Using Remote Controlled Speech Agents to Explore Music Experience in Context

Nikolas Martelaro
HCI Institute
Carnegie Mellon University
nikmart@cmu.edu

Sarah Mennicken
Spotify
sarahm@spotify.com

Jennifer Thom
Spotify
jennthom@spotify.com

Henriette Cramer
Spotify
henriette@spotify.com

Wendy Ju
Cornell Tech
wendyju@cornell.edu

ABSTRACT

It can be difficult for user researchers to explore how people might interact with interactive systems in everyday contexts; time and space limitations make it hard to be present everywhere that technology is used. Digital music services are one domain where designing for context is important given the myriad places people listen to music. One novel method to help design researchers embed themselves in everyday contexts is through remote-controlled speech agents. This paper describes a practitioner-centered case study of music service interaction researchers using a remote-controlled speech agent, called DJ Bot, to explore people's music interaction in the car and the home. DJ Bot allowed the team to conduct remote user research and contextual inquiry and to quickly explore new interactions. However, challenges using a remote speech-agent arose when adapting DJ Bot from the constrained environment of the car to the unconstrained home environment.

Author Keywords

Remote user research;
interaction prototyping; speech agents; music experience

CCS Concepts

•**Human-centered computing** → **User centered design; Interaction design process and methods;** *Ubiquitous and mobile computing design and evaluation methods;*

INTRODUCTION

Observing user interactions with products and services in the real-world is crucial to designing these products and services to be useful, usable, and meaningful [22]. As interactive systems permeate the myriad contexts of people's daily lives, there is a need to support user researchers and designers in observing and understanding how people interact with these systems in-the-wild.

DIS '20, July 6–10, 2020, Eindhoven, Netherlands.

ACM ISBN 978-1-4503-6974-9/20/07 ...$15.00.
http://dx.doi.org/10.1145/3357236.3395440

Intelligent and personalized music services and the devices they are accessed through is one such application used throughout many places and times of people's everyday lives. Music is enjoyed in the home, the workplace, the car, the gym, and on-the-go through devices such as personal computers, phones, web-connected radios, and voice assistants. Additionally, while most music services and devices are currently designed for general use, there are challenges to designing domain specific applications suited for specific environments or user needs [28]. To design music services, devices, and interactions that fit well into people's lives, music service designers need ways of understanding and exploring the varied contexts where their products will be used.

Various user research methods, such as contextual inquiry [5], experience sampling [12], and remote-user testing [16], have been employed by user researchers and designers to understand the in-context experience of their product's users. However, there are practical challenges that product and service designers and in particular, music service designers, face when trying to understand the real-world and contextual experience of listening to music. People interact with their music services in mobile spaces, in intimate spaces, and at times of the day where it would be impractical for user researchers to be in person. In addition, the experience of listening to music can be highly personal and private such that in-person observation of listening behavior can be disruptive to the user or alter their experience.

Capturing real-world system logs along with location data can give music service designers some sense of how people interact with the service in context, but usage data alone only shows a limited view of the user's full context and behavior. Log data also requires analysis post-experience and may not expose context-specific issues or opportunities for more exploratory design work that is possible when designers and researchers have a rich, in-the-moment view of the user's interaction.

The challenges of having a delayed and limited view into the real-world user experience of products and services can lead to long feedback loops between design, deployment, and evaluation of prototypes, ultimately reducing the rate that designers can learn about and improve their systems [11]. Carter et. al. suggest that "*designers and researchers must be able to close the interactive design loop, encompassing both prototyping and evaluation,*

and learn from their prototypes" [11, pg. 1]. They also suggest focusing on the development of tools and methods to support design practitioners so that their research and ideas can move to the real world. How then, can music service designers and researchers overcome the temporal and spatial challenges of understanding real-world usage of products and services while also having a rich in-context view of the user's experience and the ability to quickly test interactions?

Recently, remotely controlled conversational agents are being used to address the challenges of providing user researchers and designers with the ability to understand users in context while interacting with interactive systems. For example, Broadman and Koo of IDEO have explored how designers conversing with users remotely throughout their day through phone-based chatbots were able to conduct needfinding on and prototyping of new services for personal fitness, public services, and disease outbreak [7]. Martelaro and Ju used a remote-controlled speech agent along with data capture and video streaming systems to explore driver's experience with advanced driving assistance technologies while on their daily commutes [26]. More recent work by Martelaro and Ju more fully developed the idea of using speech agents to allow designers and researchers to do remote needfinding and interaction design exploration in context. They have also pilot-tested the use of remote-controlled speech agents to understand people's experience with music while driving [25, 27]. In addition, both Bentley et. al. [4] and Ammari et. al. [1] observe that a main use case of voice assistant usage is music listening. Taken together, these findings suggest that designers and researchers acting through a remotely controlled speech agent can be a usable method for understanding people's experience with music services in varying contexts.

To further explore how speech agents can be used as design research tools for contextually situated music services, we extend the work of Martelaro and Ju [27] with a practitioner-centric and reflexive case study of interaction researchers from Spotify using a speech agent to explore people's experience with music in two contexts: the car and the home. To do this, we modified and tested DJ Bot [25], a remote controlled speech agent and observation tool built alongside a music playing service. Our goals in testing DJ Bot were to see how professional music service interaction researchers could incorporate using a remote controlled speech agent into their user research and interaction prototyping practice. The interaction researchers were interested in a remote controlled platform, such as DJ Bot, to observe how people listen to music in context, such as the car and in their home. DJ Bot's use of a speech agent also provided a conversational way for the researchers to probe listeners in the moment.

The primary focus of this paper is to describe how the Spotify team used DJ Bot and to discuss what worked well, what didn't, and how user research through remote speech agents could be useful to other design teams. We first describe how the in-car and in-home deployments were implemented. Then we discuss the benefits of using a remote-controlled speech agent to explore in-context interaction along with the challenges of acting through a remote speech agent, modifying the method for use in the home, and incorporating the method into professional practice. We then reflect on where the method is currently applicable and where it can be further developed. Overall, our experiences suggest that music service practitioners and interactive system designers in general, can interact with users through speech agents to perform rich in-context exploration. However, considerations should be made to facilitate efficient user observation and real-time interaction prototyping.

BACKGROUND

One of the goals of the Spotify interaction research team is to understand people's music listening experience in different contexts. As of December 2019, Spotify has over 271 million monthly active users across 79 international markets [34]. While music listening behavior can be studied on a large scale given digital music services [32], quantitative analysis alone cannot capture the richness of people's interaction with their music, especially in the varied contexts where people listen. Music is consumed in many places such as the home, car, gym, and workplace, often as a secondary or background activity throughout the day [18, 19, 20, 31]. Music is also experienced on many devices; the Spotify player is currently supported on desktops, TVs, voice assistants, cars, mobile phones, and gaming consoles [34]. Aspects of personal context, such as someone's current activity, emotions, motivations, and personal preferences, also factor into music listening behavior [18, 19, 20, 31]. As music permeates so much of our lives and activities, researchers argue that listening behavior should be studied in context to understand the true nature of people's music experience [18, 19, 20, 31].

Due to the context-dependence of music listening, researchers interested in understanding listener experience to inform the design of new products and services will often engage directly with users through qualitative user research methods. Prior work on user research methods and Wizard of Oz prototyping methods can inform the development of new research methods for understanding users in context. In the following sections, we discuss these methods and describe how they inform the features music system designers may want and the limitations they may overcome by using a remote-controlled speech agent to conduct in-context user research.

User Research Methods

Much of the foundation of professional user research methods is built upon direct observations and interactions with people. Researchers often employ *rapid ethnography* to quickly learn about user needs and inform their product and service designs [29, 33]. Researchers also conduct contextual inquiry sessions [5] where they observe and interview people interacting with a technology in their real-world context. Mixing observations with user interviews is critical in understanding how users actually behave [22, 30] but can be challenging to accommodate in highly personal contexts such as one's car or home.

To overcome the challenges of not being able to physically observe people's experience in different contexts, researchers have used mobile phone-based experience sampling [12, 17] to query users about their experiences with technology in-the-wild. Experience sampling has often been used by music researchers to capture a more complete picture across the various places in which users listen to music [18, 20, 31]. While experience sampling does allow researchers to capture a user's thoughts on their experience in-the-moment, it does have some limitations. Random experience sampling can miss when people are listening to their music, limiting researcher's ability to capture user's thoughts at the right

time. Experience sampling also requires researchers to define what to explore ahead of time rather than providing a live view of a user's interactions. Even when providing the user with the ability to send images and videos of their context, such as with Carter et. al.'s Momento system [10], experience sampling often does not let remote researchers see the full user context. This can potentially lead them to miss important contextual factors that they had not thought about ahead of time. Finally, text-based experience samples can be hard for users to engage with in specific contexts such as when driving their car or cooking at home, limiting the moments that researchers can receive timely answers to their queries.

The expansion of wireless connectivity nearly everywhere has enabled researchers to conduct remote screen-sharing user observations [6] and test the usability of mobile applications in the field [9, 35]. Although the remote researcher is not co-located with the user, these remote methods can be just as effective as in-person methods and can even have positive benefits by avoiding observers biasing users [2, 8]. While these methods allow for remote researchers to engage with users in-context, it can be challenging when they are not using a screen-based interface such as a laptop or mobile phone. When looking to explore music consumption in environments like the car or around the home, we may need better ways than screen sharing for observing and engaging with users, particularly as the devices employed (e.g. smart speakers) may not have a screen nor require screen-based interaction.

Speech-based interfaces can help to overcome some of the challenges of remote observation methods that utilize experience sampling and screen-sharing. Today, music interaction is one of the most common uses of voice assistants [1, 4] and extending voice control to conversation about music is now believable for everyday users. Speech-based interaction can also allow researchers to ask questions and get answers from users who are doing other activities such as driving or folding laundry. Furthermore, using a speech agent instead of asking questions directly may reduce the researcher's influence on the overall interactive experience, helping avoid a common challenge with conducting research in-the-wild [11].

Wizard of Oz Methods for Interactive Systems

Employing a remote-controlled speech agent for user research is inspired by Wizard of Oz studies whereby a human operator simulates an intelligent system before significant resources are spent on technology development [13]. Wizard of Oz methods have long been used throughout the design process for multimedia applications, from early experience explorations to evaluations of final designs [14]. Systems such as Klemmer et. al.'s SUEDE [21] focus on exploratory investigations of speech interaction by allowing non-programmer designers to construct speech interfaces and capture data about the interaction. A major benefit to SUEDE is the focus on interaction design over speech recognition and natural language understanding capabilities, allowing designers to test earlier in the design process with more people who may not be well understood by modern speech recognition due to their accent or language.

Other systems such as Topiary [23] and DART [24] allow for Wizard of Oz prototyping of location-aware applications. Wizards using these tools would follow along near users testing an application, such as in the "Voices of the Oakland Cemetary" project, where a wizard trailing a user triggered audio cues on the user's AR interface [15]. As wireless networks have increased in bandwidth and speed, researchers are now able to blend remote user research conducted by one or more researchers with Wizard of Oz prototyping, such as in Martelaro and Ju's "WoZ Way" system for prototyping in-car interfaces [26]. Recently, Martelaro and Ju proposed a system using a remote-controlled speech agent for exploring in-context music recommendations [25] and conducted a pilot test with the system [27].

We have adapted this system to see how practicing interaction researchers at Spotify can use a remote-controlled speech agent to understand how listeners interact with music. Our goal with this case study is to explore how the method works when employed in the previously tested context of the car and to extend the method to another environment such as the home. We also aim to inform practical applications of this method by learning what real-world challenges arise when multiple researchers conduct remote user research with a speech-agent.

DJ BOT

DJ Bot is a remote-controlled speech agent embedded alongside the Spotify music application within a user's everyday life. For our case study, the interaction research team explored two contexts, the car and the home, as these are the most common places for people to listen to music [18]. DJ Bot includes video and audio streaming for remote observation, a control interface for Wizard of Oz speech interaction and music control, and datalogging. During our research, we made iterative updates to the system based on reflections from the interaction sessions and discuss the final implementations here.

Car Implementation

DJ Bot in the Car (Figure 1a) is built using the "WoZ Way" system described by Martelaro and Ju for in-car interaction prototyping and observation [26]. In our implementation, a laptop (Apple MacBook Pro) in the back of the vehicle speaks out messages sent by the remote interaction team. The laptop streams live video and audio using video chat software (Google Hangouts) via a 4G cellular router. Up to four high-definition cameras are placed around a user's vehicle and are combined with a video mixer, providing a view of the driver's face, road, and cabin. These camera streams provide the interaction wizards with a view of the road and the driver for context awareness and safety, as recommended in [26].

Participants play music on their car stereo through their phone using the Spotify mobile app. This provides a familiar interface for listening to music in the car and allows drivers to use their

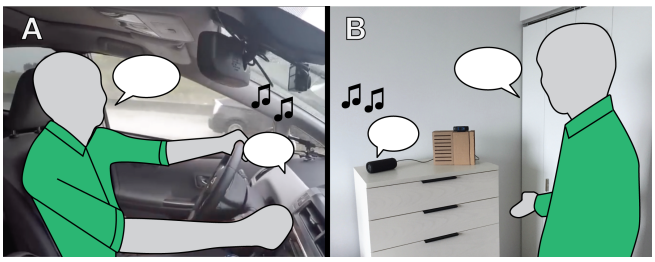

Figure 1. DJ Bot in the (a) Car and (b) Home. Remote user researchers interact with users through speech agents in two locations where music is often a common secondary activity and where voice interaction is preferable.

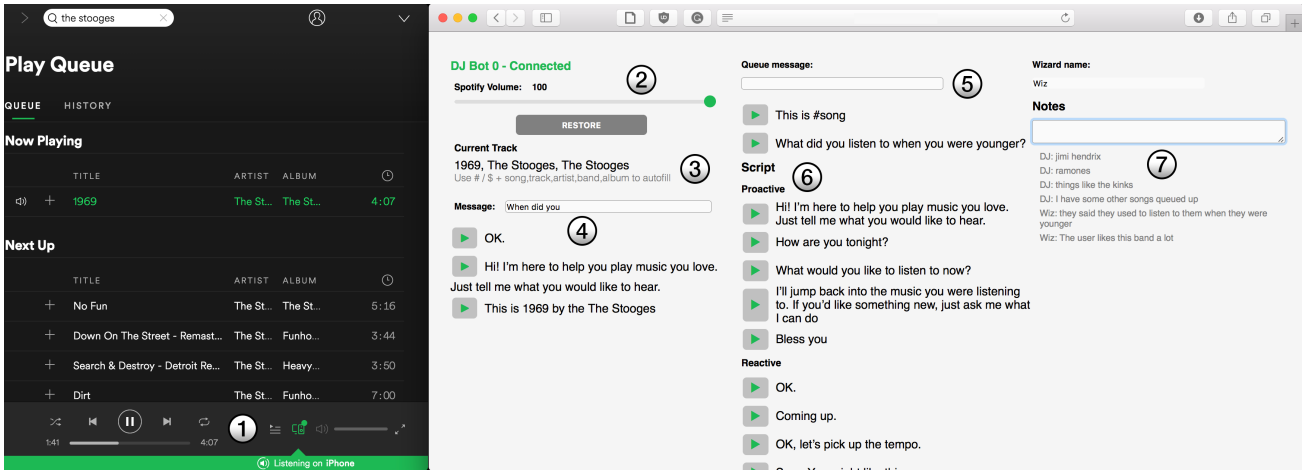

Figure 2. Remote control interface. (1) Music control through Spotify desktop app, (2) Volume control, (3) Live song data, (4) Custom message input, (5) Queued message input, (6) Scripted speech messages, (7) Collaborative notes.

vehicles' built-in controls on the dashboard and steering wheel to manually control their music. A separate speaker is used for DJ Bot's voice which is rendered using MacOS's on-board text-to-speech system. We opted to not give DJ Bot a physical form since there was no need to have any physical display beyond what was provided by the car's interface. Our intent in the car was to only have drivers interact with DJ Bot through voice as this is safer than a physical interface [3].

Home Implementation

DJ Bot in the Home (Figure 1b) uses the same underlying software as DJ Bot in the Car, however resembles a desktop radio. A small computer (Mac mini) is encased in a stylized cardboard box made to look like a radio. The use of cardboard was intended to signal that the device is a prototype and not a finished product. A single high definition webcam with microphone is mounted to the top of the box. This camera is placed prominently so that the user is fully aware they are being recorded and so that the user can easily see when the camera is on/off via a blue light. A separate speaker plays the music and any synthesized speech. This is similar to Amazon's Echo Dot, where the device contains the intelligence and the user can choose their own sound system to play through.

For the in-home implementation, we gave DJ Bot a simple physical form that evoked a radio or speaker so that users would understand that the device was intended for listening to music. We opted to not include physical control interfaces on the DJ Bot body to promote voice-based interaction and to allow users to control the music volume through the speaker of their choice. We used only one camera mounted on DJ Bot as this contains the visual field that the user needs to consider when placing DJ Bot in their home. While it would have been advantageous to place cameras throughout the user's home to understand where they were when they interacted with DJ Bot, we wanted to respect users' privacy. Furthermore, having a single camera on DJ Bot presents a more realistic capability that a smart speaker could have.

Control Interface

The Spotify team controls DJ Bot's speech using a web-based interface and the music using Spotify's desktop application. We take advantage of Spotify's "Connect" feature which allows any instance of Spotify to control another instance that is online and on the same account. The web-based control interface, shown in Figure 2, includes volume control, a free response input for speech messages, a queuing area for speech messages, pre-scripted speech messages, and an input area for notes. All messages and notes are sent from the web interface to DJ Bot using an MQTT communication broker.

The Spotify team controls DJ Bot collaboratively. One researcher controls the speech while another controls the music. The research team is geographically distributed with one member in San Francisco and another in Boston. Both team members can access the web interface. A separate audio "backchannel" allows the researchers to talk to each other during the session. Figure 3 shows a diagram of the system's communication channels. Video is recorded locally in the user's location and remotely using a screen recorder. Audio from the in-context DJ Bot interaction and the researcher backchannel is recorded on separate audio channels.

CASE STUDY METHODOLOGY

Our experience using DJ Bot was situated within a larger effort at Spotify on gaining a rich understanding of music listeners within their daily lives. The research team for this case study is made up of two academic researchers and three members of the Spotify research organization. The academic researchers assume the role of meta-researchers, exploring the interaction research team's use of DJ Bot. The interaction research team at Spotify focused on

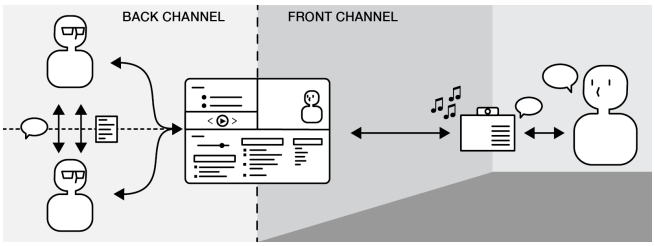

Figure 3. DJ Bot communication channels. Front channel - user interaction with DJ Bot. Back channel - interaction with control interface and conversation between researchers via voice and text notes.

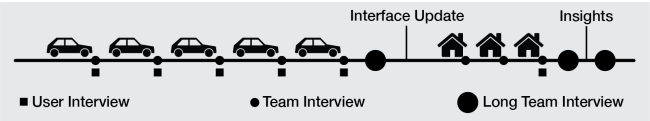

Figure 4. Research timeline. The Spotify team conducted five in-car studies with individual users and three in-home sessions with one user. A meta-researcher conducted two long interviews with the research team, then derived a set of insights to review with the research team.

using DJ Bot to observe how listeners might request music from a speech agent in context.

As this project was situated in a professional context, we utilized an iterative design approach to exploring the interactions with DJ Bot. We used observations and semi-structured interviews to understand how the research team used DJ Bot throughout the project. All interaction sessions and interviews were recorded and then transcribed by an online transcription service. Figure 4 shows our research timeline with interaction sessions and interviews labeled. We conducted five on-road sessions with DJ Bot in the Car and one week-long interaction with DJ Bot in the home. The first author, acting as a meta-researcher, conducted observations of the interaction research team as they used DJ Bot with listeners. The author took notes during these sessions and reflected on them in order to facilitate the semi-structured interviews with the Spotify team.

After each of the in-car sessions, the Spotify team conducted a short interview with the listener around their interactions with DJ Bot. The meta-researcher then followed up the listener interview with a short interview with the Spotify team focusing on their use of DJ Bot during the session. After all the in-car sessions were completed, a long interview was conducted exploring the Spotify team's experience as a whole.

Based on insights from the in-car sessions, we updated the DJ Bot control interface for use in the home. We conducted a set of pilot home studies within the group in order to improve the design of the control interface. We then engaged with one user for three interaction sessions over the course of one week. The meta-researcher conducted a short interview with the Spotify team after each interaction session. After the last in-home session, the Spotify team conducted a short user experience interview with the participant. The meta-researcher then conducted a long interview with the Spotify team focusing on their use of DJ Bot for the in-home sessions. We then conducted a second larger interview reflecting on both the in-car and in-home DJ Bot experiences. Finally, the first author reviewed the recorded sessions and interview transcripts and prepared a set of insights, benefits, limitations, and best practices to present to the Spotify team. The entire research team then collaborated to reflect on and refine these insights during the preparation of this case study.

CONTEXT 1: MUSIC IN THE CAR

The car is one of the most common places people listen to music [18] and listening is contingent on the driving activity [36]. Listening to music is also one of the only forms of entertainment that people can enjoy while driving. Since the driver can focus more on their music during a drive, it provides a prime environment for exploring how people think about and engage with their music.

The car also provides an environment where written methods of experience sampling cannot be used but where people can have a verbal conversation [36]. It should be noted that any secondary activity such as music listening or having a conversation can be distracting to the driver. To mitigate risk, we followed recommendations from previous remote in-car studies [26] by providing the DJ Bot wizards video of the driver and road and telling them to only interact with drivers during less engaging moments of driving. Drivers were reminded to drive safely, avoid interacting with their phones manually and told that they could stop interacting with DJ Bot at any time.

Users

We invited five users (3F, 2M) with a range of music tastes and driving behaviors to interact with DJ Bot. Participants were recruited through a large research university and were compensated $50 USD for their time.

1. Male, mid 20's. Graduate student. 60-minute drive on freeway during the afternoon.
2. Female, late 20's. Lab manager. 70-minute commute home during the afternoon.
3. Female, mid 30's. Lab manager. 45-minute commute home during the afternoon.
4. Female, mid 30's. Administrator. 45-minute commute home during the afternoon.
5. Male, mid 60's. Self-employed. 70-minute drive on freeway during the afternoon.

Procedure

The remote design team interacted with and observed drivers using DJ Bot during afternoon drives lasting 45–70 minutes. Three drivers drove their own vehicles on their everyday commutes from work to home. Two drivers drove a research vehicle back and forth along a scenic highway.

DJ Bot was setup in each driver's car 30 minutes before they departed. During setup, drivers reviewed and signed an informed consent form. Drivers were told to only interact with DJ Bot using speech and to first and foremost drive safely. Beyond that, drivers were encouraged to behave normally. The experimenter then turned on the cameras and asked DJ Bot to say "*hello*," signaling to the remote Spotify team that the driver was ready to leave.

The Spotify team interacted with the driver using the DJ Bot text-to-speech system and controlled the music selection. The remote interaction researchers observed the live camera feed from the road, took music requests, asked the driver questions, and answered questions from the drivers. During each session, the researchers loosely followed a prepared script with various questions about listening to music and different prompts for suggesting how to guide DJ Bot to the right music. Examples of these scripted interactions include:

- What do you like to listen to while driving?
- What did you listen to when you were younger?
- And now for a classic driving song! Go ahead and steer me until you find something you like.

- Describe something you'd like to hear right now. Like "play me something from England with guitars."

Interaction Dynamics

The DJs often used the context and user's preferences to guide their conversation. Across the five sessions, DJ Bot made an average of 70.6 utterances (SD = 25.2) while Drivers made an average of 80.8 utterances (SD = 33.0). DJ Bot spent about 4.1% (SD = 0.6%) of the time on the drive speaking, while Drivers spent on average 8% (SD=4%) percent of the time on the drive speaking. In general, drivers spoke for more of the drive than DJ Bot. However, drivers (2) and (4) spoke for less than 5% of the drive. These drivers treated the interaction in a more transactional manner, mostly making requests for songs to be played, similar to how currently available speech agents interact with users. Drivers (1),(3) and (5) spent more time speaking during their drive, speaking 12.6%, 7.8%, and 12.1%, respectively. Throughout the sessions, drivers often spent time requesting music, exploring new music, talking about music for driving and talking about memories associated with music. Of these, most of the conversation was centered around exploring new music and talking about memories of music. For example, drivers often spoke for long stretches when responding to the question "What did you listen to when you were younger?"

What worked well during the in-car sessions?

The researchers noted that it was impressive how well the remote interaction worked, both for interacting with the driver and for interacting with each other. In general, the researchers were able to work as a team to observe and interact with the driver and could coordinate their interactions in the backchannel well even though they were located in different offices. They were also able to easily tailor their questions and music recommendations based on their conversation with the driver.

Another aspect of the system that worked well for the researchers was the video of the driver's face. Due to the constrained nature of the car, the researchers could always see how the driver was responding to the questions or recommendations from DJ Bot. The team relied on the video to gauge the driver's feeling toward a song or line of questioning, helping the team to adjust their interaction with the driver in-the-moment.

Finally, many drivers enjoyed the interaction they had with DJ Bot. The research team used DJ Bot's flexibility to try out slightly more conversational styles of interaction, telling jokes or providing information about bands. For example, in one session, the researchers told a driver about an upcoming Beyoncé concert, prompting the driver to say "*I need tickets! Thank you for telling me.*" Most drivers felt comfortable having more conversation with DJ Bot and felt that the conversations were a nice addition to the music experience.

What didn't work well during the in-car sessions?

One major issue with the interactions was the impact of lost connectivity to DJ Bot in the car. While the network coverage was good throughout most of the drives, all drives experienced at least one full signal loss where the remote researchers could not see the driver or control the music. During these scenarios, drivers would often ask to change the music without any response from DJ Bot. While this led to subpar experiences where drivers could not change their music and would verbally check on DJ Bot, it did reveal to the research team that they should consider strategies to inform the driver of a connectivity failure and indicate that the speech agent would no longer respond. While these challenges made the user research hard, they also lead the team to reflect on how fully online experiences would (not) function. The research team discussed how if you lose connection during the study, or you don't understand the user and you have to recover, you will also have to deal with those connectivity and unavoidable errors in the real world.

The researchers also had challenges structuring their interaction based on how much of the drive was left. This led to the researchers not getting to all of the questions in their script and having to abruptly end their interaction once the driver reached their home or the end of the drive.

Another challenge during the in-car sessions was managing the perceived intelligence of the speech agent. In some cases, the team could not understand the drivers or had issues finding a good song, making DJ Bot seem inept. This led to the researchers focusing their efforts on finding a song and asking the driver to clarify rather than asking the driver about their connection with the music. At other times, DJ Bot was perceived as too smart, breaking the guise of a speech agent and leading people to focus on the fact that they were in a test.

Finally, there were a number of pragmatic challenges with interaction through the remote speech-agent. The music volume was often too high to hear the driver's requests or answers to DJ Bot's questions. Typing out everything DJ Bot said was slow and lead to long delays in the interaction. Capturing interesting moments and analyzing them after the interaction sessions was too time consuming. This led to the captured data not being as useful for fast analysis or for quickly sharing results with colleagues.

CONTEXT 2 - MUSIC IN THE HOME

After the in-car studies, we explored how well using a speech agent would work in a context outside of the car. Studying people in their homes would provide a different view into a user's life and is another area where people often listen to music during their day [31]. The home also presents different contextual challenges from the car, such as people moving around, groups of people listening to one device, and more primary activities that someone could be doing.

Interacting with users in their homes also allowed us to explore how well a remote-controlled speech agent would work for longitudinal studies. While the in-car sessions were one-time interactions, we planned for the in-home sessions to include multiple interaction over the course of a week. Due to the logistical challenges of planning an in-home session, the team decided to focus on interacting with one user.

DJ Bot Improvements

Based on our interactions and challenges with DJ Bot in the car, we decided to iterate on the design of the DJ Bot control interface. Specifically, the music volume was often too high to hear the driver, answering questions about the music or typing out questions was difficult and time consuming, and the difficulty of reexamining data after a session using only video or speech transcripts lead to insights being forgotten. The interface was

updated to include automatic and manual volume control, a direct connection to the Spotify desktop application data including song, artists, and album, a queue for messages to be sent in the future, and a collaborative note taking section. The interface was visually redesigned to be easier to navigate. We also improved our data logging system to save all DJ Bot speech and notes directly on DJ Bot, allowing for easy time synchronization with DJ Bot's on-board video recording.

User

We invited one user to interact with DJ Bot in her home for one week. This user was a Spotify customer and Amazon Echo voice assistant owner living in a shared apartment in a large city. She was recruited using an online user testing recruitment service and was compensated $300 USD for three one-hour interactions over a week plus a semi-structured interview after the interaction sessions.

Procedure

One member or the research team visited the user's home for 30 minutes and setup DJ Bot. The researcher explained that the goal of the study was to explore in-home interaction with music and had the participant sign an informed consent form. The participant was told that when the remote team logs in to the system, they will have access to the camera and the audio stream. With this in mind, the participant chose a location to put DJ Bot. In this case, the user decided to place DJ Bot in their bedroom looking across the room at a window.

The team planned three one-hour interaction sessions. The team emailed the user a day in advance of each session with questions for the user or tips on using DJ Bot. These included requesting relaxing or energizing music before the second session and asked about possible upcoming activities or calendar events before the third session. An hour before each session, the team emailed a reminder to the participant to ensure they were aware of the scheduled recording time. At the start of the schedule interaction session time, the team logged into DJ Bot, turned on the camera, and started data recording. The team then interacted with the user for an hour and a half. After the session, the team stopped the recording systems, turned off the camera, and sent an email confirming the conclusion of the session.

Interaction Dynamics

During the sessions, DJ Bot spoke 31, 21, and 13 times respectively. Overall, the user treated their interaction with DJ Bot in a transactional manner, requesting or skipping songs. On a few occasions, the users asked for facts about an artist. The user did try out different ways of requesting music based on the email prompts; however, there was little conversation about the music. The team noted the user's activities, which were myriad: cleaning their room, folding laundry, taking phone calls, and talking with other housemates.

What worked well during the in-home sessions?

Being able to interact with the user for a longer period of time allowed for the research team to get to know the user's taste better. This made their interactions with the user easier as time went on. The researchers also valued seeing the user's genuine reactions to errors made by DJ Bot. Because the user treated DJ Bot more like currently available voice assistants, the team was able to see the user sigh and appear displeased with poor music choices or unwanted questions.

What didn't work well during the in-home sessions?

While users in the car could only drive and listen to music, in the home, the user was often doing other tasks and not focusing as much on their music. This made interactions with the user much harder and limited the interaction that the researchers could have with the user. Pragmatically, the single video camera often made the visual feedback useless for the researchers. The user was free to move around their home and still listen to their music. This meant that the researchers often could not see what the user was doing or how she was reacting to the music or questions. This led to long, boring stretches of time where the researchers were simply looking into the user's room but not getting any feedback. Finally, the researchers also had issues hearing the user when she was too far away from DJ Bot.

BENEFITS AND CHALLENGES OF USING A REMOTE-CONTROLLED SPEECH AGENT FOR IN-CONTEXT MUSIC EXPLORATIONS

After reviewing the eight short post-session interviews (5 car / 3 home) and three long team interviews (1 post car sessions, 1 post home sessions, 1 overall), we compiled a set of benefits and challenges that the team experienced while using DJ Bot.

Interacting through a remote speech agent overcomes challenges of space and time

Overall, interacting through a remote-controlled speech agent allowed the research team to overcome the main issue of being in the same space as the user during times where they would listen to music. The research team could be in more intimate spaces such as the car and the home without needing to leave their offices. Interacting remotely allowed the team to collaborate in a way they had never been able to do before. One researcher noted "*It is, also, good for teams that, like us, are geographically not co-located. That has been a tremendous difference because if we had a Wizard of Oz set up that would just work in Boston or San Francisco, that would be much harder. This way, it made it really easy to be in three different places and still all be in it.*"

The flexibility of interacting through a speech agent is better for more exploratory work

The speech agent allowed the interaction research team to explore more open-ended ideas than they would if they were testing with an automated system or asking users questions via a survey. By acting as the speech agent, the team could work improvisationally, asking the participants questions based on the situation at hand and following up on the user's responses and requests immediately. A researcher stated, "*you can have it say whatever you want. You can have more exploratory type of conversations - surprising types of conversations.*" The team commented positively about the additional range of interactions they could have with the user: "*Depending on how exploratory [you] would want to be, you can make it more or less structured.*" The team spoke about how the early in-car sessions were open-ended and constructive. The team found that after having developed a plan and many scripted speech interactions, they could focus their inquiry more while still being flexible. One researcher likened this to jazz improvisation, where there is an underlying structure but also freedom to alter their interaction based on the situation.

Being behind a bot enables a more realistic interaction

The Spotify team discussed how using a speech agent in an everyday environment allowed for rapid exploration of the contextual factors that influence music listening. One researcher noted, "*By being remote and by having the synthesized voice, that is very useful to let the user keep the illusion of their real context. I think that is really good. Maybe that is why it is so strong to figuring out new contexts or new setups.*"

While DJ Bot arguably introduces another actor into the context, the team perceived this interaction to be less disruptive than if they were present in person. It reduced the influence an in-person experimenter would have, and, as a result, reactions were more natural. For example, during the in-car sessions, users would sing along to their favorite songs or yell at other drivers as if they were alone. During the home session, the user would audibly sigh about poorly chosen songs but did not mention their displeasure about those same songs during the post-interview with the research team. As noted by one of the researchers, "*in the second session, she sighed a lot, you know? And we were just like, 'Yeah, that doesn't seem like she's super enjoying things', like, what we just heard. This allows, really, to capture some genuine reactions, versus, you know, the socially more acceptable response that she gave to us when we were interacting with her as humans (during interviews), and not as DJ Bot.*"

Getting to know users through a speech agent can make follow-up interviews a bit awkward

The live interaction through the speech agent helped the researchers to quickly learn about the user. However, the speech agent mediating the interactions between researchers and the user made their connection with the user a bit one-sided. While the researchers felt like they knew the person after the interaction session, the user knew nothing of the researchers. One researcher noted, "*One thing that I found really interesting was, because we could see the participants, and we felt like we know so much about them, and then you see them in the interview, and that's the first time when the see you. So it felt kind of creepy to follow up on things that you observed before, because for those people for whom it worked, that they considered it being a bot, for those, it was a little bit awkward.* " This suggests that researchers should be aware that the interaction can be one-sided and then follow-up interactions might include rapport building.

Seeing the user is just as important as hearing them

Through video, DJ Bot provided information that the team could not typically capture during an in-person interaction. In the car, the video of the driver and the road provided a unique view of a driver's behavior with their music with respect to the road. The team perceived this to improve their understanding of the environmental context for driving.

In contrast, not seeing the user in the home due to the single camera and the unconstrained environment affected the team's ability to gain a broader sense of the overall home context. The researchers remarked on how they preferred seeing "*how people interact when they are sitting in their car and we can observe them the entire time"* whereas with DJ Bot in the home the researchers realized "*that sitting there trying to watch them when they are not even in the field of view might not make for the best use of two researchers time.*"

It is challenging to try to be a bot

On one hand, the Spotify team felt that their direct interactions through DJ Bot were not that different from having a conversation or doing contextual inquiry with a user. On the other hand, the team found it challenging to act like a bot and to simulate intelligent capabilities that are better done computationally. In the context of music, it's impossible to wizard the personalization of automated song selection based on sophisticated user models. One researcher remarked "*From a Wizard of Oz standpoint, you can't go with the really great answers, right? You're not smarter than the machine because you don't know the genre (of music). So that's made it interesting for me. There's this tension between I wanna do it right, and give the right recommendations, and hey, we're doing this automatic thing, as if we're pretending to be an automatic system.*" Functionally, having to type most of DJ Bot's utterances made the interaction more stilted, though most of the users did not find the delays to be a significant problem.

The team discussed how they worked to act more "bot-like" during the sessions to avoid seeming too smart or overly conversational or making their presence apparent which could potentially break the context. For example, the researchers would discuss via the backchannel what interactions would fit within the situation. In some cases, users did feel that the speech agent guise was broken, though it was not apparent that this changed people's behavior much during the session.

Interacting through a remote speech agent can allow for in-the-moment interaction design exploration

The researchers commented that one of the most interesting things about interacting through DJ Bot was that it combines in-the-moment designing of new interactions (via elicitation of user needs through the speech agent), with functionality (search, personalization etc.) that already exists. This means that it becomes possible to explore changes in how design choices may impact the reactions of people to adaptive system, in-the-moment and in different contexts. As part of another project, the researchers were exploring how different (non-)decisions made by interaction design teams could impact machine learning outcomes. The researchers suggested that conducting user research through DJ Bot could help them understand the impact that content recommendation choices and interaction design choices had on the individual. In our case study, the researchers could also see how individuals might respond to similar kinds of interaction across different contexts, helping the researchers understand what contextual factors might play into future design decisions.

Being behind a speech agent allows for collaborative reflection among the team

A recurring theme that came up in our interviews was how interacting through DJ Bot allowed the team to reflect on the interaction sessions in-the-moment and throughout their research process. The team heavily used the audio backchannel to discuss what they were seeing and learning about the user and context. The team also used the backchannel to plan their next actions based on the current situation. This allowed them to conduct very short cycles of questioning and testing followed by discussing what they saw, what they might do next, and whether their approach had to be adjusted to yield better insights. One researcher said, "*we could have this conversation about 'Are we getting this out of this? Yes? No?'*" This opportunity to directly interact with

the participant while being able to follow up with each other is like a think-aloud about their process of conversing with the user. For example, during the first in-home session, the team had developed an interaction plan, but halfway through, realized that their plan was not working out as expected. The team discussed what to do and formed a new plan which they used for the rest of the session. The team described this planning as similar to lab-based user tests, where in-person researchers talk with the observing team 'behind the glass' during breaks or after a session. By interacting through a speech agent, the research team can change course without interrupting or waiting until after a session. All members of the team could influence the interaction as the speech agent represented everyone, changing the dynamic from more traditional user research where one team member acts on the team's behalf.

Active note taking supported researcher reflection and learning

The researchers also commented on how useful the collaborative notes were while interacting through DJ Bot in the home. The notes created another communication channel where the researchers could comment on important events about the user and capture their thoughts. "*The home sessions were more collaborative than the car ones. I think we did more reflection during the sessions. We talked a lot over the notes, and the notes facilitated a lot of that.*" Participating team members could see the other researchers' thoughts and follow up with them on the voice backchannel.

In the home sessions, all the notes were sent to the DJ Bot and time synchronized with the DJ Bot speech interactions, song information, and video. With these notes, the team could quickly review the notes and plan for future sessions. The notes "*allowed us to create this timeline of how our thoughts evolve, because sometimes we're like, 'Oh, maybe next time we do that,' and then we just put it in there in the note, and then we see – we can kind of recall why this is happening, even if we don't write down what's happening.*" Furthermore, the notes also helped the researchers to capture learnings from their sessions and overcome one of the challenges with processing such a large amount of data after the interaction session. One researcher said that the notes "*just made it much easier than having to do this annoying tedious task of annotating the video file afterwards.*"

DISCUSSION

Comparison with other user research methods

Interacting through DJ Bot had similarities and differences with other types of user researcher methods. Remotely interacting with the user through a speech agent is similar to other methods of remote user testing [9] and has similar benefits such as limiting the effect of the researcher's presence on the user, as evidenced by most users showing genuine reactions to DJ Bot. Having the live video streams gave the researchers the ability to do a form of rapid ethnography [29, 33] remotely.

The work in this case study primarily builds upon Martelaro & Ju's "WoZ Way" in-car remote observation and interaction prototyping system [26]. As with this previous work, the Spotify researchers found that in the constrained environment of a car instrumented with multiple video streams, they could get a sense of the context and focus on the user and their interactions with the context and the speech agent the entire time. However, when trying to extend the "WoZ Way" method to the home, observations in the unconstrained environment were significantly less useful due to the user being out of view and focusing on other tasks beyond listening to music. Without seeing much, the value of DJ Bot as a remote ethnography method was limited.

During the in-car sessions, the researchers were able to flexibly ask the drivers questions about their music. Because the researchers were behind the speech agent, they could also freely coordinate and reflect on the user's answers to their questions. This is something that would be impossible (or at least quite rude) during an in-person team interview. In the home, the researchers focused their effort more on trying to act like a bot rather than ask the user questions. Their limited interaction and their limited view into the home made it hard to do more in-depth interviewing via DJ Bot. In this case, a regular in-person interview would have been more valuable for the researchers.

Compared to multimodal experience sampling [10, 17], having the user interact with a speech agent allowed the team to capture moments of interaction that would be impossible to capture through text based queries, especially in the car. In the home, while it would be fine to have the user interact with a mobile phone to answer an experience sample, interacting through voice still might be preferred while doing other activities, such as folding laundry. As noted by other researchers exploring experience sampling in-the-wild, seeing the participant in context was critical to understanding the user's experience [10] and without seeing the participant the value of interacting through a speech agent was greatly diminished.

Overall, the researchers felt that interacting through a remote-controlled speech agent allowed them to act in a variety of ways similar to other user research methods but with the main benefit of being remote and yet in the context. The researchers described how their use of DJ Bot as a research tool could exist on a sliding scale from highly exploratory needfinding interviews and observations to Wizard of Oz interaction prototyping.

System Limitations

Understanding a new context through a speech agent can be limited by how much the remote team can see and hear in the user's environment. In our case, the challenges of not seeing the user in the home environment led to the system being less useful than anticipated. Using only audio can mitigate privacy risks, but also reduces what a researcher can understand about the physical situation. Network issues significantly break the user experience and leave the research team without data.

Compared to experience-sampling, live interaction through a speech agent can generate a large amount of semi-structured data that can be challenging to process quickly for insights. This can increase the time required to analyze the data, making it less practical for practitioners in industry working on a fast-paced team. Still, even without reviewing the recorded data from the sessions, the in-the-moment learnings are useful for a research team. Adding notes, transcriptions, and markers of interesting events also helps the team to filter and review data more efficiently afterward.

While human operators can be intelligent and improvise, it can be challenging to recreate what a data-driven or automated system would be able to do. There are also moments where the

research team cannot figure out a reasonable response. These inconsistencies can break the user's simulated experience while interacting with the speech agent, but also highlight issues that future autonomous systems will need to handle.

Testing things in a controlled manner can be challenging when researchers need to respond to the situation at hand. Getting through a set of scripted interactions is challenging when the user or context demands that the interaction move in a different direction. In addition, there is coordination work that has to be completed quickly between the researchers when the user or context changes unexpectedly.

Considerations for using speech agents for remote context exploration

Given our experiences using DJ Bot, we suggest that using remote speech agents can be useful for in-context explorations in environments where you can closely observe the user, in spaces that are hard to access in-person, and where the flexibility of speech-based interaction allows for exploratory research. We have identified several considerations for using speech agents in remote context explorations based on our experiences:

- Capture video in places where the user can be seen interacting with their environment. What visual channels will help in giving you the full context? For example, in the car showing both the user interacting with the steering wheel as well as the road ahead.

- Set up a backchannel for interacting with your collaborators during the session. The backchannel could be voice- and/or text-based. Take time to reflect on the experience in-the-moment and afterward. What communication methods can best support reflection with your team?

- Identify a preliminary set of interaction rules. For example, should you react to context changes you can only detect as a human observer? Will you proactively talk to the participant after a specific duration of time? Allow yourself to improvise in-the-moment based on the user interaction.

- Identify what aspects of the interactive system you can easily control and what will be hard to simulate. What type of system do you want to learn about? In our case, we had to think about the accuracy and speed of giving user-specific music recommendations.

- Develop the researcher control interface iteratively. You might discover additional functionality or automation as you go along. Taking notes on what is useful can provide insights into what information or control is important for the research team and the product.

- Synchronize data streams and notes with the video. What moments of interaction are you most interested in looking at? What analysis are you planning to do? Add features that allow highlighting of specific moments to make post-analysis easier.

- Consider and respect user privacy. Only observe and interact during specified times. Frame the engagement as a conversation not as covert monitoring. Are there other ways to help the participant feel comfortable with the remote interaction?

FUTURE WORK

The challenges we experienced with DJ Bot suggest a few avenues for future work for extending the use of remote-controlled speech agents as user research tools. While we found that our adaptation of DJ Bot from the car to the home was not as successful as planned, there are still open questions about how to use a speech agent in environments like the home productively. It may be possible to use multiple cameras and simulate a speech-agent though the entire home or have multiple instances of the agent throughout the home. There are also opportunities for changing the perspective of remote-controlled speech agents from a third-person object to a first-person object worn on the user's body. This could enable better interaction in environments like the home where the user changes location frequently and allow for environments that are difficult to instrument, such as a user riding their bike or commuting on the bus.

Another aspect of future work is to explore when and how to automate certain functions of the speech agent. Using simple timed questions could allow for experience sampling interactions while still maintaining the benefits of hands-free data collection. During live interaction, adding in automation capabilities could improve the effectiveness of the researcher by allowing them to focus on the interaction rather than focusing on the mechanics of operating the speech agent. In the case of music exploration, we would look into integrating the real Spotify recommender system into the speech agent and allow for some forms of automated speech generation.

CONCLUSION

From our experiences with DJ Bot, we have found that using speech agents can be a useful way of performing problem-space exploration in-context. Overall, remotely observing users and interacting with them though a speech agent helped us to overcome the spatial and temporal challenges of conducting in-context user research. By remotely interacting through a speech agent, practitioners can conduct design research in ways that enable greater collaborative reflection and geographical diversity. While our case study explored understanding people's interactions with music in-context, we believe that other practitioners developing context-sensitive products, such as home appliances, personal robots, or even health devices, can benefit from exploring real-world contexts through speech agent systems similar to DJ Bot. Though there are challenges to be overcome in employing this method to new contexts and applications, we believe that speech agents provide a new avenue for exploring the contextually situated aspects of people's experience with technology in their everyday lives.

ACKNOWLEDGMENTS

We would like Kyu Keogh, Hamish Tennent, Rob Semmens, Dylan Moore, and Abhijeet Agnihotri for their help through this project. We would also like to thank Sarah Fox for reviewing a draft of this paper.

REFERENCES

[1] Tawfiq Ammari, Jofish Kaye, Janice Y. Tsai, and Frank Bentley. 2019. Music, Search, and IoT: How People (Really) Use Voice Assistants. *ACM Trans. Comput.-Hum. Interact.* 26, 3, Article 17 (April 2019), 28 pages. DOI: http://dx.doi.org/10.1145/3311956

[2] Morten Sieker Andreasen, Henrik Villemann Nielsen, Simon Ormholt Schrøder, and Jan Stage. 2007. What happened to remote usability testing?: An empirical study of three methods. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, NY, USA, 1405–1414.

[3] Adriana Barón and Paul Green. 2006. *Safety and usability of speech interfaces for in-vehicle tasks while driving: A brief literature review*. Technical Report. University of Michigan, Transportation Research Institute.

[4] Frank Bentley, Chris Luvogt, Max Silverman, Rushani Wirasinghe, Brooke White, and Danielle Lottrjdge. 2018. Understanding the long-term use of smart speaker assistants. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 91.

[5] Hugh Beyer and Karen Holtzblatt. 1997. *Contextual Design: Defining customer-centered systems*. Elsevier.

[6] John Black and Marc Abrams. 2017. Remote Usability Testing. In *The Wiley Handbook of Human Computer Interaction*. Wiley-Blackwell, 277–297. DOI: http://dx.doi.org/10.1002/9781118976005.ch15

[7] David Broadman and Sera Koo. 2016. Chatbots: Your ultimate prototyping tool. Available at https://medium.com/ideo-stories/chatbots-ultimate-prototyping-tool-e4e2831967f3. (Sept. 2016). Date Accessed: 2017-01-05.

[8] A.J. Bernheim Brush, Morgan Ames, and Janet Davis. 2004. A Comparison of Synchronous Remote and Local Usability Studies for an Expert Interface. In *CHI '04 Extended Abstracts on Human Factors in Computing Systems (CHI EA '04)*. ACM, New York, NY, USA, 1179–1182. DOI: http://dx.doi.org/10.1145/985921.986018

[9] Paolo Burzacca and Fabio Paternò. 2013. Remote usability evaluation of mobile web applications. In *Kurosu M. (eds) Human-Computer Interaction. Human-Centred Design Approaches, Methods, Tools, and Environments (HCI 2013. Lecture Notes in Computer Science)*, Vol. 8004. Springer, Berlin, Heidelberg, 241–248.

[10] Scott Carter, Jennifer Mankoff, and Jeffrey Heer. 2007. Momento: Support for Situated Ubicomp Experimentation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, NY, USA, 125–134. DOI: http://dx.doi.org/10.1145/1240624.1240644

[11] Scott Carter, Jennifer Mankoff, Scott R. Klemmer, and Tara Matthews. 2008. Exiting the Cleanroom: On Ecological Validity and Ubiquitous Computing. *Human–Computer Interaction* 23, 1 (Feb. 2008), 47–99. DOI: http://dx.doi.org/10.1080/07370020701851086

[12] Sunny Consolvo and Miriam Walker. 2003. Using the experience sampling method to evaluate ubicomp applications. *IEEE Pervasive Computing* 2, 2 (2003), 24–31.

[13] Nils Dahlbäck, Arne Jönsson, and Lars Ahrenberg. 1993. Wizard of Oz Studies: Why and How. In *Proceedings of the 1st International Conference on Intelligent User Interfaces (IUI '93)*. ACM, New York, NY, USA, 193–200. DOI: http://dx.doi.org/10.1145/169891.169968

[14] Steven Dow, Jaemin Lee, Christopher Oezbek, Blair MacIntyre, Jay David Bolter, and Maribeth Gandy. 2005a. Wizard of Oz Interfaces for Mixed Reality Applications. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems (CHI EA '05)*. ACM, New York, NY, USA, 1339–1342. DOI: http://dx.doi.org/10.1145/1056808.1056911

[15] Steven Dow, Blair MacIntyre, Jaemin Lee, Christopher Oezbek, Jay David Bolter, and Maribeth Gandy. 2005b. Wizard of Oz support throughout an iterative design process. *IEEE Pervasive Computing* 4, 4 (Oct. 2005), 18–26. DOI: http://dx.doi.org/10.1109/MPRV.2005.93

[16] Susan Dray and David Siegel. 2004. Remote Possibilities?: International Usability Testing at a Distance. *interactions* 11, 2 (March 2004), 10–17. DOI: http://dx.doi.org/10.1145/971258.971264

[17] Jon Froehlich, Mike Y. Chen, Sunny Consolvo, Beverly Harrison, and James A. Landay. 2007. MyExperience: A system for in situ tracing and capturing of user feedback on mobile phones. In *Proceedings of the 5th International Conference on Mobile Systems, Applications and Services (MobiSys '07)*. ACM, New York, NY, USA, 57–70.

[18] Alinka E. Greasley and Alexandra Lamont. 2011. Exploring engagement with music in everyday life using experience sampling methodology. *Musicae Scientiae* 15, 1 (March 2011), 45–71.

[19] Patrik N. Juslin and Petri Laukka. 2004. Expression, perception, and induction of musical emotions: A review and a questionnaire study of everyday listening. *Journal of New Music Research* 33, 3 (2004), 217–238.

[20] Patrik N. Juslin, Simon Liljeström, Daniel Västfjäll, Gonçalo Barradas, and Ana Silva. 2008. An experience sampling study of emotional reactions to music: listener, music, and situation. *Emotion* 8, 5 (2008), 668–683.

[21] Scott R. Klemmer, Anoop K. Sinha, Jack Chen, James A. Landay, Nadeem Aboobaker, and Annie Wang. 2000. Suede: A Wizard of Oz Prototyping Tool for Speech User Interfaces. In *Proceedings of the 13th Annual ACM Symposium on User Interface Software and Technology (UIST '00)*. ACM, New York, NY, USA, 1–10. DOI: http://dx.doi.org/10.1145/354401.354406

[22] Mike Kuniavsky. 2003. *Observing the user experience: A practitioner's guide to user research*. Morgan Kaufmann.

[23] Yang Li, Jason I. Hong, and James A. Landay. 2004. Topiary: A Tool for Prototyping Location-enhanced Applications. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology (UIST '04)*. ACM, New York, NY, USA, 217–226. DOI: http://dx.doi.org/10.1145/1029632.1029671

[24] Blair MacIntyre, Maribeth Gandy, Steven Dow, and Jay David Bolter. 2004. DART: A Toolkit for Rapid Design Exploration of Augmented Reality Experiences. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology (UIST '04)*. ACM, New York, NY, USA, 197–206.

[25] Nikolas Martelaro and Wendy Ju. 2017a. DJ Bot: Needfinding Machines for Improved Music Recommendations. In *2017 AAAI Spring Symposium Series*.

[26] Nikolas Martelaro and Wendy Ju. 2017b. WoZ Way: Enabling real-time remote interaction prototyping & observation in on-road vehicles. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. ACM, New York, NY, USA, 169–182.

[27] Nikolas Martelaro and Wendy Ju. 2019. The needfinding machine. In *Social internet of things*. Springer, 51–84.

[28] Sarah Mennicken, Ruth Brillman, Jennifer Thom, and Henriette Cramer. 2018. Challenges and Methods in Design of Domain-specific Voice Assistants. In *2018 AAAI Spring Symposium Series*.

[29] David R. Millen. 2000. Rapid ethnography: Time deepening strategies for HCI field research. In *Proceedings of the 3rd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques (DIS '00)*. ACM, New York, NY, USA, 280–286.

[30] Jakob Nielsen and Jonathan Levy. 1994. Measuring usability: preference vs. performance. *Commun. ACM* 37, 4 (1994), 66–76.

[31] Adrian C. North, David J. Hargreaves, and Jon J. Hargreaves. 2004. Uses of Music in Everyday Life. *Music Perception: An Interdisciplinary Journal* 22, 1 (Sept. 2004), 41–77. DOI:http://dx.doi.org/10.1525/mp.2004.22.1.41

[32] Minsu Park, Jennifer Thom, Sarah Mennicken, Henriette Cramer, and Michael Macy. 2019. Global music streaming data reveal diurnal and seasonal patterns of affective preference. *Nature Human Behaviour* 3, 3 (2019), 230.

[33] Tim Plowman. 2003. Ethnography and critical design practice. In *Design research: Methods and perspectives*, Brenda Laurel (Ed.). 30–38.

[34] Spotify. 2020. Spotify Company Info. https://newsroom.spotify.com/company-info/

[35] Sarah Waterson, James A. Landay, and Tara Matthews. 2002. In the lab and out in the wild: Remote web usability testing for mobile devices. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems (CHI EA '02)*. ACM, New York, NY, USA, 796–797.

[36] Kristie Young, Michael Regan, and M. Hammer. 2007. Driver distraction: A review of the literature. *Distracted driving* (2007), 379–405.