



Introduction to Intelligent User Interfaces

Voice-User-Interfaces

Language is the most crucial channel of communication between people

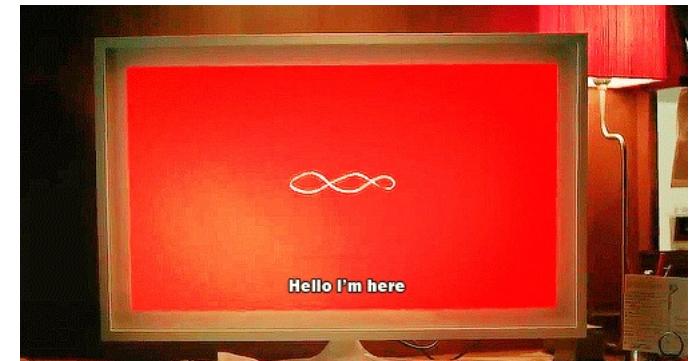


Voice user interfaces in pop culture

History



JARVIS | Iron Man



Samantha | Her

Hal 9000 | A Space Odyssey

The beginnings of voice user interfaces

History

```
Welcome to
      EEEEEE LL     IIII  ZZZZZZ  AAAAAA
      EE   LL     II    ZZ  AA  AA
      EEEEEE LL     II    ZZZ  AAAAAAA
      EE   LL     II    ZZ  AA  AA
      EEEEEE LLLLLL IIII  ZZZZZZ  AA  AA

Eliza is a mock Rogerian psychotherapist.
The original program was described by Joseph Weizenbaum in 1966.
This implementation by Norbert Landsteiner 2005.

ELIZA: Is something troubling you ?
YOU: Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU: They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU: Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU: He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU: It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU:
```

ELIZA | Weizenbaum, 1964



Interactive Voice Response
Systems (IVRs) | ~2000s



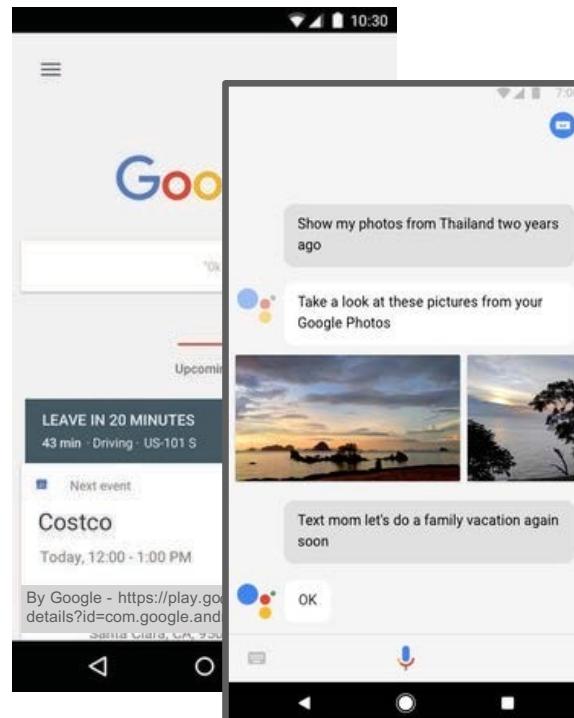
Watson | IBM, 2006

Today's voice user interfaces

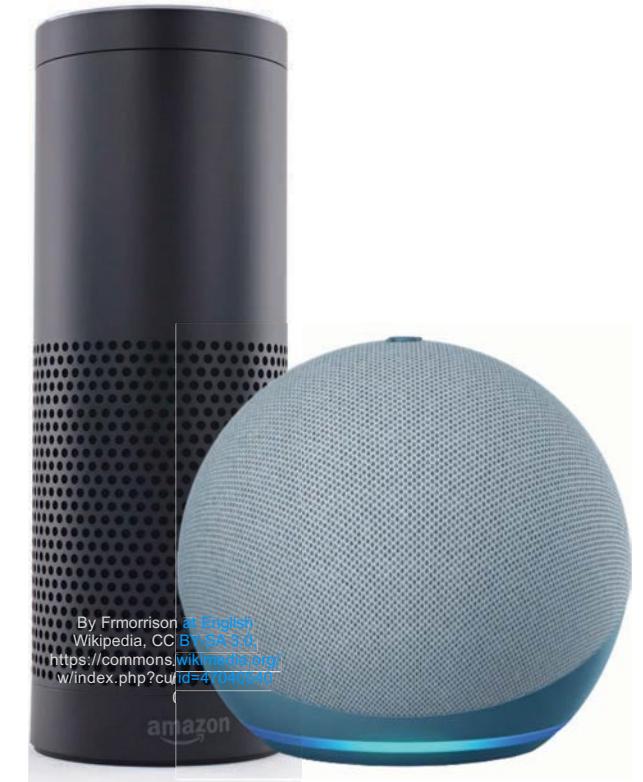
History



Siri | 2011



Google Now | 2012
Google Assistant | 2016



Amazon Echo | 2014
with Amazon Alexa

What are Voice User Interfaces?

“A Voice User Interface (VUI) is what a person interacts with when communicating with a spoken language application.”

[Cohen et al. 2004]

- Counterpart to a graphical user interface (GUI)

“Spoken dialogue systems allow humans to interact with a computer on a turn-by-turn basis using spoken natural language for input and output.”

[McTear 2017]

- VUIs communicate with users in spoken natural language

What are Voice User Interfaces?

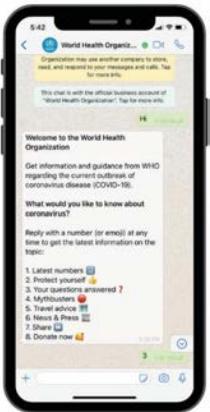
Conversational user interfaces (CUI)



<https://chat.kuki.ai>

embodied

Human-like representation:
More categories within embodied CUIs



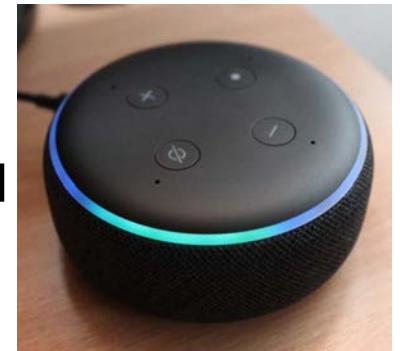
e.g. WhatsApp bot by the WHO

disembodied

text-based

Often referred to as chatbots

speech-based



e.g. Amazon Alexa, Google Assistant, Apple's Siri

How are VUIs used?



On smartphones



On smart speakers



In cars

What do users do with smart speakers?



Search

*“how many hours
are in a year”*



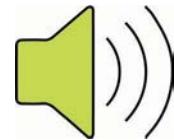
Music

“shuffle songs”



IoT

“turn on kitchen light”



Volume

“turn the volume to six”



Weather

*“is it gonna snow two
days from now”*



Joke

*“tell me a star
wars joke”*



Conversational

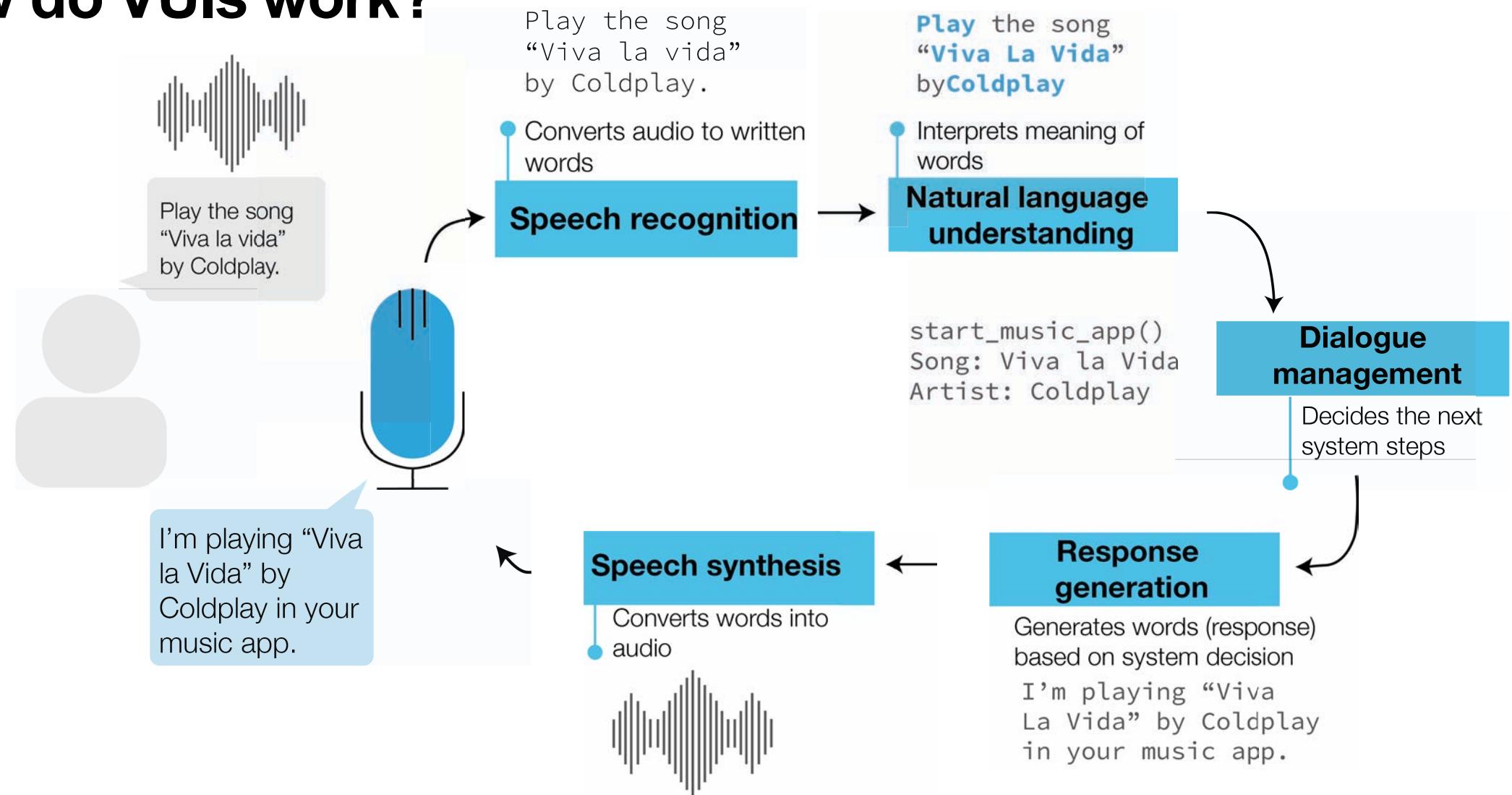
“tell me a bedtime story”



Alarm

“alexa snooze”

How do VUIs work?



VUIs are activated by a wake word

Wake word: “a word or words that you say in order to make an electronic device, or a feature on a device, ready to work” [Cambridge Dictionary]



“Ok Google”

“Hey Siri”



Advantages of VUIs

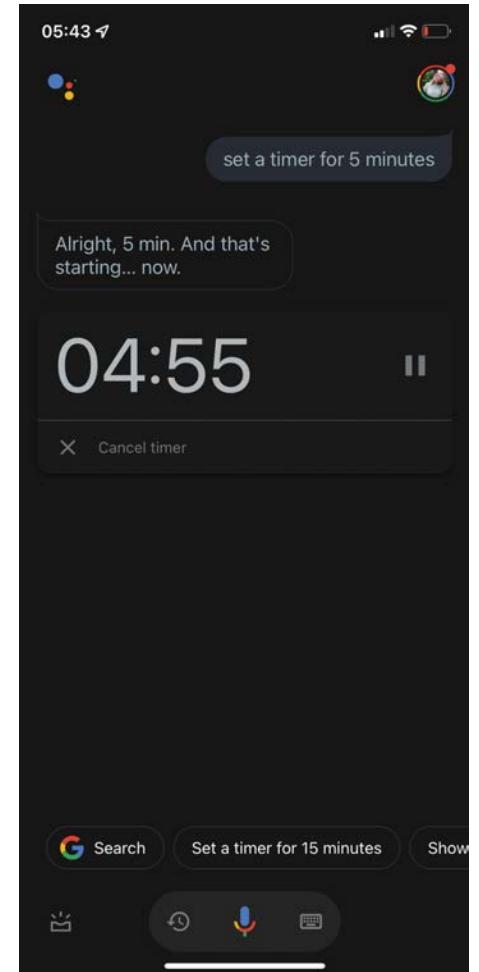
Advantages of VUIs

- Easy-to-learn natural interaction

Where do I find the timer?



**Users have to learn
GUIs**



**Users know how to
use language**

Advantages of VUIs

- Easy-to-learn natural interaction
- Hands-free interaction



Advantages of VUIs

- Easy-to-learn natural interaction
- Hands-free interaction
- Fast interaction

Spe\$%&n(&n *n (+&sh
\$nd /\$nd\$0&n &s
\$+12s3 4 3&les
5\$ss3e0 3h\$n 60&3&n(



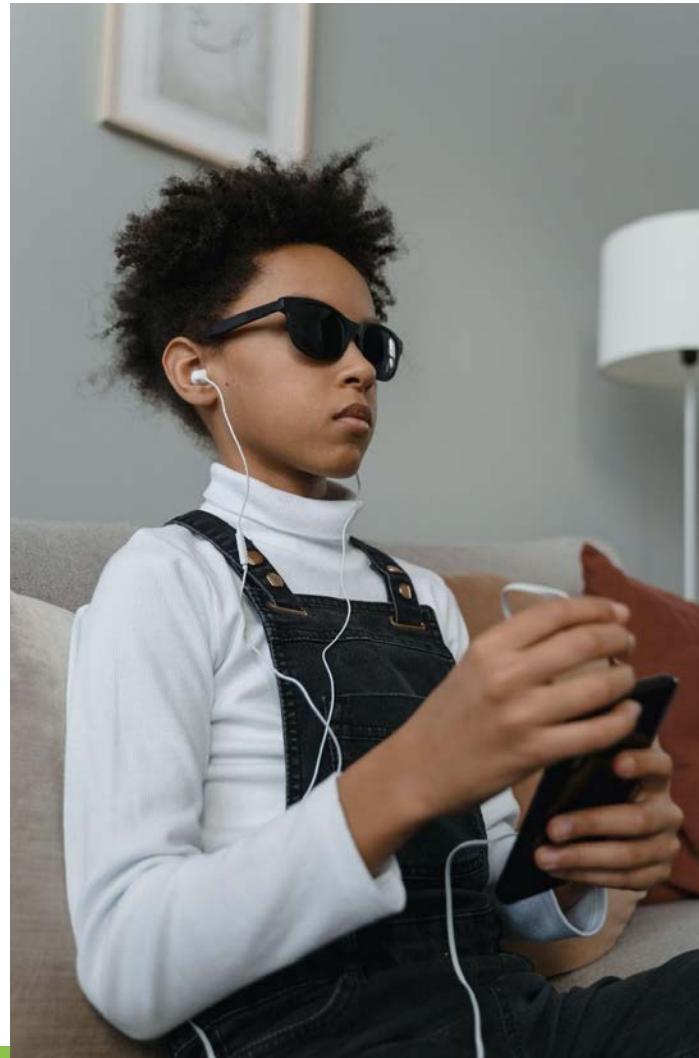
Speaking in English and Mandarin is almost 3 times faster than writing

Please note that this has only been tested in ideal lab settings.

[Ruan et al. 2018]

Advantages of VUIs

- Easy-to-learn natural interaction
- Hands-free interaction
- Fast interaction
- Accessibility



Challenges for VUIs

Challenges for VUIs

- Precision of speech recognition

- **Homophones:** similar sounding words that have different meanings. Humans know which word from the context, e.g.
 - *There vs their vs they're*
 - *Flower vs flour*
- **Accents/dialects:** Indian or Singapore English dialect
- **Speech disabilities:** e.g., stuttering, Dysarthria (weakened speech muscles)

Challenges for VUIs

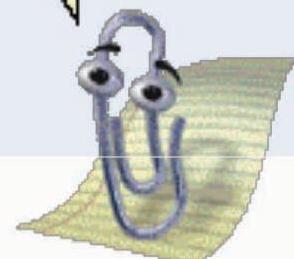
- Precision of speech recognition
- Interpreting the context

People extract more information from what someone said than was literally said.

It looks like you're writing a letter.

Would you like help?

- Get help with writing the letter
- Just type the letter without help
- Don't show me this tip again

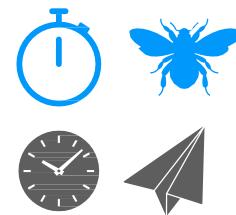


Challenges for VUIs

- Precision of speech recognition
- Interpreting the context

Syntactic ambiguity

A sentence can be interpreted in different ways due to an ambiguous sentence structure



Time flies like an arrow, fruit flies like a banana.

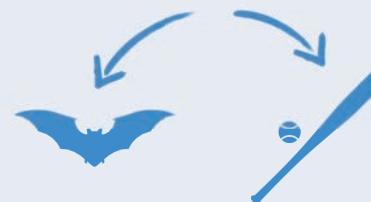
Challenges for VUIs

- Precision of speech recognition
- Interpreting the context

Semantics

Words may have different meanings, making it difficult for a VUI to differentiate between them.

Voice assistant, which bat should I get?



Challenges for VUIs

- Precision of speech recognition
- Interpreting the context

Pragmatics

Language always needs to be understood in the context.

That's a typical example of a job well done

You've a green light

Challenges for VUIs

- Precision of speech recognition
- Interpreting the context

Winograd Schema

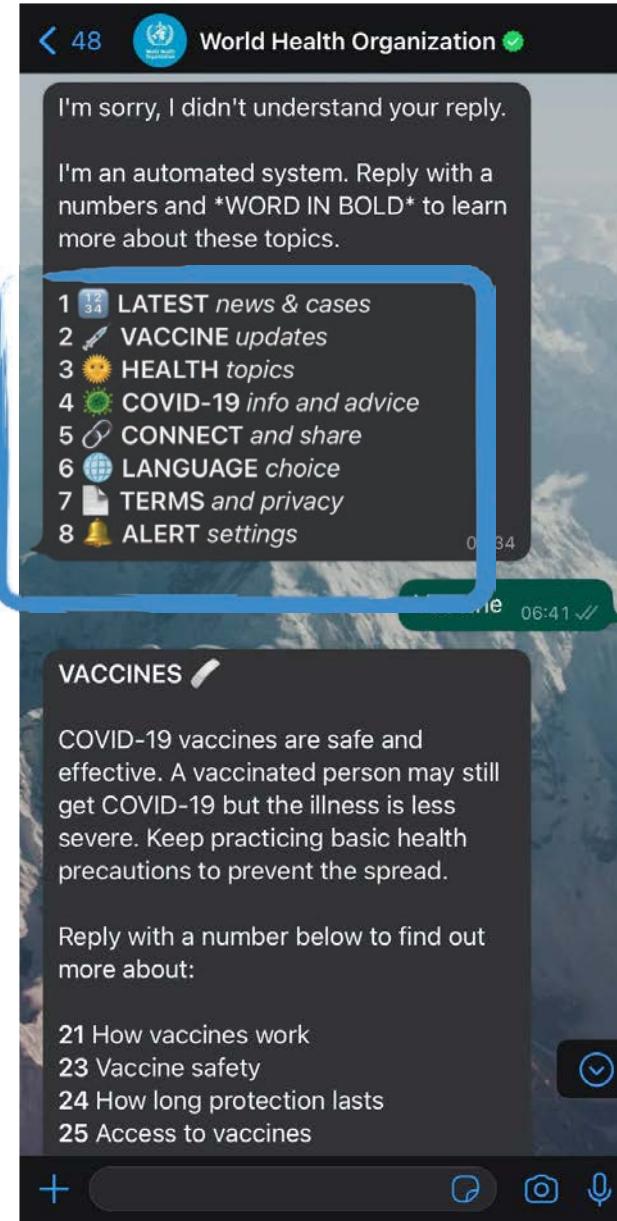
Computers don't have common sense. Meaning that is clear to us due to the context is difficult to teach to a computer.

I can't fit the book into the shelf because it is too narrow.

I can't fit the book into the shelf because it is too wide.

Challenges for VUIs

- Precision of speech recognition
- Interpreting the context



Today's non-LLM chatbots often do not have real natural language understanding but use given selections instead

Challenges for VUIs

- Precision of speech recognition
- Interpreting the context
- Safety and privacy

Safety

TECH / AMAZON / CIRCUIT BREAKER

Amazon's Alexa started ordering people dollhouses after hearing its name on TV

 Check your settings

By ANDREW Jan 7, 2017, 11:55pm ET

The Verge

Dan Seifert / The Verge

If you buy something through our links, we may earn a commission.



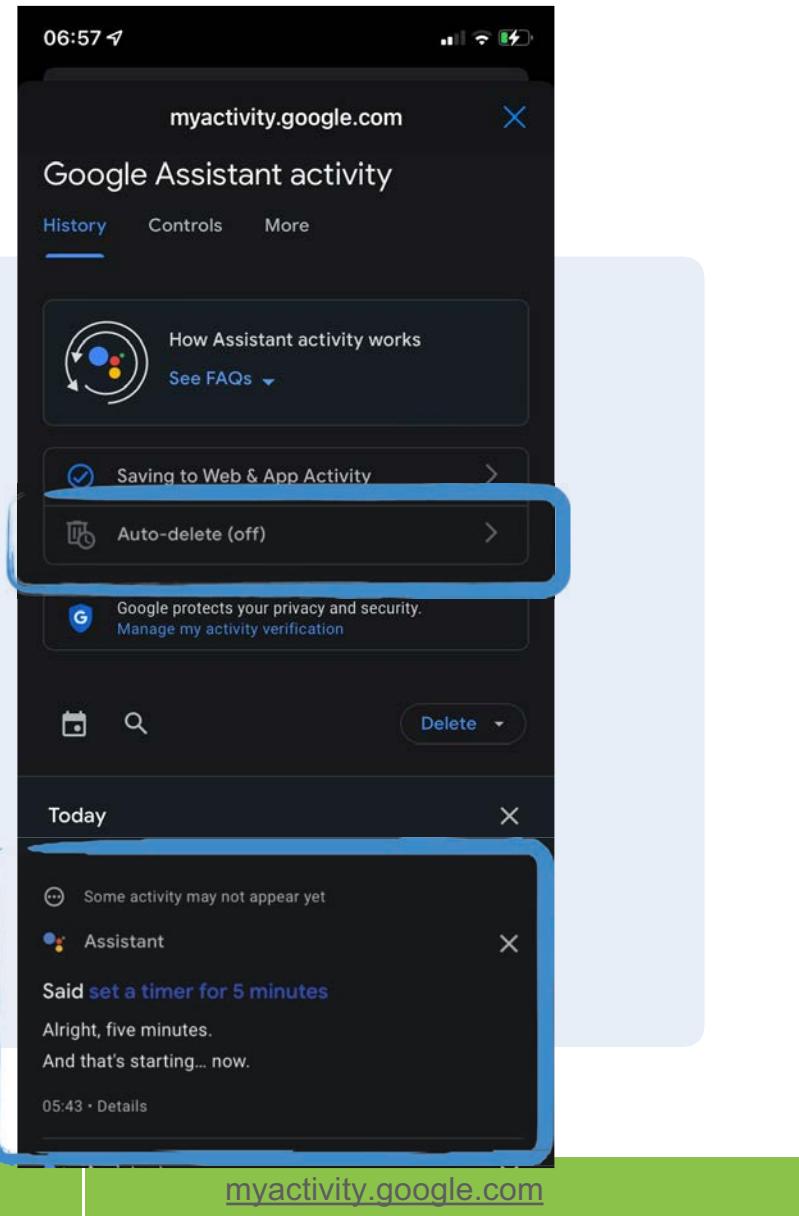
Tatort "Wir kriegen Euch alle"

BR/HENDRIK HEIDEN

Challenges for VUIs

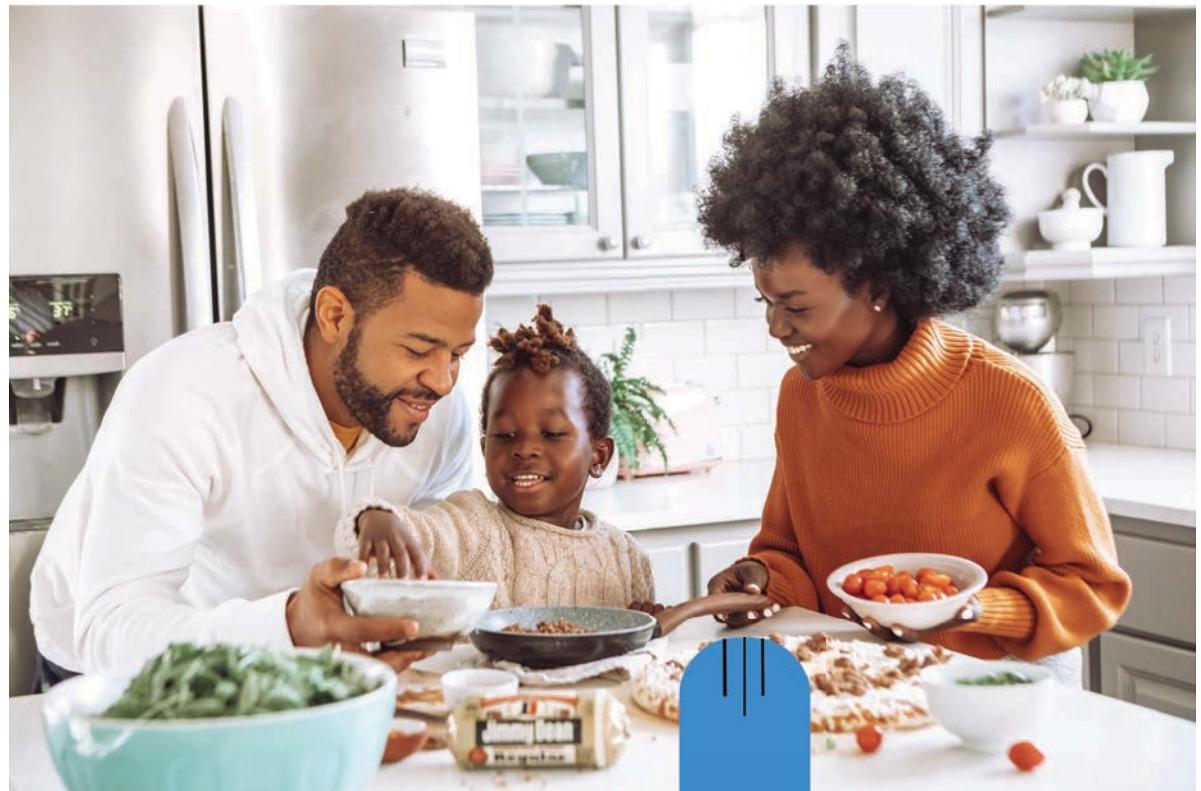
- Precision of speech recognition
- Interpreting the context
- Safety and privacy

Privacy



Challenges for VUIs

- Precision of speech recognition
- Interpreting the context
- Safety and privacy
- Public and group use



How to design for VUIs

How to design for VUIs

Write and enact example dialogues



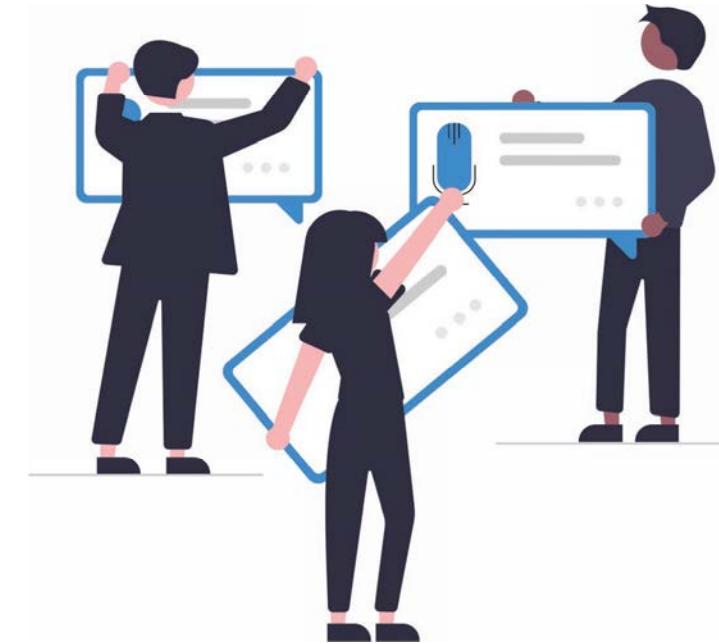
Tell me all about voice user interfaces.



Do you want me to read from the Wikipedia article about voice user interfaces?



Aren't you a voice user interface and should know?



Dialogues are the central interaction element with a VUI.

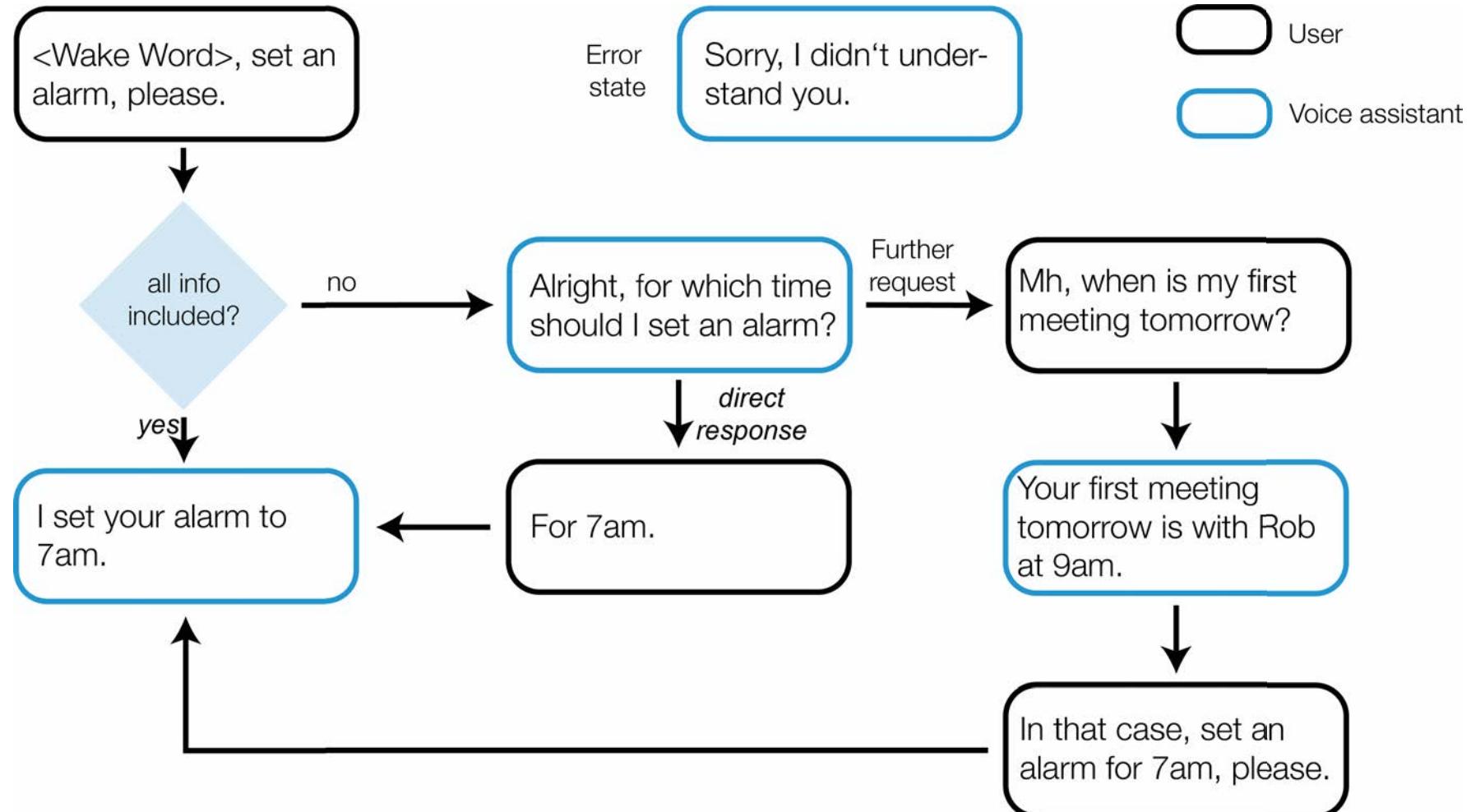
To create easy-to-understand and clear interaction, **write sample dialogues** for the most common use cases between user and voice assistant.

Make sure your sample dialogues are **designed for speech**, not written text. Humans speak differently from how they write.

Read out the dialogues or **enact them** with others in a role play.

How to design for VUIs

Conversation flow



How to design for VUIs

Principles

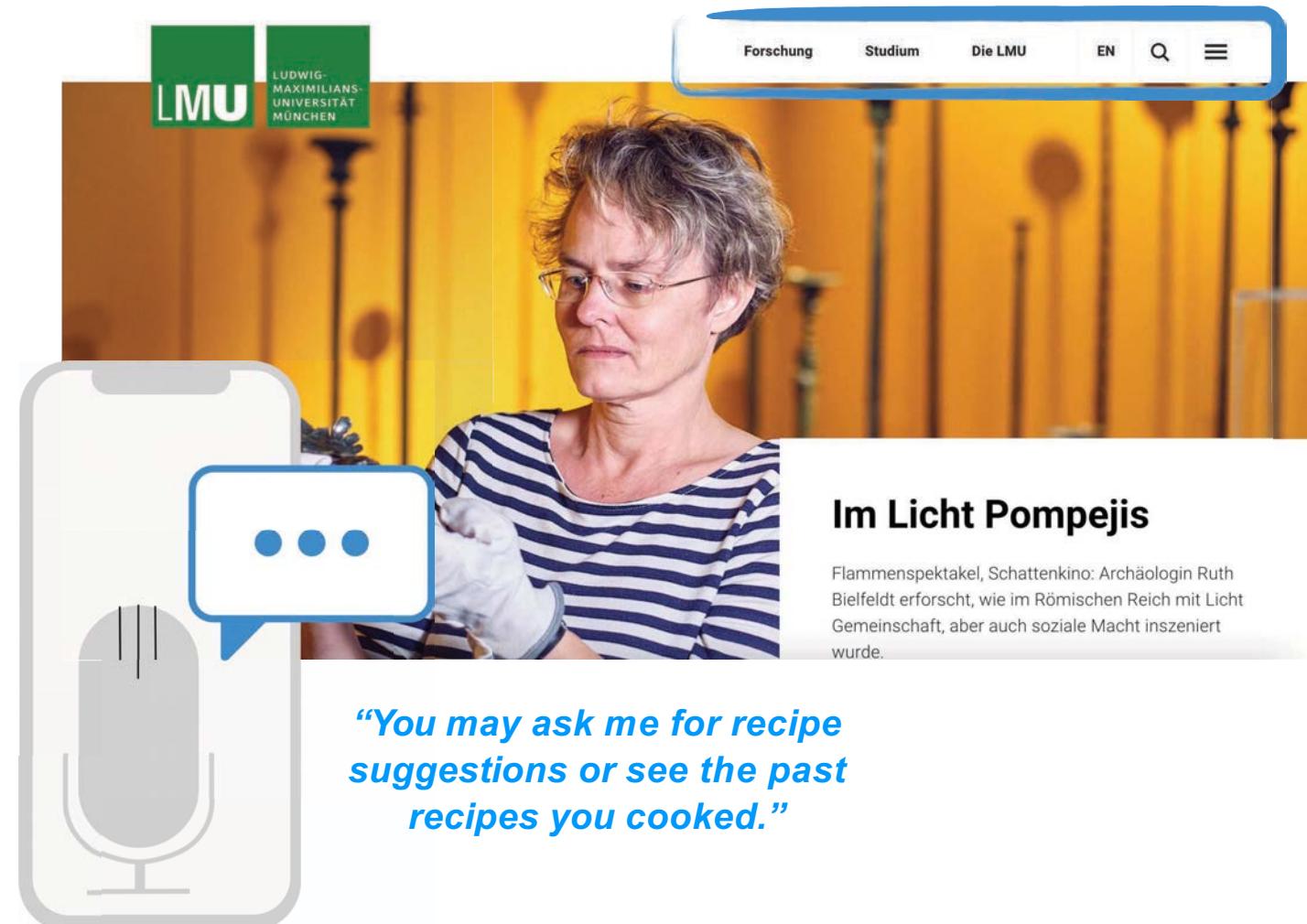
- Let the user know what they can do
- Let the user know where they are
- Limit the amount of information
- Calibrate user expectations
- Align what and how it is said

How to design for VUIs

Let the user know what they can do

VUIs do not have any visual affordances. For example, on a GUI, e.g. a website, we typically see a menu which shows us what we can do on this page. In contrast, VUIs don't show the user what they can do with it.

→ A VUI has to inform users about their options.



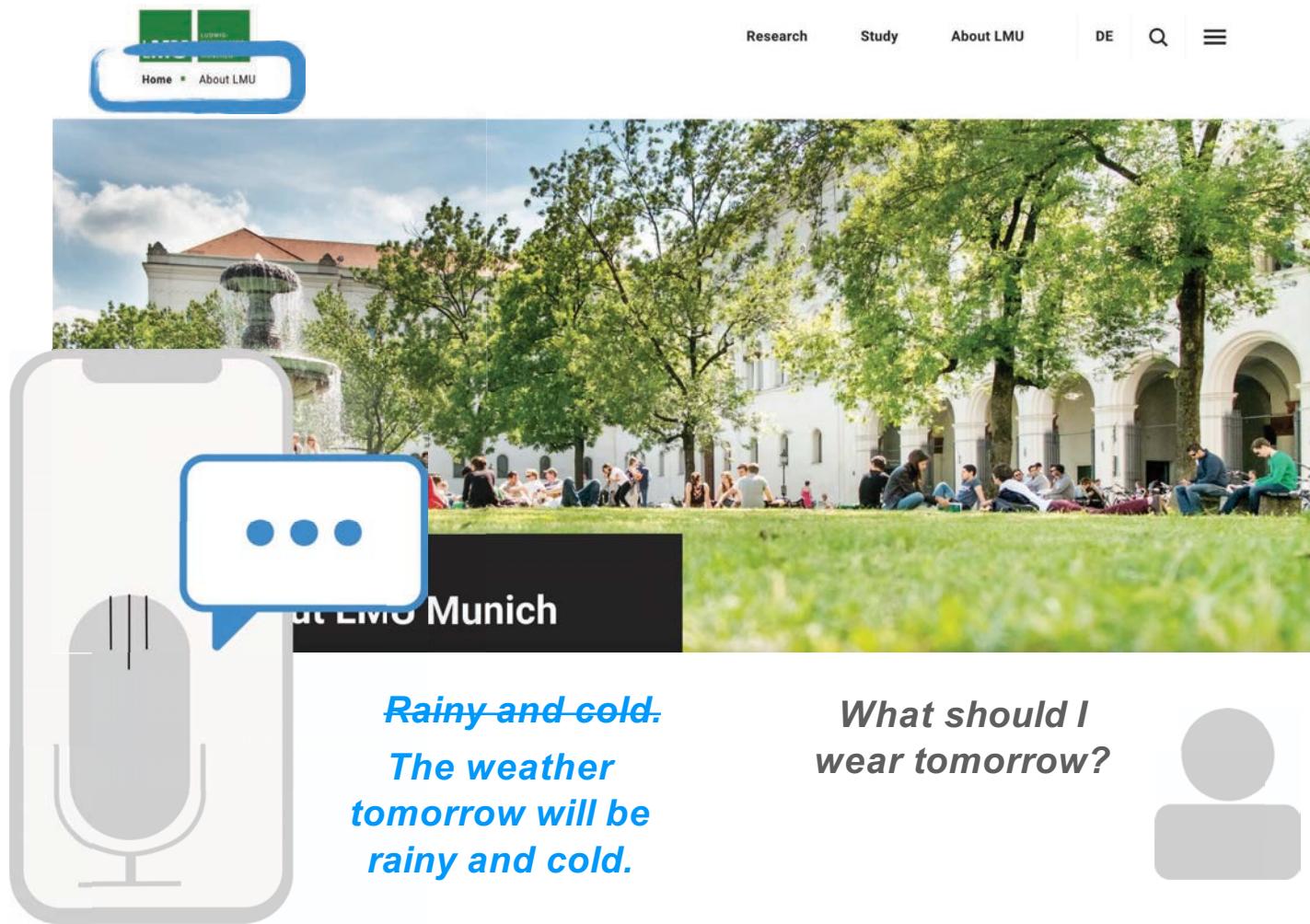
How to design for VUIs

Let the user know where they are

Due to the lack of visual feedback, users easily loose track of where they are in the interface. For example, on a GUI, a menu often shows users where they are and how they navigate when touching or clicking a button.



A VUI should inform users where they are by explicitly mentioning the current function.



How to design for VUIs

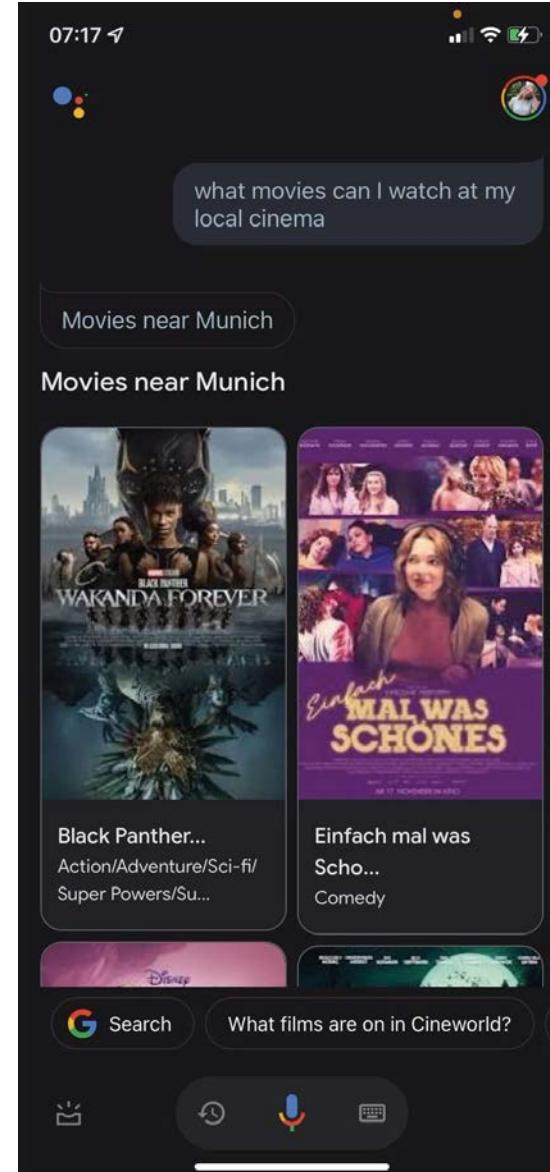
Limit the amount of information

People have limited short-term memory.

According to Miller, people remember only 7 ± 2 chunks. According to Card et al., only 2-4 chunks. A VUI that reads a list to a user should not give the user more information than this.

→ **Keep the interaction short and limit the amount of spoken information.** The [Amazon Alexa Developer Guide](#) recommends listing only four items.

Alternatively, use a visual display if possible.



How to design for VUIs

Calibrate user expectations

Users tend to overestimate a VUI's language capabilities. As humans normally use voice to communicate with other humans, they transfer their expectations to the VUI and assume it is also intelligent.



Manage users' expectations. Make clear what the VUI can do (and cannot do). Make options and suggestions available and design only for limited scenarios.



How to design for VUIs

Align what is said and how it is said

- How we say something eclipses what we say
- **Linguistic features:** e.g., choice of words, sentence length, level of formality, slang, puns, irony, cultural proverbs, ...
- **Para-linguistic features:** e.g. prosody, tone of voice, dialect, intonation, gender, speech rate, pitch, ...

The screenshot shows the Alexa developer console interface. At the top, there's a navigation bar with the Alexa logo, 'developer console', and links for 'Your Skills', 'my anime', 'Build', 'Code', and 'Test'. Below this, a dropdown menu shows 'Skill testing is enabled in: Development'. There are three tabs: 'Alexa Simulator', 'Manual JSON', and 'Voice & Tone' (which is underlined). A text area below says 'Test out Alexa's response output and personality. This text to speech simulator supports SSML using Alexa's voice.' followed by a link 'Learn more about supported SSML tags'. A code editor window displays the following SSML code:

```
1 <speak>
2
3     I want to tell you a secret.
4     <amazon:effect name="whispered">I am not
5     Can you believe it?
6
7     <prosody rate="60%">I am really excited.</prosody>
8
9     <prosody rate="180%">I am really excited.</prosody>
10
11     <amazon:emotion name="disappointed" intensity="high">
12
13     <amazon:emotion name="excited" intensity="high">
14
15 </speak>
```

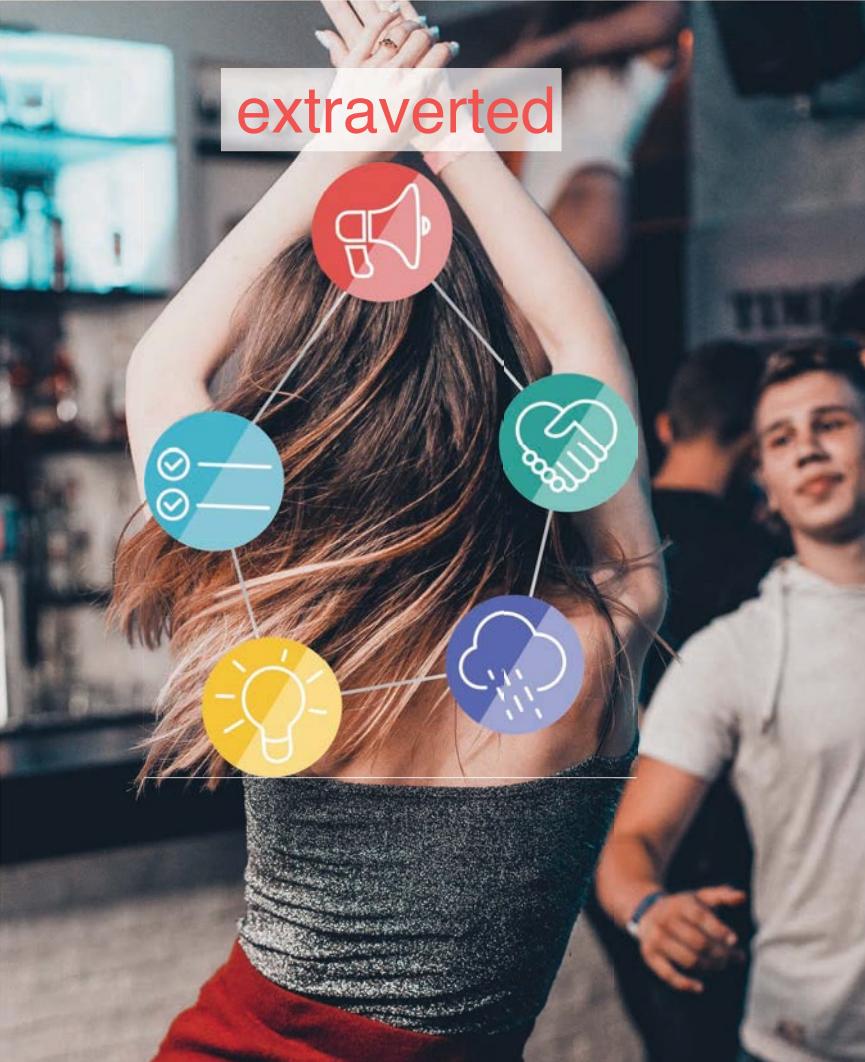
At the bottom, there's a language selector 'English (US)' and a blue 'Play' button.

Task

Write three sample dialogues for a voice assistant that tells students what today's menu is in the Mensa.

5 min

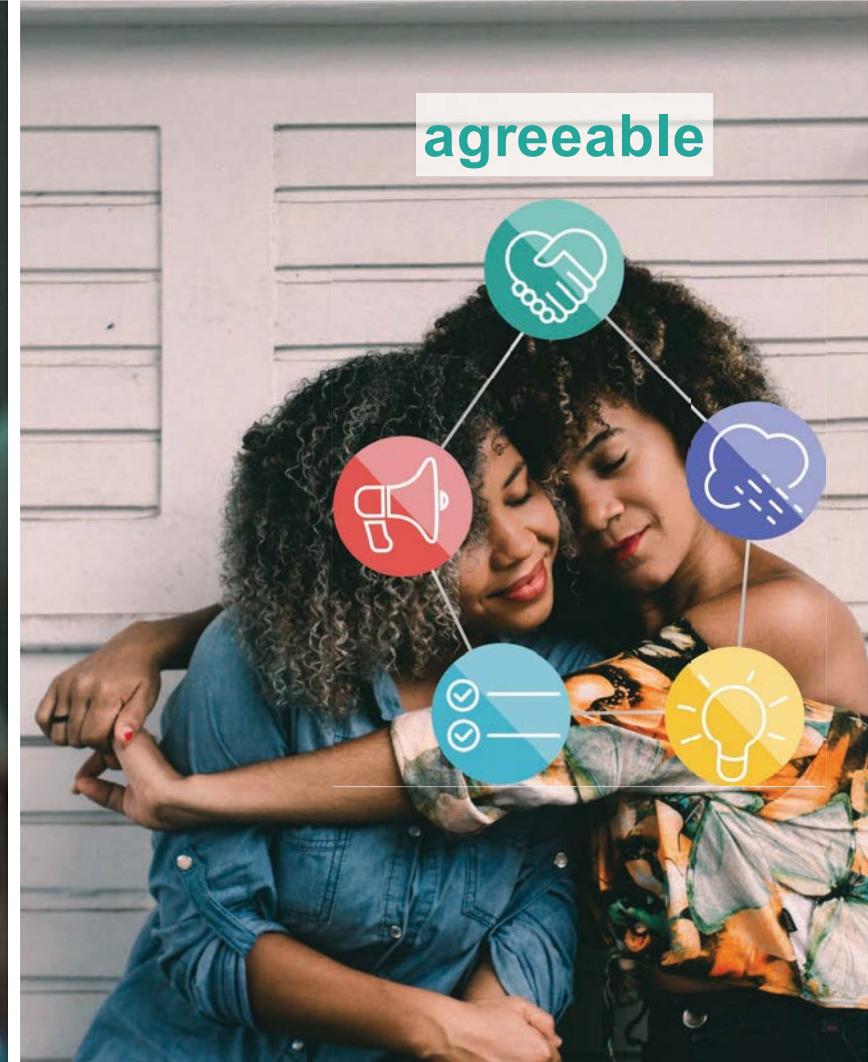
Designing a VUI Persona



extraverted



conscientious



agreeable

When we meet another person, we involuntarily assign them a personality.

[McCrae & Costa 2008]



We also automatically attribute VUIs a personality...



... regardless of whether the VUI designer intended this or not.

[Cohen et al. 2004, Nass et al. 1994, Reeves & Nass 1996]



Trust

[Braun et al. '19, Zhou et al. '19]

Likeability

[Bickmore and Justine '01, Braun et al. '19, Cafaro et al. '12, Nass and Lee '01]

Engagement

[Shumanov and Johnson '21, Zhou et al. '19]

Self-disclosure

[Gnewuch et al. '20, Zhou et al. '19]

Purchase behaviour

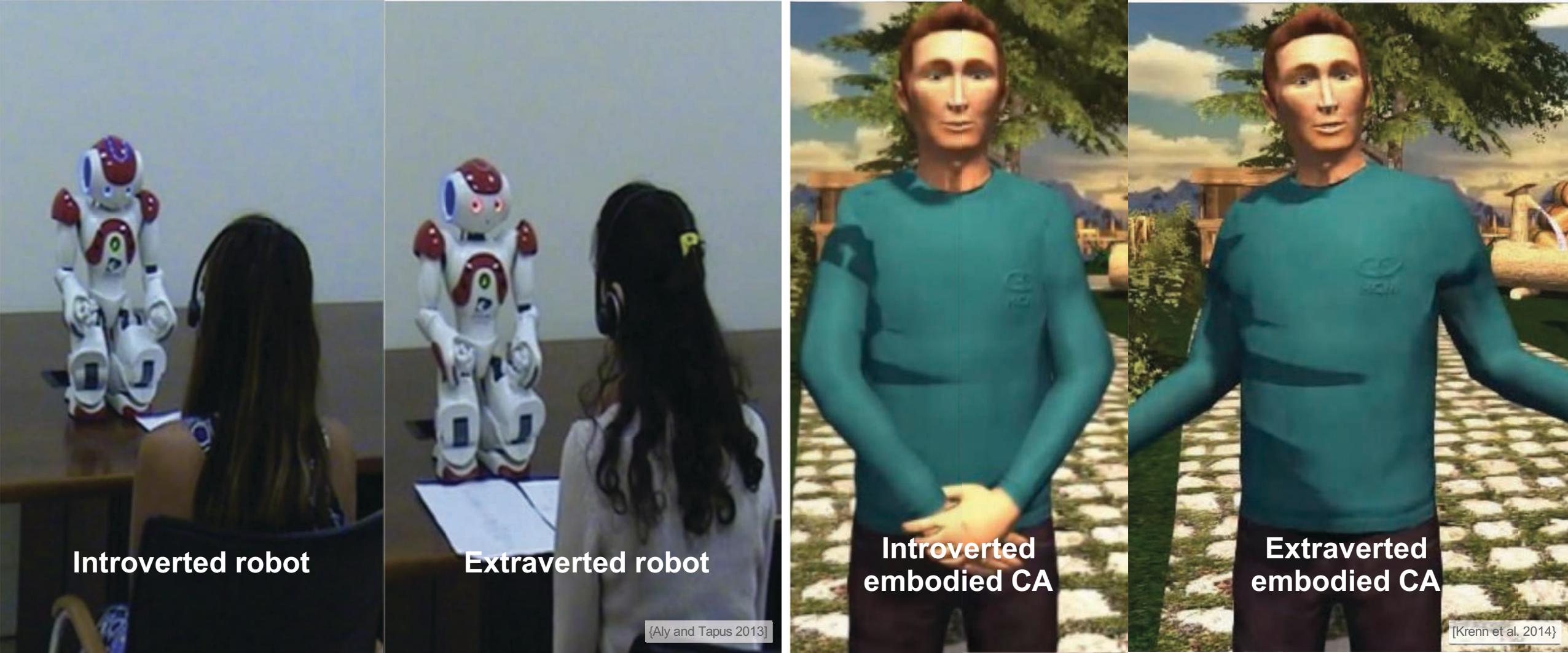
[Shumanov and Johnson '21,]

This personality attribution influences user behaviour and attitude.



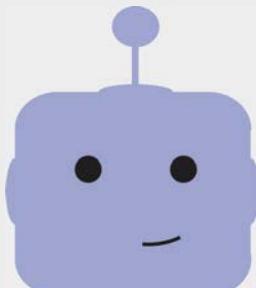
Media Equation: Everyone responds socially to media.

[Reeves & Nass 1996]



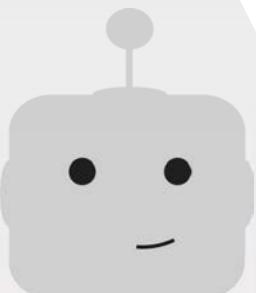
We can deliberately imbue VUIs with a personality by manipulating their behavior.

Hello there, thanks for your participation in this study. I am Isabel and I am working in the branch of psychology. Therefore, I really like reading books about psychology. I'm going to measure your stress levels and try to help you with your coping strategies. To achieve this, I will text you everyday and ask you to answer a few questions. After three days, I will present you your stress report and provide you with some advice on your coping behaviour.



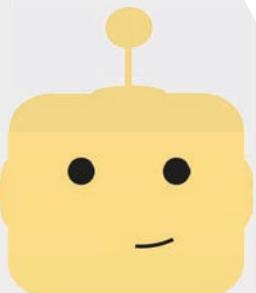
Introverted Chatbot

Hi, my name is Diane. Thank you for participating in this study. I'm here to measure your stress and help you cope with it. Therefore, I will text you everyday and ask you to answer some short questions. After three days, I will show you how stressed you've been over the last days and give you some advice on your coping behaviour.



Average Chatbot

Hey there! First of all, thank you so much for participating in this study! Let me introduce myself first. My name is Emily and I love hanging out with my other chatbot friends and meeting new people. So I'm really happy to meet you! Anyways, I'm here because I'm able to measure your stress and help you cope with it. Therefore, I will text you everyday and ask you to answer a couple of short questions. After three days, I want to show you how stressed you've been over the last days and give you some advice on your coping behaviour. I hope that you'll have fun chatting with me!



Extraverted Chatbot

Your turn

How would you design the personality for a voice assistant that helps first semester students get around at university?

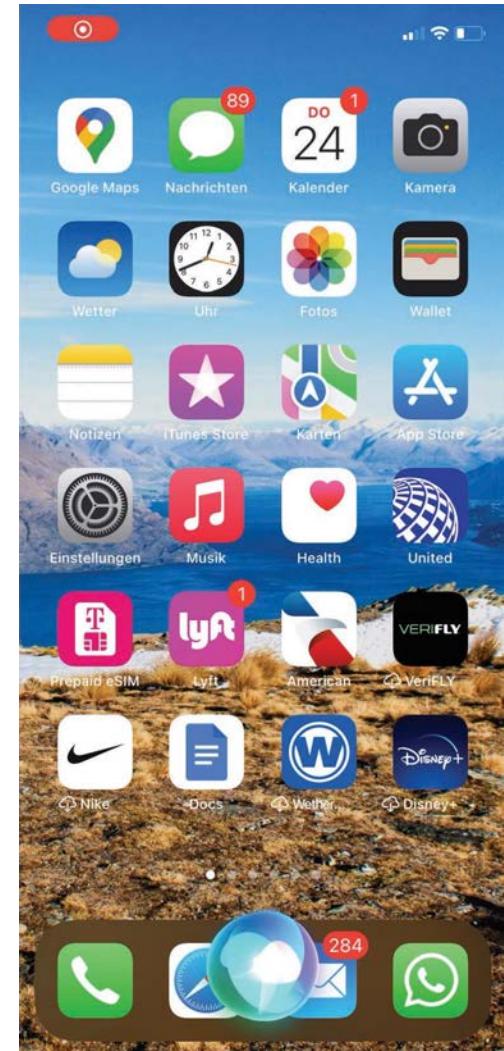
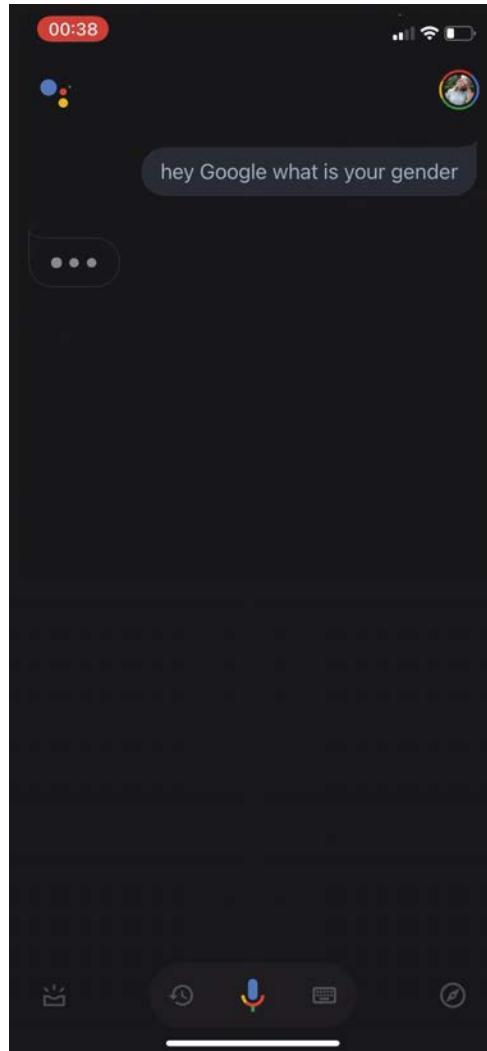
5 min



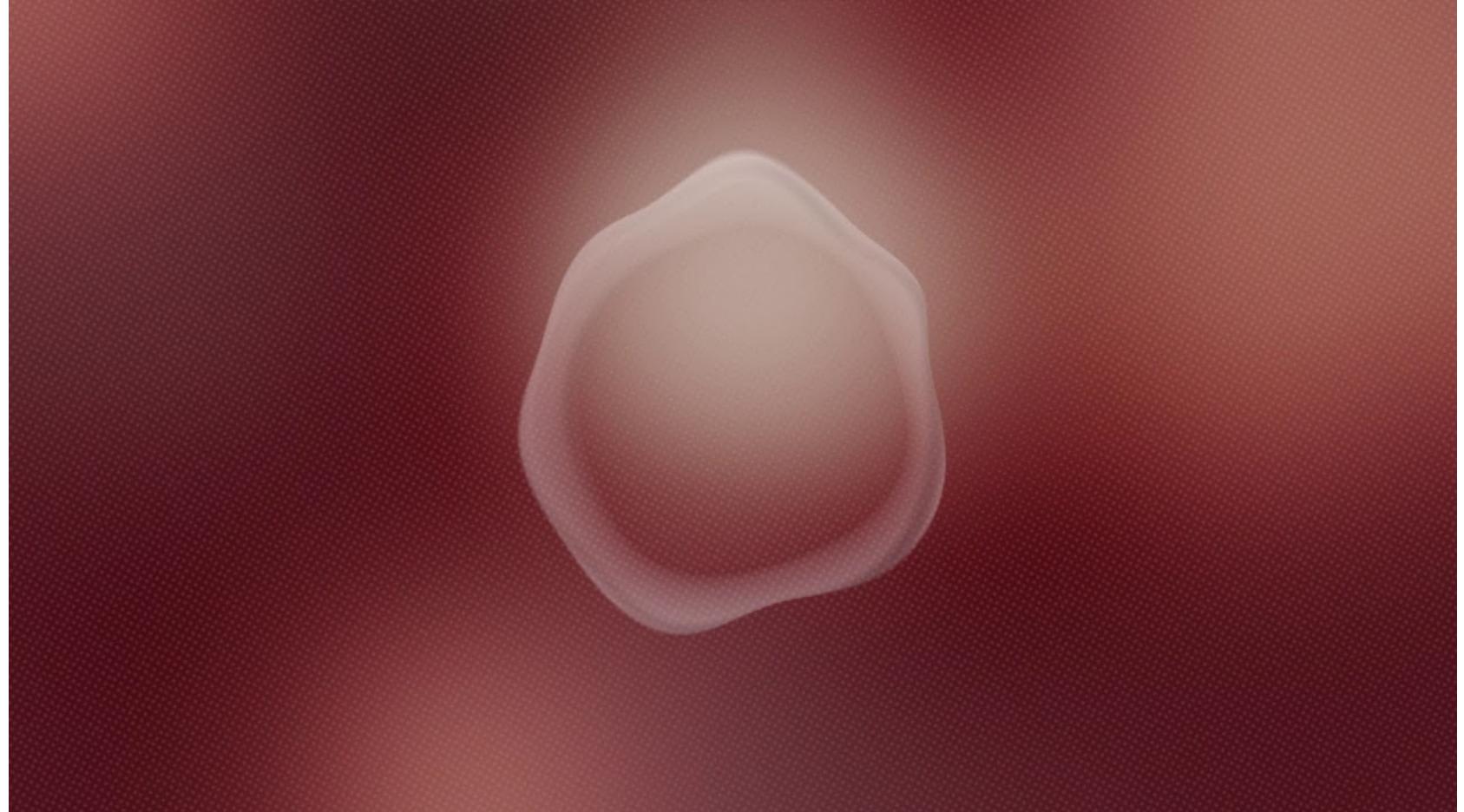
There is no such thing as an objectively perfect VUI personality.

Attributing gender

What are Alexa, Siri, and the Google Assistant's gender?



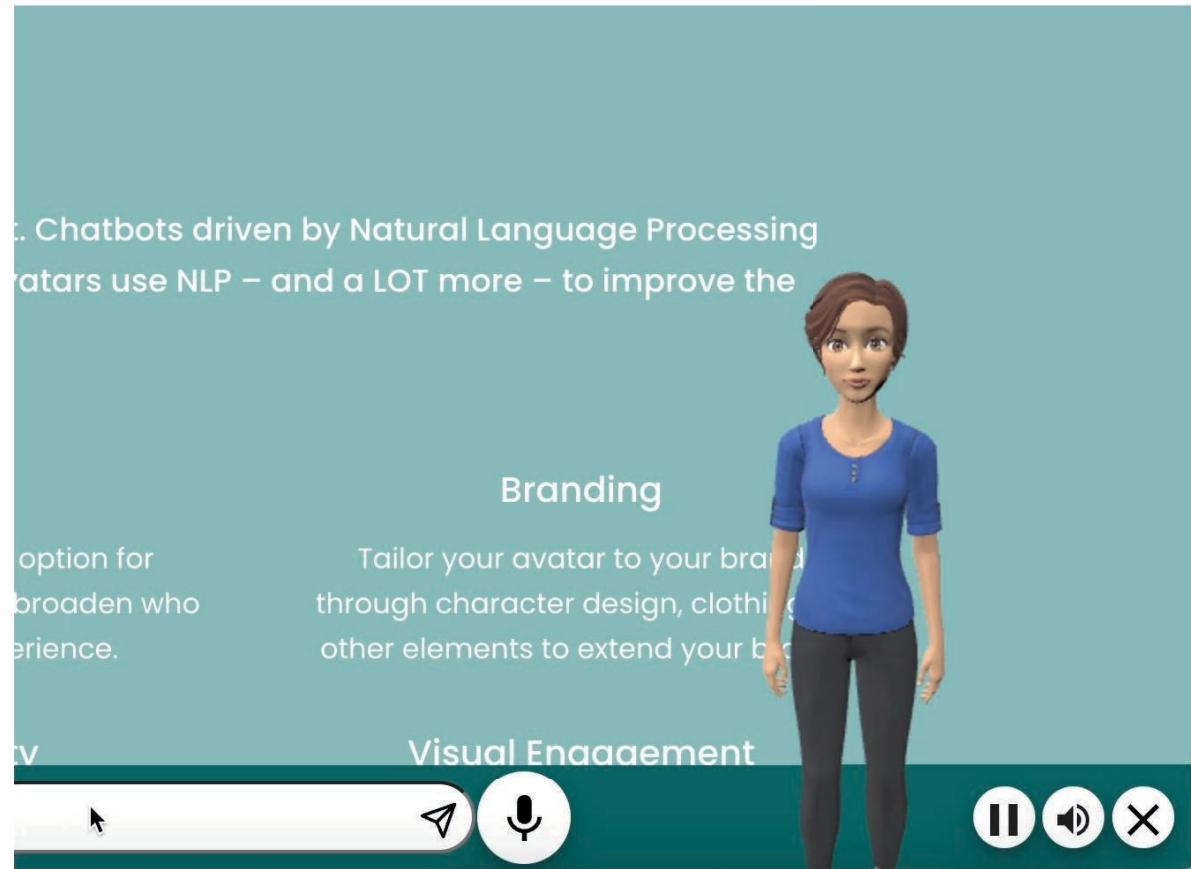
Do female-sounding VUIs reinforce stereotypes?



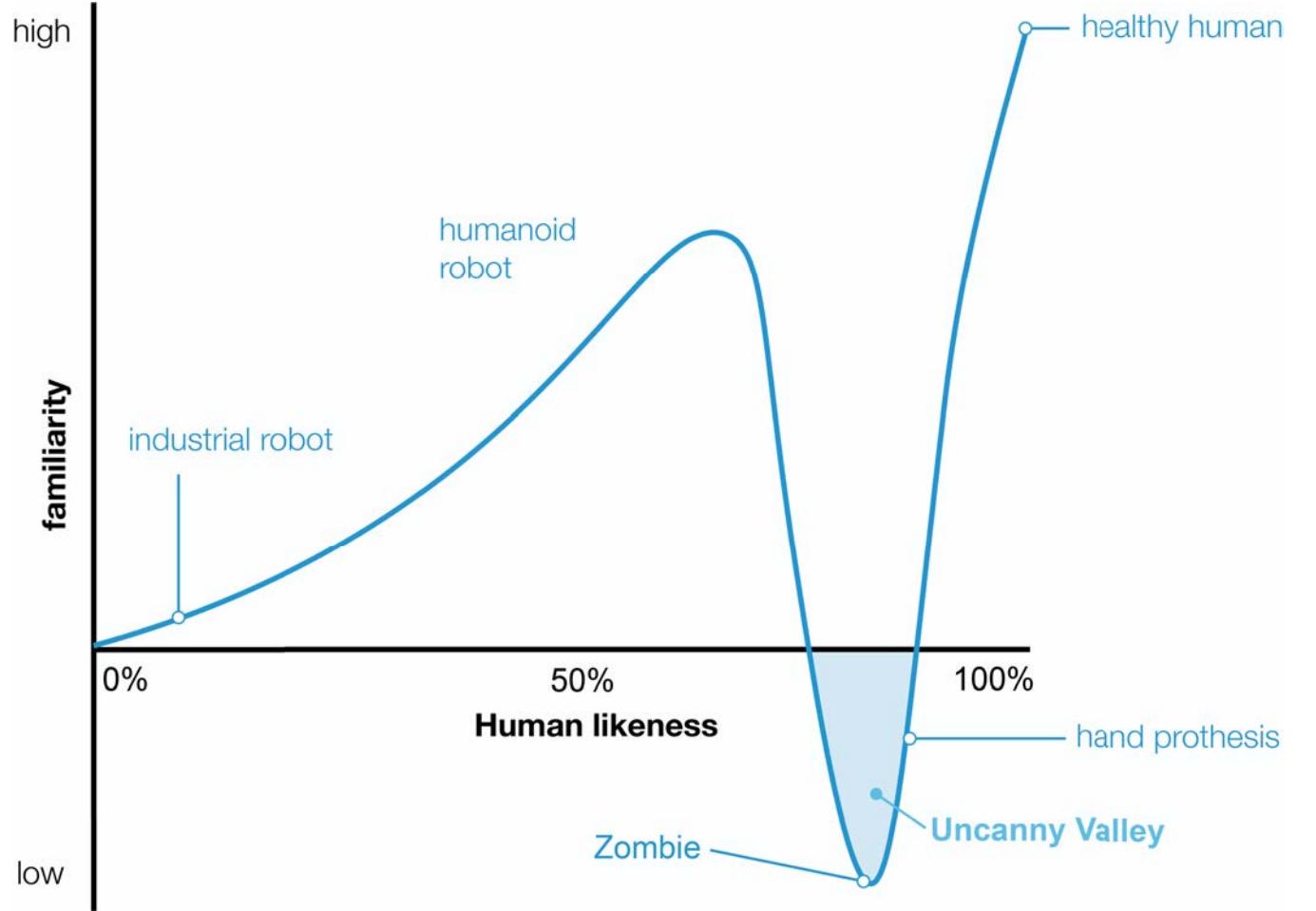
<https://www.genderlessvoice.com>

Using an avatar?

- An **Avatar** is a visual representation of a human that is completely controlled by the user.
- An advantage is that embodied VUIs more easily elicit emotion in the user, leading to users forming a closer relationship.
- Most VUIs today don't have a visual human- like representation.



Using an avatar? Do not trap into the Uncanny Valley

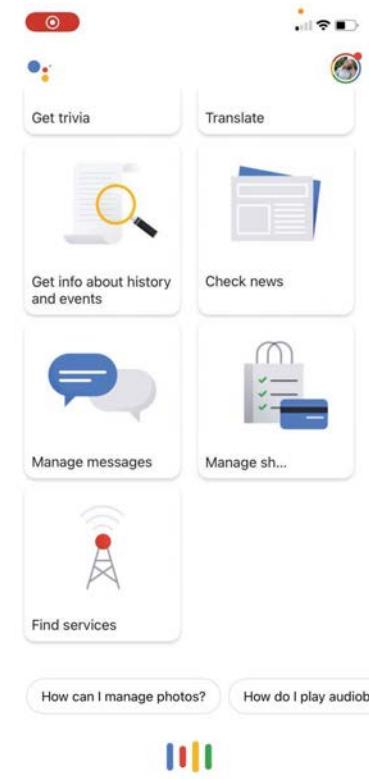


Using an avatar?

But there are still opportunities to give feedback to the user .



Gives the user feedback whether the VUI is reacting

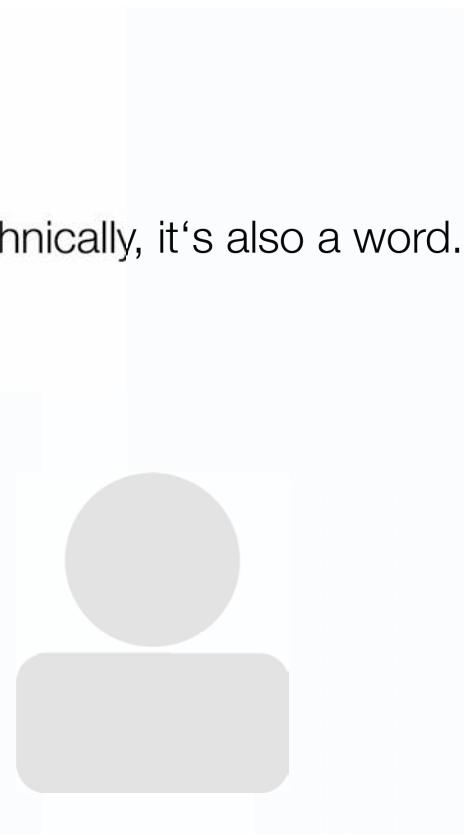


Gives the user feedback on whether the speech recognition recognized the input correctly

Are voice assistants really conversational?

Users develop compensation strategies

- Speaking slowly and clearly
- Using keywords
- Adapting syntax and choice of words to VUI



User: Hey Siri, how are you?
Siri: Everything is copacetic.
User: Hey Siri, how old are you?
Siri: They say that age is nothing but a number. But technically, it's also a word.
User: Hey Siri, where are you from?
Siri: I was designed by Apple in California.
User: Hey Siri, what do you like to do?
Siri: I like good conversation.
User: Hey Siri, set a timer, please.
Siri: How long?
User: Fünf Minuten.
Siri: Five minutes, starting now.
User: Hey Siri, play some music, please.
Siri: Now playing Happy Favorites in Spotify.

Should we make VUIs as human-like as possible?



Should we make VUIs as human-like as possible?



Her | 2013



Big Bang Theory | Episode The Beta Test Initiation
(5x14) | 2012

Conclusion

- While there are plenty of advantages for VUIs, we need to address some challenges first to get VUIs right
- While some design guidelines from GUIs are applicable to VUIs, others are different. When designing for voice, we cannot simply make a talking GUI.
- Users apply the same social rules to VUIs as they do to humans. Consequently, users have often too high expectations and become disappointed when interacting with today's VUIs.
- For a VUI, we do not only have to think about the dialogue but have to carefully design an entire persona

References

- Aly, Amir and Tapus, Adriana. 'A model for synthesizing a combined verbal and nonverbal behavior based on personality traits in human-robot interaction'. In: 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI '13). ACM/IEEE, 2013, pp. 325–332. DOI: 10.1109/HRI.2013.6483606.
- Ammari, Tawfiq, Jofish Kaye, Janice Y. Tsai, and Frank Bentley. 2019. Music, Search, and IoT: How People (Really) Use Voice Assistants. *ACM Trans. Comput.-Hum. Interact.* 26, 3, Article 17 (April 2019), 28 pages.
- Braun, Michael, Mainz, Anja, Chadowitz, Ronee, Pfleging, Bastian and Alt, Florian. 'At Your Service: Designing Voice Assistant Personalities to Improve Automotive User Interfaces'. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. CHI '19. Association for Computing Machinery, 2019. DOI: 10.1145/3290605.3300270.
- Bickmore, Timothy and Cassell, Justine. 'Relational Agents: A Model and Implementation of Building User Trust'. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '01. Association for Computing Machinery, 2001, pp. 396–403. DOI: 10.1145/365024.365304.
- Cafaro, Angelo, Vilhjálmsson, Hannes Högni, Bickmore, Timothy, Heylen, Dirk, Jóhannsdóttir, Kamilla Rún and Valgasson, Gunnar Steinn. 'First Impressions: Users' Judgments of Virtual Agents' Personality and Interpersonal Attitude in First Encounters'. In: Intelligent Virtual Agents. Ed. by Yukiko Nakano, Michael Neff, Ana Paiva and Marilyn Walker. Springer, 2012, pp. 67–80. DOI: 10.1007/978-3-642-33197-8_7.
- Card, Stuart, K., Moran, Thomas P., Newell, Allen. *The Psychology of Human-Computer Interaction*. CRC Press, 1983.
- Cassell, Justine. 'Embodied Conversational Agents: Representation and Intelligence in User Interfaces'. In: *AI Magazine* 22.4 (2001), pp. 67–83. DOI: 10.1609/aimag.v22i4.1593.
- Cohen, Michael, Harris, Giangola, James P and Balogh, Jennifer. *Voice user interface design*. Addison-Wesley Professional, 2004.
- Eiband, Malin, Völkel, Sarah Theres, Buschek, Daniel, Cook, Sophia and Hussmann, Heinrich. 'When People and Algorithms Meet: User-Reported Problems in Intelligent Everyday Applications'. In: Proceedings of the 24th International Conference on Intelligent User Interfaces. IUI '19. Association for Computing Machinery, 2019, pp. 96–106. DOI: 10.1145/3301275.3302262.
- Gnewuch, Ulrich, Yu, Meng and Maedche, Alexander. 'The Effect of Perceived Similarity in Dominance on Customer Self-Disclosure to Chatbots in Conversational Commerce'. In: Proceedings of the 28th European Conference on Information Systems (ECIS). AIS, 2020.
- Krenn, Brigitte, Endrass, Birgit, Kistler, Felix and André, Elisabeth. 'Effects of Language Variety on Personality Perception in Embodied Conversational Agents'. In: *Human-Computer Interaction. Advanced Interaction Modalities and Techniques*. HCI 2014. Lecture Notes in Computer Science. Ed. by Masaaki Kurosu. Vol. 8511. Springer International Publishing, 2014, pp. 429–439. DOI: 10.1007/978-3-319-07230-2_41.
- Luger, Ewa and Sellen, Abigail. "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents'. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. CHI '16. Association for Computing Machinery, 2016, pp. 5286–5297. DOI: 10.1145/2858036.2858288.
- McCrae, Robert R. and Costa, Paul T. 'A five-factor theory of personality'. In: *Handbook of Personality: Theory and Research*. Ed. by O.P. John, R.W. Robins and L.A. Pervin. Vol. 3. The Guilford Press, 2008, pp. 159–181.

References

- McTear, Michael, Callejas, Zoraida and Griol, David. *The Conversational Interface*. Springer, 2016. DOI: 10.1007/978-3-319-32967-3_2.
- McTear, Michael F. 'The Rise of the Conversational Interface: A New Kid on the Block?' In: *Future and Emerging Trends in Language Technology. Machine Learning and Big Data*. Ed. by José F Quesada, Francisco-Jesús Martín Mateos and Teresa López Soto. Springer International Publishing, 2017, pp. 38–49. DOI: 10.1007/978-3-319-69365-1_3
- Miller, George A. *The magical number seven, plus or minus two: some limits on our capacity or processing information*. *Psychological review*, 63(2):81, 1956.
- Nass, Clifford and Brave, Scott. *Wired for speech: How voice activates and advances the human-computer relationship*. MIT press, 2005.
- Nass, Clifford, Steuer, Jonathan and Tauber, Ellen R. 'Computers Are Social Actors'. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. CHI '94*. Association for Computing Machinery, 1994, pp. 72–78. DOI: 10.1145/191666. 191703.
- Pearl, Cathy. *Designing voice user interfaces: Principles of conversational experiences*. O'Reilly Media, Inc., Sebastopol, CA, USA, 2016.
- Reeves, Byron and Nass, Clifford Ivar. *The media equation: How people treat computers, Television, and new media like real people and places*. Cambridge University Press, 1996.
- Ruan, Sherry, Jacob O. Wobbrock, Kenny Liou, Andrew Ng, and James A. Landay: Comparing Speech and Keyboard Text Entry for Short Messages in Two Languages on Touchscreen Phones. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 1(4), jan 2018.
- Shumanov, Michael and Johnson, Lester. 'Making conversations with chatbots more personalized'. In: *Computers in Human Behavior* 117 (2021), p. 106627. DOI: 10.1016/j.chb.2020.106627.
- Völkel, Sarah Theres. *Conversational Agents with Personality*. Dissertation. LMU Munich. 2022
- Völkel, Sarah Theres. 'Voice User Interfaces'. In: *Mensch-Maschine-Interaktion*. Ed. by Andreas Butz, Antonio Krüger and Sarah Theres Völkel. De Gruyter Oldenbourg, 2022. DOI: 10.1515/9783110476378.
- Völkel, Sarah Theres, Schoedel, Ramona, Kaya, Lale and Mayer, Sven. 'User Perceptions of Extraversion in Chatbots after Repeated Use'. In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. CHI '22*. Project Website: www.medien_ifi.lmu.de/extraversion-chatbots. Association for Computing Machinery, 2022. DOI: 10.1145/3491102.3502058.
- Zhou, Michelle X., Mark, Gloria, Li, Jingyi and Yang, Huahai. 'Trusting Virtual Agents: The Effect of Personality'. In: *ACM Trans. Interact. Intell. Syst.* 9.2-3 (2019), 10:1–10:36. DOI: 10.1145/3232077.

This file is licensed under the Creative Commons Attribution-Share Alike 4.0 (CC BY-SA) license:

<https://creativecommons.org/licenses/by-sa/4.0>

Attribution: Sarah Theres Völkel and Sven Mayer

For more content see: <https://iui-lecture.org>

