

Designing Voice User Interfaces

Embedded Interface Design

with **Bruce Montgomery**



Learning Objectives

Students will be able to...

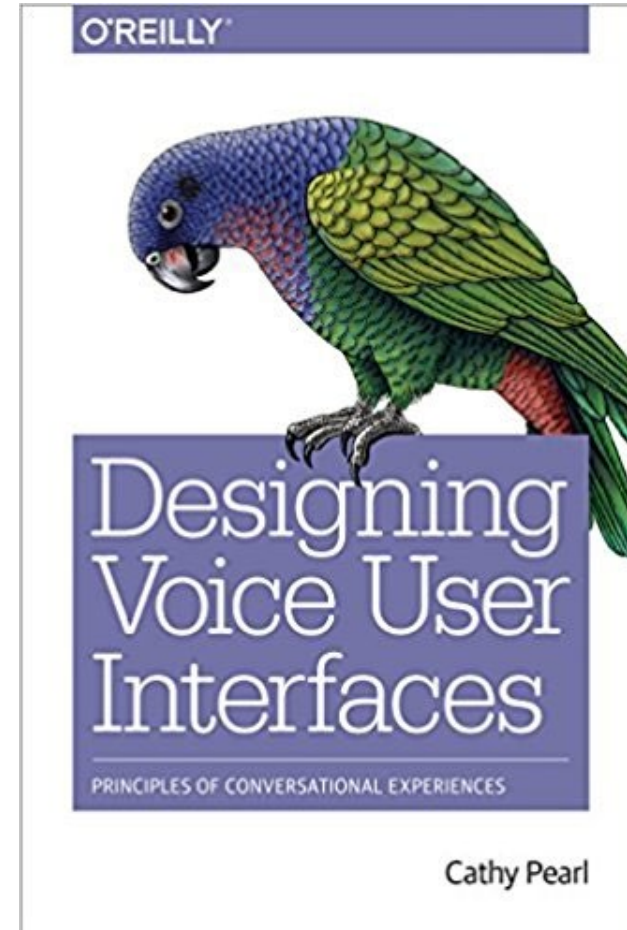
- Understand the role of VUIs in embedded product interfaces
- Recognize common practices and tools for design and development
- Consider issues with using avatars with voice interfaces



Designing Voice User Interfaces - Pearl

- Pearl, 2017, O'Reilly
- History and applicability of VUIs
- Basic and advanced design principles
- Personas, avatars, actors, video games
- Speech recognition technology
- User testing; analyzing and improving performance
- Voice enabled devices and cars

Reference [1]



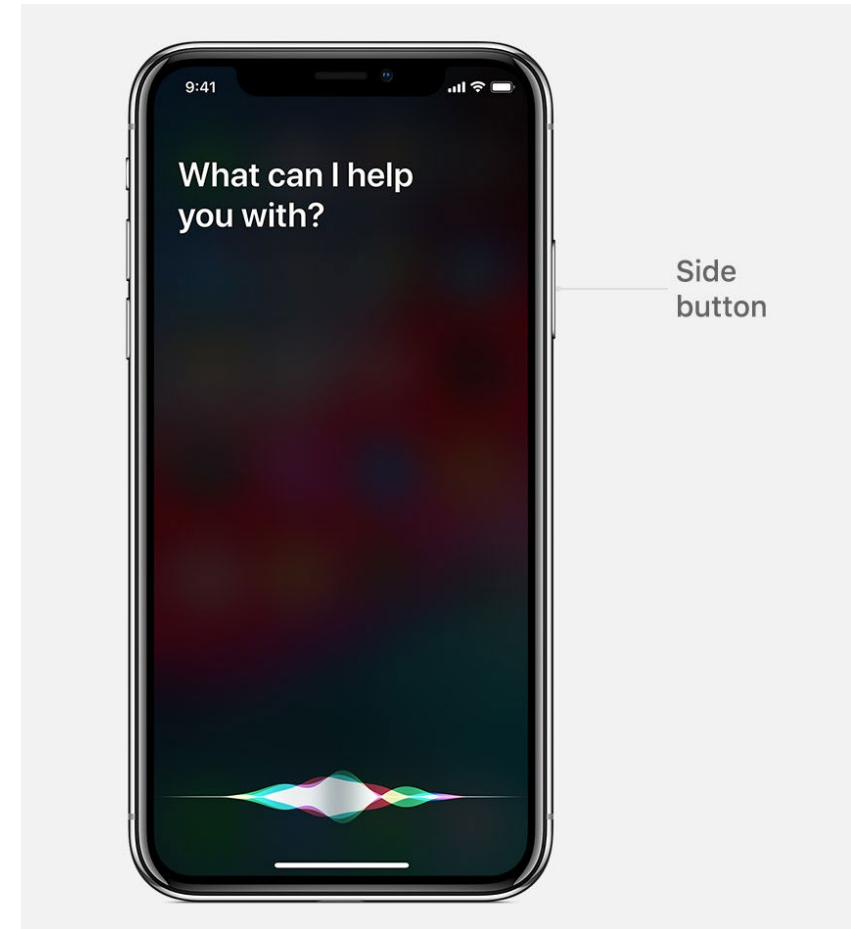
VUI History

- 1950s – first Bell Labs work on single-speaker digit recognition
- 1990s – first viable speaker-independent systems
- 2000s – interactive voice response (IVR) systems
- Now – “Second era” of VUIs – Siri (2011), Google Now/Home (2012), Cortana (2014), Amazon Echo (2014)
- Reference [1], [2]



VUI Devices Today

- Typical Devices
 - Mobile phone assistants
 - Home assistants
 - Watches/bands/earbuds
 - TV voice remotes
 - Toys and Games
 - Cars/Autonomous vehicles
 - Vehicle VUI Issues
 - Noise
 - Upgrading devices
 - Phone interfaces
 - Distracted driving
 - Shifting between functions, devices
- Reference [1], [3]



VUI Advantages/Disadvantages

- Advantages
 - Speed
 - Hands-free
 - Eliminates Physical Controls
 - Intuitiveness
 - Empathy
 - Fit for small device sizes
- Disadvantages
 - Public spaces
 - Discomfort speaking to a computer
 - Preference to type/text
 - Privacy
 - Lack of true conversational interaction
- Other?

Reference [1], [4]



Basic VUI design principles

- Not looking at IVR (Interactive Voice Response) systems design, like a verbal phone menu/task system
- Focus is on VUIs for mobile or embedded devices
 - Is a visual component available/desired?
 - VUIs often handle one-turn tasks (set an alarm, play a song)
 - Conversational design is multiple turns; requires remembering past questions and responses
 - Consider methods of acknowledging, cancelling; opposite tasks
 - When asking for information, better to provide examples:
 - Say “Tell me your date of birth, such as July 22, 1976” not “Tell me your date of birth, with the month, day, and year.”

Reference [1]



Basic VUI design principles continued

- Confirmations – what’s appropriate for which transaction
 - Multi-tiered, Explicit, Implicit, Non-speech, Generic (“Thank you”), Visual
- Conversation markers (“Got it.”, “Great.”) – humanizes transaction
- Error conditions
 - Speech detected but not recognized
 - Can be explicit (“Sorry, what was that?”) or implicit – do nothing
 - Recognized but not handled
 - Input is recognized but response is absent or incorrect
 - Recognized but incorrectly
 - Could do nothing – unexpected input – or do the wrong thing
 - Escalating error
 - “What’s the city and state” “Ocala” “Sorry, please say the city and the state” “Oh, it’s Ocala, Florida”
- Don’t blame the user
- Recognize novice and experienced users

Reference [1]



Basic VUI design principles continued

- Keep track of context (ex. pronouns)
 - “Who was the 16th president?” “Abraham Lincoln” “Where was he born” “Hodgenville, Kentucky”
- Universal issues
 - For IVRs: Repeat, Main Menu, Help, Operator, Goodbye; not so clear for VUIs
- Latency
- Disambiguation
 - “Call Bruce” “Cell phone or home phone” “Cell” “Calling Bruce Montgomery...”
- Accessibility
 - Time efficient interactions – be brief, allow interruptions
 - Context – when prompting, help tell the user how they should respond
 - Personalization over Personality

Reference [1]



Disabilities Impacting VUI Design

- Obvious: deafness, motor impairment
- Less obvious: ADHD, autism spectrum
- 10-15% of world's population has an impactful disability
- Difficulties understanding “deaf speech”
- Augmenting with visual cues for the deaf, but ensuring pure voice interaction works for the sight impaired

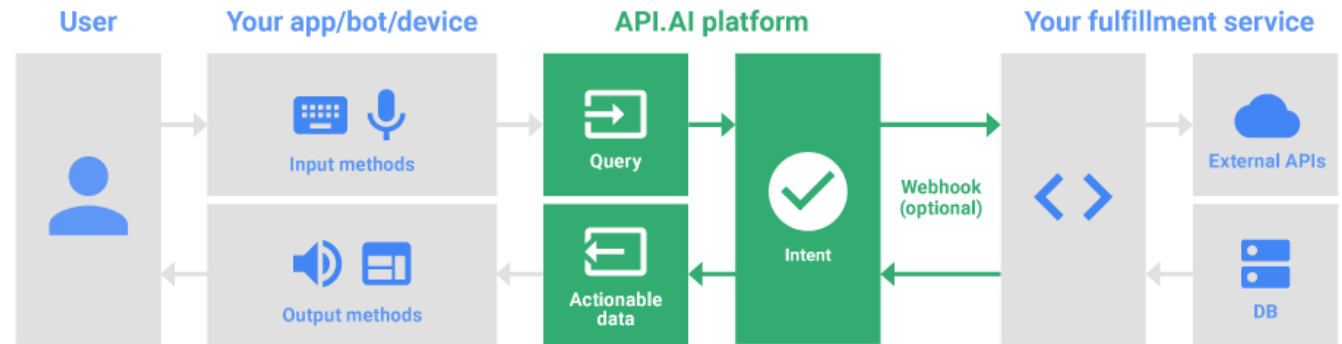
Reference [1]



Typical VUI Design Tools

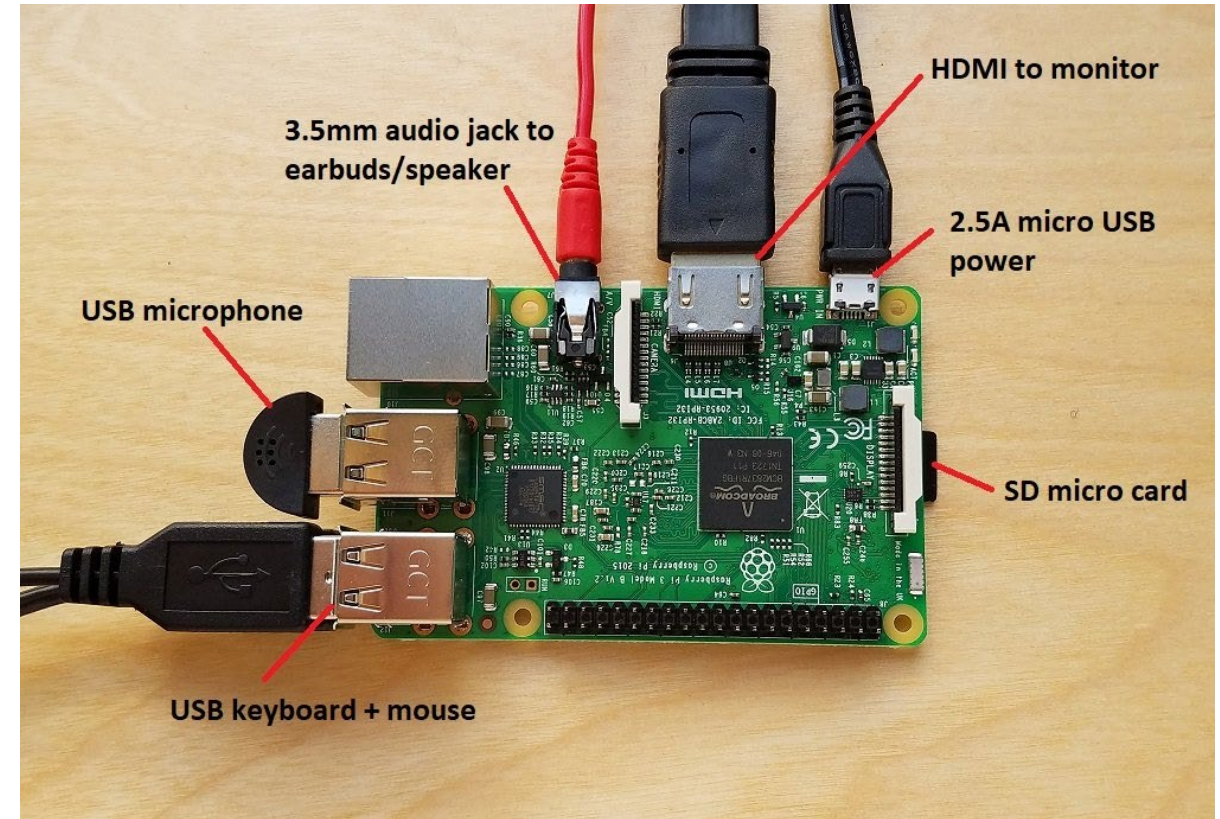
- Sample Dialogs – based on common use cases
- Visual Mock-ups – Sketches and Wireframes
- Flows and Dialog Maps – obvious for an IVR, aka callflows, not as much for a VUI
- Dialog map tools:
- Whiteboard, Tortu, Sayspring, Botsociety, etc.
- Prototyping tools:
- Google Assistant, Tincan.AI, Wit.ai, Api.ai, Nuance Mix

Reference [1], [5], [6]



Example: Alexa on a RPi

- AVS Device SDK
- RPi with speaker & microphone
- Setup for Alexa AVS [6]
- Controlling GPIO using Alexa Skills



Using avatars with VUI

- Avatars are often not appropriate based on the audience and the interactions (ex. Clippy from Word)
- Often appropriate for storytelling, video games
- Avatar types: Static photo, animated, static animation, text only, animated illustrations

Reference [1]



Figure 3-4. Photo, static avatar, illustration

It looks like you're writing a letter.

Would you like help?

☒ Get help with writing the letter

☒ Just type the letter without help

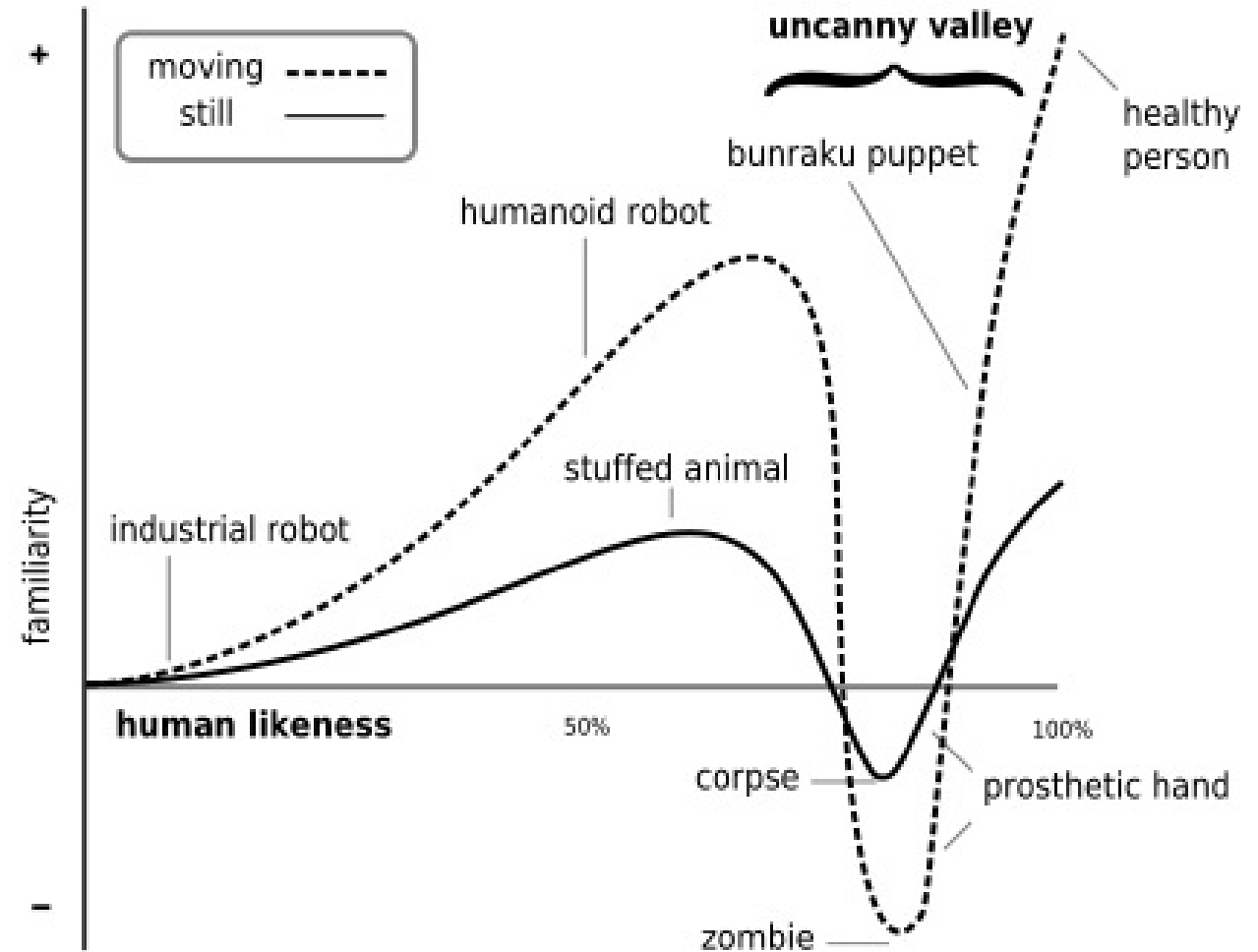
☐ Don't show me this tip again



Why avatars are hard: The uncanny valley

- Whether the image is still or moving, there is a range of visual feedback on near-human interaction that makes people uncomfortable: The uncanny valley
- One of the easiest ways to avoid this is not use human or photo-realistic avatars

Reference [1]



Speech Recognition Technology

- ASR (automated speech recognition) tools
 - Fee-based:
 - Google, Nuance, Bing, iSpeech
 - Free
 - Web Speech API, Wit.ai, Sphinx, Kaldi
 - Amazon (but only if creating skills)
- Key assessment of tools
 - Robustness of dataset/accuracy
 - Endpoint detection performance
 - Interruptions/Barge-in
 - Handling timeouts and pauses
 - Too-much-speech timeout
 - N-best matches for recognition
- Challenges
 - Noise
 - Multiple speakers
 - Children
 - Accents
 - Names, spelling, alphanumerics
 - Data privacy

Reference [1]



VUI Testing and Usability Assessment

- Testing (and Usability Assessment)
 - User testing with defined tasks
 - Likert-scale surveys for satisfaction
 - Remote, lab, and guerilla testing
 - Recording testing
 - Measuring recognition accuracy
- Measures
 - Task Completion Rates
 - Dropout Rates
 - Amount of time in the VUI
 - Barge-ins
 - High no-speech timeouts or no matches
 - Navigation or latency
- Requires good logging methods
- Tools are still catching up to technology...

Reference [1]



Figure 7-3. Google Speech Dashboard

References

- [1] Designing Voice User Interfaces, Pearl, 2017, O'Reilly
- [2] <https://www.amazon.com/Amazon-Echo-And-Alexa-Devices/b?ie=UTF8&node=9818047011>
- [3] <https://support.apple.com/en-us/HT204389>
- [4] <https://www.computerworld.com/article/3252218/collaboration/cortana-explained-why-microsofts-virtual-assistant-is-wired-for-business.html>
- [5] <http://site.clairvoyantsoft.com/google-api-artificial-intelligent-assistant/>
- [6] <https://developer.amazon.com/docs/alexa-voice-service/set-up-raspberry-pi.html>
- [7] <https://www.instructables.com/id/Control-Raspberry-Pi-GPIO-With-Amazon-Echo-and-Pyt/>

