

A watercolor illustration featuring several stylized robots and a woman. The robots are depicted in various colors and designs, including blue, red, and grey. One robot in the center has a blue body and a red head. Another robot to the right has a red body and a blue head. A woman with blonde hair, wearing a red top and an orange skirt, stands on the right side. The background is a light, textured surface.

# Human AI Interaction

Lecture 12: Voice  
[aidesignclass.org](http://aidesignclass.org)

# Today: Interacting with voice

- What are the basics of interacting with voice?
- What is likely to remain the same despite better voice recognition and generation?
- Is voice always the answer? Why not?

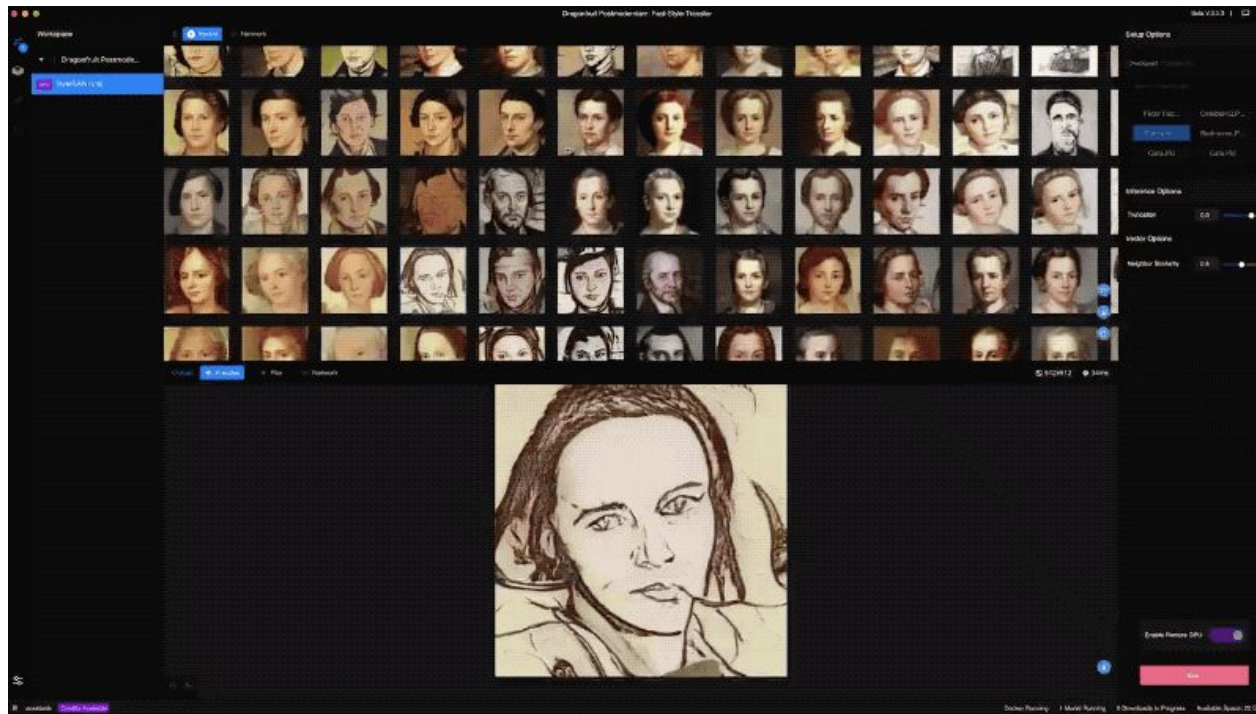
# UX for images

RunwayML

Image from [Verge](#)

What is good about this UX?

What isn't?



# Designing with voice

Voice seems like a “natural” way of interaction.

- Is it faster?
- What are the main design challenges?
- When should you (not) use voice?

# OK, Google (What should we use voice for?)

Questions:

- How compelling is this demo?
  - Would you use it?
- What feels great about it?
- What seems off?



# Voice vs. text

## Text:

Slower to produce

Faster to ingest

Can skim, search for keywords

20% or more: only basic writing skills

## Voice:

Dictation 3x faster than typing (even more on small screen)

Slower to ingest

Poor ability to skim, search

Users are widely proficient

# Designing with voice

Voice seems like a “natural” way of interaction.

- Is it faster? **Yes to speak, not to listen**
- What are the main design challenges?
- When should you (not) use voice?

# Grice's maxims

The *cooperative principle* describes how people achieve effective conversational communication in common social situations. (this is conversational cooperation, not necessarily social cooperation.)

In conversation, people follow:

- Quality: be truthful, substantiated by evidence
- Relation: say what's relevant, omit what isn't
- Manner: be clear, avoid ambiguity, obscurity
- Quantity: as informative as needed, but no more

Violations of this maxim happen for a reason.



# A haircut appointment

The promise



# A haircut appointment

“I don’t have time for spam calls,” she explains. “I’m busy enough as is.” ... “I purposely ignored those calls because it said ‘Google,’” she says...

This time, she listens intently without ever responding to the AI. “That was weird,” she says as she hangs up. “I’m a little freaked out.”



**M**yriah Q. hasn't stopped moving since the moment I entered the bar. She's got patrons seated on the sidewalk and the backyard areas, and she is pacing between opposite ends of the venue to keep up with the happy hour rush. Occasionally, she hops behind the counter to mix drinks, restart the music playlist, or

# Could we have predicted this reaction with Grice's maxims?

The *cooperative principle* describes how people achieve effective conversational communication in common social situations. (this is conversational cooperation, not necessarily social cooperation.)

In conversation, people follow:

- Quality: be truthful, substantiated by evidence
- Relation: say what's relevant, omit what isn't
- Manner: be clear, avoid ambiguity, obscurity
- Quantity: as informative as needed, but no more

Violations of this maxim happen for a reason.

# Discovery

[‘Alexa also happens to be fairly dumb.’](#) She doesn’t understand the phrase “I’m awake.””


Num. requests	Percent	Intent
30290	52.75	Search (card result, if applicable or search page)
9457	16.47	Navigate directly to a page
3031	5.28	Play music
1890	3.29	Search within a page for a query
1459	2.54	Read an article aloud
1176	2.05	Find and focus a tab
1094	1.91	Focus on a music player page
888	1.55	Close current tab
794	1.38	Perform a Google search
766	1.33	Go to the next search result

Query types we found for Firefox Voice

ALEXA, DO MORE

TECH INNOVATION

## Amazon’s Alexa isn’t the future of AI—it’s a glorified radio clock, and stupid otherwise



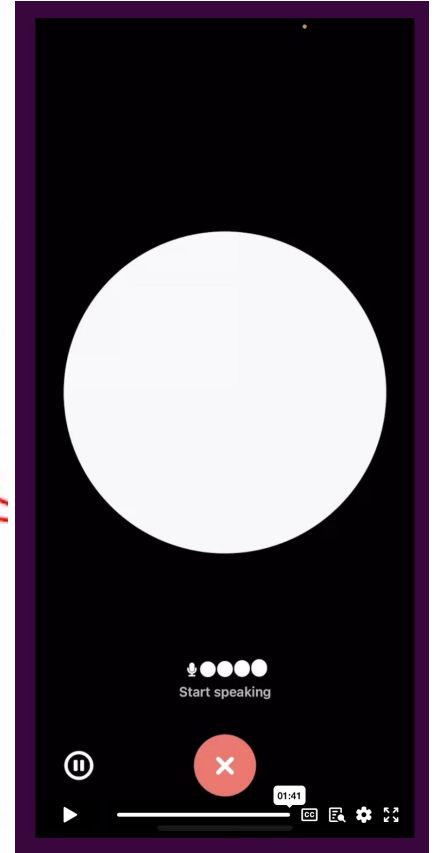
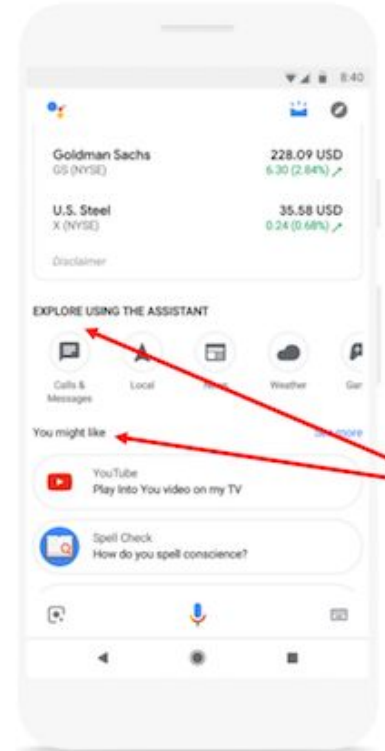
Alexa likely demonstrating her best skill: saying “I don’t know.”

Image: AP Photo/Jeff Chiu

The image shows a close-up of a person's hand touching the top of a black Amazon Echo smart speaker. The speaker has a glowing red light ring around its top edge. The background is blurred, showing what appears to be a kitchen or living area with white cabinets and a wooden countertop.

# In what ways can you help users discover what a voice interface can do?

Right: Google Assistant (clipped, from voicebot.ai), ChatGPT voice input. Below “Hey Disney!”



# Designing with voice

Voice seems like a “natural” way of interaction.

- Is it faster? **Yes to speak, not to listen**
- What are the main design challenges?
  - Discovery
  - Following the cooperative principle
- When should you (not) use voice?

# Understanding speech by understanding text?

Many systems today follow:

Speech -> text -> parse  
text to find intents -> fulfill  
intents

What could go wrong?



# Designing with voice

Voice seems like a “natural” way of interaction.

- Is it faster? **Yes to speak, not to listen**
- What are the main design challenges?
  - Discovery
  - Following the cooperative principle
  - Non-textual information is hard to capture
- When should you (not) use voice? And what should you do instead?
  -



# Put that there

One answer is  
multimodal: use  
speech **and**  
something else

When does this  
work? When not?



# Put that there

One answer is multimodal: use speech **and** something else

- Plan for errors
- Design so speech and other modes cancel each others' errors

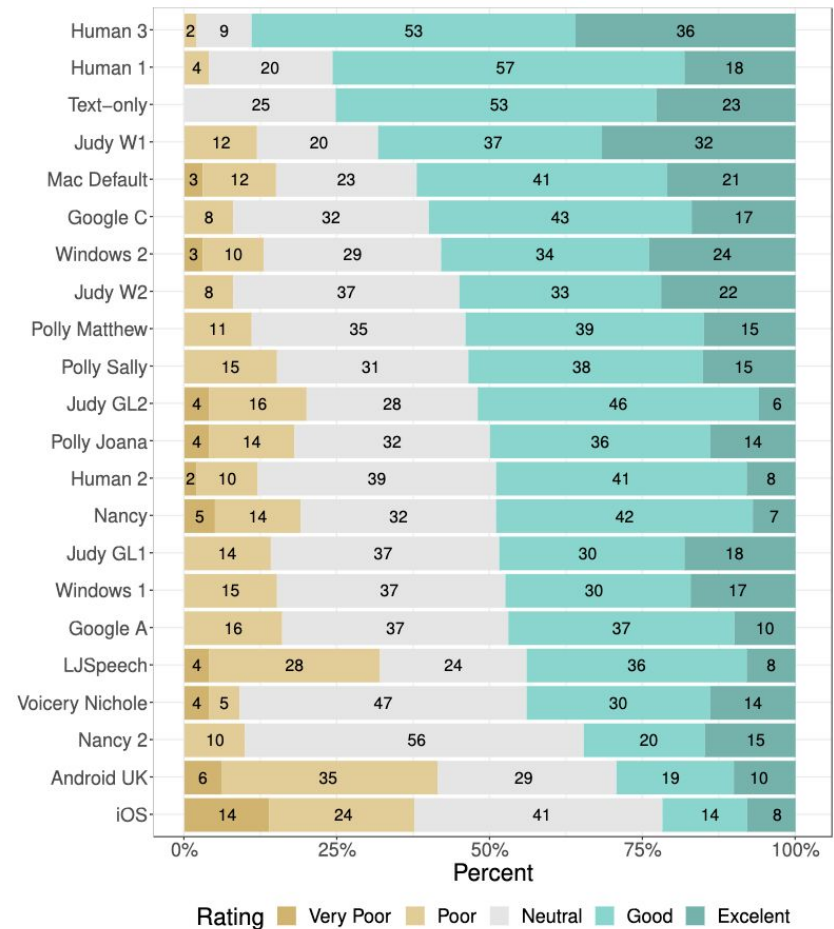


# Does it sound human?

- Many computer voices are now preferred to human voices
- (Yet no voice is universally preferred)
- Study from 2020 – things may have changed

Bottomline: Users are comfortable with machine voices

Chart [courtesy of Julia Cambre](#)



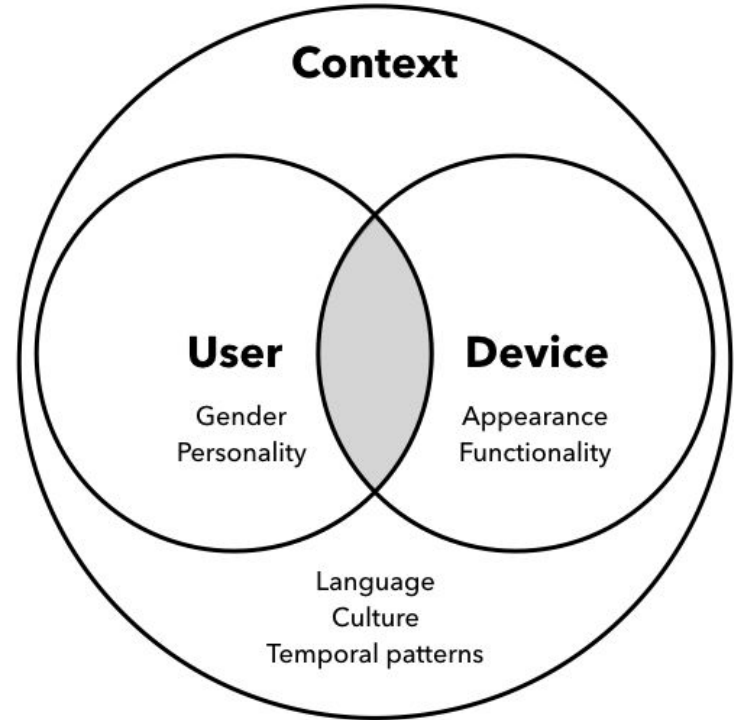
**Figure 1. Percentage of positive, neutral, and negative listening experience ratings for each voice, ordered by positive ratings.**

# Which voice is preferred?

It depends!

- on user
- on embodiment of device
- Cultural and language preferences

No one voice is universally “best”



# Designing with voice

Voice seems like a “natural” way of interaction.

- Is it faster?
  - Yes to speak, not to listen
- What are the main design challenges?
  - Discovery
  - Following the cooperative principle
  - Non-textual information is hard to capture
- When should you (not) use voice? And what should you do instead?
  - Use multimodal interfaces if possible
  - Use different modalities to compensate for each other



# Dial 911 to order pizza?

