Democratizing Design and Fabrication Using Speech

Exploring co-design with a voice assistant

Andrea Cuadra
Cornell Tech
New York, NY, USA
apc75@cornell.edu

David Goedicke
Cornell Tech
New York, NY, USA
dg536@cornell.edu

J.D. Zamfirescu-Pereira
UC Berkeley
Berkeley, CA, USA
zamfi@berkeley.edu

ABSTRACT

Recent advances in artificial intelligence (AI) and voice interfaces have made possible the digital reincarnation of the draftsperson and craftsperson, capable of designing and producing custom work "to spec" for a reasonable cost. We present findings from an experiment using a Wizard-of-Oz-backed recreation of a voice assistant-based designer; participants describe a holiday ornament of their own imagination to the voice assistant, which elicits details and offers design choices before ultimately "creating" the ornament for the participant. For 8 out of 16 participants, a video channel allows the participant to see the designed object as it is iteratively designed and refined. We offer observations about the nature of the interactions between participants and the voice assistant, as well as quantitative measures of participant-reported cognitive load and predicted and actual satisfaction and accuracy of reproduction. Participants report greater satisfaction with their ornaments and experience reduced cognitive load if they can see them being designed; their expectations are higher if they cannot. 15 of 16 participants would use a voice assistant again for a design task.

KEYWORDS

CMC, Voice assistants, Wizard-of-Oz

ACM Reference Format:
Andrea Cuadra, David Goedicke, and J.D. Zamfirescu-Pereira. 2021. Democratizing Design and Fabrication Using Speech: Exploring co-design with a voice assistant. In *3rd Conference on Conversational User Interfaces (CUI '21), July 27–29, 2021, Bilbao (online), Spain.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3469595.3469624

1 INTRODUCTION

Historically, companies and professionals had physical objects created for them by working with machinists, woodworkers, and other skilled craftspeople [4, 7, 15]. This communication generally included prototypes, drawn images, measurements, and a conversation to better understand the need and the object being made [18]. In a consumer setting, a homeowner looking to update a kitchen might have asked a cabinetmaker to create a new cabinet or piece of furniture specific to the their needs. In a professional setting, a mechanical engineer might have asked a machinist to produce a cam, gear, or other part to exacting specifications, leaving the exact implementation (tool use, etc.) up to the machinist.

Mass production and the economic returns available to scale [3] have upended this old order, and the number of cabinet makers, woodworkers, and machinists has dropped as the relative costs associated with labor for custom work leave individual craftspeople unable to compete on price with Ikea-style mass-market products [5, 7].

The 4th industrial revolution and the advent of low-cost laser cutters and 3D printers offers the opportunity to return to a more bespoke world in which physical objects are made to individual specification [19]. A skilled designer can communicate with individuals in a low-jargon manner and translate end-customer needs into a sequence of steps for those fabrication tools. A skilled designer is also expensive, however, and the economic incentive suggests that we prepare for this role to be taken on by AI in some form. Indeed, our colleagues have already begun building AI assistants for visual design: Scones [8], for example, uses modern deep learning techniques to create sketches of natural scenes through conversation with a user. In this work, we ask:

What does a human-machine co-design interaction with an AI-powered voice assistant look like?

To understand exactly what that AI should be doing, we use Wizard-of-Oz techniques in a design task to investigate how users communicate a creative design goal with a voice assistant.

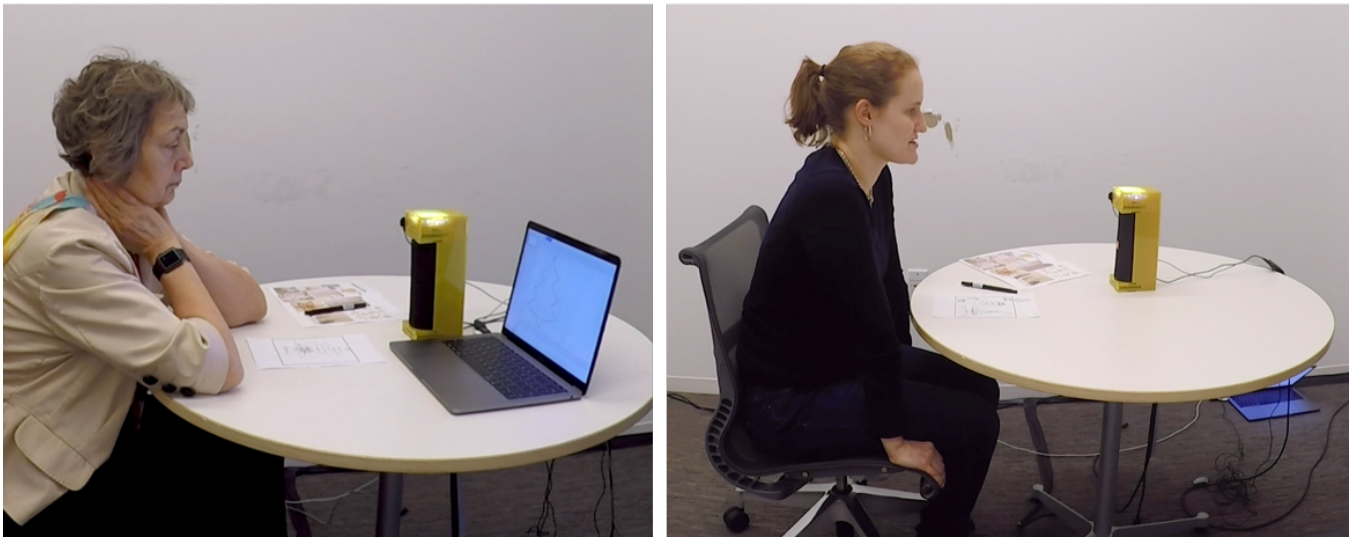
In our study, participants create a holiday ornament by explaining their desired design to a voice assistant, which is in actuality controlled by a human operator following the participant's instructions and offering suggestions and alternatives. The human operator's human-ness is disguised using a purpose-built voice synthesis tool made to look and sound like Alexa, Amazon's voice assistant. We explicitly examine the effects of one feature that could prove useful to this task: a video feed that both allows the user to observe the voice assistant's design-in-progress, and allows the voice assistant to show possible alternatives to the user. Of N=16 participants, half were randomly assigned to the **voice-only** condition; the other half were assigned to the **voice-and-video** condition and could observe work in progress via a video channel in addition to the voice channel.

We perform both an exploratory qualitative analysis and an empirical quantitative analysis of this task; in addition to asking open-ended exploratory questions aimed at providing material for future studies, we measure cognitive load and ask participants to predict and later evaluate how accurately the voice assistant understood their instructions, and how satisfied they would be—or were—with the final result.

CUI '21, July 27–29, 2021, Bilbao (online), Spain

ACM ISBN 978-1-4503-8998-3/21/07.
https://doi.org/10.1145/3469595.3469624

Figure 1: The two experimental conditions: *voice-and-video* on the left; *voice-only* on the right.



2 RELATED WORK

We rely on Clark & Brennan's theories of conversational grounding to analyze the affordances of voice assistants [2]. A screenless smart speaker-based voice assistant, such as the Amazon Echo Dot Alexa or the Google Mini Assistant, offers three main affordances for conversational grounding: *audibility*, *co-temporality*, and *sequentiality*. The inclusion of a screen adds, in a limited way, the additional affordance of *visibility*. Nowadays, many voice assistants exist in devices with screens, such as voice assistants on smartphones or the Amazon Echo Shows. Over a decade ago, Lee, Kissler, and Forlizzi, found that the language people use in communicating with robots is influenced by the mental models people hold, which could affect how a participant interacted with the voice assistant in the presence or absence of a visual channel [12]. Earlier still, Kraut *et al.* investigated the effect of a visual channel on human subject performance in a voice communication task. In that study, two participants, a "worker" and a "helper", were tasked with reproducing an arrangement of colored shapes; the helper could see the desired goal, and would communicate instructions to the worker, who would in turn place the colored shapes [9]. They found that the presence of a visual channel yields faster task completion and greater accuracy, among other effects [9].

We are also motivated by prior work investigating the relationship between visual design tasks and spoken or textual natural language, including Scones [8] (described earlier) and PixelTone [11], a multi-modal image editor that enables the selection, naming, and modification of semantically-meaningful regions of a natural image using a combination of voice and manual sketching on a touchscreen.

3 BACKGROUND AND STUDY DESIGN

In our study, we explore using a screen to characterize the effect of adding some visibility into the assistant's work in progress affects interactions with the voice assistant. We sought to answer the following research questions:

RQ1 How do humans design with a collaborative voice assistant partner?
- a) What language do they use with the assistant?
- b) Do they engage in typical design tasks like iteration, exploration, and undo?
- c) Do they possess feelings of ownership?

RQ2 What is the effect of an assistant-to-user video channel on a creative design task directed by a human, and executed by a voice assistant?
- a) Effect on user expectations.
- b) Effect on user satisfaction with outcome.
- c) Effect on user cognitive load.

Methodologically, our study draws on techniques developed in Martelaro & Ju's WoZ Way [13] study. We record audio, video, user input data (sketches), designer data, and Wizard of Oz speech and interfaces for remote observation; participants do not know ahead of time that the voice assistant is operated by a human.

3.1 Hypotheses

Our primary hypotheses are drawn largely from Kraut *et al.*'s study of visual shared spaces [9]. In that work, a visual shared space (analogous to the video channel in our study) is correlated with a $2.5x$ increase in "acknowledgement of understanding"-style utterances made by the worker [9] (analogous to the voice assistant in our study). Because of this, we hypothesize:

H1 Participants will expect their instructions to have been better understood in the voice-only condition than in the voice-and-video condition.

H1 can also be predicted by Pirolli & Card's work on information foraging [17]: in the voice-only condition, users are missing the entire dimension of visual feedback. Participants "don't know

what they don't know"—blind spots and unstated expectations can result in misunderstandings in design. Those misunderstandings, unknown before participants can see the final result in the voice-only case, are likely to inflate expectations around accuracy in understanding.

Olson & Olson's [16] field study of collaboration over distance lends further support to **H1**: the video channel not only (obviously) contributes to the "multiple channels" characteristic of collocated synchronous interactions, but the visual channel in particular also provides a modicum of the "coreference" characteristic. Greater similarity to an in-person interaction among humans along these two characteristics will result in the voice-and-video condition providing an interaction that is substantially more similar to the historic customer-craftsperson interaction described in Section 1.

Kraut *et al.* also find that results are simply **better** with a visual shared space. Following the same reasoning that leads us to believe that participants will overestimate accuracy in the voice-only condition, we hypothesize:

H2 Participants will ultimately be less satisfied with their ornament than they expect to be in the voice-only condition than in the voice-and-video condition.

H2 is due to both the expectations-of-accuracy mismatch as well as the misunderstandings caused by the lack of a visual feedback channel. With reduced ability for conversational grounding [2], we also expect satisfaction to simply be **greater** in the voice-and-video condition.

Considering the interaction in our study more broadly, we also expect participants to be positively surprised by the capabilities of the voice assistant. As described by Kwon, Jung, and Knaepper, social robots' abilities can be overestimated at first glance by users because of their ability to speak and take turns in conversation [10]. People ho have used voice assistants—our participants included—will likely be familiar with voice assistants' limited capabilities. This sets up an *expectations gap* whereby our participants will expect the voice assistant in our study to have typical, that is to say, highly limited abilities. With a human operating the assistant, however, they will find that the assistant is far more capable than expected:

H3 Participants familiar with voice assistants will be surprised by the capabilities of our Wizard-of-Oz voice assistant designer.

Lastly, drawing on insights from Olson & Olson's work on collaboration over distance [16], we expect that the addition of a video channel will make the task closer to an in-person task, and reduce cognitive load:

H4 Participants in the voice-only condition will experience a higher cognitive load [6] than participants in the voice-and-video condition.

4 METHOD

We devised a Wizard-of-Oz setup to investigate the aforementioned hypotheses. Participants were asked to design an ornament by using a Wizard-of-Oz voice assistant. We varied whether or not there was a video channel showing the voice assistant's design in progress over the course of the task. All aspects of the automated behavior were controlled or authored by a researcher during the experiment. This included both the verbal responses, the light animations on the Wizard-of-Oz device, and the drawings generated for laser cutting. The study lasted about an hour. We describe our participants, procedure, and equipment below. This protocol was deemed exempt by the Internal Review Board of Cornell University under Protocol #1810008354.

4.1 Participants

A total of 16 participants (7 men, 9 women) completed the study. Five participants were recruited from the local community via a senior center mailing list and eleven participants were recruited from the university community. All participants reported an interest in "making" and had no recent experience using a laser cutter. All participants were more than 18 years old.

4.2 Procedure

Participants were asked to enter a study room to interact with a voice assistant. A researcher sat in that room to observe the interactions and communicate with the wizards via text message if needed. The researcher stayed out of the participant's field of view to avoid distractions. A camera was set up next to the participants to keep a primary record of the interactions. Two Wizards sat in a "hidden" room controlling the audio and visual channels of the voice assistant. Participants were told that the study has 5 parts:

- **Part 1:** Participants look over an "idea book" of ornaments as examples of what the voice assistant can do, and draw their design on a piece of paper. (The voice assistant cannot see this drawing.)
- **Part 2:** Participants order their ornament from the voice assistant, initiating the conversation with the phrase "Okay Computer, make me an ornament." For half the participants, a screen shows the voice assistant's design-in-progress.
- **Part 3:** Participants fill out a questionnaire and tell the interviewer how they thought the activity went, while the ornament is fabricated in the nearby campus maker studio.
- **Part 4:** Participants receive the ornament.
- **Part 5:** Participants fill out a post-completion questionnaire, react to the resulting ornament, and are debriefed.

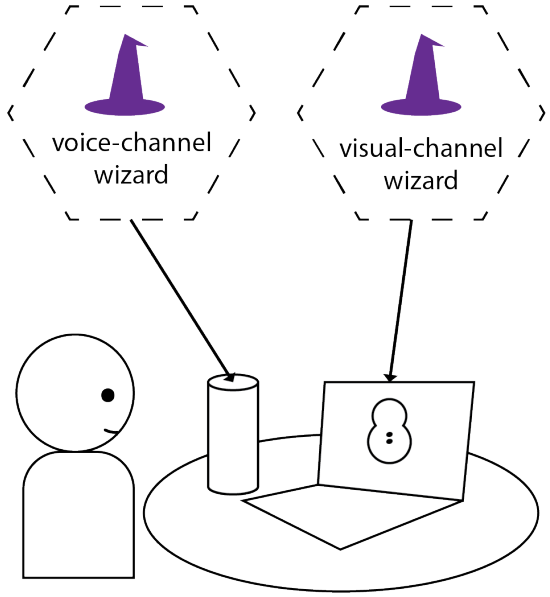
4.2.1 Communication with the wizards. A small omni-directional microphone, directly connected to the Wizard-of-Oz voice assistant in the study room, provided audio to the human controllers in the hidden room. The audio stream was shared using the appear.in videoconferencing service running in Google Chrome on a 2018 13" MacBook Pro computer. The audio of the computer in the study room was set to silent, so that the participant could not hear the Wizards' interactions.

In the condition where the wizard could visually show the current drawing, this laptop was also used to display the drawings; a shared Google Slides document was used for this purpose. Figure 1 shows the study room setup for each of the two conditions.

4.3 Equipment

The interaction between the participant and voice assistant was recorded from a third-person profile perspective, as a conversation between two people would be recorded. The video was recorded

Figure 2: Participant interacting with the Wizard-of-Oz voice assistant. The setup includes two behind-the-curtain wizards, one controls the audio channel, and one controls the visual channel.



on a GoPro Hero 5 Camera with 1080p at 24 fps and camera angle wide. The images in figure: 1 are stills from two separate participant recordings.

The Wizard-of-Oz voice assistant had an audio channel and a visual channel, as shown in Figure 2. The audio channel (and visual cues associated with it) was controlled by one wizard. The visual channel, which contained the drawings as they were made, was controlled by another wizard. The setup for each channel is described in detail below.

4.3.1 Audio Channel. The Wizard-of-Oz voice assistant was a JBL speaker encased in a custom, laser-cut box made out of yellow acrylic, with a small ring of digital LEDs for visual feedback to the participant. The ring of LEDs had four different animations that signaled the different states to the user, mimicking the states signaled by common voice assistants such as Siri:

Idle: A slow pulsing light as a sign that the system is on.
Listening: A white light reacting to human voice volume.
Thinking: A spinning circle of light indicates a busy assistant.
Speaking: A colorful reaction to the computer voice audio.

The voice assistant was controlled through a modified version of Martelaro's WoZ-Way setup [13]. A website allowed a wizard to produce computer-synthesized speech and play it directly to the Wizard-of-Oz voice assistant in the study room. Light settings **Listening** and **Thinking** were activated via foot-switches that were operated by the voice wizard.

The level of agency or decision making for the voice wizard was relatively low: The main tasks were to react to the requests from the participant and to relay questions from the drawing wizard to the participant. The voice wizard essentially acted as a mediating layer between the drawing wizard and the participant.

4.3.2 Visual Channel. The drawing wizard used a 13" MacBook Pro to draw the ornaments in Sketch, a vector drawing composition computer program. In the voice-and-video condition, the wizard would periodically copy the current drawing to the shared Google slides document for the participant to see.

During the active part of the experiment ("part 2" in the description in §4.2), the drawing wizard was in constant communication with the voice wizard to clarify the strategy as well as to communicate with the participant via the Wizard-of-Oz voice assistant.

4.4 Materials

The primary quantitative data were recorded using two printed questionnaires. The first questionnaire included two items surrounding expectations (measured with a five-point Likert scale), and a NASA Task Load Index (NASA-TLX) to measure cognitive load. The questions surrounding expectations were:

Q1: How accurately do you think the voice agent understood your introductions
Q2: How satisfied do you expect to be with your voice-agent-made ornament?

The NASA TLX instrument we utilized included six items to measure mental demand, physical demand, temporal demand, performance, effort, and frustration, and used a 20-point scale ranging from very low to very high. This questionnaire was used after the interactions with the voice assistant, but before the participant saw the laser-cut ornament. The second questionnaire, administered after the participants had received their ornament, included two questions:

Q1: How accurately do you think the voice agent understood your introductions?
Q2: How satisfied are you with your voice-agent-made ornament?

5 RESULTS

All participants successfully interacted with the Wizard-of-Oz voice assistant, and created personalized ornaments. The interactions themselves included multiple conversation turns, and the resulting ornaments looked very similar to the drawings participants had made, regardless of the presence or absence of a visual channel (see Figure 3 for examples of initial drawings, assistant-produced designs, and fabricated ornaments). The video channel allowed the participant to see the designed object as it was iteratively drawn and refined. During the interviews, participants mentioned being impressed by the voice assistant's abilities, yet seemed surprised at the end of the study when we explained there was a human controlling the voice assistant. This suggests that the Wizard-of-Oz setup was successful. We report both qualitative and quantitative results in more detail below.

5.1 Qualitative findings

In addressing **RQ1**, "How do humans design with a collaborative voice assistant partner?" we were particularly interested in the subjective experience of designing using a voice assistant. To derive these qualitative insights, we reviewed the interactions and

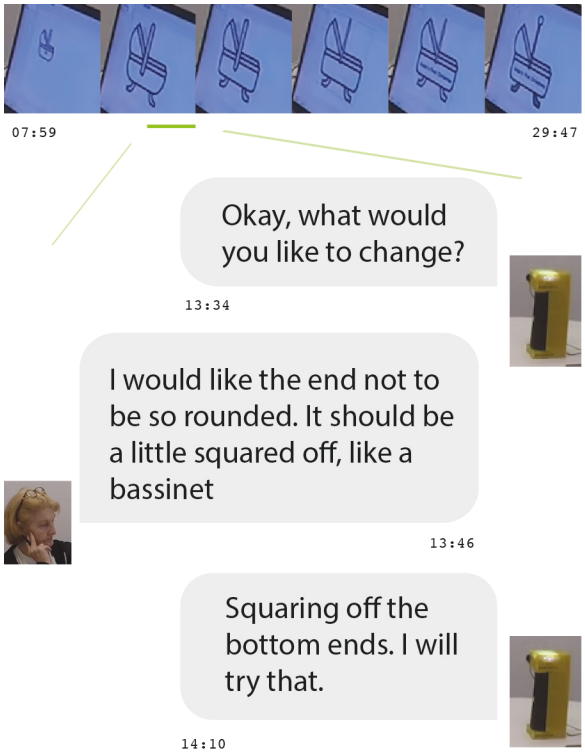
Figure 3: Each row above represents the three stages of a particular participant's design process. On the left is the sketch each participant created for reference; in the middle is the design as created by the design wizard; on the right is the final fabricated ornament for each participant.

discussed the elements that stood out the most. These are described below.

In the interviews we conducted between task completion and receiving the fabricated ornament, some participants indicated initially feeling that they "*didn't make this ornament*" as they were mostly "*ordering it from an assistant*"—however, after holding the physical ornament, they felt as though they had actually "*made the ornament.*" This is a relevant finding as it suggests that people can "make" things using voice, the most natural form of human communication. Some participants, in particular those who had experience with vector drawing tools, expressed that they would rather use those instead of the voice assistant due to the increased control and decreased time to get to their desired outcomes. On the other hand, participants also cited the constraints imposed, and that the assistant "*made decisions*," as helping them get past the initial period of "*I don't know where to start, or what to make.*" In the video condition, participants made adjustments to their designs, which embodied many of these initial decisions that the voice assistant made, once they saw the first version of their request on the screen. See Figure 4 for an example of an interaction between a participant and the voice agent adjusting a design.

After the exercise, most participants (15 out of 16) seemed very excited when we hypothetically suggested trying the exercise again.

Figure 4: Excerpt of a participant expressing design adjustments to the voice assistant. The timestamps are in minutes:seconds from the moment that the participant started addressing the voice assistant (e.g., by saying, "Okay Computer, make me an ornament.")



They noted that now that they "*knew*" how the making voice assistant worked, they could express what they wanted better. We found a few recurring blind spots, or instances of participants not considering a relevant component of the design, concerning ornament size or material—perhaps because participants were primed by the sketch and visual interface (in the voice-and-video condition) to consider the design as an outline, not as a completed whole. Even though participants in general had strong ideas of what they wanted, a few expressed difficulty in communicating their ideas effectively to the machine. In some instances, participants did not *think* to express an expectation, and in others they did not know *how to* express it. For example, one participant wanted a material from the idea book that she did not know the name of. She kept on indicating how she wished the materials were labeled in the idea book.

All participants had at least some minimal experience with voice assistants, and as a result typically came in with very low expectations of what the voice assistant would be able to understand and do. Participants wanted to know from the outset what the assistant's capabilities were, and they were only able to discover those by interacting with the assistant. Most were ultimately surprised by what the assistant was capable of—the assistant was actually a human, so this was expected. We wonder how much they felt they knew the assistant's capabilities after a few iterations of the exercise in the study, and a future study could examine this question.

Finally, none of participants asked to "undo" something they made, even though a few redrew their initial designs, and even though "undo" is probably one of the most used features in CAD software that designers and makers use.

One particular threat to ecological validity that bears mentioning relates to what Clark & Brennan [2] refer to as *Delay Costs*: the Wizard-of-Oz method we used resulted in large latencies between the participants' utterances and the voice assistant's responses; several participants cited this delay in explaining why they didn't try to correct the voice assistant's mistakes or improve upon the design in the interactive part of the study. Even though this delay was exaggerated in our study, it is still present in current voice assistants—an important consideration to take into account when designing conversations with machines that may respond more slowly than a human.

5.2 Quantitative findings

To examine the effects of the presence or absence of a voice channel on the measures outlined in **RQ2**, we ran two-tailed paired t-tests on each of the ten dependent variables we measured to determine the statistical significance of our results. Despite noticing trends in our data as depicted in the bar graphs of our results, we did not find any statistically significant differences. This may be due to our relatively low sample size.

The means of the five-point Likert-style rankings across predicted and actual reports of satisfaction and accuracy as shown in Figure 5, trend towards supporting **H1-H2** and **H4**. After the activity concluded, participants expected their instructions to have been better understood in the voice-only condition than in the voice-and-video conditions, by a small margin: 4.875 (out of 5) for voice-only compared with 4.75 for voice-and-video, in support of **H1**. This suggests that there was more expectation alignment in the condition with a visual channel. Participants also were on average more satisfied with their ornaments in the voice-and-video condition, reporting satisfaction of 4.75 compared with 3.75 in the voice-only condition, a trend that supports **H2**.

Lastly, the trends we observed from our NASA TLX resultssuggest mental demand and frustration are greater in the voice-only condition, while performance is greater in the voice-and-video condition, supporting **H4**.

We also ran ANOVA tests to determine the variance in satisfaction and understanding expectations versus reality. These calculations were performed on measurements before receiving the ornament (after interacting with the voice assistant), and after receiving the ornament. Four ANOVA tests were implemented, one for each question for each condition. We were not able to reject our null hypotheses in any of the cases, suggesting that there was not enough of a difference in expectation versus reality in either condition in this limited sample set to achieve statistical significance.

6 DISCUSSION

This study explores human-machine co-design through a voice assistant, and characterizes what these interactions may look like.

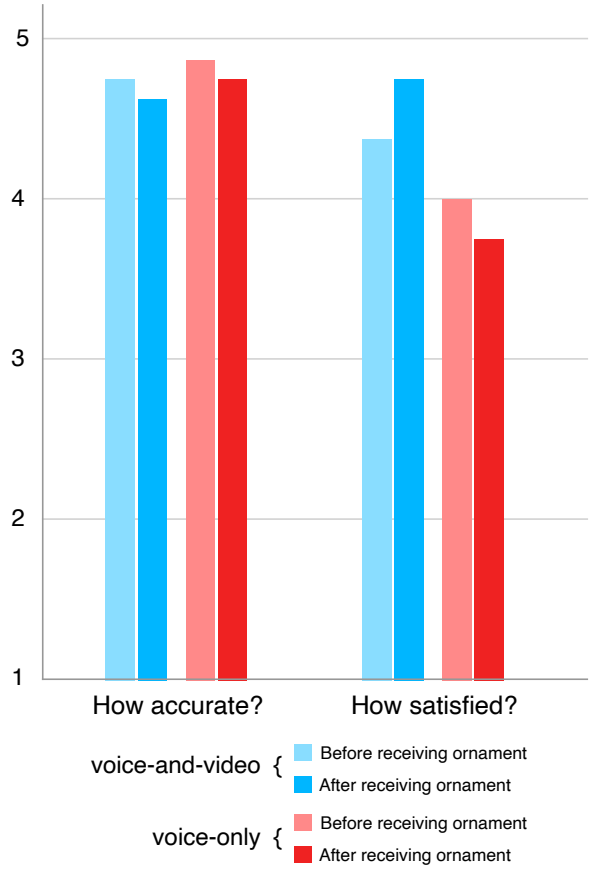


Figure 5: Differences in perceived accuracy and satisfaction between the voice-and-video (blue, *N*=8) and voice-only (red, *N*=8) conditions, both before (lighter shade) and after (darker shade) the participant received the ornament.

We used a novel, two-wizard method that can be replicated in future Wizard-of-Oz-style voice assistant studies, potentially with more participants. Our method can also be used to conduct studies somewhat remotely. One of the two wizards was "in" the Wizard room remotely via Zoom [16], yet we still constructed a common "information space" (as described by Zhang *et al.* [20]) by using an experimental setup with multiple communication channels. We characterized human-voice assistant interactions during a co-design "making" task, and uncovered important needs that should be addressed in future designs of similar interactions.

We identified the need for a standard list of questions that must be asked in order to create wholesome designs: dimensions, material type, color, etc. Additionally, we found the value in setting clear expectations, and biasing towards action for interactions with a voice assistants; nevertheless, this needs to be strategically planned to avoid information overload at any point in the interaction. For example, it was much easier for participants to adjust something that was "guessed" based on what the participant said than for them to try to explain even a small design from scratch. For example, saying "*I want a snowman*" was much easier than saying "*I want three circles stacked on top of each other, with an arc in the bottom half of the top circle and two smaller inset circles in the top half of the top circle.*"

Designers of voice assistants engaged in visual design tasks will need to strike a balance between the amount of data necessary for people to know what the possibilities are, and the amount of time and other physical or attentional resources it would take to communicate that information. For example, at some point in the interaction, the voice assistant could encourage the user to undo an action they did not like. By doing so, the interaction would more purely reflect user preferences, and could feel smoother and more natural. In the same vein, participants reacted positively to the assistant making decisions, or offering suggestions for them. They felt it was more like a "collaboration". Future work may examine automatic decision-making on behalf of the voice assistant, and see if participants react equally as well when a human is not behind the curtain. As these questions begin to be addressed, we move closer to creating more inclusive "making" machines, like the voice assistant explored in this work.

6.1 Limitations

This study had a few limitations related to scope. First, the sample size was very small for the quantitative analysis, which may have prevented us from observing statistically significant differences given relatively small effect sizes. Second, substantial delays between conversation turns led participants to limit the number of turns out of a desire to reduce awkward waiting, rather than satisfaction with the current state. Third, we did not adhere to a strict protocol for voice wizard communication, so the effects we observe could be due to variability in conversation; for example, we did not encourage discussion of materials or scale in the voice-only condition, but these did came up in the voice-and-video condition and likely mediated satisfaction. Despite these limitations, we were able to characterize a speculative version of human-machine co-design interactions, and were able to uncover relevant directions for future work. In future work, the strategy of using probes [1], which have the sort of flexibility work at this stage requires, could be employed to imagine and evaluate human-voice assistant co-design scenarios. For example, McGregor and Tang effectively employ probes to explore the role speech agents could take in meetings [14].

7 CONCLUSION

In this exploratory study, we present findings from an experiment using a Wizard-of-Oz-backed recreation of a voice assistant-based designer; participants described a holiday ornament of their own imagination to the voice assistant, which elicits details and offers design choices before ultimately "creating" the ornament for the participant. We relate observations about the nature of the interactions between participants and the voice assistant, including expectations of the technology and reactions to components such as delays. In future work, we hope to present the result of additional video coding and video analysis. For example, we could measure interaction task length and see if our findings are consistent with Kraut et al.'s [9] observed effect of a visual channel on human subject performance time in a voice communication task. Similarly, we could measure number of conversation turns, and see if there were any differences between conditions. Here, we also report quantitative

measures of participant-reported cognitive load and predicted and actual satisfaction and accuracy of reproduction. Though we did not find statistically significant differences between conditions, the trends observed suggest that participants report greater satisfaction with their ornaments if they can see them being designed, but have higher expectations if they cannot; cognitive load is reduced in the *voice-and-video* condition. 15 of 16 participants said would use a voice assistant again for a design task.

REFERENCES

[1] Kirsten Boehner, Janet Vertesi, Phoebe Sengers, and Paul Dourish. 2007. How HCI interprets the probes. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 1077–1086.

[2] Herbert H Clark, Susan E Brennan, et al. 1991. Grounding in communication. *Perspectives on socially shared cognition* 13, 1991 (1991), 127–149.

[3] Lizabeth Cohen. 2004. A consumers' republic: The politics of mass consumption in postwar America. *Journal of Consumer Research* 31, 1 (2004), 236–239.

[4] Cathy Lynne Costin. 2001. Craft production systems. In *Archaeology at the millennium*. Springer, 273–327.

[5] Tom Fisher and Julie Botticello. 2018. Machine-made lace, the spaces of skilled practices and the paradoxes of contemporary craft production. *cultural geographies* 25, 1 (2018), 49–69.

[6] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.

[7] John Atkinson Hobson. 1919. *The evolution of modern capitalism: a study of machine production.* Walter Scott Publishing Company.

[8] Forrest Huang, Eldon Schoop, David Ha, and John Canny. 2020. Scones: towards conversational authoring of sketches. In *Proceedings of the 25th International Conference on Intelligent User Interfaces*. 313–323.

[9] Robert E Kraut, Darren Gergle, and Susan R Fussell. 2002. The use of visual information in shared visual spaces: Informing the development of virtual co-presence. In *Proceedings of the 2002 ACM conference on Computer supported cooperative work*. ACM, 31–40.

[10] Minae Kwon, Malte F Jung, and Ross A Knepper. 2016. Human expectations of social robots. In *Human-Robot Interaction (HRI), 2016 11th ACM/IEEE International Conference on*. IEEE, 463–464.

[11] Gierad P Laput, Mira Dontcheva, Gregg Wilensky, Walter Chang, Aseem Agarwala, Jason Linder, and Eytan Adar. 2013. Pixeltone: A multimodal interface for image editing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2185–2194.

[12] Min Kyung Lee, Sara Kiesler, and Jodi Forlizzi. 2010. Receptionist or information kiosk: how do people talk with a robot?. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*. ACM, 31–40.

[13] Nikolas Martelaro and Wendy Ju. 2017. WoZ Way: Enabling real-time remote interaction prototyping & observation in on-road vehicles. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 169–182.

[14] Moira McGregor and John C Tang. 2017. More to meetings: challenges in using speech-based technology to support meetings. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*. 2208–2220.

[15] Nicholas Negroponte. 1975. *Soft architecture machines*. MIT press Cambridge, MA.

[16] Gary M Olson and Judith S Olson. 2000. Distance matters. *Human–computer interaction* 15, 2-3 (2000), 139–178.

[17] Peter Pirolli and Stuart Card. 1995. Information foraging in information access environments. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM Press/Addison-Wesley Publishing Co., 51–58.

[18] Ellis A Van Den Hende and Jan PL Schoormans. 2012. The story is as good as the real thing: Early customer input on product applications of radically new technologies. *Journal of Product Innovation Management* 29, 4 (2012), 655–666.

[19] Julia Walter-Herrmann and Corinne Büching. 2014. *FabLab: Of machines, makers and inventors.* transcript Verlag.

[20] Zhan Zhang, Aleksandra Sarcevic, and Claus Bossen. 2017. Constructing Common Information Spaces across Distributed Emergency Medical Teams. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 934–947.