Giving Voice to Silent Data: Designing with Personal Music Listening History

Jordan Wirfs-Brock
University of Colorado
Boulder, CO, USA
wirfsbro@colorado.edu

Sarah Mennicken
Spotify
San Francisco, CA, USA
sarahm@spotify.com

Jennifer Thom
Spotify
Boston, MA, USA
jennthom@spotify.com

ABSTRACT

Music streaming services collect listener data to support personalization and discovery of their extensive catalogs. Yet this data is typically used in ways that are not immediately apparent to listeners. We conducted design workshops with ten Spotify listeners to imagine future voice assistant (VA) interactions leveraging logged music data. We provided participants with detailed personal music listening data, such as play-counts and temporal patterns, which grounded their design ideas in their current behaviors. In the interactions participants designed, VAs did not simply speak their data out loud; instead, participants envisioned how data could implicitly support introspection, behavior change, and exploration. We present reflections on how VAs could evolve from voice-activated remote controls to intelligent music coaches and how personal data can be leveraged as a design resource.

Author Keywords

voice assistants; co-design; participatory design; personal informatics; music; speculative design

CCS Concepts

•**Human-centered computing** → **User centered design;**

INTRODUCTION

Roughly a quarter (22%–26%) of U.S. adults now own a voice-enabled smart speaker such as the Amazon Echo or Google Home [6, 11]. In recent years, these devices have added increasingly complex functionality, including conversational capabilities and system intelligence that incorporates contextual information and user histories. Further, research on voice assistants (VAs) suggests that they can and should be designed to learn and retain user context [4]. Yet these design principles are not reflected in one of the primary uses of smart speakers: playing music [1, 4]. People typically access music via voice in a simple, transactional manner, such as requesting to play specific songs or to change the volume. These type of transactional interactions do not invite reflection on the music content or on the listener's behavior.

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

ACM ISBN 978-1-4503-6708-0/20/04 ...$15.00.
http://dx.doi.org/10.1145/3313831.3376493

The data required for a more contextualized, reflective VA music experience does exist. Music streaming services, such as Apple Music, Deezer, Pandora, and Spotify, log a plethora of data (what users play, when they play it, how they search for it) to optimize services and offer personalized content. We refer to this logged listening history as *silent data* — silent because it is rarely exposed to users in a meaningful way. Even with visual interfaces, it is difficult for people to build an understanding of what data is being logged about them and how that data affects the offered content. While exposing users to their data can increase the intelligibility of autonomous systems [30, 3], it can also easily overwhelm them if it is not presented adequately [28]. These challenges of data intelligibility and transparency are further compounded in audio-only voice interfaces because data has traditionally been presented in highly visual formats. While methods for communicating data via audio do exist, such as sonification [19], these methods are not prevalent in VAs and are unfamiliar to most users. There is more to making data meaningful to end-users than adapting the mode of presentation — we must also ensure that it supports users in making sense of their data.

We explore ways that we might give voice to this *silent data* to design more contextualized, personalized voice experiences around music. We draw inspiration from recent work that encourages the HCI community to "investigat[e] metadata as a design resource" [8] and extend this to include end-users. We also invoke the traditions of speculative design and personal informatics to shed light on how people want to engage with their logged data. We conducted workshops with ten Spotify listeners to imagine interactions with a near-future voice-enabled music service through a series of pair-based design activities that included guided visualization, dialog mapping, and Wizard of Oz prototyping. In the workshops, we provided participants with detailed personal music listening history data — such as top artists and songs, inferred genre preferences, and temporal listening patterns — which grounded their design ideas in their current listening behaviors.

Our research goals were to (1) engage end-users and their personal data in a design process as a method for doing research around voice interfaces and personal informatics, and (2) to uncover insights about future, untapped user needs. Music is a promising domain to pursue these goals because of the prevalence of music listening on smart speakers as well as the strong ties between music, identity, and nostalgia [18]. Further, people enjoy music both privately and in groups [33], thus focusing on music allows us to examine personal and

social contexts for data-inspired interactions. We conducted the design workshops in pairs so that participants could make sense of their own listening history data in the context of another person's and share and discuss their data-inspired voice interaction ideas. In this paper, we contribute empirical findings describing how personal data, through voice interactions, can lead to more active, meaningful music listening practices. We close by discussing further opportunities for researchers to engage with silent data.

RELATED WORK

Our research explores how personal data can be used to improve the design of VAs. We briefly state current limitations in voice interactions, particularly with respect to playing music, and discuss concepts from personal informatics and design methodologies that informed our work.

Voice Assistants

The open-ended scope of current VAs, such as Apple's Siri or Amazon's Alexa, is simultaneously exciting and paralyzing. Smart speaker owners can ostensibly ask their VA anything: to coordinate their to-do lists, to provide information on demand, even to keep them company with jokes and conversations. At the same time, VA users face the challenge of not knowing what to ask or how to ask it [32, 15] — users settle on what they will use the device for in the first few days and rarely change this use or explore new functionalities [4]. Even though users could potentially converse with a VA on open-ended topics, the most common functionality users settle into is playing music, which can be accomplished most reliably through transactional requests (e.g. "Play the *Hamilton* soundtrack") [4, 1].

To understand users' goals and frustrations with VAs, Luger and Sellen [25] employed semi-structured interviews and found that, "users had poor mental models of how their [VAs] worked," due to a lack of system feedback — users did not know what the system was and was not capable of. While users found playful responses from VAs engaging, this playfulness set unrealistic expectations of the system's capabilities. They propose that the design of VAs should selectively reveal the limits and capabilities of the system's intelligence as one possible way to bridge this expectation-reality mismatch.

In addition to helping to correct users' mental models of VAs, we suggest that system intelligence can also potentially be revealed to users in the form of personal data. In the case of music listening, this system intelligence might include a conversation with the user about the VA's knowledge of the user's musical taste. We present this knowledge — gleaned from logged music listening data — to users through a variety of design activities in order to elicit user needs and values.

Personal Informatics

Personal informatics refers to technologies, systems, and practices through which people "collect and reflect on personal information" [24] to better understand themselves. People's methods and motivations for personal informatics vary: they can track automatically or manually [24], with digital or analog tools [2, 26]; they track in order to change their behavior, to gain insights into their identities, and even for the inherent rewards of tracking itself [38, 27]. The practice of self-tracking, and the data collected through this practice, serves multiple purposes, acting simultaneously as "tool, toy, and tutor" [27] helping the user to accomplish goals, to play, and to learn. In recent years, personal informatics in HCI has taken a turn toward *lived informatics*, which considers how people experience self-tracking in their everyday lives [13, 38].

Lived informatics focuses on how people make sense of their own data. In *Dear Data*, Lupi and Posavec [26] collected mundane data about their lives — how many doors they walked through, how many times they said thank you — and created and shared hand-drawn visualizations on post cards. A reaction to the rise of automated data logging, *Dear Data* emphasizes that personal data is meaningless without interpretation and reflection: "It is only by adding personal context that you get closer to real meaning. We shouldn't expect any app to tell us something new about ourselves." Kim et al. [23] were inspired by this work to make "DataSelfie," a platform to create visualizations of survey data. They suggest that "designing a personalized visual vocabulary" might be a critical feature in our ability to use data to inspire reflection. Data representations can also help express our identities: Elsden et al. [12] organized a speed-dating event where participants created and exchanged personal data profiles, demonstrating how data can be used to foster conversations and interpersonal relationships. Design work has shown how personal metadata can inspire new modes of interaction: Chronoscope [8] invites users to navigate through their sprawling photo archives using a timestamp-based kaleidoscope wheel.

For many people, music consumption plays several important roles in their lives, such as helping them manage self-identity, mood, and interpersonal relationships [18]. Data about music listening behavior has been employed as a design resource to encourage personal reminiscence. Gulotta et al. [16] revealed snippets of fabricated data of the music listening history of one's relatives in a speculative calendaring web application to elicit reflections about multi-generational relationships. Using metadata about user listening logged by last.fm, Odom et al. [34, 35] built physical music players for exploring alternative interactions based on temporal information.

In these examples, metadata about music listening served as an input to prototypes designed to generate insights about emotional responses to music. We extend this line of work by using metadata about music listening as a design material in a workshop context with end-users to help people express their desired interactions with an existing system, VAs.

Prototyping and Design Methods

Participatory design seeks to democratize design by involving end-users in every stage of the process, from ideation to refinement [44]. A full participatory design process can build long-term relationships with users for mutual learning [37], or create hybrid spaces for designers and users to co-create [31]. Our methods, which engage end-users in design activities, draw inspiration from participatory design but are not a full deployment of it. We also invoke reflective design, which asserts that, "Technology that monitors and reports on user activity or experiences should be carefully designed to avoid making the technology, rather than the user, the final authority

on what the user is doing [39]." We apply this sentiment to the design applications of logged user data in streaming music services.

In recent years, designers have begun to apply participatory methods to VAs. In workshops with sighted and visually impaired secondary-school students, Metatla et al. [29] showed how non-digital methods such as body-storming and physical prototyping with LEGO blocks can be incorporated into a voice design process. In workshops with teenagers, Fitton et al. [14] used paper and digital prototypes together in sequence, a method we adapt in our study.

Participatory design methods can also be a valuable way to uncover user needs independent of the goal of creating a particular product or artifact. Wun et al. [43] conducted workshops where participants created autobiographical data visualizations using stamps they carved from potatoes and sponges and derived insights about the flexibility and expressiveness of lo-fidelity authoring tools. Similarly, Huron et al.'s [22] work on how novices construct data visualizations with "tangible tokens" revealed how people think and talk about data representations. Design research can also apply speculative design practices. Cheon et al.'s [9] work using "futuristic autobiographies" shows how speculation can be a tool to both envision the specific details of human-robot interactions and to explore inter-participant disagreements as a way of interrogating and exposing values [10].

We focus on eliciting how users envision meaningful interactions with a VA that is intelligent enough to engage their personal data. By leveraging aspects of participatory and speculative design methods — particularly around ideation, refinement and reflection — as tools to uncover user needs around emerging technology, we interrogate values around human-computer interactions with personal data that do not yet exist but are within the realm of possibility.

METHOD

We conducted five 90-minute in-lab design workshops with two participants each. The ten participants were recruited in the metro-Boston area of the U.S. (5 female and 5 male, aged 24-55; see Table 1). Two researchers (a facilitator and a note-taker) were present in each session. Participants were compensated for their time with a $150 gift card.

Drawing on the *extreme user* method [20], we sought out participants who are highly engaged music listeners, frequent VA users, and familiar with personal data collection. As extreme users tend to be early adopters of emerging technology, they are well suited to support our goal of eliciting ideas for future interactions. We recruited listeners of Spotify, a streaming music service, based on logged behaviors inferred from their account history. To find highly engaged users, we filtered for people who recently and regularly use a variety of features (i.e. creating playlists, following artists). We further selected users who had listened to music via a voice-activated speaker regularly in the past four weeks. Via email, we sent a screener survey that included a question about whether the person does any self-tracking, such as using an activity-tracking app, keeping track of their media consumption or other personal behavior, or simply keeping a journal. All but one participant self-reported doing at least one of these activities. During recruitment, we informed participants that researchers and another participant would be able to view their music listening history data. Participants were given the option to leave the study if they objected to this. All participants agreed to share their personal data before consenting to the study.

Session	*Pseudonym*	*Age*	*Voice Assistant*
1	Adam	25	Amazon Echo
1	Aaron	24	Google Home
2	Brianna	27	Amazon Echo
2	Bob	55	Amazon Echo + Dot
3	Clara	27	Google Home
3	Chloe	27	Amazon Echo
4	Diana	26	Google Home
4	Daisy	27	Google Home
5	Eric	47	Amazon Echo Dot
5	Elliot	26	Amazon Echo, Sonos

Table 1. Participant Details

Design Workshop Activities

The participants were guided through a set of activities (Figure 1) that we selected and refined through a series of internal pilots with colleagues. We chose this specific set of activities because they provided varying levels of structure that appealed to different thinking styles and transitioned participants from a designer to a user perspective. Each workshop began with a short explanation that the goal of the session is to engage personal listening history data through imagining near-future voice interactions. The facilitator instructed participants not to be limited by the capabilities of their current VAs.

The first activity was an *ideation activity* using a list of participants' five top played songs of all time and the respective play-counts (e.g., *#1 "We Own The Night" by Dance Gavin Dance – 59 plays*) as a discussion prompt. We let participants share initial reactions, then asked them to pick a song from the list that is particularly memorable or significant. Next, focusing on this song chosen from listening history data, we led a *guided visualization* based on the concept of futuristic autobiographies [9]. We asked the participants to close their eyes as we narrated prompts that asked them to think about interacting with a future, ideal VA in five to ten years around this specific song. The full script (in our supplementary materials) included prompts asking the participant to imagine the details of the scene around them, what they say (if anything) to the VA, what the VA says (if anything) back, and what would happen if this interaction were to continue for a long time.

In the next activity, *card-based dialog mapping*, the participants mapped out the interaction they had visualized on paper using cards to structure their thinking. The facilitator gave participants a set of cards (included in our supplementary materials) labeled: *Context (When? Who? Where? What are you doing?)*; *Person says*; *Voice Assistant says*; *Music plays (a song or playlist)*; *Sound plays (a sound effect or noise)*; and a blank wild card. We instructed them to use the cards to document the interaction they had just imagined in the guided visualization. After they were done mapping the envisioned

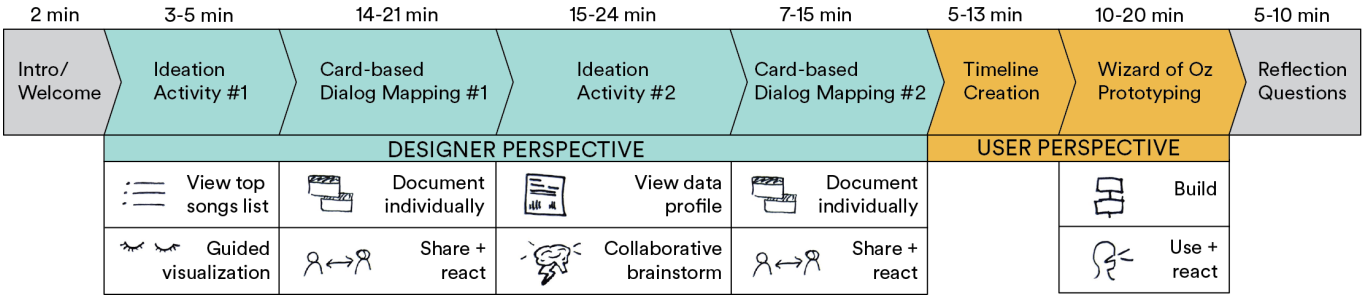

Figure 1. Overview of design workshops, starting with activities that encourage participants to take a designer perspective and finishing with activities that encourage them to take a user perspective. We include activity duration ranges, as the individual activities varied in length from session to session.

experiences with cards (example in Figure 3), each participant shared what they had created, and the other participant was invited to ask any follow-up questions.

For the next *ideation activity*, the facilitator gave each participant a customized one-page personal data profile (Figure 2), allowed them to react to it, and led a short discussion about what information on the sheet each participant would and would not want to share with other people. We conducted the workshops in pairs so that participants could look at their data profiles side-by-side and discuss whether this made them think about their own data differently. This allowed them to put

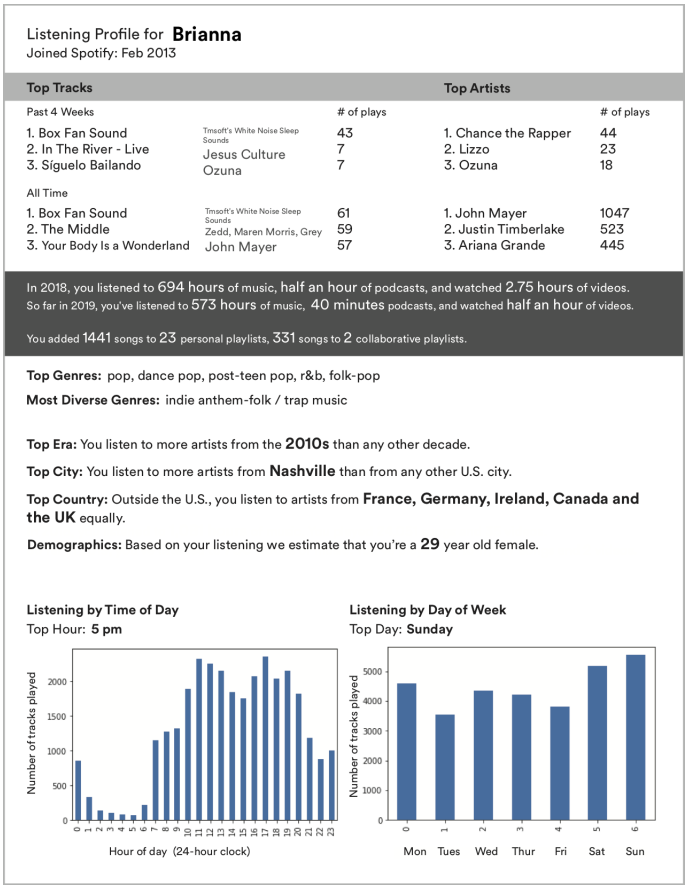

Figure 2. Example of a data profile showing information such as top songs and play-counts, hours listened, and top genres and categories.

their data in the context of someone else's, reflect on the differences, and discuss alternate interpretations and values about how to use their data. Using the data profiles as a prompt, the facilitator led a *collaborative brainstorming session* about how a VA in the future might use this personal listening data. Each participant then chose one idea that emerged from the brainstorm and did another *card-based dialog mapping* activity to document the details of this interaction.

Next, participants were asked to put all of the voice interactions they had generated during the previous activities onto a *timeline* to illustrate how these events might unfold over time unfold in someone's life over days, weeks, months, even years. The facilitator prompted the participants to think about what might trigger each of these interactions, how frequently they might happen, and what temporal changes (weekday versus weekend, seasonality, etc.) might affect when they occur.

For the last activity, the facilitator constructed *Wizard of Oz prototypes* of the participant's voice designs on-the-fly. The participants instructed the facilitator about what to build in a simple prototyping environment that was able to queue and play music, record and play sound effects, and generate text-to-speech utterances. The participants then used the prototypes as the facilitator "drove" the interaction.

We closed the workshops with a short, semi-structured discussion about how the participants felt about the workshop activities and about the voice experiences they had designed.

Analysis

All workshops were video recorded and transcribed. We analyzed the discussions that arose during the design workshops by performing an inductive analysis using an affinity diagramming technique [5]. To do this, we converted the session transcripts into 551 individual notes. Two researchers iteratively clustered these notes into categories until reaching agreement, then clustered the categories into high-level themes. The result is a set of high-level themes and sub-themes that are mapped back to quotes from the design workshop transcripts, a subset of which we describe in our findings.

FINDINGS: DATA AS DESIGN RESOURCE

In the workshops, although participants used their personal listening history data as a brainstorming tool, data was rarely the focus of the interactions participants designed. Instead, viewing their data caused participants to reflect on their identities as music listeners. Out of this data-motivated self-awareness,

participants imagined future voice interactions that are tied back to their current listening habits. In these interactions, personal listening data served as a trigger for activities such as content exploration, social sharing, and behavior modification.

Personal Listening Data and Identity

Some participants saw dimensions they valued about their own identities reflected back to them in their data. After reading on his data profile that he listens to more artists from New York than any other U.S. city, one participant reflected,

> So I've had a very, very connected relationship with that place. My family's from there. So, well my mom's side is from there. So I think having that kind of idealized in my head, I like all of the, you know, Lou Reed...The Strokes, and I'm just still into New York artists. - Elliot

Another participant took pride in the fact that his inferred "listening age" was much younger than his actual age: *"I'm kind of old and it thinks that I'm 35, so I'm like cool. Try to be youthful." - Bob.* Participants were also proud of their tendency to explore music from other countries (Brianna) or the fact that they make lots of playlists (Chloe). Logged listening behavior was a kind of digitally arbitrated proof, or as one participant put it, *"It's not all just talk. [The data] validates it." - Clara.*

Alternatively, in some cases, the data conflicted with participants' prior beliefs about their music listening behaviors and challenged their prior notions about their own identities. They had to reconcile these conflicts by coming up with plausible explanations for the mismatch by questioning the authority of the data or by updating their intuitions about their behaviors. For example, Brianna's data profile included "trap music" as one of her genres, *"I'm like, really, 'trap music?' I'm sure it's a side effect of the fact that I'm a teacher and I've played a lot of weird requests for my students."* Several participants also discussed how their data reflected the listening behavior of other people who have access to their streaming music account, like significant others (Daisy, Chloe) or children (Eric).

Participants reflected that personal listening history data could tell you *some*thing — just not perhaps what they might have expected. For example, participants' most frequently played songs were not necessarily their favorite songs. Top songs by play-count could be sleep music: *"I listen to the first one almost every night. It's like my lullaby" - Diana*. Eric's top songs list was filled with an album he has been using as sleep music for the past seven years. Another participant, Bob, expressed a subtle ambivalence about one of his top songs, *"God, that's a great song. But I don't think that's so reflective of me."* The high play-count did not match his mental model of his own musical taste preferences, so he postulated that it might have been listed on his profile because it is "obscure" and "weird" — even though the song was there solely because of the number of times he has played it.

Instead of reflecting musical taste, Adam noted that high play-counts could simply reflect his work patterns:

> I think it's less about favorite songs and more quite frequently when I'm listening to a new song, I'll put it on repeat and then I might walk away from my desk...if that

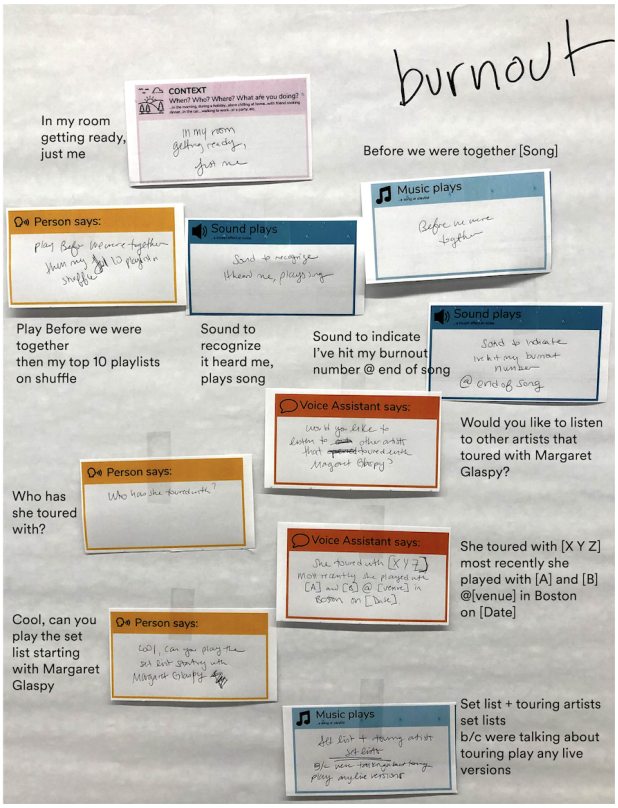

Figure 3. In the card-based dialog mapping activity, Clara designed this interaction to support listening to her current obsession, Margaret Glaspy, without getting burnt out.

> happens multiple times over the course of a work day, that ends up being a ton of repeated songs. - Adam

Further, viewing someone else's data brought new dimensions to how people interpreted their own. Aaron self-identified as a music nerd; Adam, on the other hand, did not consider himself a "musically-inclined person." These identities were challenged when they compared their data profiles and saw that Adam had listened to twice as many hours of music in 2019 as Aaron. This comparative context motivated Adam, who associates listening to music with doing work, to work less and go outside more. Aaron, however, was motivated to "gamify" his data and listen to more music.

These examples illustrate a process of negotiation, interpretation, and sense-making: As participants looked at their data profiles, they put them into the context of their own memories, identities, and mental models of their listening behaviors in order to determine what insights each data point might reveal.

Data Grounds Futuristic Thinking in Current Experiences

The futuristic voice interactions our participants conceptualized are grounded in their current listening behaviors. For example, when Clara looked at her personal data profile, she was intrigued by the differences between her top songs from the past 28 days and her top songs of all time. This highlighted an important dimension of her listening behavior — she goes through phases where she is obsessed with an artist, then listens to that artist until she reaches the point of burnout:

> Margaret Glaspy has helped me a lot these past four weeks. And I've listened to her songs so many times that I was just complaining to all my friends...."I can't listen anymore because I'm going to hate it. And I don't want to hate it, I want to still love it." - Clara

So she imagined a VA that could help her manage this behavior and extend the pre-burnout "still love it" phase:

> I wish it was like when I hit five listens in one day, it was like, "You've listened five times. Here's another artist that opened for her." Or if it gave me other context where I could get distracted from her a little bit. - Clara

This futuristic interaction — where a VA is aware of her personal threshold for burnout, tracks her listening accordingly, notifies her with a sound effect, and offers options for adjacent music to explore (Figure 3) — is grounded, through the play-count data, to her current listening behaviors and addresses an existing problem.

In the case above, the data drew the participant's awareness to a specific kind of temporal listening behavior. In other cases, the data drew participants' attention to past memories or events, leveraging the nostalgic potency of music to ground people in real-world needs and experiences. For example, Eric chose, from his all-time top songs list, a song that had been the favorite of a friend who had passed away. He imagined being able to tell a VA that this song had been one of his friend's favorites; then, whenever another one of his friend's favorite songs played randomly, he could repeat this process, building a playlist in memoriam of that friend organically, leveraging natural reminiscences in situ rather than sitting down and racking his brain for all of his friend's favorite songs. Reflecting on identity, through data, led to awareness of a specific problem, which the participant addressed through an imagined futuristic voice interaction.

Personal listening history data reminded participants of specific events and memories, such as concerts or trips. It also made them aware of habits around how and when they listen to music, as well as deviations from those habits. As one participant put it, looking at the data profile, *"made me feel really like, you know, kind of introspective. Like I've been listening to this, I didn't even realize that I was listening to something so many times and that it's...part of my musical memory, now"* - *Diana.* Looking at meta-data about the music they listen to, such as genres, where artists are from, and when songs were released, gave participants the language to describe their own listening preferences — language they could then use to access new music experiences on VAs.

The personal data sheets included information about artists' geographic origins, which listeners are often not consciously aware of. Once they became aware of this meta-data — in the context of their own music listening histories — this previously silent data became a part of the participants' language. Elliot used this new, meta-data-derived language to support sense-making about his behavior; from there, he envisioned new experiences where artist origin location is a dimension of exploration: *"We've noticed you've listened to a lot from Detroit. Would you like to visit Detroit? And here's some places to go in Detroit, and here's some music venues."* The data gave participants the language to explore themselves as they simultaneously explore interactions with a new technology.

Data as a Trigger, not a Focus

Some of the voice interactions designed by the participants used personal data explicitly, such as when Chloe imagined a VA that would play a celebratory noise when she added the 100th song to a playlist and prompt her to share this playlist with friends. Data was also explicit in the interaction to avoid burnout (Figure 3). However, while each of these interactions explicitly references a data point, the data is not the point. Instead, the most important elements of these voice experiences are the actions that the data triggered, such as facilitating sharing, reflection, or music exploration.

In some interactions, personal listening history data was not explicitly stated by a VA, but was more subtly and implicitly embedded as a meaningful action that was inspired by the data. For example, Aaron imagined being able to conjure up additional information to enhance the inherent nostalgia of music by asking a VA, *"When and where did I first listen to this?"* which would then transition into other relevant information about a song, such as the movies in which it was featured. Or, after considering that music had location data connected to it (where an artist is from, where an artist recorded a song), Elliot imagined a VA that could create a playlist generated from the itinerary of an upcoming road trip.

Music exploration and discovery is a through-line in these imagined voice experiences. Participants felt that cultivating a diverse music repertoire was important, as Diana said, *"I would want to explore more because I would want to evolve my music personality."* But exploration is challenging, as Brianna put it, *"Now I like do a lot of the thinking myself. I'll take out my phone and find the song that I want....So I'm doing all the work."* Many of the voice interactions facilitated exploration by using personal listening history data to give participants the vocabulary to explore, provide context behind recommendations, and to ease them into related content:

> I don't go like, "Oh I want to listen to Brit pop." I usually don't think like that. But that would be an interesting way to navigate to more music. - Bob

> I would want [a recommendation] to be as specific as possible and something interesting that I might keep in my brain, I wouldn't want it to just be a throw-away fact of like, "This band is similar to this one." I want it to be like, "Oh 58% of people who listen to one of your favorite bands also listen to this band." - Adam

> [A VA might say,] "This song was number three on the top 40 playlist on July 22, 1986, would you like to hear the top 10 songs from that time?"...It's kind of nostalgic, right...and I want it to suggest it to me. - Eric

In these examples, data can enable and support meaningful interactions around music exploration.

FINDINGS: DESIRED VOICE INTERACTIONS

Through the various design activities and the use of personal data, our participants described the desired functionalities, roles, and specific interactions they want future VAs to offer.

From Remote Control to Music Coach

Current interactions with VAs are similar to using a remote control, where users must give specific instructions for playback or request specific content. In our activities, however, participants imagined future voice-enabled music services that could leverage deep knowledge about music, its cultural context as well as an understanding of lyrics or musical parts. This included the ability to play a short preview clip of a song and jump straight to the chorus, skip backward to relevant sections of the song, recite lyrics, or provide information about the meaning behind a song. In many of the experiences participants imagined, this musical intelligence combined with intelligence about the listener to form enhanced conversational interactions. Participants also wanted the assistant to proactively support them in expanding their musical personalities, echoing behavior change goals that are prevalent in other areas of personal informatics. Existing music search through VAs provides support for cases where users have some specific content in mind, and operate under the expectation that users to know how to request it [21]. Yet we found that people often simply did not know exactly what music they wanted to listen to, or what experience they were in the mood for. Thus, they described a collaborative process for identifying needs along with the voice assistant, treating it like a music coach, rather than simply commanding it. In some cases, these conversational interactions helped listeners figure out exactly what they wanted to listen to — the dialog was a tool to facilitate intention focusing:

> I think then it would be either a conversation between me and my device, you know, being like here's something new and then it, did you like that? And I'd be like, yes or no. - Brianna

> ...the conversation we can have is that [the VA] can kind of educate me, and we can go back and forth with what artists are coming out? What genres? - Daisy

In these situations, the human and the VA collaborate to figure out what music to play in a specific moment through "tasting" recommendations refined by real-time user feedback.

The design ideas our participants generated used the affordances of voice interactions to address short-comings of music streaming apps, which rely on predominantly visual interaction modalities. One participant imagined creating a hands- and eyes-free, road-trip playlist. While driving, he might say,

> "Play some songs for the moment, play some excited happy songs." And the [VA] says, "I think I know a good one, here's 'Sins Of My Youth."' It starts with that but...leads into a place of other songs that are similar in terms of like the mood....And then it offers to share the playlist with everyone who's in the car. - Elliot

Asking a VA to send the playlist to his friends leverages the in-the-moment nature of voice interactions to bookmark content for the future. Several participants mentioned that they often come across an interesting new artist or track, but cannot explore it fully in that moment. They described a VA that could use a more holistic understanding of their life and routines to time- and location-shift music discoveries by reminding them about it at a more appropriate time when they can use visual modalities or devote more attention to exploring the content in detail.

Collaboration for a Long-term Relationship

Some participants felt that, currently, they have to do the hard work of finding music that they might enjoy. Ideally, they wanted the VA to use its understanding of music to reduce the effort on the user's end. However, they also expressed that they would be willing to put in the work needed to instruct the assistant up front, with the expectation that it will pay off in the long-term and add value to the interaction. They imagined teaching the VA about their preferences by tagging context, such as with a mood descriptor, or explaining their preferences, *"I feel like it'd be really easy to be like, 'I like that, but that song makes me feel really happy' or 'I'm feeling uplifted"' - Brianna.* Through this kind of ad hoc labeling, participants imagined being able to teach the VA about their own language for categorizing and making sense of music.

In addition to explicitly providing the VA with information, participants also envisioned that the VA could ask for help when it cannot make a definite decision with the information it has. For example, Eric imagined an experience where he wanted the VA to extend a listening session that he enjoyed by saying simply, *"More of this."* However, he acknowledged that the VA might not get it right and could potentially suggest content that might not fit the situation. But he described how the VA should implicitly realize that it did not understand his desires if he repeatedly says, *"Skip"* or *"Next,"* and then proactively ask for clarification so that it could more accurately interpret his request. This is in contrast to the way people currently provide context about their music preferences, such as by taking time to build playlists or by specifying user settings.

Examples like these demonstrate the desire for relationship-building with a VA that is situated, in-the-moment, and adaptable. This is a notable departure from how most people interact with music services now, but as one participant reflected, these changes could develop gradually:

> It would feel weird initially, but as, again, and you're talking about in the future as, I've already started to have a relationship with this device that feels comfortable for me and I think this is just that advanced. - Elliot

VAs open up an opportunity for light-weight, on-the-fly interactions around music listening to build a relationship and specify preferences in the moment in which they are relevant and at the forefront of the user's mind.

DISCUSSION

We offered music listeners an unfamiliar material for exploration and creation: *silent data*. This data was silent because it was hidden within the system (collected about users but not shared with them) and because it is traditionally not communicated in voice-only interaction modes. In this section, we discuss opportunities for how personal data can be used to help people explore new technologies, how participatory methods could be used to approach designing systems that log user data, and how designers might use personal data as a resource for creating more context-aware VAs.

From Silent Data to a New Vocabulary for Exploration

Through focusing on silent data, participants gave themselves a voice by using their personal listening history data to create new *vocabularies for exploration*: for exploring the capabilities of VAs and for exploring their own listening behaviors.

This simultaneous internal (personal behavior-focused) and external (technology interaction-focused) exploration suggests design opportunities for voice-enabled music services: Providing users with personal data could be a way to scaffold them through the process of feature discovery. Personal data, which is grounded in a user's lived experience, can be the bedrock for a shared language between a user and a system. The use of a shared language and terminology could allow users to move beyond transactional interactions to context-aware, relationship-based interactions.

An example from our design workshops was how several participants used artist origin data (e.g., "You listen to more artists from Detroit than any other U.S. city") as a way to access new music content and envision new voice experiences that engaged this metadata. This concept might also be used more broadly to help people explore new ways to use unfamiliar technologies. For example, data about Internet-of-Things use and energy consumption in homes has been notoriously difficult for users to understand due to its complexity, variety of sources, and multi-user nature [41]. In the field of pervasive intelligibility, data has commonly been used in retrospective interviews to better understand or contextualize user behavior and activities. Asking pairs or group of participants – possibly people sharing a home – to collaboratively design future smart home interfaces that center around specific aspects of logged data can be another, more generative way to drive user-centric innovation. Another use case could be presenting and designing with data collected by social networks or media platforms to elicit different understandings of privacy and the ways in which people might want their values reflected in services. The process of looking at their data, describing it through a collaborative sense-making process, and then designing or imagining with it could support users in articulating their needs.

Toward User-centered Design of Data-logging Systems

Intentional self-tracking co-exists with unintentional (often unconscious) surveillance: As we track data about ourselves, data about us is being tracked by third-parties. In these situations, data collection happens without end-user reflection, thus people are denied the opportunity to truly incorporate that data into their lived experience. Researchers have called attention to the challenges posed by the participatory nature of data collection [40], but the HCI community can do more to extend this participatory nature to the *design* of data logging systems. Putting users close to raw data could help researchers and system designers figure out what they can do with the data that might ultimately be meaningful to users. This may also help address a common pitfall in data collection: The way that data is collected often does not line up with the way that it will ultimately be used.

In our study, by exposing participants to a range of personal data points in the context of design workshops, we were able to let them discover which data resonate with them and to envision how that data might be used in meaningful ways. Consider the example of play-counts: A naive assumption might be that play-counts could be a proxy for a listener's favorite music or representative of a listener's musical identity [17]. But in our study, we found that high play-counts represent all kinds of behaviors — sleep music, study music, a child's favorite song — thus the concept of "favorite" is more nuanced than a simple play-count. This discrepancy between metric and meaning was heightened because participants were being asked to *create* with this data point. This type of insight could help refine the metrics collected by media streaming services such as Spotify, Netflix, or YouTube and how they are used internally. Exposing silent data to end users can be a way of iteratively prototyping the data logging process.

This approach complements other user research methods, such as interviews where researchers show users their personal data [7]. Whereas interview methods are well-suited for retrospection and reflection about what data means, data-inspired design activities evoked forward-focused, generative discussions about what users can *do* with data.

Voicing Silent Data Through Interaction Design

As participants imagined interacting with VAs around their personal listening history data, silent data was most valuable as design resource when users first imbued it with their own context. We suggest updating Lupi and Posavec's assertion that, "We shouldn't expect any app to tell us something new about ourselves" [26] to: We shouldn't expect an app to tell us something new about ourselves without first guiding it with our own human interpretation. Digital systems can give voice to silent data, but this process depends on users' initial interpretations to make it meaningful. VAs offer one such way for users to provide this interpretive guidance to services that accumulate silent data. A music service that operates more like a coach than a remote control reflects this concept: Before offering meaningful feedback, a coach must first get to know the athlete.

The metaphor of a coach updates previous HCI models for VAs, falling somewhere between a "virtual butler" [36] and an "artificial companion." [42] By exploring voice interactions through a highly specific context, music listening and discovery, we affirmed Luger and Sellen's findings that "effective [VA] use requires ongoing work and investment." [25] Our findings extend this idea: The "work" Luger and Sellen describe consists of users learning a VA's features, whereas the work our participants imagined also included teaching a VA about their memories, interpreting the meaning of data, and developing user-specific shared languages.

There are concrete ways designers can incorporate this collaborative user and agent "work" into music services and voice interactions: Participants' interest in annotating music on-the-fly presents opportunities for a VA to expose its system intelligence while building a relationship with the user. Imagine an interaction where a user asks a VA to, "Play more music like this." A VA might respond, "Ok, how would you describe this song in your own words?" Or, after playing an algorithmically selected song, a VA could ask users why they *think* it chose that song, then respond with the reasoning behind the

recommendation. These types of interactions, while building system contextual intelligence, could also tease out a user's preferences (some might enjoy these interactions, whereas others might find them intrusive and ignore them). They also ease users into a more collaborative, less transactional relationship. While VAs provide a hands- and eyes-free way to build this type of music coach relationship, these interactions could also be performed by a visual or multi-modal music streaming service.

CONCLUSION

As the personal data that various technologies collect about people becomes richer and more prevalent, HCI researchers are faced with the challenge of deciding if and how to incorporate that data into the user experience. The concept of silent data, and the voice we can give to it by leveraging logged data for inspiration, suggests a new approach to representing user data. The participants of our design workshops viewed their personal data not as something simply to be presented back to them by a voice assistant, but as something that could be more generally interpreted to spark meaningful interactions and music exploration. Further, our participants viewed personal data as a foundation upon which they could build long-term relationships with a context-aware, voice-based music coach.

ACKNOWLEDGMENTS

We would like to thank Sam Way, Jean Garcia-Gathright, Glenn McDonald and Owen Heneghan for their feedback and input into this work as well as all of the people who piloted our workshops.

REFERENCES

[1] Tawfiq Ammari, Jofish Kaye, Janice Y Tsai, and Frank Bentley. 2019. Music, Search, and IoT: How People (Really) Use Voice Assistants. *ACM Transactions on Computer-Human Interaction (TOCHI)* 26, 3 (2019), 17.

[2] Amid Ayobi, Tobias Sonne, Paul Marshall, and Anna L Cox. 2018. Flexible and Mindful Self-Tracking: Design Implications from Paper Bullet Journals. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 28.

[3] Victoria Bellotti and Keith Edwards. 2001. Intelligibility and accountability: human considerations in context-aware systems. *Human–Computer Interaction* 16, 2-4 (2001), 193–212.

[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. 1999. Contextual design. *interactions* 6, 1 (1999), 32–42.

[6] Ava Mutchler Bret Kinsella. 2019. *Smart Speaker Consumer Adoption Report*. Technical Report.

[7] Praveen Chandar, Jean Garcia-Gathright, Christine Hosey, Brian St. Thomas, and Jennifer Thom. 2019. Developing Evaluation Metrics for Instant Search Using Mixed Methods Methods. In *Proceedings of the 42Nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'19)*. ACM, New York, NY, USA, 925–928. DOI: http://dx.doi.org/10.1145/3331184.3331293

[8] Amy Yo Sue Chen, William Odom, Ce Zhong, Henry Lin, and Tal Amram. 2019. Chronoscope: Designing Temporally Diverse Interactions with Personal Digital Photo Collections. In *Proceedings of the 2019 on Designing Interactive Systems Conference*. ACM, 799–812.

[9] EunJeong Cheon, Stephen Tsung-Han Sher, Šelma Sabanović, and Norman Makoto Su. 2019. I Beg to Differ: Soft Conflicts in Collaborative Design Using Design Fictions. In *Proceedings of the 2019 on Designing Interactive Systems Conference*. ACM, 201–214.

[10] EunJeong Cheon and Norman Makoto Su. 2018. Futuristic autobiographies: Weaving participant narratives to elicit values around robots. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 388–397.

[11] NPR Edison Research. 2019. *The Smart Audio Report*. Technical Report. Somerville, NJ.

[12] Chris Elsden, Bettina Nissen, Andrew Garbett, David Chatting, David Kirk, and John Vines. 2016. Metadating: exploring the romance and future of personal data. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 685–698.

[13] Daniel A Epstein, An Ping, James Fogarty, and Sean A Munson. 2015. A lived informatics model of personal informatics. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 731–742.

[14] Daniel Fitton, Janet C Read, Gavin Sim, and Brendan Cassidy. 2018. Co-designing voice user interfaces with teenagers in the context of smart homes. In *Proceedings of the 17th ACM Conference on Interaction Design and Children*. ACM, 55–66.

[15] Anushay Furqan, Chelsea Myers, and Jichen Zhu. 2017. Learnability through adaptive discovery tools in voice user interfaces. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 1617–1623.

[16] Rebecca Gulotta, Alex Sciuto, Aisling Kelliher, and Jodi Forlizzi. 2015. Curatorial Agents: How Systems Shape Our Understanding of Personal and Familial Digital Information. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 3453–3462. DOI:http://dx.doi.org/10.1145/2702123.2702297

[17] David Hargreaves, Dorothy Miell, and Raymond Macdonald. 2002. What are musical identities, and why are they important. *Musical Identities* (01 2002).

[18] David J. Hargreaves and Adrian C. North. 1999. The Functions of Music in Everyday Life: Redefining the Social in Music Psychology. *Psychology of Music* 27, 1 (1999), 71–83. DOI: http://dx.doi.org/10.1177/0305735699271007

[19] Thomas Hermann, Andy Hunt, and John G Neuhoff. 2011. *The sonification handbook*. Logos Verlag Berlin, Germany.

[20] Lars Erik Holmquist. 2004. User-driven innovation in the future applications lab. In *CHI'04 Extended Abstracts on Human Factors in Computing Systems*. ACM, 1091–1092.

[21] Christine Hosey, Lara Vujović, Brian St. Thomas, Jean Garcia-Gathright, and Jennifer Thom. 2019. Just Give Me What I Want: How People Use and Evaluate Music Search. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 299, 12 pages. DOI: http://dx.doi.org/10.1145/3290605.3300529

[22] Samuel Huron, Sheelagh Carpendale, Alice Thudt, Anthony Tang, and Michael Mauerer. 2014. Constructive visualization. In *Proceedings of the 2014 conference on Designing interactive systems*. ACM, 433–442.

[23] Nam Wook Kim, Hyejin Im, Nathalie Henry Riche, Alicia Wang, Krzysztof Gajos, and Hanspeter Pfister. 2019. DataSelfie: Empowering People to Design Personalized Visuals to Represent Their Data. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 79.

[24] Ian Li, Anind Dey, and Jodi Forlizzi. 2010. A stage-based model of personal informatics systems. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 557–566.

[25] Ewa Luger and Abigail Sellen. 2016. Like having a really bad PA: the gulf between user expectation and experience of conversational agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 5286–5297.

[26] Giorgia Lupi and Stefanie Posavec. 2016. *Dear data*. Chronicle Books.

[27] Ben Lyall and Brady Robards. 2018. Tool, toy and tutor: Subjective experiences of digital self-tracking. *Journal of Sociology* 54, 1 (2018), 108–124.

[28] Sarah Mennicken, Jo Vermeulen, and Elaine M Huang. 2014. From today's augmented houses to tomorrow's smart homes: new directions for home automation research. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 105–115.

[29] Oussama Metatla, Alison Oldfield, Taimur Ahmed, Antonis Vafeas, and Sunny Miglani. 2019. Voice User Interfaces in Schools: Co-designing for Inclusion With Visually-Impaired and Sighted Pupils. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 378.

[30] Bonnie M Muir. 1994. Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics* 37, 11 (1994), 1905–1922.

[31] Michael J Muller. 2007. Participatory design: the third space in HCI. In *The human-computer interaction handbook*. CRC press, 1087–1108.

[32] 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. DOI:http://dx.doi.org/10.1145/3290605.3300277

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

[34] William Odom and Tijs Duel. 2018. On the Design of OLO Radio: Investigating Metadata As a Design Material. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 104, 9 pages. DOI: http://dx.doi.org/10.1145/3173574.3173678

[35] William Odom, Ron Wakkary, Jeroen Hol, Bram Naus, Pepijn Verburg, Tal Amram, and Amy Yo Sue Chen. 2019. Investigating Slowness As a Frame to Design Longer-Term Experiences with Personal Data: A Field Study of Olly. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 34, 16 pages. DOI:http://dx.doi.org/10.1145/3290605.3300264

[36] Sabine Payr. 2013. Virtual butlers and real people: styles and practices in long-term use of a companion. In *Your Virtual Butler*. Springer, 134–178.

[37] Toni Robertson and Jesper Simonsen. 2012. Challenges and opportunities in contemporary participatory design. *Design Issues* 28, 3 (2012), 3–9.

[38] John Rooksby, Mattias Rost, Alistair Morrison, and Matthew Chalmers. 2014. Personal tracking as lived informatics. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 1163–1172.

[39] Phoebe Sengers, Kirsten Boehner, Shay David, and Joseph'Jofish' Kaye. 2005. Reflective design. In *Proceedings of the 4th decennial conference on Critical computing: between sense and sensibility*. ACM, 49–58.

[40] Katie Shilton. 2012. Participatory personal data: An emerging research challenge for the information sciences. *Journal of the American Society for Information Science and Technology* 63, 10 (2012), 1905–1915.

[41] Yolande A.A. Strengers. 2011. Designing Eco-feedback Systems for Everyday Life. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 2135–2144. DOI: http://dx.doi.org/10.1145/1978942.1979252

[42] Yorick Wilks. 2004. Artificial companions. In *International Workshop on Machine Learning for Multimodal Interaction*. Springer, 36–45.

[43] Tiffany Wun, Lora Oehlberg, Miriam Sturdee, and Sheelagh Carpendale. 2019. You say Potato, I say Po-Data: Physical Template Tools for Authoring Visualizations. In *Proceedings of the Thirteenth International Conference on Tangible, Embedded, and Embodied Interaction*. ACM, 297–306.

[44] John Zimmerman and Jodi Forlizzi. 2014. Research Through Design in HCI. In *Ways of Knowing in HCI*, Judith .S. Olson and Wendy.A. Kellogg (Eds.). Springer, New York, NY, 167–190.