant Alexa as well as Spotify's application for desktop computers.

2.1 Design and Tasks

We used a mixed design with the interface as a within-subject factor and participants' gender as a between-subject factor. Participants were asked to interact with the same commercial music service using a VUI and a GUI. We balanced the order of the conditions to avoid sequence effects.

Participants were asked to search for ten artists and ten songs using each interface. In addition, they performed ten additional tasks, such as saving a song in a playlist or activating the shuffle function, with both interfaces. Artists and songs were derived from the top 40 artists and songs of the hot 100 billboard artists and song charts from June 2019.

Springer and Cramer showed that current commercial VUIs are not able to interpret all spoken instructions correctly [33]. Using GUIs, however, users can be required to type artists' names and song titles, which can also be error-prone. To make the two interfaces comparable, we presented the artists and songs participants were asked to search for using different modalities. When using the VUI, artists and songs were presented using a printed list (see Table 1 for an example). When using the GUI, we provided the artists and songs through speech by the voice from Amazon's Alexa. The lists of the artists and the songs, as well as the order of GUI and VUI, were arranged using Latin squares. Figure 1 shows the study setup in which a participant searched for music with a VUI and Figure 2 shows the study setup of the search with a GUI.

2.2 Measures

We collected quantitative objective and subjective measures to determine effects on usability. We also collected qualitative data to determine potential reasons that help to explain the quantitative measures. We determined the effectiveness and efficiency when using each interface by measuring the task success rate and task completion time when retrieving songs and artists. Task completion time was determined by measuring the time participants needed to search for a song or artist from starting the search to the time the music is played by the device. The task success rate was determined by the number of attempts per task. If the participant managed to play the artist or the song on the first try, we counted it as a full success. If the song or artist was not played until the second or third attempt, it was determined to be a partial success. After three attempts, the task was classified as a failure and the participant continued with the next one. We have summed up the overall result for each song and artist for each device by giving each successful task the value 1 and each partially successful task the value 0.5.

We used two standardized questionnaires to quantify participants' subjective impression. To measure subjective usability, we used the System Usability Scale [6]. We used the NASA Task Load Index (NASA-TLX) questionnaire to assess the perceived taskload [15]. We used the Raw NASA-TLX [14], the NASA-TLX without the additional weighting process.

Qualitative feedback was collected through interviews and taking notes throughout the experiment. After using each interface, participants were asked to provide feedback in addition to the quantitative questionnaires. They were asked to indicate what the overall satisfaction was, what they liked and what they disliked about each interface. Furthermore, we also asked about additional features they would recommend for each interface.

2.3 Apparatus

Depending on the condition, we provided participants with a VUI or a GUI. For the VUI, we provided the participants with an Amazon Echo Dot 3rd generation with the pre-installed Amazon Alexa (Alexa) Voice. The Amazon Echo was the best-selling stationary SPA at the time of the study and the Amazon Echo Dot 3rd generation was the one with the latest software. For the GUI, we used a standard ASUS laptop with a 15.6-inch screen running Microsoft Windows 10.

Artist	**Song**
Thomas Rhett	Truth Hurts by Lizzo
Billie Eilish	Pop Out by Polo G Featuring Lil Tjay
BTS	High Hopes by Panic! At The Disco
Khalid	Con Calma by Daddy Yankee Featuring Snow
Lil Nas X	Middle Child by J. Cole
Post Malone	Eastside by benny blanco, Halsey & Khalid
Ariana Grande	Earfquake by Tyler, The Creator
Luke Combs	Act Up by City Girls
Ed Sheeran	Going Bad by Meek Mill Featuring Drake
Halsey	Pure Water by Mustard & Migos

Table 1: Sample list of artists and songs provided to the participants.

For both conditions, we used Spotify as it was the most widely used streaming service[3] when conducting the study. We used the premium version of Spotify to run the service without advertisements and to access all the functions necessary for the study. Basic users accounts do not provide the ability to search and listen to all tracks and users can only skip a small number of songs. We preinstalled Spotify on the laptop and activated the Spotify Skill on the Echo Dot.

2.4 Procedure

We conducted a pilot study to test the procedure and avoid technical problems during the study. The actual study took place in a quiet office. The study took about 45 minutes per participant. After welcoming a participant, we explained the general aim of the study and the procedure. Afterward, we asked them to provide informed consent and sign an informed consent form. Participants filled a demographic questionnaire. Besides age, gender and background, we also asked about their experience with Spotify and smart speakers, such as Amazon's Echo.

Before using each interface, we provided instructions on how to use the respective device. After participants familiarized with the device, they were asked to search for 10 artists while we measured the number of attempts and the task completion time. After searching the artists, they were asked to search for 10 songs while we again measured the number of attempts and the task completion time. Finally, we asked them to perform ten additional tasks, such as saving a song in a playlist or activating the shuffle function. After completing the tasks with a device, we asked them to fill the quantitative questionnaires and to provide qualitative feedback. After completing the procedure with the first device participants completed the same procedure with the second device.

2.5 Participants

We recruited 20 participants (10 female, 10 male) via social media and e-mail. We intentionally recruited both inexperienced and experienced Alexa and Spotify users to get a balanced sample. Participants were between 22 and 31 years old (M = 25.3, SD = 2.56). Five of the male and four of the female participants had a technical background. They either studied or had a degree in computer science or a related discipline. They were all native German speakers.

Four participants stated that they use an Amazon Echo device often or very often, seven rarely or very rarely interact with such a device, and nine participants had no experience with an Amazon Echo device. Ten participants used the Spotify streaming service often or very often, four participants used it occasionally and six participants never used Spotify for music streaming.

3 RESULTS

We used two-way mixed ANOVAs to determine effects of the interface on task completion time, task success rate, SUS and NASA TLX. To determine gender effects, we looked for interface×gender effects which could imply that one interface is equally suited for male and female users while the other is not. We did not test for a main effect of gender as we do not consider it relevant to determine

[3] https://www.statista.com/statistics/798125/most-popular-us-music-streaming-services-ranked-by-audience

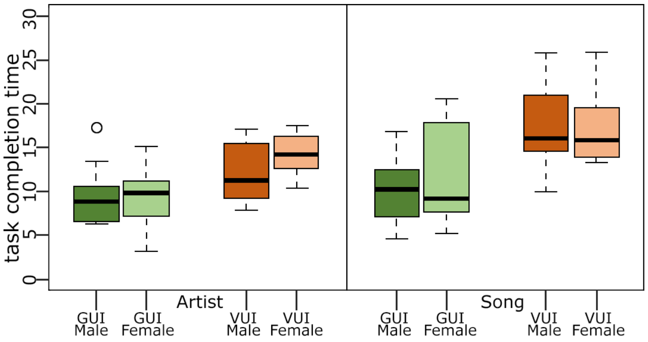

Figure 3: Average task completion time to search for an artist or a song in seconds.

general differences between our female and male participants. We used Shapiro-Wilk tests to ensure that the data is approximately normally distributed and Levene's test to ensure the homogeneity of variances.

3.1 Task Completion Time

As we did not limit the time to search for an artist or song but only limited participants to three unsuccessful attempts. With up to two minutes, some attempts took much longer than the average task completion time. We removed outliers by discarding all task completion times that were more than three standard deviations from the mean. Thus, we removed 16 out of 800 trials that took longer than 51.26 seconds when analyzing the task completion time.

Figure 3 shows the average task completion time when searching for artists and songs. The statistical analysis revealed a significant effect of the interface on the task completion time when searching for artists $F(1, 20) = 15.99$, $p = .001$. Participants were faster using the GUI (M=9.36, SD=3.36) than with the VUI (M=13.20, SD=3.18). The analysis also revealed a significant effect of the interface on the task completion time when searching for songs $F(1, 20) = 14.95$, $p = .001$. Again, participants were faster using the GUI (M=10.81, SD=4.82) than with the VUI (M=17.40, SD=4.45). We found no significant interface×gender effect, when searching for artists $F(1, 20) = 0.74$, $p = .402$ or songs $F(1, 20) = 0.26$, $p = .615$.

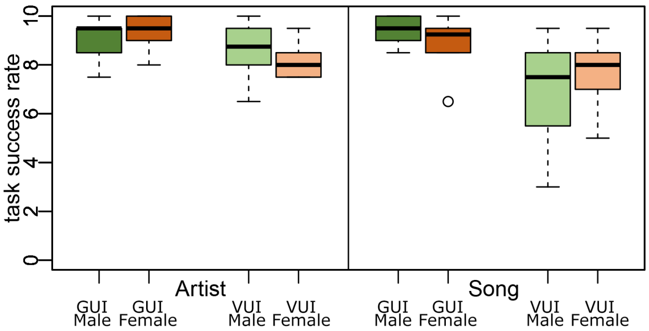

Figure 4: Average task success rate when searching for an artist or a song.

3.2 Task Success Rate

Figure 4 shows the task success rate when searching for artists and songs. The analysis revealed a significant effect of the interface on the task success rate when searching for artists $F(1, 20) = 11.65$, $p = .003$. Participants had a higher task success rate when using the GUI (M=9.23, SD=0.79) than when using the VUI (M=8.40, SD=0.95). The analysis also revealed a significant effect of the interface on the task success rate when searching for songs $F(1, 20) = 17.78$, $p = .001$. Participants had a higher task success rate when using the GUI (M=9.20, SD=0.86) than when using the VUI (M=7.25, SD=1.79). We found no significant interface×gender effect, when searching for artists $F(1, 120) = 0.06$, $p = .812$ or songs $F(1, 20) = 0.13$, $p = .726$.

3.3 System Usability Scale

Participants rated the usability of each interface using the SUS. Figure 5 shows the average SUS for both interfaces. The analysis revealed a significant effect of the interface on the SUS, $F(1, 20) = 34.85$, $p < .001$. Participants provided a higher SUS rating for the GUI (M=86.6, SD=10.6) than for the VUI (M=55.4, SD=21.6). We found no significant interface×gender effect $F(1, 20) = 0.03$, $p = .863$.

3.4 NASA Task Load Index

Participants rated the perceived task load using the NASA-TLX. Figure 6 shows the average NASA-TLX for both interfaces. The analysis revealed a significant effect of the interface on the NASA-TLX, $F(1, 20) = 17.64$, $p = .001$. Participants provided a lower NASA-TLX rating for the GUI (M=21.60, SD=16.01) than for the VUI (M=43.05, SD=22.72). We found no significant interface×gender effect $F(1, 20) = 1.51$, $p = .235$.

3.5 Qualitative Data

We collected qualitative feedback through open questions and a short interview after a participant used one of the interfaces. When asked about the GUI, P1, P3, P4, P6, P7, P8, P12, P13, P14, P16, and P17 positively acknowledged that the GUI is "intuitive, fast and state of the art". P3, P9, and P20 especially appreciated the hints and P5 also fancied the auto-correction of results provided by the GUI. However, P5, P8, and P20 criticized that the GUI is

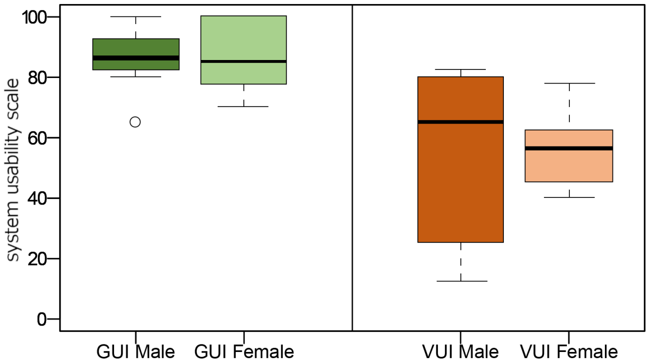

Figure 5: Average System Usability Scale (SUS) scores for both interfaces. Higher scores indicate higher usability.

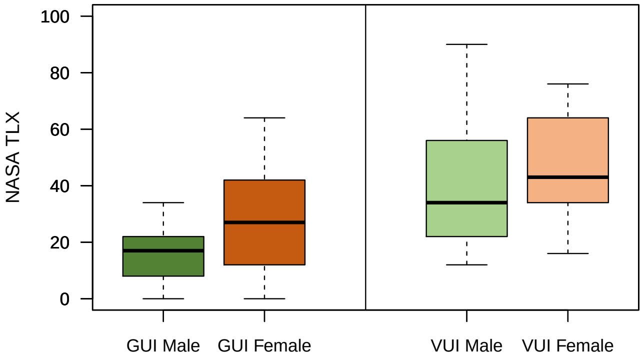

Figure 6: Average NASA-TLX scores for both interfaces. Higher scores indicate higher task load.

too confusing and too overloaded. The sometimes unclear voice of Alexa which was used to read the names of the songs and artists was criticized by P4, P6, and P7. In addition, P8 complained that some functions are only accessible with a right-click and P1 found some functions unclear and suggested apart from this that the GUI could be improved by adding keyboard shortcuts. Alongside this P3 lacked a better recognition of different spellings, P5 wanted to add songs to playlists faster, P16 missed a button to queue a song and also P17 missed a button to add a song directly to a playlist.

When asked about the VUI, seven participants highlighted that it is easy to play a song with the VUI. P1 especially appreciated the hands-free usage of the VUI, P7 the little physical effort when interacting with the VUI and also P6, P9 and P11 mentioned that also unclear pronounced terms were understood. P6 also stated that the device has sensible answers to a variety of different questions, P13 appreciated the instant feedback and P20 liked that the device can also be used during other activities. Nevertheless, P1 and P3 criticized the voice of the VUI. According to P1, P3, P5, P7, P11 and P16, the voice was sometimes difficult to understand and also the partial pronunciation of e.g. English numbers was considered insufficient by P7. Furthermore, the recapitulation of commands was criticized by P6 and P17 stated that it is annoying to say the wake word "Alexa" before each interaction. Additionally, P2, P9, P12, P14, P15, and P20 lacked common Spotify features that they could use with a GUI and therefore stated that it is unclear what the device is able to do. Accordingly, P1, P2, P3, P4, P16, P20 stated that the VUI should provide the most common functions. For example, P1, P2, and P3 missed a function to like a song, P1, P2, P3, P4, and P20 missed the function to add songs to a playlist and P3 also indicated that queuing a song would be a useful feature.

4 DISCUSSION

To compare the usability of GUIs and VUIs while considering participants' gender as an additional factor, we conducted a controlled experiment with 20 participants. Our results consistently show that the GUI provides higher usability, at least in the context tested in the study. Participants were faster and more effective when using the GUI. Consequently, the GUI received higher usability and lower task load rating. We, also consistently, revealed no interface×gender effects on any of the measures.

Previous work suggests that the commands entered with a VUI are closer to natural language than commands entered with a GUI [11]. This can, at least partially explain why participants were consistently faster using the GUI. Using the VUI, activation words by the participants as well as courtesy words by participants and the VUI increased the task completion time. The VUIs' lower usability is further amplified by confusing responses from the VUI. Statements such as "You can't buy songs on Spotify." or "You can't give 'Like' ratings on Spotify." left participants confused, because they were able to use these functions on Spotify with the GUI. In line with Dingler et al. [8] we also highlight the importance of a consistent experience when supporting fundamentally different interaction modalities. Spotify is currently available on diverse platforms, including cars, smart speakers, mobile devices and desktop experience. While services should provide consistent functions across devices, it is at least necessary to avoid misleading statements.

Previous work discussed that women are less well understood by commercial devices for more than a decade [22]. Similarly, Tatman showed that [34] algorithms used by Google understand males better than females. In contrast to previous work, our results do not support the assumption that algorithms necessarily understand men better than women. Using representation learning the current approach for speech recognition [7, 16], there are no fundamental reasons why a system should understand specific demographics better than others. If a system understands men better than women, developers simply have to add more samples from women to the training set.

From an optimistic perspective, our results suggest that the speech recognition used by the Amazon Echo got improved over the last years and has no gender bias anymore. The study, however, was conducted in a calm office environment and participants were sitting directly in front of the device, creating optimal conditions for speech recognition. Under such optimal conditions, speech recognition is potentially not challenging anymore. Under more challenging conditions, however, biases could still manifest. Consequently, we suggest that more research on bias-free speech recognition is needed. Future work should not be limited to users' gender but, in line with recent work [3], should also take other demographic characteristics into account.

We conducted a study with 20 participants. While the sample size provided sufficient power to reveal differences between the interfaces, statistical power when aiming to reveal gender effects is certainly limited. Our results enable future work to determine the appropriate sample size for such a comparison. Our study focused on music search, the most common but specific use case. Future work should compare the usability of VUIs and GUIs using further use cases. An important direction for future work is to consider other situations. We did not consider situations where the user uses the hands or eyes for a primary task. Examples for such situations include interacting with a VUI while driving but also while cooking or in the bathroom. In these situations, using a GUIs that requires text entry through a keyboard is certainly not usable.

5 CONCLUSION

VUIs are becoming ubiquitous. They are not only integrated into many interactive systems for end-users. With smart speakers, they also sparked a new class of interactive systems. The increasing popularity of smart speakers and VUIs, such as Amazon's Echo, Google Home or Apple's HomePod, requires to reflect on how these systems integrate into the larger landscape of interactive systems. We conducted a study to investigate the effects of participants' gender and the interaction modality when controlling music playback. With a balanced sample of female and male participants, we measured the task completion time, the task success rate, usability, and workload while using a GUI or a VUI.

Our results show that interacting with music services is more effective and efficient when using a GUI. The design of the VUIs causes some challenges such as a higher mental effort due to the missing display. Entering abbreviations and unknown words is more difficult. GUIs offer users help through suggestions and autocompletion which is especially helpful when searching for music. Today, design guidelines are largely optimized for GUIs. Thus, further work is necessary to fully establish consistent design guidelines for VUIs. One participant put it this way: "Alexa should rather suggest what can be done than always just say what is not possible".

The knowledge gained in our study opens up the possibility of taking a closer look at how others GUIs such as smartphones and tablets and others VUIs such as Google Home or Apple's HomePod work in the context of music searches. In addition, the constructive feedback and the desired features of the participants should be considered when implementing the next update or the next generation of music streaming services and commercial VUIs. Another direction is eliciting and comparing interface alternatives with potential users [17]. Finally, we believe that, considering the warnings by researchers [1, 13, 21, 26], the media [4, 9], and even policy makers [5] should be taken seriously. We need empirical investigations that determine the effects of interacting with female-gendered smart speakers on users' bias.

REFERENCES

[1] Rachel Adams and Nora Ni Loideain. 2019. Addressing Indirect Discrimination and Gender Stereotypes in AI Virtual Personal Assistants: The Role of International Human Rights Law. In *Annual Cambridge International Law Conference 2019, New Technologies: New Challenges for Democracy and International Law.* SSRN, 21. https://doi.org/10.2139/ssrn.3392243

[2] Frank Bentley, Chris Luvogt, Max Silverman, Rushani Wirasinghe, Brooke White, and Danielle Lottridge. 2018. Understanding the Long-Term Use of Smart Speaker Assistants. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 3, Article 91 (Sept. 2018), 24 pages. https://doi.org/10.1145/3264901

[3] Su Lin Blodgett and Brendan O'Connor. 2017. Racial Disparity in Natural Language Processing: A Case Study of Social Media African-American English. *CoRR* abs/1707.00061 (2017), 4. arXiv:1707.00061 http://arxiv.org/abs/1707.00061

[4] Ian Bogost. 2018. Sorry, Alexa Is Not a Feminist. *The Atlantic* (2018). https://www.theatlantic.com/technology/archive/2018/01/sorry-alexa-is-not-a-feminist/551291

[5] Jennifer Breslin and Rachel Pollack. 2019. *I'd blush if I could: closing gender divides in digital skills through education.* UNESCO Programme and Meeting Document, Chapter The rise of gendered AI and its troubling repercussions, 85–145. https://unesdoc.unesco.org/ark:/48223/pf0000367416.locale=en

[6] John Brooke. 1996. SUS: A quick and dirty usability scale. *Usability Evaluation in Industry* 194 (1996), 189–194.

[7] G. E. Dahl, D. Yu, L. Deng, and A. Acero. 2012. Context-Dependent Pre-Trained Deep Neural Networks for Large-Vocabulary Speech Recognition. *IEEE Transactions on Audio, Speech, and Language Processing* 20, 1 (Jan 2012), 30–42. https://doi.org/10.1109/TASL.2011.2134090

[8] Tilman Dingler, Rufat Rzayev, Alireza Sahami Shirazi, and Niels Henze. 2018. Designing Consistent Gestures Across Device Types: Eliciting RSVP Controls for Phone, Watch, and Glasses. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18).* ACM, New York, NY, USA, Article 419, 12 pages. https://doi.org/10.1145/3173574.3173993

[9] Editorial. 2019. The Guardian view on female voice assistants: not OK, Google. *The Guardian* (Jun 2019). https://www.theguardian.com/commentisfree/2019/jun/24/the-guardian-view-on-female-voice-assistants-not-ok-google

[10] Yang Gao, Zhengyu Pan, Honghao Wang, and Guanling Chen. 2018. Alexa, My Love: Analyzing Reviews of Amazon Echo. In *2018 IEEE SmartWorld, Ubiquitous Intelligence Computing, Advanced Trusted Computing, Scalable Computing Communications, Cloud Big Data Computing, Internet of People and Smart City Innovation.* IEEE, 372–380. https://doi.org/10.1109/SmartWorld.2018.00094

[11] Ido Guy. 2016. Searching by Talking: Analysis of Voice Queries on Mobile Web Search. In *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '16).* ACM, New York, NY, USA. https://doi.org/10.1145/2911451.2911525

[12] Florian Habler, Valentin Schwind, and Niels Henze. 2019. Effects of Smart Virtual Assistants' Gender and Language. In *Proceedings of Mensch Und Computer 2019 (MuC'19).* ACM, New York, NY, USA, 469–473. https://doi.org/10.1145/3340764.3344441

[13] Charles Hannon. 2016. Gender and Status in Voice User Interfaces. *Interactions* 23, 3 (April 2016), 34–37. https://doi.org/10.1145/2897939

[14] Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 9 (2006), 904–908. https://doi.org/10.1177/154193120605000909

[15] Sandra G. Hart and Lowell E. Staveland. 1988. *Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research.* Vol. 52. Elsevier, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

[16] Geoffrey Hinton, Li Deng, Dong Yu, George Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara Sainath, and Brian Kingsbury. 2012. Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. *IEEE Signal Processing Magazine* 29, 6 (2012), 82 – 97. https://doi.org/10.1109/MSP.2012.2205597

[17] Fabian Hoffmann, Miriam Ida Tyroller, Felix Wende, and Niels Henze. 2019. User-Defined Voice Commands, Display Interactions and Mid-Air Gestures for Smart Home Tasks. In *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia (MUM 2019).* ACM, New York, NY, USA, 7. https://doi.org/10.1145/3365610.3365624

[18] Jiepu Jiang, Wei Jeng, and Daqing He. 2013. How Do Users Respond to Voice Input Errors?: Lexical and Phonetic Query Reformulation in Voice Search. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '13).* ACM, New York, NY, USA, 143–152. https://doi.org/10.1145/2484028.2484092

[19] Jie Kang, Kyle Condiff, Shuo Chang, Joseph A. Konstan, Loren Terveen, and F. Maxwell Harper. 2017. Understanding How People Use Natural Language to Ask for Recommendations. In *Proceedings of the Eleventh ACM Conference on Recommender Systems (RecSys '17).* ACM, New York, NY, USA, 229–237. https://doi.org/10.1145/3109859.3109873

[20] Clare-Marie Karat, Christine Halverson, Daniel Horn, and John Karat. 1999. Patterns of Entry and Correction in Large Vocabulary Continuous Speech Recognition Systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '99).* ACM, New York, NY, USA, 568–575. https://doi.org/10.1145/302979.303160

[21] Alison Duncan Kerr. 2018. Alexa and the Promotion of Oppression. In *Proceedings of the 2018 ACM Celebration of Women in Computing (womENcourage '18).* ACM, New York, NY, USA. https://womencourage.acm.org/2018/wp-content/uploads/2018/07/womENcourage_2018_paper_54.pdf

[22] Nicole Kobie. 2019. Voice assistants may be less likely to understand women. *New Scientist* 242 (05 2019), 15. https://doi.org/10.1016/S0262-4079(19)30858-9

[23] Cole Christopher Maita. 2018. An Exploratory Study on Consumer Perceptions of Amazon Echo, Alexa, and Smart Speakers. https://libres.uncg.edu/ir/asu/f/Maita_Cole%20Spring%202018%20Thesis.pdf

[24] Chelsea Myers, Anushay Furqan, Jessica Nebolsky, Karina Caro, and Jichen Zhu. 2018. Patterns for How Users Overcome Obstacles in Voice User Interfaces. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18).* ACM, New York, NY, USA, Article 6, 7 pages. https://doi.org/10.1145/3173574.3173580

[25] Chelsea M. Myers, Anushay Furqan, and Jichen Zhu. 2019. The Impact of User Characteristics and Preferences on Performance with an Unfamiliar Voice User Interface. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19).* ACM, New York, NY, USA, Article 47, 9 pages. https://doi.org/10.1145/3290605.3300277

[26] Chidera Obinali. 2019. The Perception of Gender in Voice Assistants. In *Proceedings of the Southern Association for Information Systems Conference.* 6. https://aisel.aisnet.org/sais2019/39

[27] Jennifer Pearson, Simon Robinson, Thomas Reitmaier, Matt Jones, Shashank Ahire, Anirudha Joshi, Deepak Sahoo, Nimish Maravi, and Bhakti Bhikne. 2019. StreetWise: Smart Speakers vs Human Help in Public Slum Settings. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19).* ACM, New York, NY, USA, Article 96, 13 pages. https://doi.org/10.1145/3290605.3300326

[28] 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 (CHI EA '17)*. ACM, New York, NY, USA, 2853–2859. https://doi.org/10.1145/3027063.3053246

[29] Katyanna Quach. 2019. Voice recognition tech is naturally sexist. https://www.theregister.co.uk/2018/03/14/voice_recognition_systems_are_naturally_sexist/

[30] Valentin Schwind and Niels Henze. 2018. Gender- and Age-related Differences in Designing the Characteristics of Stereotypical Virtual Faces. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '18)*. ACM, New York, NY, USA, 463–475. https://doi.org/10.1145/3242671.3242692

[31] Valentin Schwind, Pascal Knierim, Cagri Tasci, Patrick Franczak, Nico Haas, and Niels Henze. 2017. "These Are Not My Hands!": Effect of Gender on the Perception of Avatar Hands in Virtual Reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 1577–1582. https://doi.org/10.1145/3025453.3025602

[32] Alex Sciuto, Arnita Saini, Jodi Forlizzi, and Jason I. Hong. 2018. "Hey Alexa, What's Up?": A Mixed-Methods Studies of In-Home Conversational Agent Usage. In *Proceedings of the 2018 Designing Interactive Systems Conference (DIS '18)*. ACM, New York, NY, USA, 857–868. https://doi.org/10.1145/3196709.3196772

[33] Aaron Springer and Henriette Cramer. 2018. "Play PRBLMS": Identifying and Correcting Less Accessible Content in Voice Interfaces. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173870

[34] Rachael Tatman. 2017. Gender and Dialect Bias in YouTube's Automatic Captions. In *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing*. Association for Computational Linguistics, Valencia, Spain, 53–59. https://doi.org/10.18653/v1/W17-1606