talking to computers

jen spatz

A thesis presented in partial fulfillment of the requirements for the degree Master of Industrial Design in the Department of Industrial Design of the Rhode Island School of Design, Providence, Rhode Island.

By Jen Spatz

Approved by Master's Examination Committee:

Charlie Cannon / Department Head, Industrial Design, RISD
Full time Faculty Advisor

Ryan Bardsley / VP, Artificial Intelligence, AI Cure
Adjunct Faculty Advisor

Elly Nattinger / Experience Engineer, Google
External Advisor

With guidance from
Tim Maly / Assistant Professor, RISD

Abstract

A popular belief amongst UX designers is that the more voice user interfaces (i.e. Alexa, Siri, Google Assistant) speak and behave like people, the more functional they will be. But, conversational mimicry is not the only way a screenless computer can communicate information. The scope of sounds humans can interpret, manipulate, and make is broad. This project seeks to identify ways designers can mine this domain for interaction cues that promote a deeper understanding of digital content and the systems that deliver it.

Introduction

Metaphors that inform the design of graphic user interfaces (hereafter GUIs), are clear: cursors touch, folders contain, and hyperlinks redirect. Stylistic attributes of these tools have evolved, but their functionality remains consistent.

Voice user interfaces (hereafter VUIs) do not make use of the interaction cues with we know. It's unclear how Alexa, Google Home, Siri, and Cortana curate Internet content and it's unclear exactly how we should relate to them. Are they knowing godlike entities, therapists, assistants, or storage systems? Should we address them in full sentences, with inflection, or in commands prioritizing search-able terms?

Formally, VUIs are paperweights that speak— and this affords less obvious interaction than does the digital desktop. Instead of a predictable system of containers and doorways, VUIs introduce helpful disembodied voices. New users don't learn their capabilities by clicking around and exploring, but rather by cautiously mimicking behaviors witnessed in advertisements and other homes.

Often times VUIs get it right and users get just what they are looking for: the weather forecast, an organized list, a reminder, a favorite song, or an obscure fact. But, are convenience and recreation really the most meaningful use for non-visual technologies?

VUIs can revolutionize more than multitasking and game-play. They prioritize listening over looking and have the potential to more deeply connect users to the sense of sound. VUIs can advance the experience of text, data, ideas, and options. But, conversation isn't

A stand brings Alexa up to eye-level, 2018

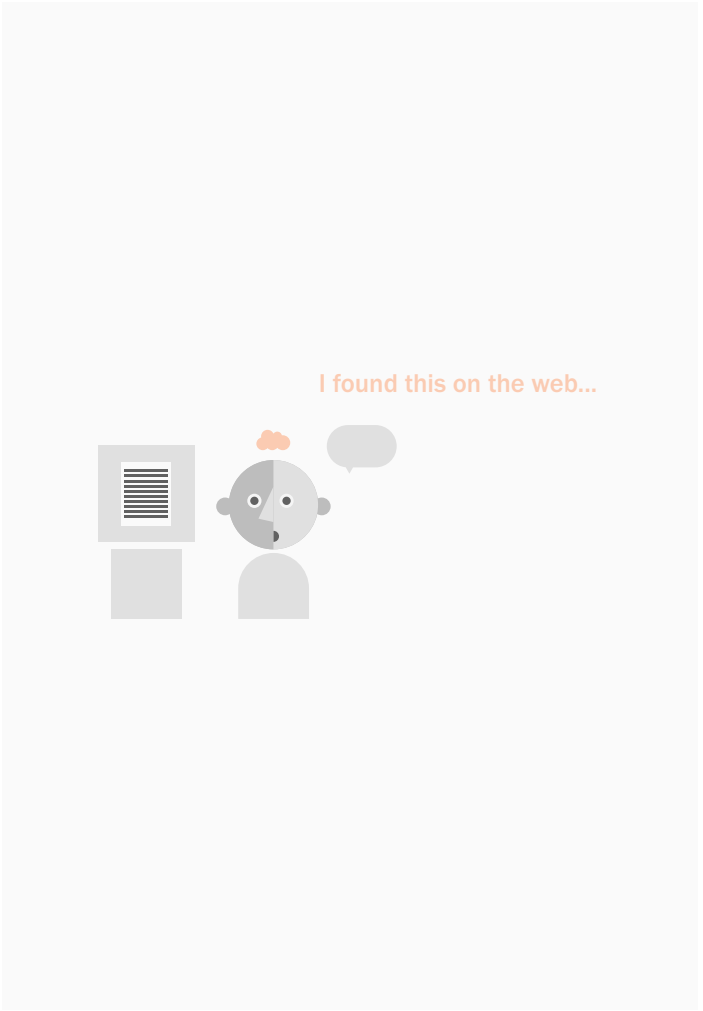
ideal for this. Its difficult to review, explore, and interpret information delivered during a conversation.

The scope of sounds we hear and produce is rich with possibilities for interaction design. Utterances, manufactured sound, residual sound, ambient sound, song etc. can and do communicate. This thesis seeks to identify ways designers may begin to mine this domain for symbols and interaction cues that promote a deeper understanding of information and the systems that deliver it.

Conversation

Conversation, the informal exchange of ideas by spoken word, rarely delivers only information. It is most appropriate when there is something more we seek: (even if unconsciously) companionship, a sense of personality, intelligence, style, maybe even a closer glimpse at a stranger's face. There is always excess in human to human conversation, and this excess brings us closer. On the other hand, the excess present in artificial conversation facilitates only inefficiency and pretense. A VUI's jokes and colorful language don't help users to better understand it's search processes, sources, or knowledge bank. Rather, they implicate the user in theater.

The most efficient way to use a VUI is to role-play, interacting with it as though it were a very obedient, very particular human who responds only when addressed by name. There is no escaping the absurdity of this. While use cases range from individuals plainly trying to treat their VUIs like people for the purpose of companionship or humor to individuals trying to get things done in the most efficient way possible, in any case, the user needs to speak to a counter-top object as though it were a living thing.



Mental Model

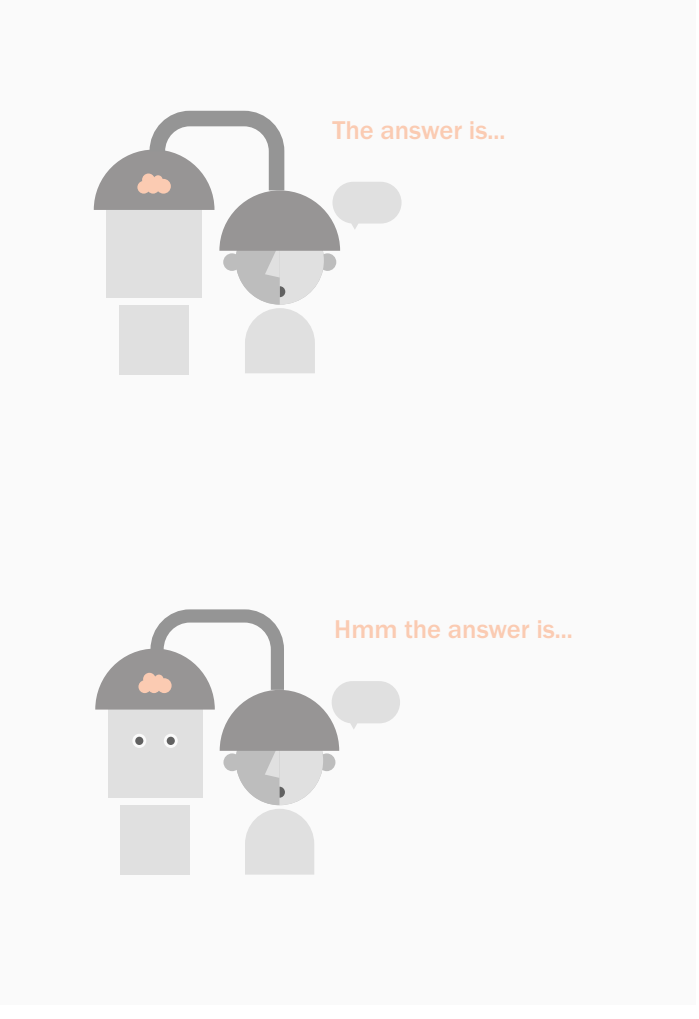
While most users do not actually believe that their devices are conscious, they do think of them as such. If one were to think of a VUI as a sea of ones and zeros, or a mechanical instrument, speaking to it would be counterintuitive. We think of VUIs like conscious entities because we need to treat them like conscious entities.

But, what kind of conscious entities exactly? Sometimes the content they deliver seems to originate from a human and sometimes it seems to originate from a computer with a human voice.

- When they say: “I found this on the web for Herman Melville,” it seems as though they are remote humans sitting in front of computer screens, reading “the web.”
- When they say: “Hmm, I don’t know how many dot’s are in Seurat’s “Bathers at Asnières,” it seems like their intelligence is not emanating from a screen but from a human-like mind.
- When they say: “Three plus three equals six, they seem stoic and emotionless, like a calculator.

It’s difficult to bond with a VUI the way a child might a stuffed animal, action, figure or doll because we as users cannot maintain a clear vision of what exactly it is.

Owners don’t develop comprehensive awareness of VUI capabilities because they can’t well anticipate their responses. Statistics show that roughly 68 percent of VUI owners use their devices less than twice daily and that the most popular cases include asking a



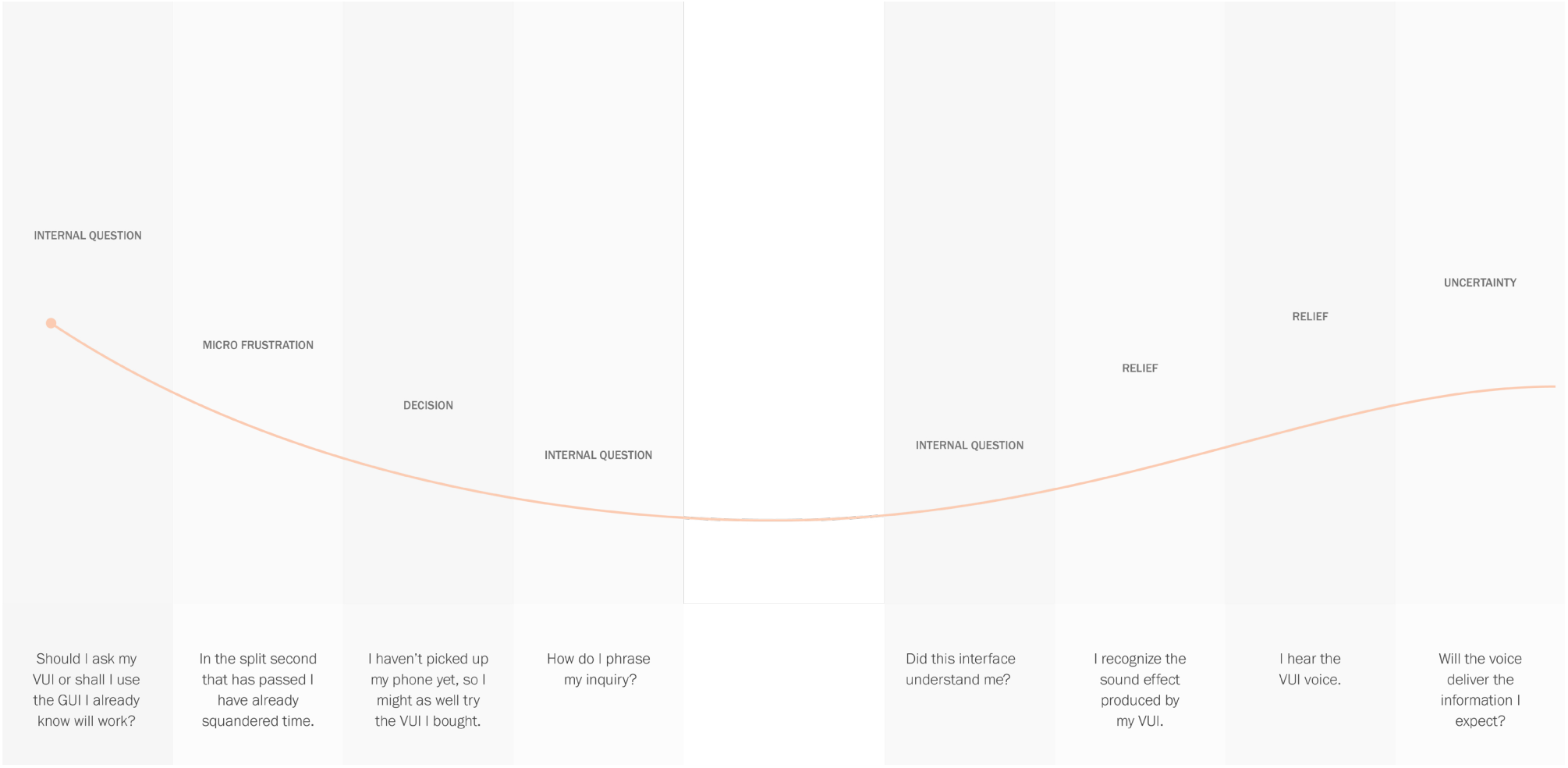
question, listening to music, and setting a timer. For the majority, VUIs facilitate the same actions again and again; for the majority, VUIs are expensive voice activated buttons.

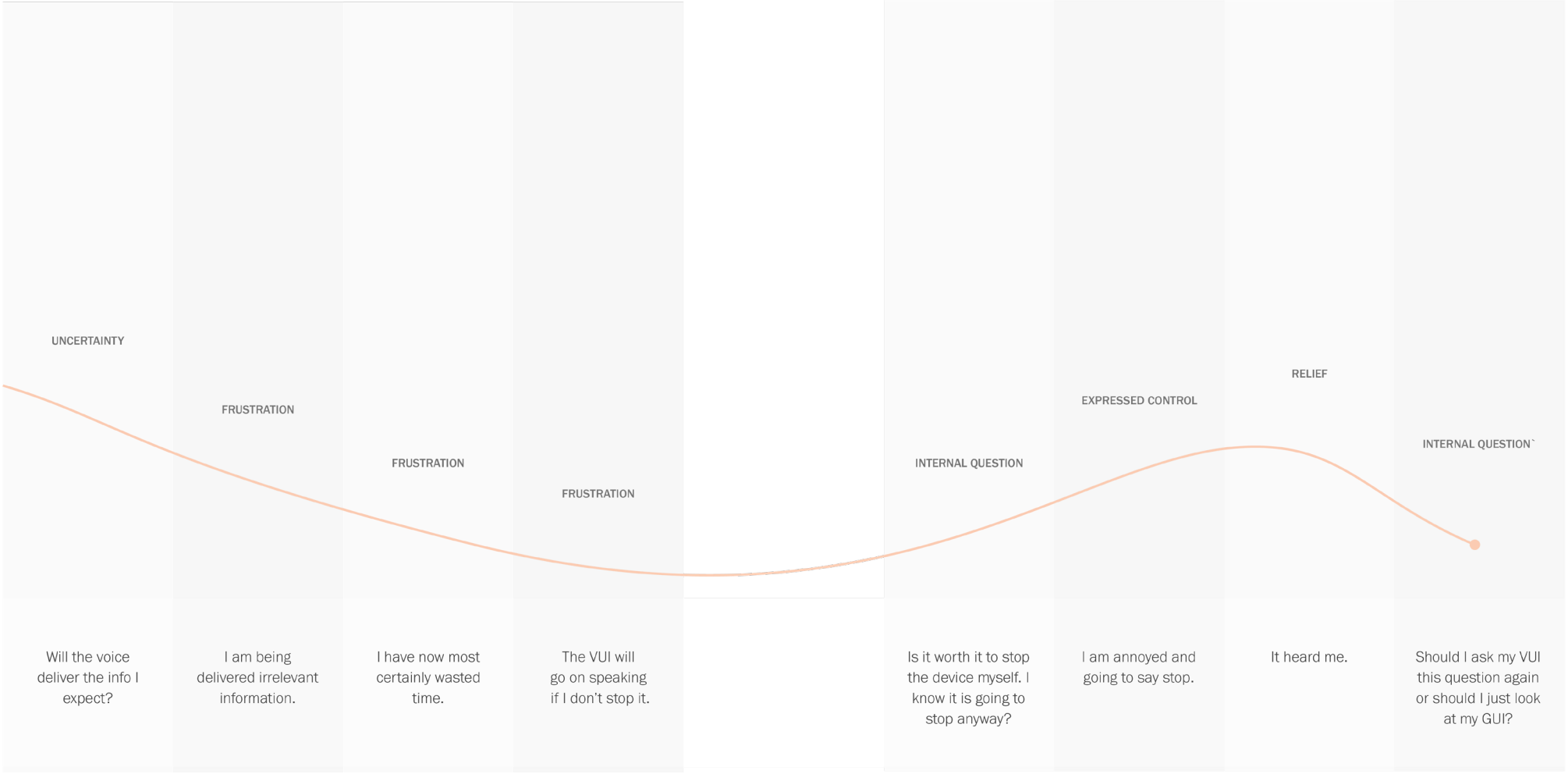
Without visual technology, we might be adept at using VUIs to perform complex actions. But, alas screens abound and every time an individual elects to use a VUI for an unfamiliar task, she forgoes the certainty of a predictable device.

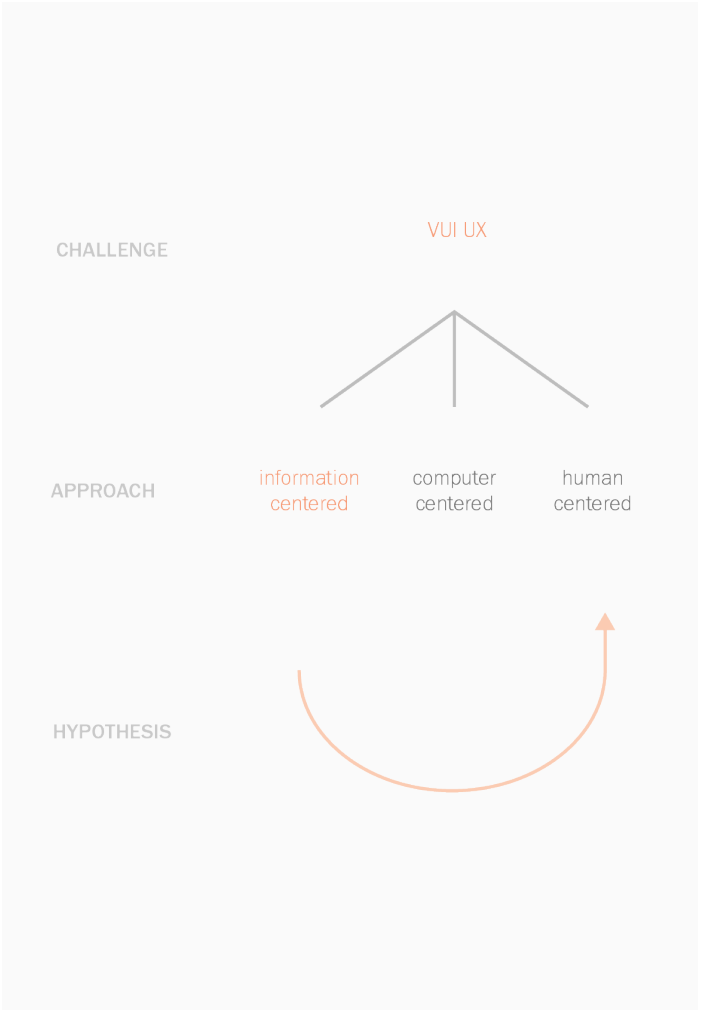
Cognitive Load

Cognitive load refers to the exhaustion of the working memory. GUI designers note that complex overly detailed visual environments have high cognitive loads. The high cognitive load associated with VUI use is due to questions that arise in silence rather than the confusion of a chaotic screen. It is nearly inevitable that some of a user's cognitive energy, before and after articulating a command, is devoted to measuring the VUI's speed and functionality relative to that of the familiar GUI in their pocket or on their desk.

User experience (UX) designers seek to lower cognitive load. VUI UX designers trust that the more VUIs understand and emulate human behavior, the more efficient and useful they will be. Since Siri's early days, VUIs have seen major advances- the tonality of Alexa's voice has matured, the Google Home can multi-task— answering questions out of context whilst navigating a list of instructions, and all assistants are able to record ideas, tell jokes, and answer personal questions (with a bit of snark). But is making VUIs more human really going to help users understand them better?


INTERNAL QUESTION
Should I ask my VUI or shall I use the GUI I already know will work?
MICRO FRUSTRATION
In the split second that has passed I have already squandered time.
DECISION
I haven't picked up my phone yet, so I might as well try the VUI I bought.
INTERNAL QUESTION
How do I phrase my inquiry?
INTERNAL QUESTION
Did this interface understand me?
RELIEF
I recognize the sound effect produced by my VUI.
RELIEF
I hear the VUI voice.
UNCERTAINTY
Will the voice deliver the information I expect?


UNCERTAINTY
FRUSTRATION
FRUSTRATION
FRUSTRATION
INTERNAL QUESTION
EXPRESSED CONTROL
RELIEF
INTERNAL QUESTION`
Will the voice deliver the info I expect?
I am being delivered irrelevant information.
I have now most certainly wasted time.
The VUI will go on speaking if I don't stop it.
Is it worth it to stop the device myself. I know it is going to stop anyway?
I am annoyed and going to say stop.
It heard me.
Should I ask my VUI this question again or should I just look at my GUI?



Approach

Conversation design is human centered design. The human behavior of rhythmically speaking, listening, and thinking in step with another person is taken as a nonnegotiable around which to shape a new technology.

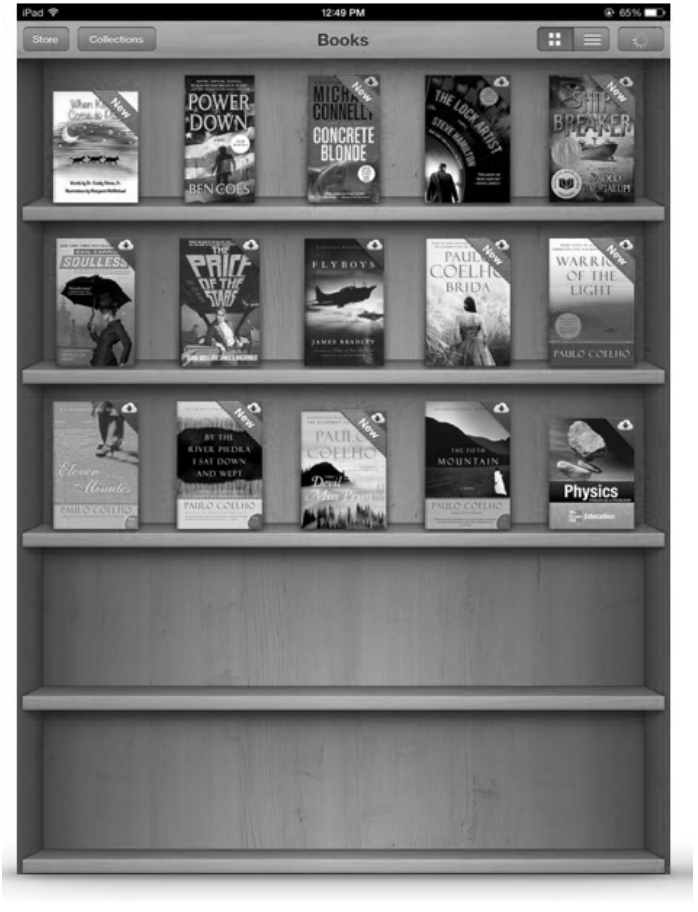
VUI designers are considering the way that people interact with people. But, the purpose of this technology, is not be to simulate a human, but to function as a dynamic portal to applications, appliances, and ideas. VUI designers need to consider the way that people interact with ideas.

The goal posts for VUI research and development need to be shifted away from the manufacture of believable unpredictable personalities and towards dynamic interactions with information.

Skeuomorphism

Skeuomorphism took shape in the 1980s. One of its earliest proponents was Steve Jobs of Apple. The idea was simple: computer interfaces would be more intuitive if they metaphorically referenced physical objects used in everyday life like folders and trashcans[1]. Today a small number of designers are beginning to question the skeuomorphic nature of voice user interfaces. Amazon's Bert Brautigam says, "The articulation of the metaphor of a human assistant and the way voice assistants mimic humans is literal. Just as buttons look literally like buttons on the skeuomorphic visual interface, the voice assistant that sounds literally like a human is a skeuomorphism."[2]

1. "Skeuomorphism Is Dead, Long Live Skeuomorphism | Interaction Design Foundation." Accessed May, 2018. https://www.interaction-design.org/literature/article/skeuomorphism-is-dead-long-live-skeuomorphism.
2. Brautigam, Bert. "The New Skeuomorphism Is in Your Voice Assistant – UX Collective." Accessed May, 2018. https://uxdesign.cc/the-new-skeuomorphism-is-in-your-voice-assistant-3b14a6553a0e.



Apple's Books interface exemplifies skeuomorphism

The first personal computers were heavily skeuomorphic. Icons that looked like representational versions of real world objects abounded. Sometimes skeuomorphism served the purpose of translating functionality and sometimes it was an aesthetic flourish as in the design of the first books interface. But, ultimately, much of the excess that early skeuomorphic interfaces introduced was striped away, leaving behind minimal graphic representation and an understandable system for interacting with information (flat design). Our metaphors remained intact. Buttons still activate, folders still contain, and cursors still touch. They simply look a bit less glossy, more confident in being themselves rather than representations of objects from another world.

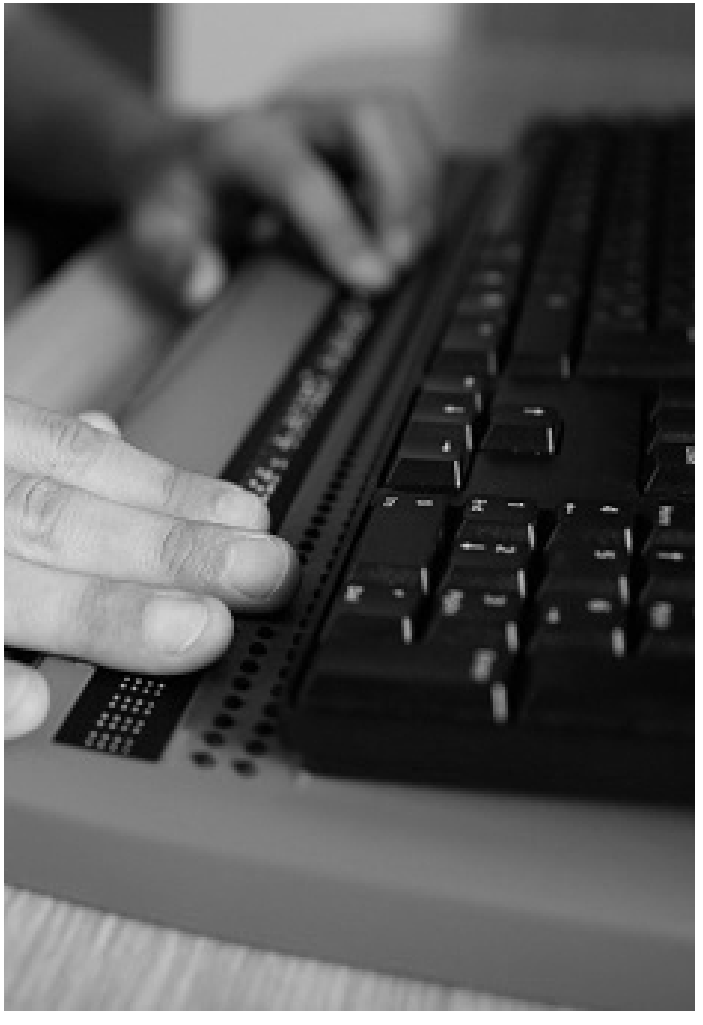
Skeuomorphism in GUI is largely regarded as passé, a fad that no longer reflects our advances, interests, and style. The simpler an interface the more beautiful. However, while heavy handed, skeuomorphism is beautiful too. Skeuomorphism means helping humans understand.

History shows that skeuomorphism usually evolves. Graphic metaphors evolved from pixelated representations to high fidelity representations, to high fidelity minimalistic representations with very subtle or no relationships to the physical objects that originally inspired them.

How will VUI skeuomorphs evolve? Is this something designers should simply accept as natural and inevitable or is it an occurrence that they should conciously anticipate and guide? When the excess of conversational skeuomorphism subsides will we relate to the VUI computer in ways that are unexpected and transformative?

Research

Many designers regard the blind and visually impaired as an extreme user group because they are often excluded from the benefits of mainstream products and services. However, in the case of non-visual technologies, the blind are not the excluded, but the experts. I looked to employees of Perkins School for the Blind and members of the National Federation of the Blind to better understand the nuances of navigating the Internet without vision.



A refreshable braille display translates digital content in conjunction with a screen-reader, 2018

Screen-readers

Screen-readers are software programs that translate screen text to synthesized speech or braille. The user sends commands by pressing combinations of keys on a keyboard or making gestures against a touchscreen.

Although the learning curve associated with screen-readers is steep, they provide users a valuable means of darting around digital content. An important distinction between screen-readers and VUIs is that screen-readers deliver context as a byproduct of their functionality and VUIs eliminate it completely.

Design writers and accessibility advocates call VUIs a revolution for the blind and visually impaired community. However, these devices do not serve to replace the technologies users employ to explore lengthy digital content such as social media, blogs, articles, and inboxes. VUIs provide assistance in task management and execution, but, despite their growing popularity, sighted individuals still use screens and non-sighted individuals still use screen-readers.

Cory holds his phone close to his face to hear his screen-reader, 2017

Cory at Work

In late October, I visited The Perkins School for the Blind. There, I met with Cory Kadlik, 26 years old, a blind technology enthusiast, working in the school's talking book library. During my visit, I sought to learn what a professional environment that did not rely on visual information might sound, feel, and look like.

I anticipated a noisy office and was surprised to discover that, even without his ear buds, my new friend's technology was not aggressive. I watched Cory take calls, use his computer, telephone, VUI, and refreshable braille display for 2 hours. At a relatively low volume, Cory's desktop screen-reader sounded like low pitched insect buzz. The high speed speech was decipherable to me only if I carefully focused my attention it.

At one point, I watched as Cory impressively took a call on his cellphone and simultaneously referenced an instruction manual with his screen-reader. It seemed as though he was scanning the manual for answers. I asked Cory about this when he hung up and he explained that indeed he was not listening to every word articulated by the screen-reader. Just as a sighted person may not digest every printed word before her, listening at very high speeds, it is possible to get an impressionistic sense of text.

Superpowers

A common misconception is that the ability of a blind person to interpret synthesized speech at high speed is a superpower of the blind. "Neuroscientist, Uri Hassan examines how the brain processes sped-up speech. He points out that even at normal speed, most people don't catch every word. According to Hassan, brain responses become slower when we speak slowly and faster when we speak quickly. Comprehension only starts to break down around two times the normal speed."[1]

1. Shafrir, Doree. "Meet The People Who Listen to Podcasts Crazy-Fast." Accessed May, 2018. https://www.buzzfeed.com/doree/meet-the-people-who-listen-to-podcasts-at-super-fast-speeds?utm_term=.tmJBAL0nYn#.ufboxKz8r8.

Perkins School for the Blind

A guide at the Perkins School for the Blind discusses tactile symbols, 2018

At the Perkins School for the Blind I learned about the extensive history of the country's oldest institution for blind and visually impaired individuals. There I encountered a tactile museum, teaching tools, writing implements, and accessible furniture.

Among the objects my tour guide highlighted were a set of blocks that looked like strange vocabulary cards. Each presented a swatch of material or small object that corresponded to a word and icon. The guide explained that these tools were intended to indicate important ideas for illiterate blind children.

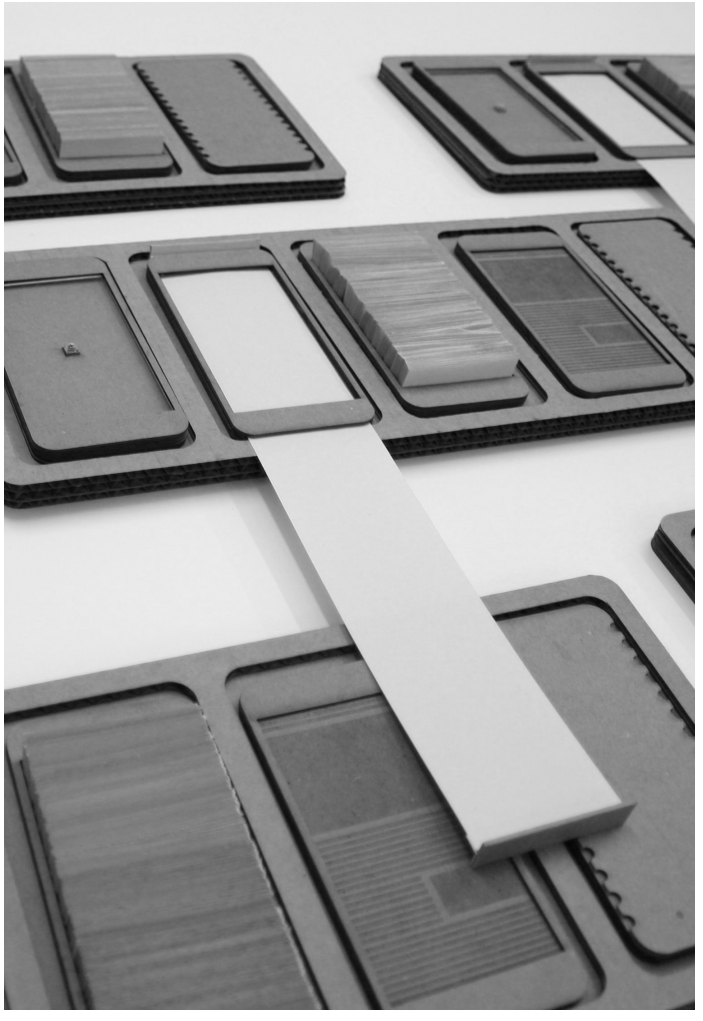
The card's representational cues seemed sloppily misaligned—tissue representing sadness because it's an object associated with the experience of sadness, an x representing "no" symbolically, and an ipad representing technology formally. But, all of these forms qualified as symbols of the same family. How was this clear to blind children? How might one know that they were experiencing a symbol and not a functional object?

The answer was obvious: because of the plaques they were attached to. The size shape and texture of this form qualified the contents as symbolic.

I noticed some other nonvisual qualifiers like these plaques at Perkins. For example, upon visiting the tactile museum I noted that the room itself qualified all objects within as educational and descriptive.

This experience inspired me to begin questioning nonvisual qualifiers. I suspected that there might be a way to use them to represent information like text length, style, and format.

The Tactile Museum at the
Perkins School for the Blind, 2018



Research props designed for National Federation of the Blind meeting, 2018

National Federation of the Blind

Granted the opportunity to present on some of my emergent ideas at a National Federation of the Blind community meeting, I set out to produce a scheme of tactile materials to aid in my discussion of nonvisual symbols and metaphors. I used a series of chipboard and paper models to prompt questions about nonvisual digital environments. I wondered to what extent the desktop metaphor with it's papers, folders, and graphics could be translated to an audio only experience?

The consensus amongst the group was that responsive motions like that of a sliding sheet of paper or bouncing icon was absent from screen-reader interactions . But the concepts of paper, buttons, and rectangular layouts were crucial to navigation. Throughout the discussion community members referred to these screen elements as such even though they perhaps did not see the skeuomporphic representations of them that I did.

Spatial Relationships

One group member explained that she relies on the grid layout of her applications to envision their distribution across her screen.

> *On my phone I remember most of the icons. Before, when I was a novice i would just swipe swipe swipe [through linear lists of icons], but now I kind of know roughly where everything is. I would say about 80 percent of the time i'm correct.*

Audio Qualifiers

At one point the conversation turned to qualities of sound within the Iphone's screen-reader, VoiceOver. The group members discussed whether or not sound could represent qualities of graphic items better than synthesized speech.

National Federation of the Blind members interact with research props, 2018

> *Using VoiceOver* [screen-reader on the iphone] *proficiently isn't for everyone, I think there are probably people who are totally blind and still don't use it as much as they could. Even though it's all language you're still going through a translation process, you know, like when you hear that string of labels that's attached to whatever item it is that you're touching, you're still processing it to figure out it's name, what part it refers to, what it can do to it, what part of this refers to what kind of item it is, like if it is a link, a button, or a text item....obviously this would be different to do based on sound, but I would be open to it.*

Synthesized Speech

The group members agreed that some aspects of sound within the VoiceOver environment blended into the experience and didn't even register as sound for them. In regards to choosing from the many available voices, one member commented,

> *I'm really not actually listening to what the voice sounds like, It's not a part of my experience. I just want to hear the the information, any voice would do, even if it's robotic.*

Ear-cons

The group members noted that the term used to refer to descriptive sounds is "ear-con." Ear-cons represent specific events, information, and feedback to the user.

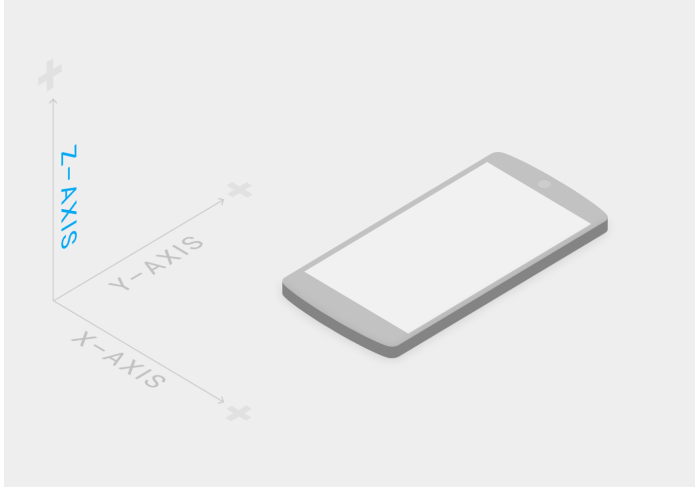
National Federation of the Blind
community meeting, 2018

Experiments

Designing for nonvisual nonconversational experiences raises the challenge: how do we observe this type of interaction when it doesn't yet happen? Gestural interface designers mine the range of human movements that occur in daily life. But, with the exception of conversational VUIs, there are few examples of humans conciously using nonverbal sound to interact. The following course of research and experimentation seeks outiline a space for exploring the potential of nonspeech sound as a medium for interaction design.

The following explorations and experiments seek to achieve 4 things:

1. To observe the ways that humans unconsciously use sound.
2. To observe the frustrations and limited nature of conversational VUI as used for advanced or lengthy procedures
3. To ideate on ways that sound might be perceived as part of a system rather than as musical or notifying
4. To ideate on ways that the use of the human voice might be divorced from notions of conversation



Illustrations from Google Material Design, 2018

Audio Material Design

Google materials has proposed a visual language for application design that honors the physics of paper in space. Designers who adhere to Google's standards are able to design interfaces that are not only aesthetically pleasing, but cohesive and understandable.

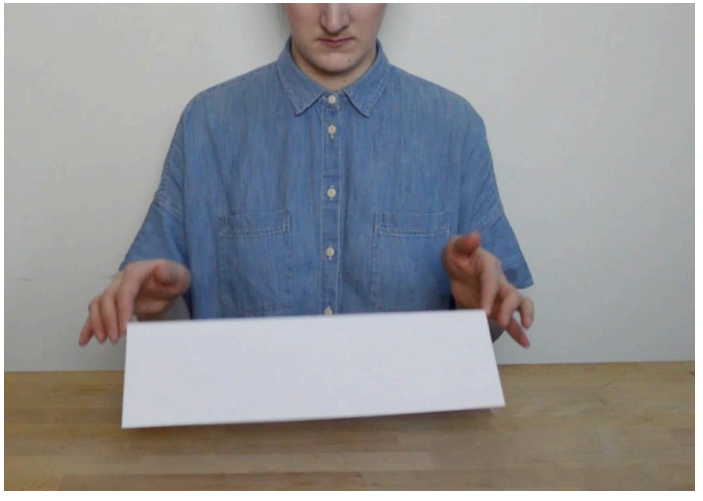
Questions: Noticing that an alarm clock buzzer, a jingle, and the woosh! of an email delivery all seem to reference different imagined spaces, I questioned if one might standardize a sound library by confining sound making to a limited set of physical objects.

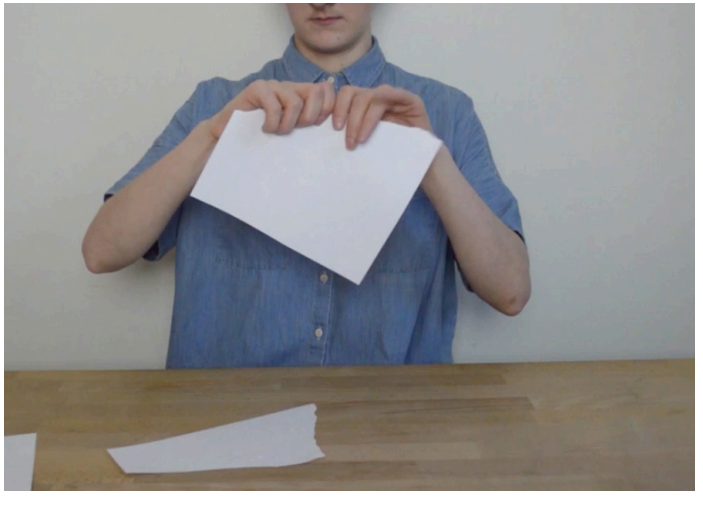
Method: In this experiment I set out to exhaust the range of sounds one can produce with paper. My actions ranged from meaningful operations (like crumpling, cutting, and tapeing paper) to abstract movements more sensitive to the physical properties of paper (like sweeping the table with it or dropping on an edge.)

Observations:

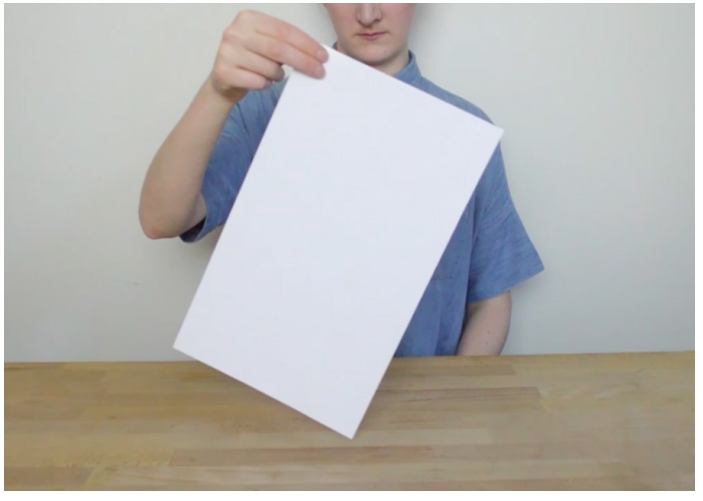
- While not every action produced a recognizable or seemingly related sound, every sound was believable as something that had occurred in physical space.
- The variability of the volume produced by certain actions seemed to eclipse some of the more subtle changes in sound.
- Sounds that emanate from a single set of materials don't necessarily register as related.
- It is more difficult to recognize subtle differences in sound quality when numerous variables are changing.

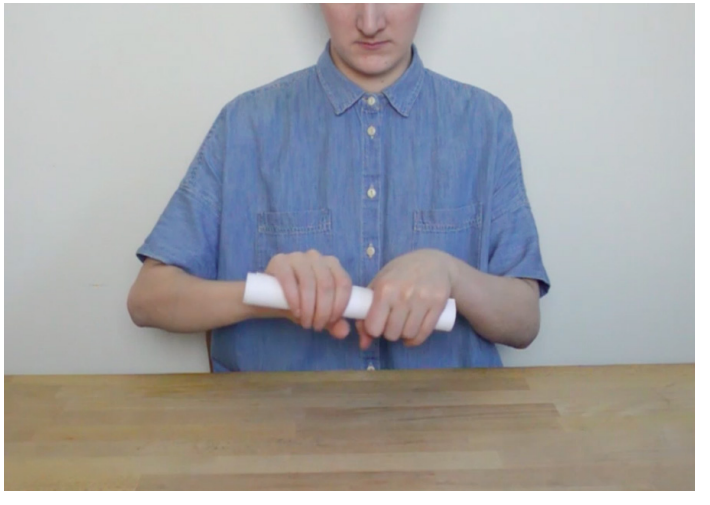
Insight: Restricting communication to a limited palette can result in a surprising constellation of sounds.





Using paper to produce a scheme of sounds, 2018





Using paper to produce a scheme of sounds, 2018

A blindfolded participant
uses his hands make sounds, 2018

Sound Charades

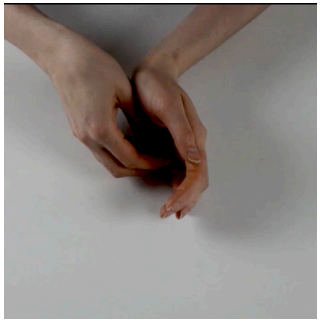
Questions: Can non-speech sound alone deliver content? Sound effects accompany motion in video games and cartoons, music delivers emotional context during movies, alarms signify events. To what extents our understanding of meaningful sound dependent on corresponding actions or effects?

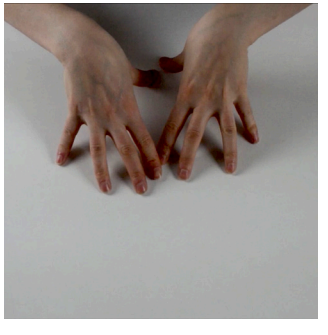
Method: During this challenge I instructed participants to use only the table surface before them and their own hands to create sounds representing nouns, adjectives, and verbs (handed to them on post-it notes.)

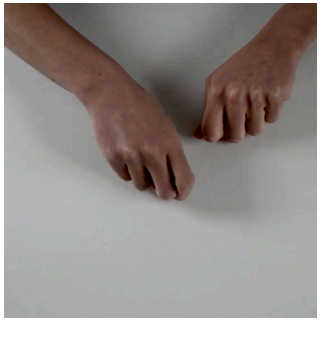
Observations

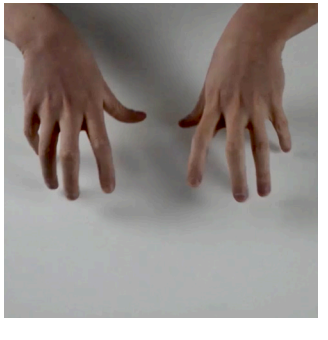
- Participants' sound-making strategies varied greatly. Sometimes they anthropomorphized their fingers using them like puppets walking across the surface of the table. Sometimes they sought to create feelings associated with words. One participant rubbed his hands together vigorously to represent "hot." Sometimes they sought to mime events associated with words, one participant mimed the motions of eating a meal with cutlery to represent lunch. Sometimes they used symbolic gestures. One participant represented 2:30 with a sequence of two knocks and a sort of karate chop (perhaps in attempt to represent half).

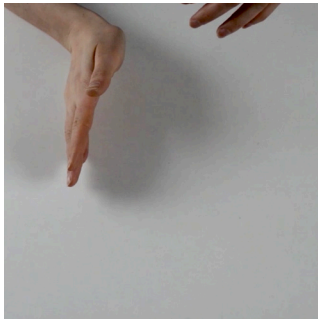
Insight: We do not have a standardized language for communicating content with sound, we are much more familiar with gesture and mime.

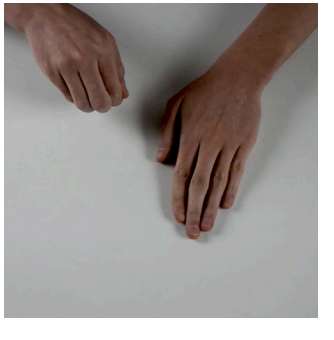


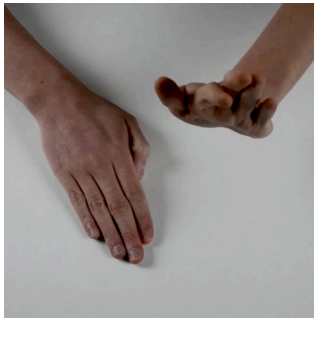












Experiment participants use their hands to make sounds , 2018

A student demonstrates VUI use to her classmates, 2018

School Workshops

In late March, I led a series of workshops at a New York City elementary school. I spoke with small classes of 1st-5th graders about VUIs and multi-modal communication. My goal was to get a sense of how younger generations think about robots, artificial intelligence, and interactions with technology.

Overall I found that students had a difficult time thinking about robots that are not personified. However, I noticed that they responded strongly to the aspects of Alexa that were least human. For example, the changing colors of Alexa's LED ring, accidental awakenings, and Alexa's "I hear you" sound effect were mentioned more frequently than the actual experience of communicating with *her.*

Students speak to Alexa,
2018

Troubleshooting Alexa

Questions: Generally, if a VUI isn't performing to a user's standards or seems ill-suited for executing a task, she can simply consult a GUI instead. But, what if this were not an option? What does troubleshooting with a VUI look like? When intuitive question asking doesn't work, what does?

Method: During this experiment I challenged 5th grade students to ask Alexa, who the oldest person in the world is, where they live, and what their age is. While this seems like a fairly straightforward line of inquiry, it is not. In order to uncover the name of the oldest person in the world, one needs to specify their interest in the oldest person *currently* living. In order to find out where this person resides, one needs to accurately repeat the name of the individual in question- Nabi Tajima.

Observations

- Students were reluctant to interrupt Alexa. Only after numerous identical wrong answers did students cut her off.
- When students felt most engaged in the conversation they often forgot to use the wake word (Alexa.)
- Students seemed unsure of where to look while listening, often staring into my face, or out the window.
- It is easiest to interrupt a VUI when it seems mechanical.
- It's difficult to stay mentally engaged when using a VUI for long periods of time, once it begins communicating an irrelevant idea the user stops paying attention.
- The more conversational an interaction between person and VUI becomes the more likely the person is to speak casually and confuse the VUI.

Insight: The existing conversational models for voice user interfaces often result in user disengagement and confusion.

Alexa, who is the oldest person live... um... person living?

Here's something I found on Wikipedia: After the death of Gertrude Baines on September 11th 2009, Ray became the oldest person living in the United States.

How old was he? How... I mean Alexa how old was he?

silence

Alexa, how old is the oldest person living?

The oldest person that is a person who resides at the world's age is 117 years 7 months and 19 days old

Where does....

[yelling] Alexa! Where does the oldest person living live?

Here's something I found on Wikipedia: She was the world's 7th oldest living person, the 2nd oldest living person in the United States and the oldest living resident of California at the time of her death.

Alexa, where d... is the oldest person living currently live

Here's something I found on Wikipedia: She was the world's 7th oldest living person, the 2nd oldest living person in the United States and the oldest living resident of California at the time of her death.

Alexa, how many I mean i'm sorry [interrupted]

Sorry, I don't know that.

Alexa, how old is the oldest person living?

The oldest person that is a person who resides at the world's age is 117 years 7months and 19 days old

Alexa where does the oldest person living live?

Here's something I found on Wikipedia: She was the world's 7th oldest living person, the 2nd oldest living person in the United States and the oldest living resident of California at the time of of her death.

[yelling] Alexa! Where... I mean... Where does the oldest human live?

Sorry i'm not sure.

Alexa, where does the oldest human being live?

Sorry, I don't know that.

Alexa, where does the oldest human being currently live?

Heres something I found on Wikipedia: Human remains at the site are the oldest that have been found in Washington and at the time were the oldest set of remains found in North America.

Alexa, how old is the oldest person living?

The oldest person that is a person who resides at the world's age is 117.

Al... Alexa, where does the person who's 117 live?

Sorry i don't know that.

Alexa, where does the oldest person alive live?

Heres something I found on Wikipedia: Ethel Lang was a British supercentinarian who at the time of her death was the oldest living person in the United Kingdom, the 2nd oldest living person in Europe after Emma Morano of Italy and the 9th oldest living person in the world.

Alexa, how old is the oldest person living and where do they live?

Here's something i found on Wikipedia: Arbella Ewing was, at the time of her death at age 114, the 3rd oldest living person in the world behind Edna Parker of United States and Maria deJesus of Portugal, the 2nd oldest person in the United States and the oldest living African American.

Alexa, what is the name of the oldest person living?

Here's something I found on Wikipedia: As the oldest living person is usually a woman, a sequence of oldest living women would be nearly redundant to this list.

Alex- Alexa, what is the name of the oldest living person?

Here's something I found on Wikipedia, as the oldest living person is usally a woman, a sequence of oldest living women would be nearly redundant to this list.

Who is the oldest person living and what is their name?

[yelling] Alexa! Who is the oldest person living and what is their name?

Here's something I found on Wikipedia: Manuela Fernández Fojaco was the 5th oldest living person, the oldest person in Europe and the oldest living person in Spain at the time of her death, at the age of 113 years, 202 days.

Who is the oldest living person?

[yelling] Alexa! Who is the oldest living person?

I'm Alexa and I can provide information, music, news, and more [interupted]

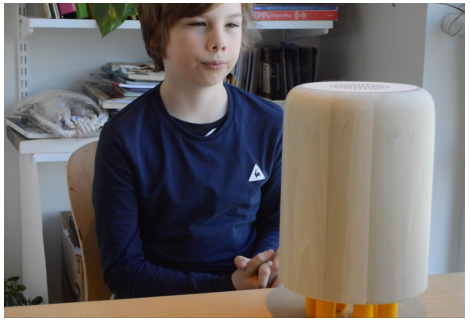
[yelling] Alexa, who is the oldest person living?

The oldest person is Nabi Tajima.

Where does Nabi live?

Alexa! Where does Nabi live?

Sorry i'm not sure about that..



Students speak to a prop VUI, 2018

Audio Abstractions

Question: I was curious about how the students might respond to a VUI that communicated with sound rather than language.

Method: So, I introduced them to “Mo,” a prop VUI concealing a blue-tooth speaker. Students were told that Mo is not as smart as Alexa and can only respond to yes or no questions. In reality, I was controlling Mo with my mobile device, playing one of four sounds after students asked questions. With four sounds, it was less clear to the students what each particular sound symbolized.

Observations

- As soon as it became clear that Mo’s responses would not be obvious, (the sounds were intentially difficult to identify as positive or negative) the students began quizzing Mo with questions they clearly knew the answers to: does 2 plus 2 equal 4?
- Almost all of the students eventually concluded that the slightly lower pitch sounds meant no.
- The students assumed that they needed to activate Mo with a wake word.

Insight: When communication is restricted to abstract sounds feedback loops are accelerated and engagement is sustained.

A prop VUI conceals a
bluetooth speaker, 2018

A series of documented conversations, 2018

Observing Conversation

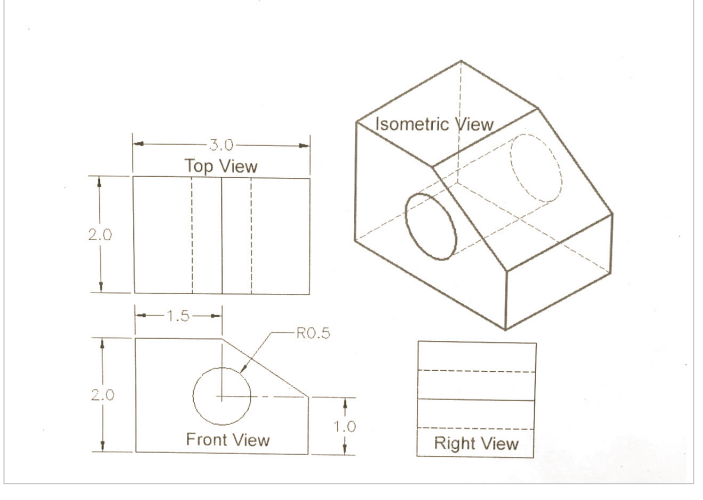
Questions: How does non-speech sound inform casual conversation? What are the subtle cues that help us to understand the needs of those we communicate with?

Method: Study 5 minute conversations between pairs of people.

Observations

- I noticed that laughter functioned as a fail proof silence filler, and a means of communicating attention, encouragement, and approval. In some cases, repeated bouts of laughter from the same individual seemed to suggest a performer/spectator dynamic. In some cases laughter served as a respectful way to break the stress of eye contact. I noticed laughers looking upwards, closing theirs eyes, and bending forward as if to momentarily leave the presence of the other individual.
- I also noticed that explicit turn taking was absent from most conversations. Strong agreement was often expressed when one party spoke simultaneous to the other party saying things like “right right right” or “yes.” These words were staccato and certain when articulated over the speaker, but slower and more questioning when articulated through silence.
- Words like “um” “uh” “erm” “ah” proved elastic. While generally speakers seemed to use them to say “i’m thinking “ the melody that carried this sound and sentiment varied drastically. Based on how prepared they were to respond.

Insight: Non verbal vocalizations allow us to augment the flow of communicated information.



Experiment participants discuss the results of their collaboration , 2018

Order of Explanations

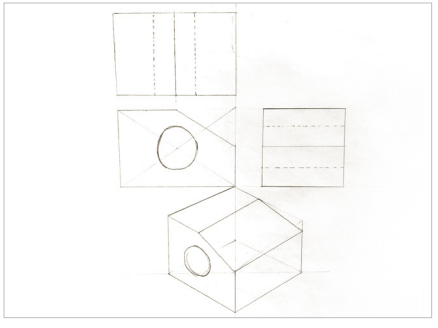
Questions: When conveying information in a crunch what dictates the order and cadence of our dialogue? How do we ascertain whether the listener has correctly understood? How do we troubleshoot communication errors?

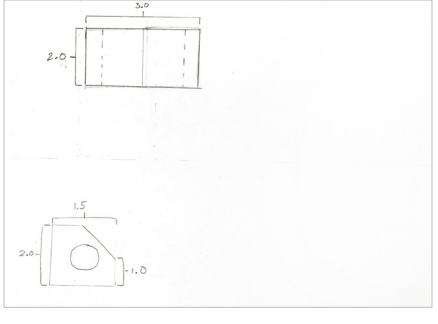
Method: During this exercise participants worked together to recreate an isometric projection. One participant served as the viewer and the other, the artist. The viewer was tasked with observing and explaining the image while the artist was tasked with recreating it according to their description. Each pair was allotted 10 minutes.

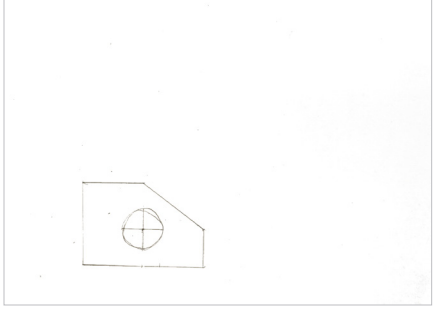
Observations

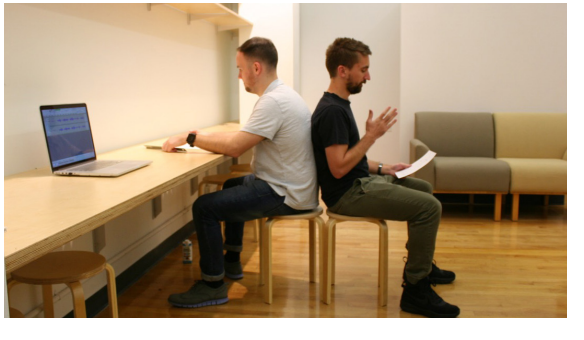
- Viewer participants tended to give instructions in two part assertions- first, making a statement about an objective fact e.g. " there is a shape in the bottom left hand corner," then giving an instruction e.g. " draw a two inch straight line one inch above the bottom edge of the page."
- There were many occasions when the team members became confused and collectively decided to begin a given part of the task over from the beginning.
- In every case the viewer described the nature of the drawing e.g. isometric projection but neglected to describe the actual object it depicted.
- The most successful team was the most closely acquainted previous to the experiment.
- It's difficult to teach or explain without oscillating between speaking descriptively and didactically.
- Its much easier to start fresh than to troubleshoot an instructional miscommunication

Insight: The ability to construct and navigate within contextual mental models is essential to the comprehension of non visual information.

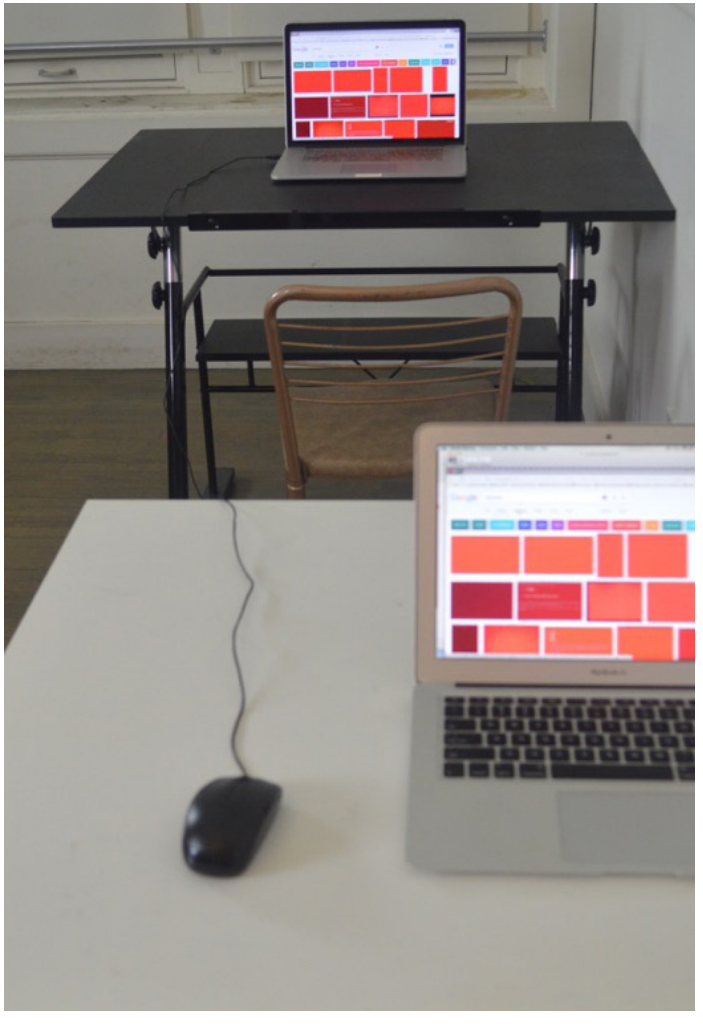








Participants' isometric drawings show
a range of completion levels, 2018



A screen share and elongated mouse cord connect participants' computers, 2018

Hands Off

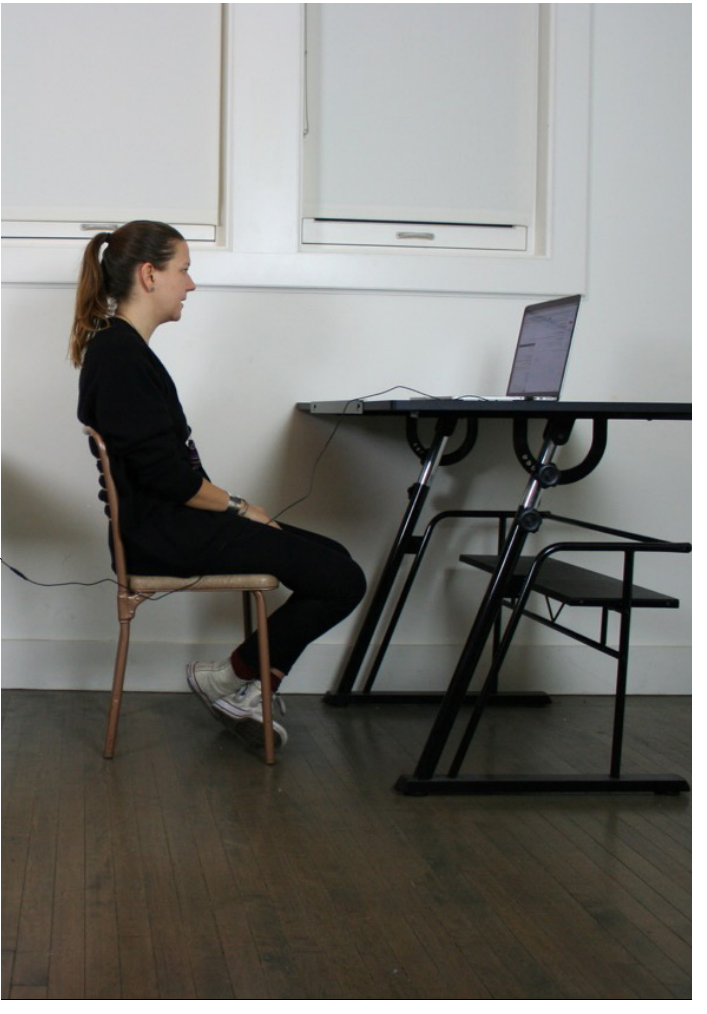
Questions: Computers extend the human body into digital space. Our fingertips indirectly touch the screen when we use a mouse and keyboard. But can our voices participate?

Method: In this experiment two participants shared screens. One went about her digital chores by dictating actions to the other. A computer mouse with an elongated cord connected the speaker's computer to the operator's desk.

Observations

- The speaking participant used her hands to gesticulate as she gave the typing/clicking participant instructions.
- She seemed uncertain about how close to sit to the computer screen sometimes peering into it closely at other times watching it almost as though it was a television.
- The silences when the speaker was reading seemed very still and uncomfortable, as though something had gone wrong.

Insight: Physical interface components anchor users to the task at hand.



A participant dictates commands, 2018

A participant works in a modeling program without the use of his keyboard, 2018

Human Command Line

Questions: Command lines and search bars eliminate the need to spatially locate digital tools. In some cases this facilitates work in design environments that feature many options. However, the command line requires the designer to relocate his or her eyes from the work they are engaged in to the the text entry space. I was curious about whether a vocal command line might elimate this oscilation thereby creating a more fluid modeling experience.

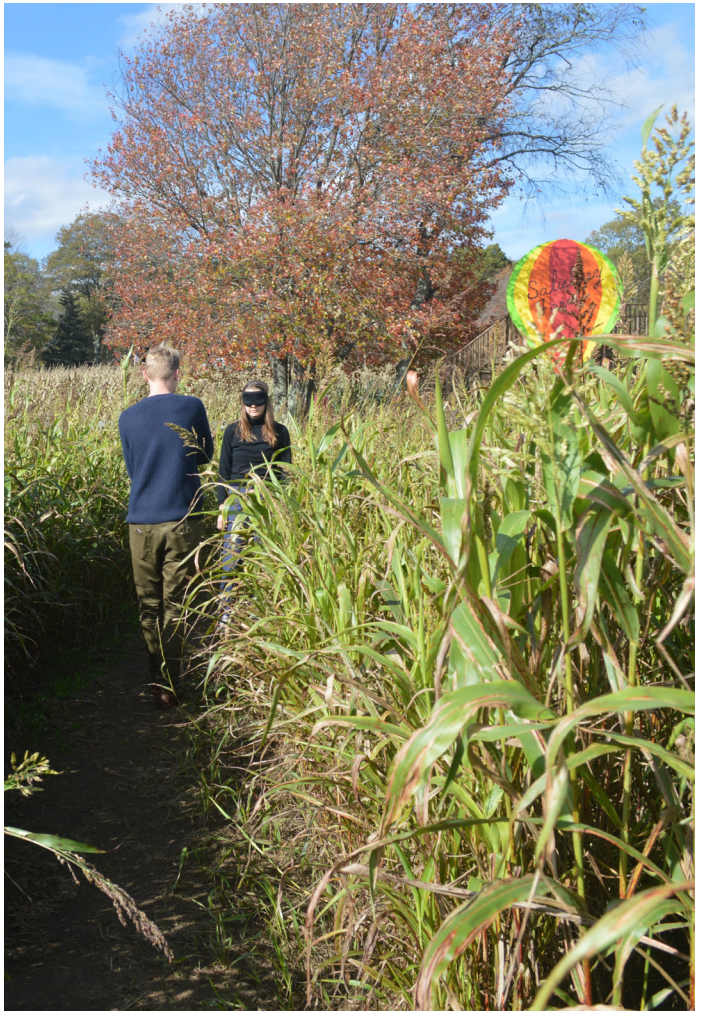
Method: Two participants with advanced working knowledge of Rhino shared screens, one in charge of typing tool names into the command line the other responsible for making all design decisions and calling out the tool names.

Observations

- This dynamic works well for typing tools into the command line. But, it is less helpful for keyboard qualifiers that need to be held down in conjunction with mouse clicks i.e. option and shift.
- The modeling participant's glance stayed fixed on his work for the extent of the experiment
- The modeling participant felt that his experience more closely related to the the act of building a model in physical space.

Insight: The ability to perform visual tasks in parallel with a voice controlled interface can create a more optimized and integrated workflow.

One participant uses a modeling software without a keyboard,
One participant types words into the command line, 2018



A blindfolded participant navigates through a corn maze, 2018

Navigating with Sound

Questions: How can sound facilitate spatial navigation? With a limited palette of sounds, what cues are most crucial to understanding direction and danger. How intuitively can sound systems be learned?

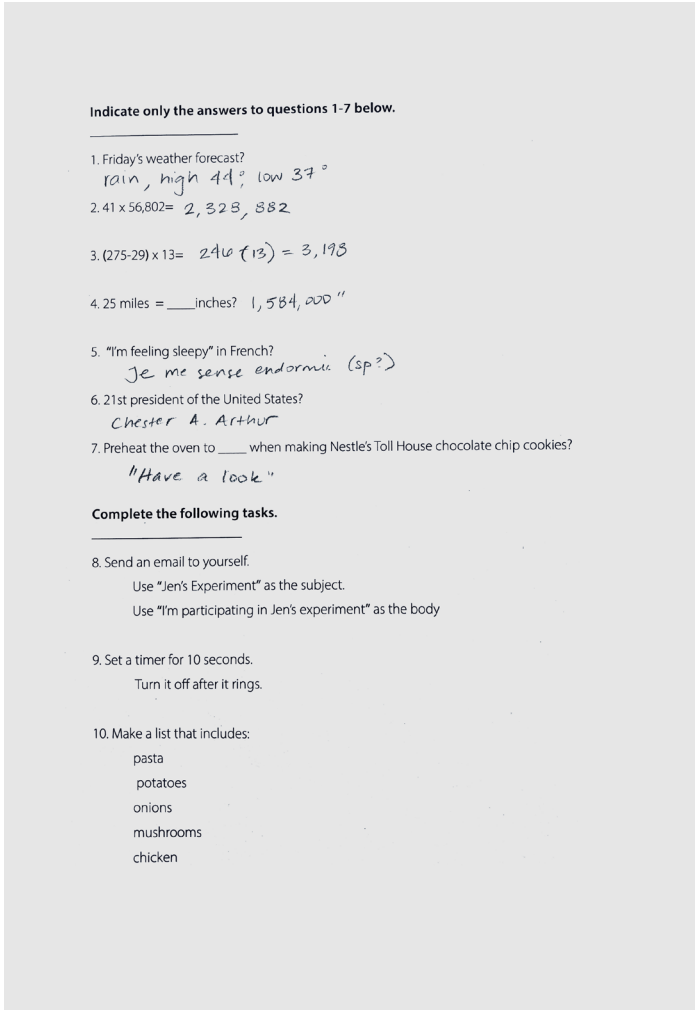
Method: Participants devised a system for communicating navigation instructions without speech. The constant jingling of coins in a glass container served as a directional guide and the clank of a chain link necklace indicated trouble ahead.

Observations

- The sound making tools were not used to indicate concrete ideas, like left and right.
- The system seemed organic.
- The constant sound of the jingling coins ensured that the walker never needed to question whether or not the leader was near by.

Insight: Repetition and reinforcement allow us to build complex and nuanced symbolic languages from even the most minimal range of sounds.

A blindfolded participant follows
the sound of a rattling chain, 2018



Indicate only the answers to questions 1-7 below.

1. Friday's weather forecast?
   rain, high 44°, low 37°
2. 41 x 56,802= 2,328,882
3. (275-29) x 13= 246 (13) = 3,198
4. 25 miles = ____inches? 1,584,000"
5. "I'm feeling sleepy" in French?
   Je me sense endormi (sp?)
6. 21st president of the United States?
   Chester A. Arthur
7. Preheat the oven to ____ when making Nestle's Toll House chocolate chip cookies?
   "Have a look"

Complete the following tasks.

8. Send an email to yourself.
   Use "Jen's Experiment" as the subject.
   Use "I'm participating in Jen's experiment" as the body
9. Set a timer for 10 seconds.
   Turn it off after it rings.
10. Make a list that includes:
    pasta
    potatoes
    onions
    mushrooms
    chicken

A quiz challenging participants to use their VUIs in unfamiliar ways, 2018

VUI Obstacle Course

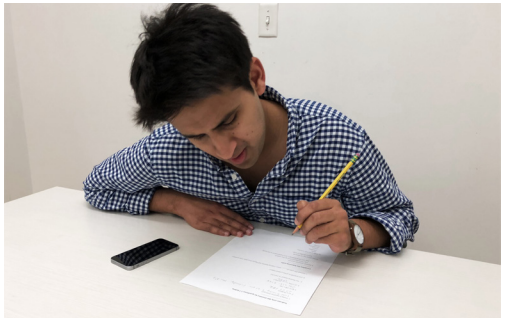
Questions: Do new users understand the nuances of VUI interaction?

Method: During this obstacle course, participants used the VUIs on their own phones to answer questions and complete tasks. 4 participants owned iPhones and used Siri. 2 participants owned androids and used Google Assistant.

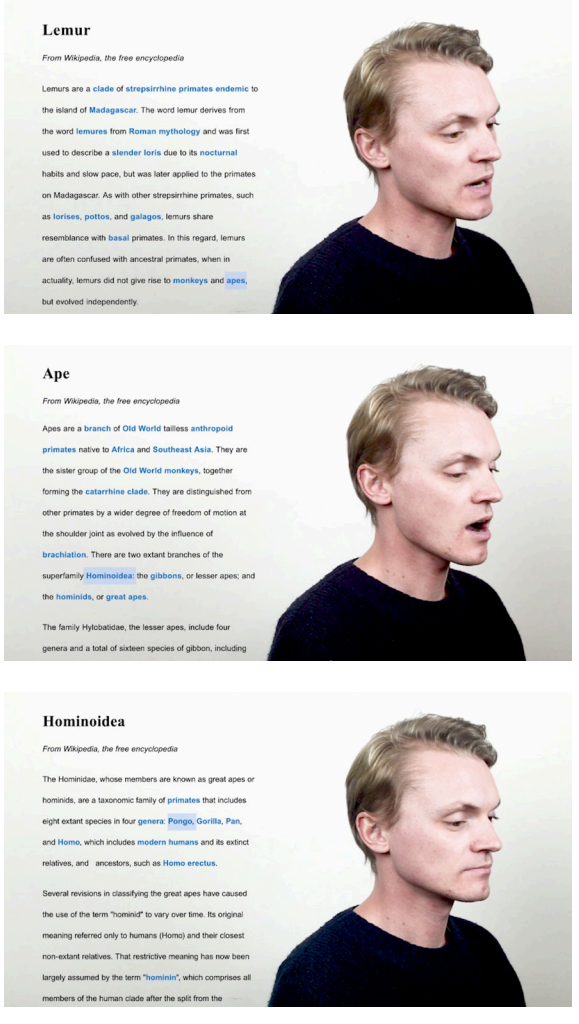
Observations

- Siri uses sound effects to indicate the metaphorical movement of a microphone between user and virtual assistant. Due to poor wifi and participant hesitation these sound effects often occurred at unexpected moments.
- Siri users did not realize that it was unnecessary to say "Hey Siri" after the "i'm listening" sound had played.
- Periods of silence seemed to suggest that the VUI was thinking when it had simply not heard the question asked.
- Participants often second guessed the phrasing of their questions after beginning to ask them.

Insight: Observing strict conversational formalities when interacting with voice user interfaces can create unnecessary obstacles in accessing information.



Participants answer quiz questions
using Siri and Google Assistant, 2018



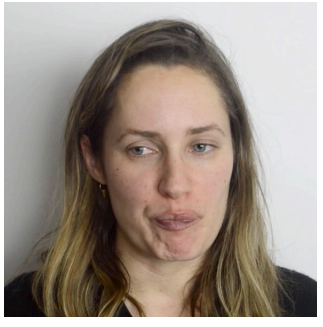
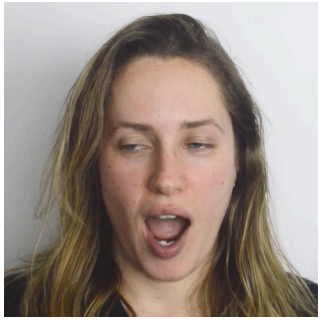
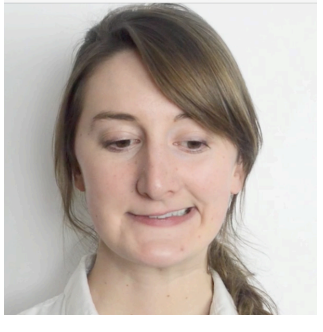
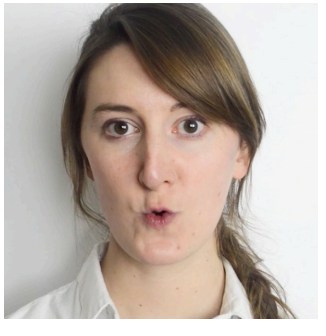
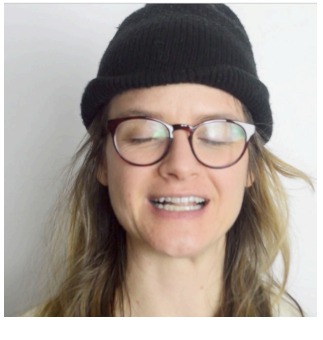
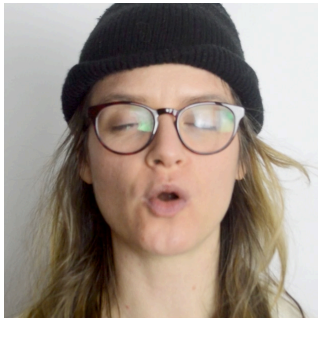
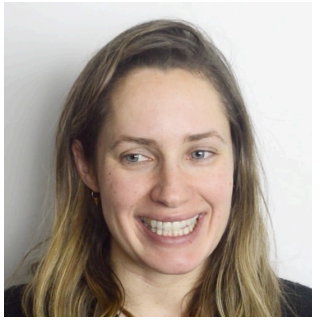
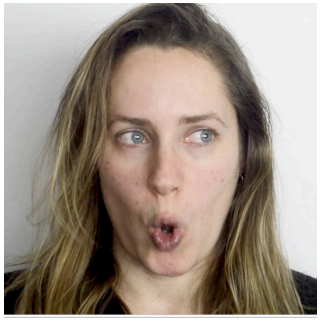
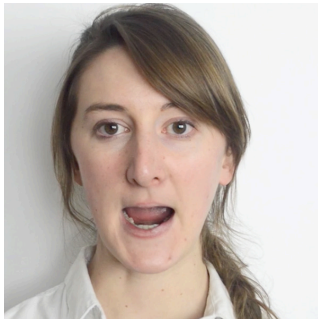
An video depicts a person selecting hyperlinks by making click sounds with his mouth, 2018

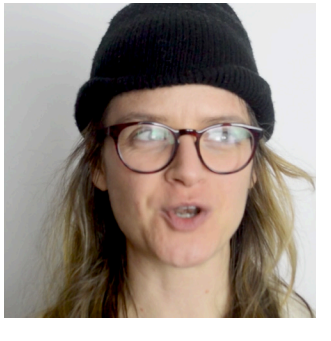
Mouth Clicks

Questions: What is the most efficient way to interrupt speech? How can the human voice function like a finger pressing a button? Can we make sounds without drawing breath? Is the cognitive impact of sound making different than the cognitive impact of speech.

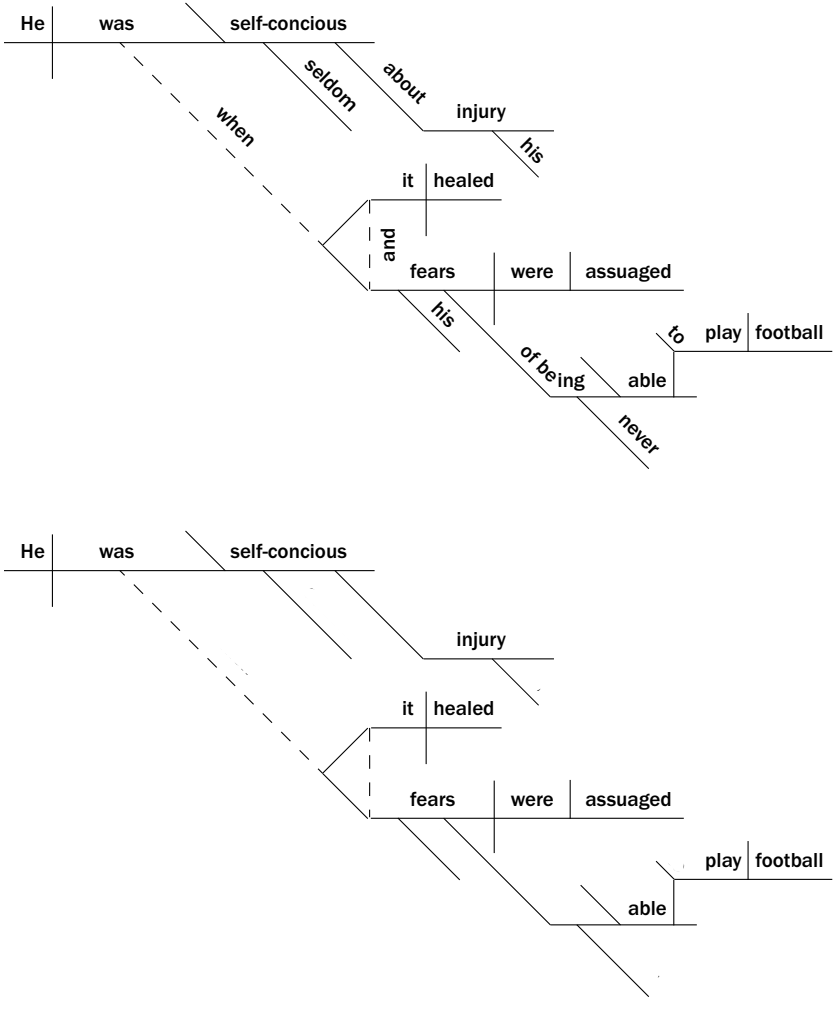
Method: During this exploration I asked 3 individuals to produce as many monosyllabic non-speech sounds as they could think of. I then used animation to showcase the use of these sounds as selection tools.

Observations:

Participants explore monosyllabic sounds, 2018



A sentence diagram displays a sentence from a popular novel

Sentence Diagrams

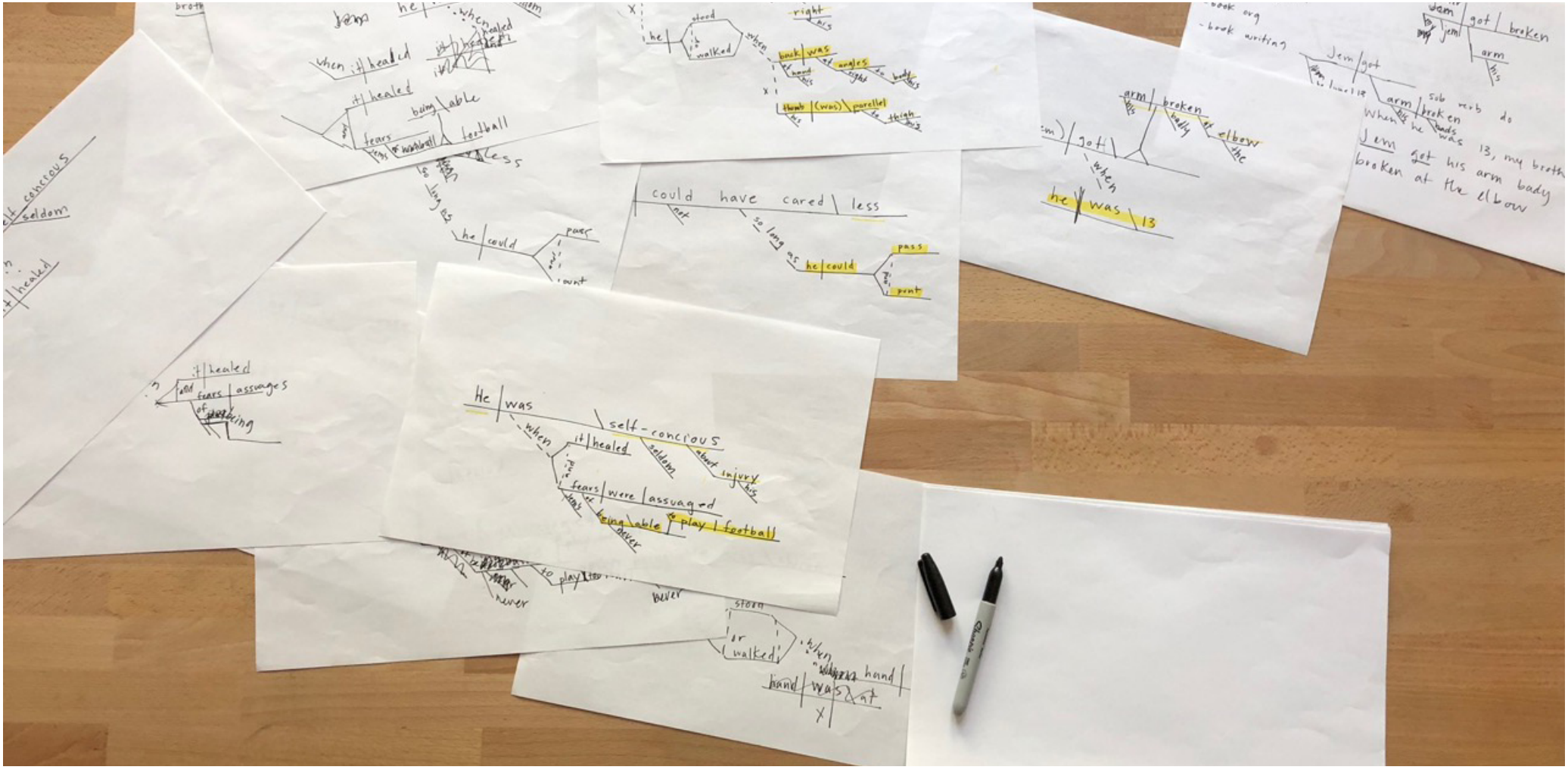
A sentence diagram is a pictorial representation of the grammatical structure of a sentence. Sentence diagrams help readers to parse out key words that capture the meaning of a sentence. These words (subject, verb, and direct object) are scribed on the horizontal while supporting descriptive and connective words are scribed on the diagonal.

Method: I created sentence diagrams for the first chapter of a popular novel to effectuate the incremental filtration of nonessential words from this text until I was left with only the subject and verb of each sentence.

Observations

- In general the essence of each sentence remained in tact until I removed the direct object. For example, “he closed” is an incomplete idea leaving the reader with the question “what exactly did he close?” However, “he closed box” answers this question with the direct object.
- When negative descriptive words such as never, seldom, and not were featured in sentences, filtering away these words confused the meaning of the sentence. For example , “he didn’t close the box” becomes “he close box” when the descriptive and connective words are filtered away.

Insight: Ideas can be distilled into short groupings of words that carry meaning even though they don’t sound like conventional language.



A sea of papers showcase hand-drawn sentence diagrams, 2018

Deliverables

Finally I set out to envision aspects of the systems my insights and observations seemed to be pointing towards. My goal was not to work out a fully functional interface, but rather to create some food for thought—a sort of prop to help researchers and designers consider alternatives to conversational interfaces.

Scenerio: Jane opens and explores a saved document containing her biology reading assignment.

Interaction: She uses a scroll wheel to scan and navigate the content— speeding past information, adjusting her position, and changing the verbosity in each sentence.

Scanning

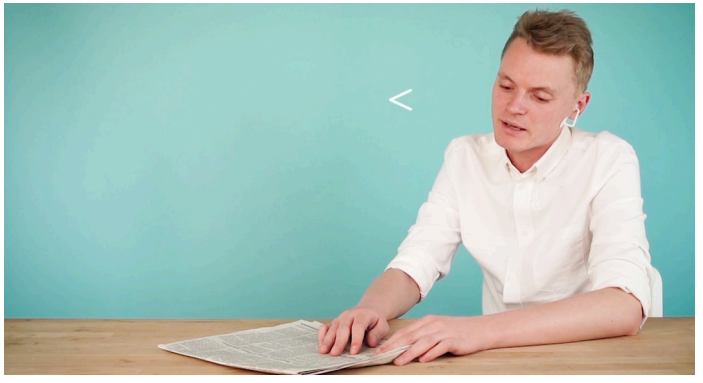
This vignette depicts a user changing qualities of synthesized speech on the fly. Similar to visual reading, she does not carefully dissect each sentence, rather she uses different tools to get an overall feel for the content.

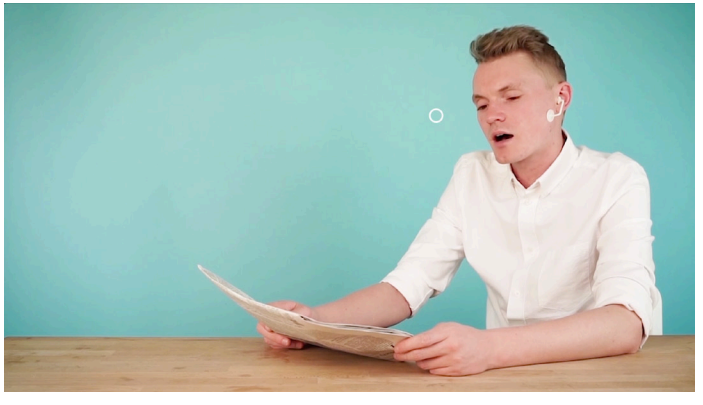
Scenerio: Alex searches through her emails from Paul. She is looking for a particular message.

Interaction: sound cues help Alex anticipate the information she is about to encounter. Lists are prefaced by a succession of staccato sounds corresponding to list length. Bodies of text are preceded by melodies.

Symbolic Sounds

My research suggested that abstractions are useful ways to qualify and describe different classes of information. In this vignette I sought to describe what it might be like to interpret sounds that quickly and impressionistically give users a sense of just how much information they are about to experience.





Scenario: Alan comes across an intriguing artist while reading the Sunday times. He falls into a Wikipedia rabbit hole by looking up his name.

Interaction: sound cues identify hyperlinks within webpages. Alan uses monosyllabic utterances and commands to select, interrupt and pause the text.

Nonvisual Navigation

Finally, I explored the possibility of a user interrupting a VUI with little effort. I imagined monosyllabic sounds as interaction tools, aiding a user in making selections and navigating forwards and back.

Conclusion

Before the VUI, digital sounds were activated and deactivated. We turned music and podcasts on or off; we set alarms and timers; we agreed to receive push notifications. Sound was triggered, but we did not interact with it. Recent advances in technology facilitate more complex operations with more complex content.
This prompts two questions:

1. How can audio interactions complement our present behaviors?

2. How can audio interactions complement our future behaviors?

Designers and researches must thoroughly investigate each of these questions to understand the full potential of audio interfaces like the VUI.

1. Observing Present Behaviors

As mentioned earlier in this report, human centered VUI designers have flagged the human behavior of engaging in conversation a nonnegotiable around which to design voice interactions. Companies like Google and Amazon have outlined key ideas that serve to guide designers in applying the principles of human to human conversation to human to VUI interaction.

But conversation is just one of many user behaviors that may apply to VUI design. In my research and contemplation, I have noted that non-mannered speech, monosyllabic utterances, and even the principles of GUI design might well be applied to VUI design.

The visual language which communicates the desktop metaphor is largely irrelevant to the design of VUIs. However, the structural system it represents should not be regarded as such. With the advent of the personal computer, common objects like trashcans, folders, and notepads were digitized. Today these digital tools are possibly more ubiquitous than the physical tools from which they originated. These mechanisms are familiar, fast, and useful. Designers must question whether or not some of their operations, absolved of their visual signifiers, might be integrated into a non-visual interface.

2. Imagining Future Behaviors

For thousands of years information has been scrawled, typed, and programmed into rectangles. This is what we're used to. Static words on paper and screens give users agency to scan, skip, and reread. But what if these operations, had been possible utilizing speech synthesis and sound from early days? Perhaps we would use our senses quite differently.

The mastery of screen-readers by the blind and visually impaired demonstrates that humans are capable of using their ears to perform (what many believe to be) the work of their eyes. Designers and researchers need to consider this and other untapped human capabilities when shaping the future of the VUI. Users may one day become adept at interacting with systems through sound interpretation and synthesized speech comprehension.

Paths Forward

Ultimately this thesis isn't about using sound and speech as interaction tools for the mere purpose of engaging in whimsical experimental design. Rather it is about questioning the conversational nuances that slow down and confuse the exchange of information between VUI and person. My goal is to not to encourage the design of an unusual interface that performs in entertaining ways but rather to promote a system of non-visual interaction that doesn't leave the user waiting for an unknown entity to slowly deliver the answer.

It is clear that VUI advancements are aimed at developing computer personalities that are indistinguishable from human personalities. This controversial future should not be the only path forward. Designers need to begin questioning how we can use VUIs to teach users, rather than give them the answers. As the first popular household AI it is crucial that VUIs set the tone for a future in which advancements in machine intelligence beget advancements in human intelligence.

Bibliography

Articles:

Brautigam, Bert. “From Human-Centered Design to Design for Enhancing Human Ability.” The Next Web. Accessed May, 2018. https://thenextweb.com/guests/design-for-enhancing-human-ability/.

Brautigam, Bert. “The New Skeuomorphism Is in Your Voice Assistant – UX Collective.” Accessed May, 2018. https://uxdesign.cc/the-new-skeuomorphism-is-in-your-voice-assistant-3b14a6553a0e.

Bush, Vannevar. “As We May Think - The Atlantic.” Accessed May, 2018. https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/.

Cooke, Henry. “Prototyping for Voice: Methodology - BBC R&D.” Accessed May, 2018. https://www.bbc.co.uk/rd/blog/2017-06-voice-ui-user-interface-children-drama.

Eifler, M. “The Office of the Future | Elevr.” Accessed May, 2018. http://elevr.com/the-office-of-the-future/.

Hendron, Sara. “Investigating Normal. | Abler.” Accessed May, 2018. https://ablersite.org/investigating-normal/.

Lockton, Dan. “Exploring Qualitative Displays and Interfaces .” Accessed May, 2018. https://medium.com/@danlockton/exploring-qualitative-displays-and-interfaces-b97ebfa908cc.

"Our Friends Electric - Superflux." Accessed May, 2018. http://superflux.in/index.php/work/friends-electric/#.

Shafrir, Doree. "Meet The People Who Listen to Podcasts Crazy-Fast." Accessed May, 2018. https://www.buzzfeed.com/doree/meet-the-people-who-listen-to-podcasts-at-super-fast-speeds?utm_term=.tmJBAL0nYn#.ufboxKz8r8.

"Skeuomorphism Is Dead, Long Live Skeuomorphism | Interaction Design Foundation." Accessed May, 2018. https://www.interaction-design.org/literature/article/skeuomorphism-is-dead-long-live-skeuomorphism.

Vartanian, Hrag. "Want a Smart Home Assistant? Invite This Artist to Watch You for Three Days Instead." Accessed May, 2018. https://hyperallergic.com/417839/lauren-mccarthy-smart-home-assistant-interview/.

Victor, Bret. "A Brief Rant on the Future of Interaction Design." Accessed May, 2018. http://worrydream.com/ABriefRantOnTheFutureOfInteractionDesign/.

Books:

Burns Florey, Kitty. Sister Bernadette's Barking Dog: The Quirky History and Lost Art of Diagramming Sentences. Melville House Publishing, 2006.
Lakoff, George, and Mark Johnson. Metaphors We Live By. Chicago: University of Chicago Press, 1980.

Images

Page 24 / Books Interface- Pinterest.com
Page 28 / Refreshable Braille Display- Wikipedia.com
Page 44 / Google Material Design- https://material.io

all other images are owned by the author