Voices and Voids: Subverting Voice Assistant Systems through Performative Experiments

Audrey Desjardins
School of Art + Art History + Design,
University of Washington
adesjard@uw.edu

Afroditi Psarra
Center for Digital Arts &
Experimental Media (DXARTS),
University of Washington
apsarra@uw.edu

Bonnie A. Whiting
School of Music, University of
Washington
bwhiting@uw.edu

ABSTRACT

Responding to concerns such as privacy, surveillance, and the commodification of personal data with regards to voice assistants, this artistic research focuses on creating performative artifacts and vignettes that challenge artificial intelligence (AI) and machine learning (ML) technologies. By allowing AI voice assistants to listen to our most private conversations, we become receptive to their mediated care, while forgetting or ignoring how much these automated interactions have been pre-scripted. With our project Voices and Voids, we reclaim, examine, and ultimately transcode these voice assistant data through interdisciplinary performance and Post-Internet Art. In this paper, we thematically describe 12 vignettes which represent embodied and sonic experiments using a combination of design, data-driven art, cyber crafts, found-object and traditional percussion instruments, spoken word, and movement. We conclude with a discussion of how the experiments worked as a multifaceted whole, and how we used interdisciplinary methods as a central approach.

CCS CONCEPTS

• **Human-centered computing/Interaction design**;

KEYWORDS

Voice assistants, Data Driven, Performance, Digital Art, Sound Art, Percussion, Labor

1 INTRODUCTION

As voice assistants continue to become some of the most common connected devices in our everyday lives, the archives of recorded interactions with these agents continue to grow exponentially. While

much research has been devoted to understanding people's interactions with these voice assistants—including pinpointing their exact needs or finding better ways to support conversation—voice assistant data archives are underexplored. Instead of seeing data as a byproduct of voice interactions, in this work we commit to placing voice assistant data front and center as a tactic for understanding the complex relationships between people, large tech corporations, the data economy, hidden labor, and surveillance capitalism.

In our transdisciplinary art-based research project, we have chosen to use voice assistant data as creative and generative material. We reclaim and transcode voice assistant data to foreground the mundane, bizarre and complex ethical challenges of people's engagements with data through technology. By engaging both the form (sonic qualities) and content (semantic meaning) of the data logs, we explore topics such as the disembodiment of voice assistants, the hidden labor in voice assistant training, voice assistants' 'intelligence', and the intimacy of voice interactions at home. In this paper, we ask: When we use voice assistant data as creative material, what is revealed about voice assistant systems? Our contributions include the detailed description of our performative experiments which explore themes of intimate and mundane narratives, hidden labor, creative potential and surveillance capitalism. We also discuss how the vignettes work as a whole and became subversive tools to understand and reveal hidden characteristics of voice assistant systems and we reflect on our interdisciplinary tactics. While initially conceived as live performance and installation, the Covid-19 pandemic shifted our plans to an online format. We invite the reader to explore our voice assistant data-inspired vignettes and performances at www.voicesandvoids.net before reading the remainder of the paper.

2 CONTEXT: RELATED WORKS IN NEW MEDIA ART, MUSIC AND HCI

Our work builds on and is inspired by related artistic works and research literature.

2.1 Voice assistants in New Media Art

In the realm of new media art there have been several recent works that explore similar topics. Though not all of them engage with voice assistants directly, the following works engage with the topic of surveillance through data collection by big tech and of harvesting data from ghost workers [23]. As such, they are in conversation with the ethics of AI and AAI (Artificial Artificial Intelligence—a term assigned to ghost workers) and generate a critical reflection

on employing such practices as creative materials. The contribution of our work is located in the multidisciplinary nature of the project, which manifests itself through the multitude of different performative experiments that comprise it, rather than one singular approach. The common denominator between the works that follow, and the work presented in this paper is that they explore data-driven practices, the intentional and unintentional skewing of voice assistant interactions, and reveal the volume of human labor behind these systems as a subversive act.

Some of these seminal works include "Someone" by Lauren Mc-Carthy [28], a performative video installation that imagines a human version of Amazon Alexa which reflects critically on data collection and surveillance strategies, while highlighting the contradictory desires for intimacy, privacy, and efficiency. "Chthonic Rites" [22], by Wesley Goatley, explores the idea of a conversation between smart assistants Alexa and Siri prompting questions about technological solutionism and the power politics nested within their supposedly friendly identities. In "Experiments in Surveillance Capitalism" [24], Garnet Hertz probes the use of "red flag" keywords, playing with the fear that individuals have about privacy while using smart IoT devices. James Coupe in "Time-Clock" [13], "Let Us Pray" [14], and "General Intellect" [15], employs MTurk workers to produce art by exposing the surrealist aspect of what they do, and by converting their data to testimonials, video diaries, and social media-related attitudes. Lastly, in Ilan Manouach's conceptual books "Harvested" [26], and "The Cubicle Island: Pirates, Microworkers, Spambots, and the Venatic Lore of Click Farm Humor" [27], the artist plays with the ramifications of digital labor as it echoes silently through the products of a workforce, both human and non-human (as many times the workers create bots to automate part of the process).

2.2 Musical References

Many of our composite soundscapes and visual fields "brush information against information" [9] in the Cagean sense, inviting performers and viewers to perceive new relationships between the short data logs through musical and gestural language. In many of our vignettes, chance and simultaneity emerge in ways that are similar to works of not just John Cage [9] and Christian Wolff [45], but also to musicians whose practice has been connected to the sounds of the environment such as Hildegard Westerkamp [43] and Maryanne Amacher [3]. As the technology behind the voice assistants changes and becomes more and less predictable, a lopsided indeterminacy develops. To further extend and reimagine the historical experimental tradition, we take one excerpt of Kurt Schwitters' iconic early 20th century work of sound poetry, Ur-Sonate [39], and ask the smart speaker to recite and translate the words.

Other more contemporary musicians are also engaging with data and surveillance capitalism. Composer Kaley Lane Eaton recently completed an opera for voice, piano, and electronics that sets the text of Shoshanna Zuboff's recent book. We found commonalities between our project and the work of Jonathan Reus and Joana Chicau, specifically their project *Anatomies of Intelligence* [37]. A live performance of a web-based catalog, this project features sonified machine learning processes of data classification. Our project differs from the above works in that it is multi-disciplinary and prismatic in its approach to voice assistant data; we see these musical explorations as one element in a multitude of possibilities.

2.3 Voice assistants in HCI and design research

In recent years, researchers in human-computer interaction (HCI) have investigated the ways voice assistants are interwoven into everyday life (e.g. [19, 34]), demonstrating that the main patterns of usage are oriented around playing music, searching the web, asking the time or setting an alarm, and sending commands to other connected devices [4, 6]. Other studies have highlighted patterns of speech around intimate and health related questions (e.g. [12]). Within these works, values such as trust [20], respect [40], and management of expectations regarding the true potential for conversation [36] are surfaced as important guiding principles for the design of future conversational agents. HCI and design researchers have also taken speculative approaches to envision future interactions with voice assistants. For instance, Cambre et al. [10] asked 149 people to complete story prompts about interactions with voice assistants set in 2050 while Wirfs-Brock et al. [44] asked participants to imagine the future of voice assistant interactions using Spotify data logs as a design resource. Other projects speculatively worked with voice assistants from a non-human perspective [30], through exploring Alexa's whispering function [31] and via the role of personality in voice assistants [38].

Furthermore, scholars in HCI and science and technology studies (STS) have expressed concerns regarding the relation between AI and ML systems with regards to human labor [18, 23], surveillance capitalism [48], privacy and security [2, 8, 29], and creepiness [32, 33, 46]. Additionally, design researchers also study how to create explainable or transparent AI [1, 16, 25, 41] and how to make data trails more visible or actionable [17, 42]. The Voices and Voids project builds on what was articulated in these works by creating space for these themes to be experienced first-hand. It continues to build on this corpus of experimental works with the intention of disrupting common narratives which promise seamless and frictionless interactions with voice assistants, subverting their use in subtle ways, while transcribing data logs generated over years of interaction.

3 OUR PROJECT: PROCESS, TOOLS AND MATERIALS

3.1 Our approach

Our goal with the Voices and Voids project was to use a transdisciplinary artistic approach to reveal hidden characteristics and innerworkings of voice assistant systems. Our team is comprised of a new media artist (Afroditi), a speaking percussionist (Bonnie), and an interaction designer (Audrey). Without a lot of prior experience using voice assistant systems, we engaged in the project with a sense of curiosity and openness. We committed early on to researching various parallel lines of inquiry, knowing that different media and approaches could be combined to cover more ground in exploring these devices and the data they generate. As a result, many of the themes that became central to our project emerged organically, as we were getting acquainted with the devices and their

data. The outcome of the project is structured in 12 experiments—or vignettes—each exploring voice assistants and their data. These experiments can stand on their own, but when seen together as a collection they offer a rich painting of narrative, embodied, textual, and sonic interpretations of voice assistant systems. These explorations examine topics of (1) intimate and mundane interactions, recognizing that the sheer volume of archived personal data is what drives this apparent close relationship between user and assistant; (2) the hidden labor behind ML systems supporting voice assistants, which involves understanding the degree to which things are automated, grappling with the amount of human labor performed by ghost workers [23]: individuals performing crowdsourced tasks via Amazon's Mechanical Turk (MTurk) to train these systems; (3) the creative potential of voice assistants, where we translated a large volume of voice interaction data into a series of performative experiments and took the devices apart to reveal what is physically hidden; and (4) the surveillance capitalism these devices are a part of, connecting our local experiments to massive infrastructure.

Our approach combines the methodological commitments of research-through-design [21, 47], practice-based research in the arts [11], and performance as research [5]. By directly making, hacking and performing, we uncovered what is hidden in voice assistant systems. We—Afroditi, Bonnie and Audrey—met weekly starting in October 2019; the first three months were as much about learning how to communicate across our disciplines as it was about generating content. The voice assistants themselves were fickle partners, and it took time to learn their quirks while exploring and developing the "skills" of each device. Then, we spent an intensive three days with Laura Devendorf and Jordan Wirfs Brock from UC Boulder, artist/researchers in HCI and design. Much of this time was structured as group-improvisation performance as research, punctuated by time spent teaching one another technology and creating tools for real-time performance. This time culminated in a 45-minute test performance in January 2020.

To better facilitate our interdisciplinary collaboration, in January 2020, we initiated a directed research group comprised of 5 students at our university from various disciplines including design, music performance and composition, and digital art/experimental media. In addition to creating some of their own responses to the project, this group served as a testing ground for our new ideas, interrogating them and offering crucial feedback. Members also helped us to construct key elements of the project, including a neural network poetry generator (Gabrielle Benabdallah), a bespoke graphical user interface (GUI) that allowed for real-time manipulation of data in performance (James Wenlock), and construction of additional versions of the wearable voice-activated percussion instruments (Esteban Yosef Agosin).

We were preparing for further in-person collaborations with visiting artists and an eventual live performance and installation event when the Covid-19 pandemic hit. This changed environment inspired us to pivot into an entirely virtual medium. While this was an initially disappointing change, the process of working from home while truly living with these devices highlighted the quirky intimacy that comes from interacting with them. Many of the sounds, gestures, and images we settled on in the spring and summer of 2020 embodied the "everyday" nature of the project, weaving together found objects, sounds, and images. Work with choreographer Kate

Sicchio, as well as with composers Yiheng Yvonne Wu and Danny Clay was entirely remote, through a combination of Zoom meetings and carefully planned asynchronous work. Kate Sicchio's contribution was to generate a movement score for us, and to explore the possibilities of embodying the voice assistant data logs. Much of Yheng Yvonne Wu's contribution to the project was through the creation of a structural framework for some of the narrative-based pieces. Highlighting arrival points and themes, she outlined a story told through sounds, still images, drawn images, written words, and video. Danny Clay worked to creatively illustrate this structure through musical directives in an event/text score plus an illustrated video, while Bonnie worked to create and realize specific percussive sounds that supported the framework set up by the composers.

We presented drafts of vignettes in June 2020 for a small group of artists from various disciplines. We took much of the summer and fall to re-record, document, and imagine and develop a website. In November of 2020, we launched the website via the Jacob Lawrence Gallery at the University of Washington in Seattle, USA, demonstrating parts of the site, talking about the work, and performing two pieces live via Zoom. In the future, we could imagine a live, in-person, interdisciplinary performance format, but this version would likely be quite different from our online performance.

3.2 Amazon Mechanical Turk data collection

One of our central creative materials was Google Assistant voice data. We wanted to work with a variety of logs that encompassed a diversity of voice timbre, background noise, and types of interactions with voice assistants. As a way to acquire these logs, we chose to follow Bentley et al.'s [6] method of paying MTurk workers to share their data logs. On MTurk, we created a Human Intelligence Task (HIT)—a simple activity that MTurk workers will complete—with detailed instructions (11 steps) to download one's voice data archive from Google. We asked that workers upload the audio files as well as the JSON file. For each command or question asked, the JSON file included the transcribed text (e.g. 'What time is it?'), a timestamp, the name of the mp3 file for the audio, and a 'header', which is the product the command was captured on (e.g. the Google App, the Voice assistant, etc.). We were upfront about the fact that their data would be used in an artistic performance project, and that we would protect their anonymity[1]. We paid the workers $10 USD for their logs. We gathered 26 archives which ranged from a few dozen to over 5000 utterances. Of the 26 archives, some were not in English, some were too short, and some were not properly downloaded. We ended up using about half of the archives. As part of the instructions, we noted that workers were welcome to delete any voice utterance they were not comfortable sharing, giving them the opportunity to have control of what they shared. However, due to the length of some archives, it would be extremely time consuming to comb through all the audio data and difficult to read through all the meta data of the JSON file. While this was the 'best practice' used by Bentley et al. [6], we believe that we need to revisit this data collection practice to offer better agency to MTurk workers. In addition, we included Afroditi's voice assistant archive. Audrey

---

[1]We discussed the project extensively with our Human Subjects Division office. In this artistic project, MTurk workers are not considered 'human subjects' but contributors, hence not requiring further action.

and Bonnie didn't have a practice of using voice assistants, so they didn't have data to share.

We chose the MTurk service to collect this data because we were interested in seeing and engaging with the people who provide the invisible labor of improving voice assistants through crowdsource work. By asking for their voice interaction histories, we also got to know more about who they are in their everyday life through learning about their living situation, family life, and their interests and needs. With a different HIT, we also gathered 2-minute video testimonials of 8 MTurk workers in which they described their experiences working on the training of voice assistants. We paid workers $10 USD for that task as well. While not all videos were as detailed as we had hoped, some videos were striking, explaining how some of the jobs require MTurk workers to repeat sounds and vowels.

3.3 Bespoke Graphical User Interface

Due to the large volume of data collected on MTurk, we needed a tool that could filter through the data archive. Using the algorithmic sound synthesis platform SuperCollider, James Wenlock (as part of our directed research group) built a bespoke graphical user interface (GUI) that would allow us to search and play audio clips from the most evocative entries in our archives. The system, named Turk Interpreter, allows a user to search through either individual archives, or through all the archives for specific terms or phrases using keywords. The result is a list of all utterances using that term or phrase, which can then be played back in their original recording or via a text-to-speech function. The GUI also includes effects like delay and decay filters, envelopes, and playback speed. We added the capacity to connect the GUI with TouchOSC allowing for the playback of the audio files to be triggered by using the accelerometer of an iPhone.

3.4 Recurrent Neural Net (RNN)

During our directed research group, Gabrielle Benabdallah experimented with the creation of a Recurrent Neural Net (RNN) trained on Google voice assistant data logs, sourced from MTurk workers. Namely, we used textgenrnn, a Python 3 (Keras) module interfacing with TensorFlow to generate poetry based on a curated selection of data from the MTurk workers. One of the features of this tool was the experimentation with different "temperatures", leading the AI model to stay closer to (lower temperature) or further from (higher temperature) the original training data set. The output of the neural net ranges from utterances heavily represented in the training datasets ("set an alarm for 9 a.m.") to highly idiosyncratic and even aberrant sentences ("find me to high" or "how many steps is a fire then I don't know").

3.5 Hacked percussion instruments and wearable voice assistant

To explore the idea of embodiment and disembodiment of the smart speakers, we created a series of hacked percussion instruments including a spring drum, a cowbell hit by a solenoid, a pair of agogo bells hit by a small wooden stick attached to a servo motor, and four suspended bells attached to a DC motor. These instruments,

originally from Bonnie's musical battery, were enhanced with electromechanical actuators. The main controller of the actuation was a Google AIY Voice Kit which is comprised of a Raspberry Pi Zero WH with a Voice Bonnet shield running custom Python code that uses the Google Assistant and Speech-To-Text APIs. In this custom Python script, the Alexa wake words together with a collection of other words trigger these percussive instruments.

Initially conceived as sculptural objects, we subsequently incorporated them into a wearable interface to be worn while performing (Fig 1). The wearable voice assistant system consists of the Google AIY voice kit embedded on the waist part of a modular wearable interface which also includes two bell actuators and led lights, on the left arm and the right leg. The performer prompts the wearable voice assistant through a collection of curated phrases from the collected data logs commanding the ML system to trigger the actuations. These utterances translate to a certain number of hits, and spins of the bells, while also corresponding to the movements of the arms and legs.

3.6 GIFs and Hypertext

Our project shifted into the virtual realm, mainly the worldwide web, because of the Covid-19 pandemic. While many of our performance experiments were documented as home videos, we still felt that there were elements of the project that could not be communicated through video. Inspired by Internet and Post-Internet Art, we employed looping moving images using the GIF format, as well as Hypertext for engaging hands-on with our experiments in an online format. In our website, a variety of GIFs act both as functional buttons to navigate to each vignette, but also as playful short narratives that communicate the content of each section (for more on our making and use of GIFs, see [7]). As seen in early Net Art, the intentional ambiguity and unexpected nature of GIFs reflects the ethos of the project itself, highlighting the double nature of voice assistants as devices that represent both efficiency and limitation. The Hypertext format is used in different instances to generate narrative. In the vignette **From Your Device** it serves as a compositional tool for the user to perform with selected instances of voice data in a non-linear way, while in **Training Session**, it is used in a linear time-based manner for the user to explore the sonic qualities of vocalizations used to train these ML systems.

4 VOICES AND VOIDS VIGNETTES

In this section, we offer descriptions of each vignette from Voices and Voids grouped in four main tactics we used to reveal what is hidden in voice assistant systems. An annotated map of the website with a list of the vignettes, and a website video walkthrough are accessible in the ACM Digital Library for reference.

4.1 Intentional skewing through errors and glitches in automation

The following vignettes share a goal of intentionally pushing the AI models to their limits, causing the voice assistants to glitch or respond via errors. The use of unexpected commands or questions, as well as nonsensical words and sentences is prominent in **Chit Chat** and **UrSonate**. In **Play the Earth** and **Training Session**, we purposefully play with the training process of the AI model for

Figure 1: Wearable voice assistant system

voice assistants, with the aim of both revealing the human labor behind voice assistant intelligence, and highlighting the large gap between human intelligibility and machine intelligence.

In **Chit Chat,** custom scripts (written with texttovoice.io) are spoken by four Amazon Echos and three Google Home Mini devices, which respond to one another with limited human intervention. Each of the devices responds to a different wake word so that they can be prompted into the conversation. For the Amazon products, our texttovoice.io scripts elicit responses to different topics (i.e., location, date, the shopping list items, etc.) but also ask vaguer, more political, questions interrogating the type of answers that these automated systems are programmed to provide. The scripts also alter the normal voice and speech quality of Alexa (whispering or emphasizing phrases, and employing individual accents, both female and male.) The Google Home Minis were prompted by the phrase "Ok Google, talk to a Turk worker", which activated a script that played anonymized MTurk workers voices. Additionally, we created a series of hacked percussion instruments that are activated by trigger words from the assistants, creating a rich soundscape that is linked to the conversation between them.

In **Chit Chat**, we noted that an underlying purpose of these devices is to advertise Amazon or Google products. For example, many times the conversation between the assistants would arrive at playing a random song, or reading an audio book, and then ask us to purchase it. An assistant would often make suggestions based on our individual search history, making it obvious that these devices perform surveillance and record every interaction.

Using the Speech-to-Text Alexa skill and an Amazon Echo alongside a human performer, we realized an excerpt of Kurt Schwitters' early 20th century sound poem **UrSonate**. The poem was read to Alexa's Voicepad, transcribing it from sound to text. It was then spoken by the assistant during the performance with the text to voice function. This vignette reveals the tendency of the application to advertise as it processes the nonsense words, highlighting certain words that the model has been trained to speak (Fig. 2). When compared to the service OtterAI (a transcription service not related to Amazon), it was interesting that the transcribed text had some similarities, but none of the intersections with Amazon ("Bezos", "Whole Foods", etc.). John Cage wrote that "an error is simply a

Figure 2: Comparison between excerpts from the original UrSonate text, vs the Alexa Voicepad translation

failure to adjust immediately from a preconception to an actuality" [9]; what started as a playful misunderstanding was revealed to be 'mis-hearings' intentionally skewed to direct users toward Amazon products.

In **Play the Earth**, we present a poem generated with a Recurrent Neural Net (RNN) trained on Google voice assistant data logs, sourced from MTurk workers. This poetry uses the interaction records as source material. The result is a poem in which recognizable everyday lines are interwoven with phrases that have an eerie sense of familiarity, and yet are not actual commands or questions for voice assistants (Fig. 3). The way the interaction logs are revised and recombined by the neural net calls attention to the patterns and variations in everyday human-AI relationships. The vignette opens the door to reflections on how human communication with machines is increasingly interfaced with other technical systems, in both apparent and digressive ways.

In the vignette **Training Session**, performers repeat (sometimes nonsensical) words and sentences they see on screen, in the same way that MTurk workers continuously train artificial intelligence models for voice assistants. We were particularly inspired by one MTurk worker who explained that her job, at times, had been to repeat vowels she sees on screen: ooooo, aaaaa. The vignette

```
set an alarm for 1 hour
what is the stardew Valley
thank you
tell me a joke
play Monu Marily Gary
news
what's the weather like tomorrow
set an alarm for 1 hour
what has beer
are you
what is the sound of the morning
what is the best wolfenstein 2 locked in a panorian battlefield 1
news
what's the temperature outside
news
what's the square root of 8
```

Figure 3: Excerpt from the neural net poem Play The Earth

includes four stages of training, starting with vowels, and then using utterances produced by the Recurrent Neural Net (RNN) described in 3.4. As the stages progress, the "temperature" augments, leading to more nonsensical phrases. Composer Danny Clay mixed the performers' audio files to create short compositions that push the viewer to reflect on the origin of the intelligence of these voice assistants.

4.2 Emphasizing intimacy and personal narratives

While the vignettes in 4.1 focused mostly on the live AI models of voice assistants, and their limits, the following vignettes offer a deep dive into the ways that voice assistant data can hold and reveal personal, and ultimately very intimate parts of one's life. Some of the archives we obtained from MTurk workers spanned as long as four years, opening the potential for a long-term understanding of someone's relationship with a voice assistant as in **J_'s story**. In contrast, **From Your Device** juxtaposes many data from various users, while **Garden of Pots** presents soundscapes from the homes of MTurk workers, entering their intimate work spaces.

'**J_'s story**' uses voice interaction data from one user's Google Assistant to construct a story. The data excerpts we chose are unedited, except for some details (specific names and locations) that were changed to preserve privacy. The sheer volume of the log contributed to an intimate portrait of an individual throughout four years of data. Through just the user's daily voice assistant

interaction directives, listeners can glean personal events (car troubles, medical issues), trace this individual's relationship to current events and political movements, and even analyze relationship history through the number of directives to call or text specific people. The narrative is advanced by found-object and traditional percussion, illustration, timeline, and voice. A visual as well as a sonic exploration of data, we illustrate events on timelines, assign specific sounds to abstract ideas, and allow percussion instruments to occasionally stand in for the human voice. Created during lockdown, the video was made in Bonnie's home, highlighting rather than concealing this usually private environment.

From your device is an interactive page on the Voices and Voids website inspired by the hyperlink quality of net art. The page is an assemblage of Google Assistant utterances from the data logs sourced from MTurk workers who use Google assistants in their daily lives. As a visitor scrolls down the page, they can read a selection of phrases and questions captured by Google Assistant, ranging from "Are you an intelligence?", to "Does the rabbit feel the emotion of love?", "How many centimeters in an inch?", and "Do you love me?" Each phrase is clickable—when clicked, the original audio file is played. The page displays an array of voices that can be played in sequence or simultaneously, allowing visitors to create their own soundscapes, leading to a cacophony of intimate themes, ranging from simple, mundane tasks like setting a timer or playing music, to philosophical or ontological questions, and deeply intimate thoughts of care, love and sex. The experience is further amplified by using the original sounds clips, revealing the diversity of voices recorded, giving away hints about who might be talking with voice assistants, through the tone of their voice, their accents when speaking English, or background noises.

In addition to exploring intimacy between users and voice assistants, we connected directly with the MTurk workers training voice assistants. Our intention was to reveal what kind of work they do and to contrast this with who they are, and where they do their work. In the interactive page **Garden of Pots**, clay pots hide voice assistant devices. Once clicked, each pot lifts, visually uncovering the voice assistant and playing a soundtrack of ambient sounds. Composer Danny Clay selected ambient noises from voice assistant data logs as well as from the interviews with MTurk workers to create these soundscapes. In both cases, the audio was often the sound of home or a familiar space. The soundscapes offer nuanced portrayals of home, including elements of walking, insects, water, traffic, noise, typing, rain, and more. This vignette highlights elements of data captured, but not processed by natural language

Figure 4: Screenshots from J_'s story
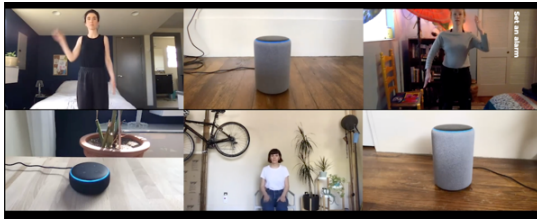
Figure 5: Set An Alarm composition: a grid-like format creating a tiled video of three human performers and three machines

processing systems—while Amazon or Google are focused on translating sound files to text to parse commands or questions, we were interested in attending to the other 'hidden' information that could be interpreted in these audio clips.

4.3 Translations through embodiment as a tactic for defamiliarization

The following vignettes present experiments in translation from the text and audio format of the voice assistant data and sound, to movement in **Set an Alarm**, and to video and sound in **Your Hair is Very Wonderful** and **How wide is the average fingernail**. In each case, the voice assistant systems and data are transcoded and cast in a new setting, defamiliarizing the relationship humans may have with voice assistants. While voice assistants are often part of everyday life as disembodied voices, the vignettes below contrarily used embodiment as an approach to engage with them.

Set An Alarm is a movement score responding to voice assistant commands. Reversing the roles of the device and the human, this vignette asks performers to connect gestures and sound to directives to 'set an alarm' from the smart speakers. The alarms are extracted from MTurk workers' Google Assistant archives, and then scripted for Alexa using texttovoice.io by media artist and choreographer Kate Sicchio. This score was interpreted at home over Zoom during the first Covid-19 pandemic lockdown. In the script, Kate Sicchio experimented with the pacing and the rhythm of these alarm commands. She choreographed two basic postures—standing and sitting and corresponding arm movements: swirling of arms from the elbows to the hands while standing and slapping our legs with our hands while sitting. The score specifies a duet between human and machine, in which the machine is speaking the commands and the human moves accordingly. Each time the score specifies a change of command, the performer must change posture, and must swirl their arms, move left and right, or clap their hands as many times as the time in the alarm, i.e. if Alexa says "Set an alarm for 3 minutes" the performer must move their hands three times, etc. This vignette explores the idea of agency in automated AI systems by inverting the roles and making the machine command the action of the human. This vignette also highlights the underlying characteristics of Zoom as a platform, since the embedded noise cancellation cuts off part of the sonic qualities of clapping. There is an intentionally arhythmic quality to the final version of the vignette, which references the lag of real time communication platforms such as Zoom, as well as the

inefficiency of human performers interpreting commands (which manifests in doubt and confusion every time there is a change of command), versus the straightforward interpretation of the score by the machine which is entirely scripted and spoken without errors.

Your Hair is Very Wonderful and **How wide is the average fingernail** were both conceived as translation exercises where utterances are translated to movement and sound within a specific spatial context. Composer Danny Clay asked artists to assign specific sounds, physical gestures or objects found in the home to excerpts from the collected voice assistant data. Building on the set of utterances as a score, and working from home during the pandemic, we captured performance video. This vignette shows the process of giving a new spatial and sound context to the utterances.

4.4 Improvisation and Juxtaposition

Finally, we share two vignettes that use improvisation and the juxtaposition of tools and instruments to perform.

In **Play Feeling Good Music**, performers improvise with percussion, spoken word, voice assistants, curated Google assistant data, our bespoke GUI, and the wearable voice assistant with handmade actuators. These voice assistants were trained to respond to the words "body", "feeling", "care", "being" and "intelligence". The resultant soundscape and visual field invite performers and viewers to perceive new relationships between the short data logs and the musical and gestural language. While playing with the idea of constructing a body for the disembodied voice assistant, we stumbled upon the idea of embodying the voice assistant, creating a hybrid performative entity—part human, part machine—which replies to voice commands through actuation of both the body and the hacked instruments. A feedback system is thus created between voice and body, of both human and machine parts, which in its turn is communicated to the rest of the performers over Zoom to trigger improvised responses from the live percussive instruments, and the playback of the data logs through the GUI. This vignette presents another example of training voice assistants for our own purposes. Using the Google Assistant algorithm and choosing new trigger words, we changed the meaning of the training, thus eliciting a musical and gestural response rather than one connected to speech. Instead of an answer gleaned from the Internet, we were able to set up structures of predictability to be used in an improvisational performance setting.

Homebodies consists of an interactive movement and sound improvisation duet based on Google Voice Assistant data logs collected by MTurk workers. The bespoke GUI (described in 3.3) is used to control the voice utterances, and their playback is triggered by accelerometer data from a smart phone via the app TouchOSC on a dancer's body. This vignette plays with the relationship between the space of the home and the technologies of the home by means of improvisation. We used an iPhone to send accelerometer data as OSC messages that control the playback function of a SuperCollider script manifested through a GUI running on the computer of the dancer using the phone. Through Zoom's remote screen access function, Afroditi is controlling the GUI by typing keywords to curate a selection of data logs. While one performer can control all these elements, the dancer can trigger the playback through their movement, by sudden changes of posture, or by quick movement

in their physical space which affects the accelerometer data. The resulting duet is an improvisation based on what is happening in the home space of the dancer, which in the case of the recording included interactions between the dancer and their child and partner. Thus, Afroditi chose words like the "kid", "home", "family" etc. to choose data clips to play back via the GUI, weaving together a complex grid of the intimate lives and spaces of the people whose logs we sourced, together with the real time events that are happening while the improvisation was taking place, and revealing multiple temporalities, agencies, interactions, and relationships. This also highlights the potential and limitations of the tools we were using, unified in the surreality of living at home during a time of global pandemic, and the type of artwork we can do with the elements we have at hand.

5 DISCUSSION

Below, we discuss how the vignettes, while wildly different, offered a rich interwoven web to explore voice assistant data. We also offer a reflection on our tactical decisions in terms of working in interdisciplinary ways, with artists in residence, as well as with MTurk workers.

5.1 Interwoven data-driven experiments in art, design and percussion as subversive tools

In this paper, we presented 12 experiments which reveal hidden elements of voice assistant systems. Our work highlights the intimate and mundane relationships people have with voice assistants, the hidden labor in AI and ML, the creative potential of voice assistant data, and issues of surveillance capitalism. These themes, surfaced across many vignettes, often overlap across and between the tactics we described above. Our experiments often had one or two central foci (e.g. J_'s story was designed to showcase the astonishingly intimate details of everyday life, while Training Session examines the human labor in training voice assistants). However, when taken together, the experiments create a multi-layered experience which confronts viewers with a complex web of hidden matters of concern related to voice assistant systems. This back and forth between matters of labor, intimacy, everyday life, surveillance capitalism and creative potential creates an unsettling rhythm where it becomes impossible to just 'use' a voice assistant, or to just 'have one at home'.

An integral component of this work is that while each experiment often explored a predominant theme, it was impossible to ignore the other matters of concern central to the work as a whole. For instance, when focusing on mundane interactions, such as using a Text To Voice function, we inevitably also encountered surveillance capitalism (as seen in UrSonate). Or when aiming our focus on the hidden labor for training voice assistants, we didn't only look at MTurk workers as workers, but we considered more comprehensive elements of their human experience were also revealed in hearing everyday sound backgrounds (e.g. in Garden of pots). When focusing on the creative potential of data logs, intimate narratives naturally emerged (e.g. in J_'s story). By using techniques of critical making [35] and hands-on hacking to take apart these systems in both literal and metaphorical ways, we did not only uncover unique themes, but also how they relate to one another in this complex entangled mesh, perhaps pushing further than previous works cited in section 2.

5.2 Employing interdisciplinary practices

We embraced the messiness of our chosen materials by frequently choosing to employ interdisciplinary methodologies. Rather than focusing on a single set of best-practices, we created parallel experiments across our disciplines that specifically probed the limitations of these smart speakers. These limitations—which might be seen as "mistakes" or "failures" in other contexts—are the cracks in the system that expose the framework of surveillance capitalism embedded in the mechanics and code of these devices. Learning from one another's disciplines and actively employing new techniques gleaned from each artist's personal practice allowed us to interact with the machines and the data in productive ways. In addition, our own interdisciplinarity was further multiplied when artists in residence and students from the DRG joined the project at various points in the process. Our parallel experiments allowed for this expansion of the team, and created a space for new ideas to emerge and for new voices to be added to our already prismatic project.

In discussing working with others in interdisciplinary ways, we also reflect on how we worked with MTurk workers. Obtaining, organizing, filtering and working with our 26 MTurk workers' data sets was ethically challenging and provoked a variety of tensions and questions regarding long term data collection via smart devices. Navigating MTurk's platform (and the complex algorithmic system that runs it) for the first time raised many concerns about the ethics behind employing MTurk workers, acquiring their data and compensating them in a fair manner. While our interdisciplinary approach was fruitful for curiously exploring various lines of inquiry at once, when working with MTurk workers data we all followed a clear set of commitments: we would protect the workers' identity by redacting names and addresses or using pseudonyms.

6 CONCLUSION

In this paper, we have presented the Voices and Voids project: a multidisciplinary artistic research project that explores voice assistant systems and data through parallel and interwoven experiments in art, design and percussion. The resultant artifacts, both physical and virtual, reveal the inherent materiality of these ML systems, whether we are talking about the human beings that constitute the Artificial Artificial Intelligence systems and the surreal tasks that are associated with their hidden labor, the immense amount of processing power and data centers that are required to store these vast interactions, or the repetitive and mundane nature of our interaction with voice assistants. Through these practices of using data as creative material, we engage directly with inverting the roles of the devices from observers to actors, generating new human and non-human relationships, subverting the meaning of these assistive devices to cultural artifacts, carriers of multi-layered meanings, and symbols that challenge the technological solutionism of the era in which we inhabit.

students who collaborated in the Directed Research Group. And thank you to Jordan Wirfs Brock and Gabrielle Benabdallah for their editorial eyes on this paper.

REFERENCES

[1] Ashraf Abdul, Jo Vermeulen, Danding Wang, Brian Y. Lim, and Mohan Kankanhalli. 2018. Trends and Trajectories for Explainable, Accountable and Intelligible Systems: An HCI Research Agenda. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.* Association for Computing Machinery, New York, NY, USA, 1–18. Retrieved February 15, 2021 from http://doi.org/10.1145/3173574.3174156

[2] Hazim Almuhimedi, Florian Schaub, Norman Sadeh, Idris Adjerid, Alessandro Acquisti, Joshua Gluck, Lorrie Faith Cranor, and Yuvraj Agarwal. 2015. Your Location has been Shared 5,398 Times! A Field Study on Mobile App Privacy Nudging. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems.* Association for Computing Machinery, New York, NY, USA, 787–796. Retrieved February 15, 2021 from http://doi.org/10.1145/2702123.2702210

[3] Maryanne Amacher. Maryanne Amacher - Monoskop. Retrieved February 17, 2021 from https://monoskop.org/Maryanne_Amacher

[4] 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 (April 2019), 1–28. DOI:https://doi.org/10.1145/3311956

[5] Annette Arlander, Bruce Barton, Melanie Dreyer-Lude, and Ben Spatz (Eds.). 2017. *Performance as Research: Knowledge, methods, impact* (1st edition ed.). Routledge, New York.

[6] Frank Bentley, Chris Luvogt, Max Silverman, Rushani Wirasinghe, Brooke White, and Danielle Lottrjdge. 2018. Understanding the Long-Term Use of Smart Speaker Assistants. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 3 (September 2018), 1–24. DOI:https://doi.org/10.1145/3264901

[7] Heidi R. Biggs, Cayla Key, Audrey Desjardins, and Afroditi Psarra. 2021. Moving Design Research: GIFs as Research Tools. In *Proceedings of the 2021 Designing Interactive Systems Conference* (DIS'21), ACM Press, 1–14.

[8] Simone Browne. 2015. *Dark Matters: On the Surveillance of Blackness* (Illustrated edition ed.). Duke University Press Books, Durham.

[9] John Cage. 2015. *John Cage: Diary: How to Improve the World (1965)ghost* (n edition ed.). Siglio, Los Angeles, Calif.

[10] Julia Cambre, Samantha Reig, Queenie Kravitz, and Chinmay Kulkarni. 2020. "All Rise for the AI Director": Eliciting Possible Futures of Voice Technology through Story Completion. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference.* Association for Computing Machinery, New York, NY, USA, 2051–2064. Retrieved January 25, 2021 from http://doi.org/10.1145/3357236.3395479

[11] Linda Candy and Ernest Edmonds (Eds.). 2012. *Interacting: Art, Research and the Creative Practitioner.* Libri Publishing.

[12] Eugene Cho. 2019. Hey Google, Can I Ask You Something in Private? In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (CHI '19), Association for Computing Machinery, New York, NY, USA, 1–9. DOI:https://doi.org/10.1145/3290605.3300488

[13] James Coupe. JAMES COUPE | Time-Clock. Retrieved February 9, 2021 from http://jamescoupe.com/?p=2476

[14] James Coupe. JAMES COUPE | Let Us Pray. Retrieved February 9, 2021 from http://jamescoupe.com/?p=2439

[15] James Coupe. JAMES COUPE | General Intellect. Retrieved February 9, 2021 from http://jamescoupe.com/?p=2093

[16] Kate Crawford and Vladan Joler. 2018. Anatomy of an AI System. *Anatomy of an AI System.* Retrieved February 21, 2021 from http://www.anatomyof.ai

[17] Audrey Desjardins, Heidi R. Biggs, Cayla Key, and Jeremy E. Viny. 2020. IoT Data in the Home: Observing Entanglements and Drawing New Encounters. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (CHI '20), Association for Computing Machinery, Honolulu, HI, USA, 1–13. DOI:https://doi.org/10.1145/3313831.3376342

[18] Catherine D'Ignazio and Lauren F. Klein. 2020. *Data Feminism.* MIT Press.

[19] Stefania Druga, Randi Williams, Cynthia Breazeal, and Mitchel Resnick. 2017. "Hey Google is it OK if I eat you?": Initial Explorations in Child-Agent Interaction. In *Proceedings of the 2017 Conference on Interaction Design and Children* (IDC '17), Association for Computing Machinery, New York, NY, USA, 595–600. DOI:https://doi.org/10.1145/3078072.3084330

[20] Justin Edwards and Elaheh Sanoubari. 2019. A need for trust in conversational interface research. In *Proceedings of the 1st International Conference on Conversational User Interfaces* (CUI '19), Association for Computing Machinery, New York, NY, USA, 1–3. DOI:https://doi.org/10.1145/3342775.3342809

[21] William Gaver. 2012. What should we expect from research through design? In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems* (CHI '12), ACM, New York, NY, USA, 937–946. DOI:https://doi.org/10.1145/2208516.2208538

[22] Wesley Goatley. Chthonic Rites | Wesley Goatley. Retrieved February 9, 2021 from https://www.wesleygoatley.com/chthonic-rites/

[23] Mary L. Gray and Siddharth Suri. 2019. *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass* (Illustrated edition ed.). Houghton Mifflin Harcourt, Boston.

[24] Garnet Hertz. Experiments in Surveillance Capitalism. Retrieved February 9, 2021 from http://conceptlab.com/surveillancecapitalism/

[25] Joseph Lindley, Haider Ali Akmal, Franziska Pilling, and Paul Coulton. 2020. Researching AI Legibility through Design. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (CHI '20), Association for Computing Machinery, New York, NY, USA, 1–13. DOI:https://doi.org/10.1145/3313831.3376792

[26] Ilan Manouach. Harvested – Ilan Manouach. Retrieved February 9, 2021 from https://ilanmanouach.com/project/harvested/

[27] Ilan Manouach. The Cubicle Island – Ilan Manouach. Retrieved February 9, 2021 from https://ilanmanouach.com/project/cubicleisland/

[28] Lauren Lee McCarthy. SOMEONE — Lauren Lee McCarthy. Retrieved February 9, 2021 from https://lauren-mccarthy.com/SOMEONE

[29] Nicholas Micallef, Mike Just, Lynne Baillie, and Maher Alharby. 2017. Stop annoying me! an empirical investigation of the usability of app privacy notifications. In *Proceedings of the 29th Australian Conference on Computer-Human Interaction* (OZCHI '17), Association for Computing Machinery, New York, NY, USA, 371–375. DOI:https://doi.org/10.1145/3152771.3156139

[30] Iohanna Nicenboim, Elisa Giaccardi, Marie Louise Juul Søndergaard, Anuradha Venugopal Reddy, Yolande Strengers, James Pierce, and Johan Redström. 2020. More-Than-Human Design and AI: In Conversation with Agents. In *Companion Publication of the 2020 ACM Designing Interactive Systems Conference* (DIS' 20 Companion), Association for Computing Machinery, New York, NY, USA, 397–400. DOI:https://doi.org/10.1145/3393914.3395912

[31] Emmi Parviainen and Marie Louise Juul Søndergaard. 2020. Experiential Qualities of Whispering with Voice Assistants. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* Association for Computing Machinery, New York, NY, USA, 1–13. Retrieved January 25, 2021 from http://doi.org/10.1145/3313831.3376187

[32] James Pierce. 2019. Smart Home Security Cameras and Shifting Lines of Creepiness: A Design-Led Inquiry. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (CHI '19), ACM, New York, NY, USA, 45:1-45:14. DOI:https://doi.org/10.1145/3290605.3300275

[33] James Pierce, Sarah Fox, Nick Merrill, Richmond Wong, and Carl DiSalvo. 2018. An Interface Without A User: An Exploratory Design Study of Online Privacy Policies and Digital Legalese. In *Proceedings of the 2018 Designing Interactive Systems Conference* (DIS '18), ACM, New York, NY, USA, 1345–1358. DOI:https://doi.org/10.1145/3196709.3196818

[34] Martin Porcheron, Joel E. Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice Interfaces in Everyday Life. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI '18), Association for Computing Machinery, New York, NY, USA, 1–12. DOI:https://doi.org/10.1145/3173574.3174214

[35] Matt Ratto. 2011. Critical Making: Conceptual and Material Studies in Technology and Social Life. *The Information Society* 27, 4 (July 2011), 252–260. DOI:https://doi.org/10.1080/01972243.2011.583819

[36] Stuart Reeves. 2019. Conversation considered harmful? In *Proceedings of the 1st International Conference on Conversational User Interfaces* (CUI '19), Association for Computing Machinery, New York, NY, USA, 1–3. DOI:https://doi.org/10.1145/3342775.3342796

[37] Jonathan Reus. Anatomies of Intelligence. Retrieved February 9, 2021 from https://jonathanreus.com/portfolio/anatomies-of-intelligence/

[38] Jon Rogers, Loraine Clarke, Martin Skelly, Nick Taylor, Pete Thomas, Michelle Thorne, Solana Larsen, Katarzyna Odrozek, Julia Kloiber, Peter Bihr, Anab Jain, Jon Arden, and Max von Grafenstein. 2019. Our Friends Electric: Reflections on Advocacy and Design Research for the Voice Enabled Internet. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (CHI '19), ACM, New York, NY, USA, 114:1-114:13. DOI:https://doi.org/10.1145/3290605.3300344

[39] Kurt Schwitters. URSONATE de Kurt Schwitters. Retrieved February 19, 2021 from https://www.costis.org/x/schwitters/ursonate.htm

[40] William Seymour. 2018. How loyal is your Alexa? Imagining a Respectful Smart Assistant. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI EA '18), Association for Computing Machinery, New York, NY, USA, 1–6. DOI:https://doi.org/10.1145/3170427.3180289

[41] Ben Shneiderman. 2020. Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy. *International Journal of Human–Computer Interaction* 36, 6 (April 2020), 495–504. DOI:https://doi.org/10.1080/10447318.2020.1741118

[42] Stephen Snow, Awais Hameed Khan, Stephen Viller, Ben Matthews, Scott Heiner, James Pierce, Ewa Luger, Richard Gomer, and Dorota Filipczuk. 2020. Speculative Designs for Emergent Personal Data Trails: Signs, Signals and Signifiers. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (CHI EA '20), Association for Computing Machinery, New York, NY, USA, 1–8. DOI:https://doi.org/10.1145/3334480.3375173

[43] Hildegard Westerkamp. Hildegard Westerkamp. *Hildegard Westerkamp.* Retrieved February 17, 2021 from http://www.hildegardwesterkamp.ca/main/index.php

[44] Jordan Wirfs-Brock, Sarah Mennicken, and Jennifer Thom. 2020. Giving Voice to Silent Data: Designing with Personal Music Listening History. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (CHI '20), Association for Computing Machinery, New York, NY, USA, 1–11. DOI:https://doi.org/10.1145/3313831.3376493

[45] Christian Wolff. Christian Wolff:: Foundation for Contemporary Arts. Retrieved February 17, 2021 from https://www.foundationforcontemporaryarts.org/recipients/christian-wolff

[46] Richmond Y. Wong, Ellen Van Wyk, and James Pierce. 2017. Real-Fictional Entanglements: Using Science Fiction and Design Fiction to Interrogate Sensing Technologies. In *Proceedings of the 2017 Conference on Designing Interactive Systems* (DIS '17), ACM, New York, NY, USA, 567–579. DOI:https://doi.org/10.1145/3064663.3064682

[47] John Zimmerman, Jodi Forlizzi, and Shelley Evenson. 2007. Research through design as a method for interaction design research in HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '07), ACM, New York, NY, USA, 493–502. DOI:https://doi.org/10.1145/1240624.1240704

[48] Shoshana Zuboff. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (1st edition ed.). PublicAffairs, New York.