Voice User Interface (VUI)

Human Computer Interaction

Fulvio Corno, Luigi De Russis

Academic Year 2019/2020

Amazon Echo

Hands-On

Voice User Interfaces

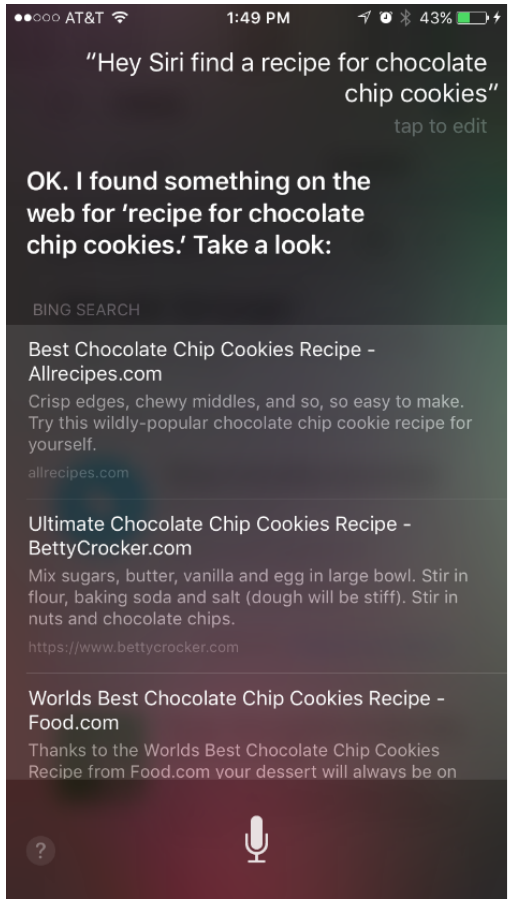
- Voice User Interfaces (VUIs) allow the user to interact with a system through voice or speech commands
  - primary advantage: hands-free, possibly eyes-free interaction
- Voice User Interfaces or Conversational User Interfaces?
  - "*which mimics a conversation with humans*"
  - "conversational" applies to both text-based chatbots and VUIs
- Contemporary VUIs can be divided in:
  - screen-first systems
  - voice-only systems
  - voice-first systems

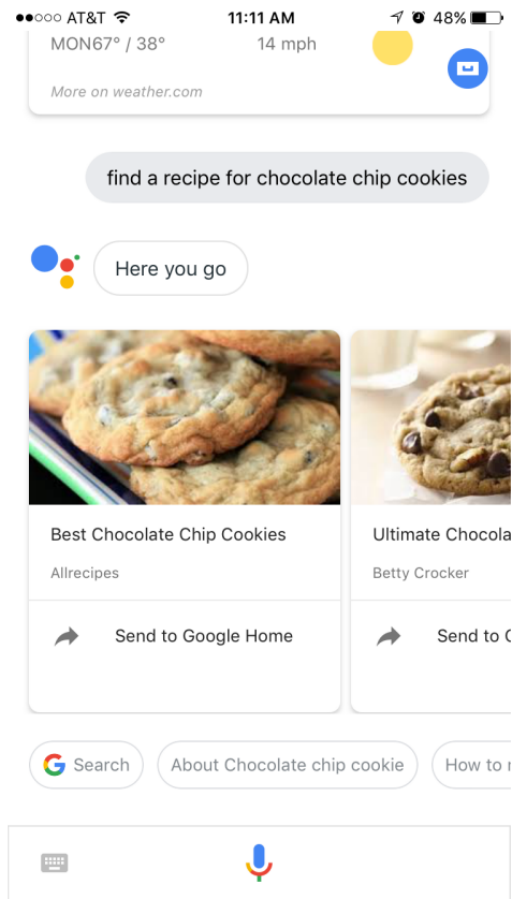
Screen-First Devices

- Most of contemporary voice interaction happens on screen-first devices
  - smartphones, mainly
- Impressive speech recognition and language processing features
  - but overall experience is fragmented
- Main limitations
  - missing functionality
  - poor use of screen space while speaking
  - missing affordances

Missing Functionality and Affordances

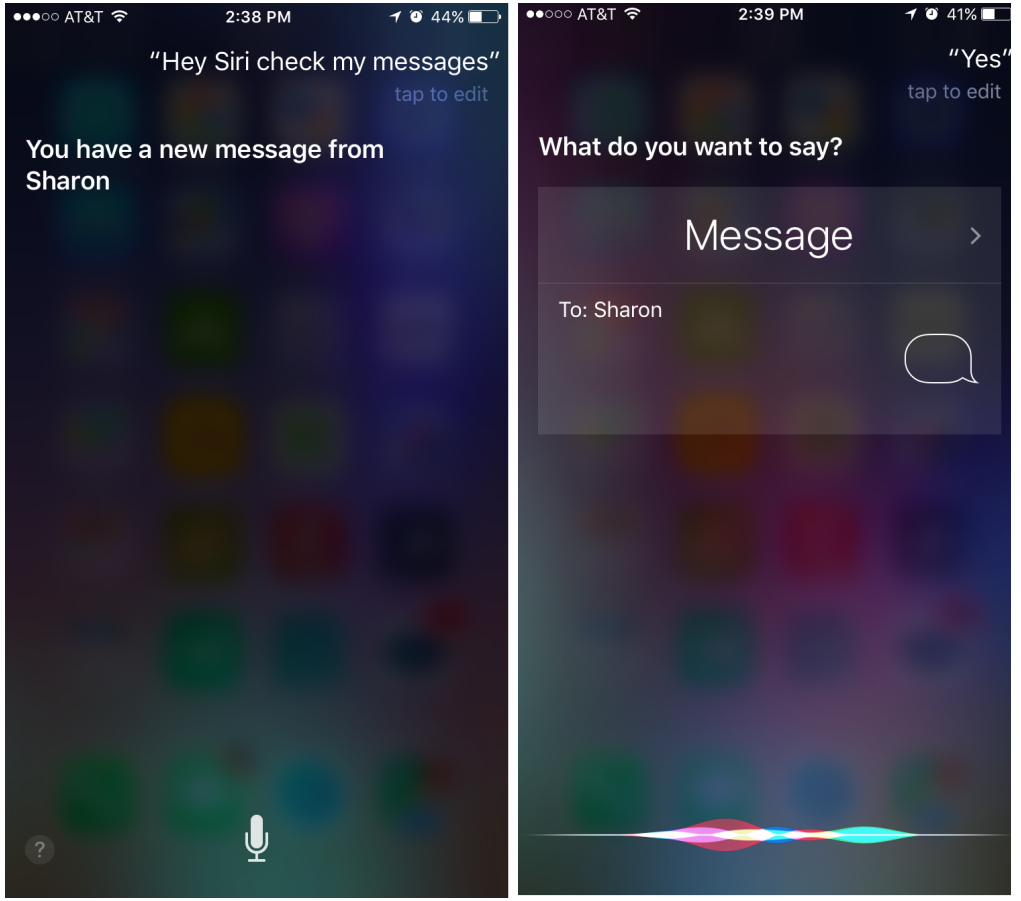
- Users can start a task via voice, but subsequent steps require them to use the touchscreen
- Visual affordances are missing (or poor)
  - Siri omits several visual affordances (e.g., it does not show that people can edit a text message before sending it)
  - Google Assistant is better in this





Poor Screen Space Use

- Tasks with some support for multi-step voice input exhibit a screen design:
  - totally different from the "normal" GUI version
  - which limits the information available to the user



Voice-Only Devices

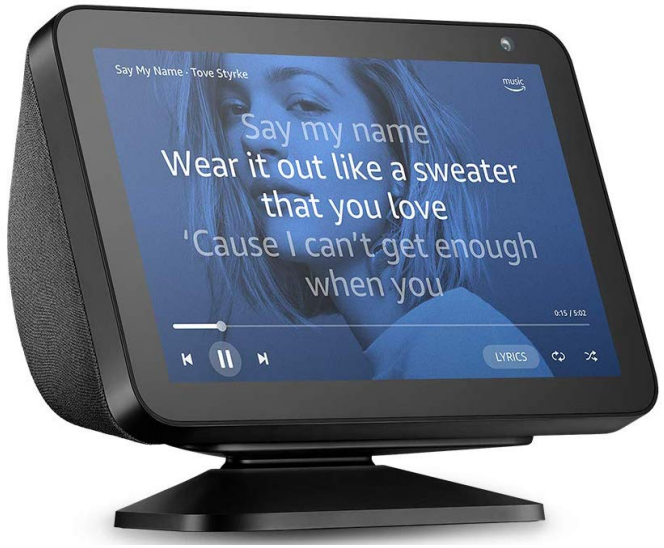
- No visual display at all
  - like the Amazon Echo we "tested" before
  - audio is for input **and** output (plus some "feedback lights")
  - hands-free operation
- Quite good accuracy in speech recognition
  - if you do not mix different languages in a sentence
  - auditory signals are the only used cues (no visual affordances)

Voice-Only Devices: Limitations

- They are quite prolix in the answers
- You have to know what to say!
- Some operations are "challenging", e.g.,
  - once a timer is set up, the user can only *ask* how much time is left
  - getting a weekly weather forecast is a… memory test
- Some actions are not allowed nor expected, e.g.,
  - you cannot insert your wifi password, vocally
  - you cannot hear about all the available (and installable) skills

Voice-First Devices

- Voice-only devices… with a screen
- A system which primarily accept user input via voice commands, and **may** augment audio output with visual information
  - no differences from the "voice" perspective
  - GUI is less capable than the one in screen-first devices
- Typically, the display is a touch screen
  - but it rarely provides buttons or menus
  - the focus is still on voice



Designing (New) VUIs

Background, process, and guidelines

Designing Voice User Interfaces

- Voice interaction between people and devices is analogous to learning a foreign languages
  - both for users and designers/developers
- Easily learnt through **immersion**
  - voice-first devices have an advantage in this
- Successful examples on voice-first devices:
  - sequential numbering of search results
  - randomly show new speech commands
  - voice-accessible interactive (visual) content
- Beware: people often have unrealistic expectations
  - they think a VUI as a "natural conversation partner"

Designing Voice User Interfaces

- To design a VUI, you firstly need to have a clear picture of
  - who is communicating, i.e., who are your users
  - what they are communicating about, what they will ask about, i.e., what their needs are
- Then, you can write some **sample dialogs** and sketch **a diagram of the conversation flow**
  - both convey the flow that the user will actually experience
  - you can also informally experiment with and evaluate different strategies
    - e.g., is it better to confirm a user's request with an implicit confirmation or an explicit one?
- Focus on the **spoken conversation** before considering any visual element
  - imagine to work with a voice-only device

Basic Conversational Frames

- **Controlling**: specifying a goal with means of achieving it
  - "Play Radio Deejay from TuneIn"
- **Delegating**: asking for an outcome without specifying how to achieve it
  - "Play some jazz music"
- **Guiding**: discussing the means of achieving a goal
  - "I want to hear some music, how should I do it?"
- **Collaborating**: mutually deciding on goals between both participants
  - "What should we do?"

Currently adopted by contemporary VUIs

Guidelines for Designing Voice User Interfaces

- Provide users with information about what they can do
  - if the user asks something that does not make sense or it is not possible, provide them with the available options
    - for instance, a weather app can say "You can ask for today's weather or a weekly forecast"
  - an "exit" strategy must be always present and available
    - e.g., Alexa's "stop"

Guidelines for Designing Voice User Interfaces

- Where am I?
  - users must be told which functionality (or part of it) they are using
  - for example:

User	Today's weather forecast	
VUI	Today's weather forecast is rainy	Rainy

- Use non-auditory feedback, if possible
  - some visual feedback may be useful, e.g., a light or a message on screen, to let user know that the system is listening

Guidelines for Designing Voice User Interfaces

- Express intentions in examples
  - in providing sample of speech commands, present the full intention
    - "you may ask: What is the weather in Turin, tomorrow?"
    - "you may say: Play Radio Deejay from TuneIn"
- During execution, leverage on default settings, additional information, or just ask for missing pieces
  - e.g., the location for the weather forecast can be retrieved by GPS or TuneIn can be set as the default for radio broadcasts

Guidelines for Designing Voice User Interfaces

- Limit the amount of information
  - Keep the delivered information **brief**!
    - e.g., Amazon recommends not to list more than 3 different options each time
  - If you have more options, either group them or find another way to accomplish the same goal

Other Guidelines

Amazon's and Google's Suggestions for Designers and Developers

Amazon's Guidelines for Alexa

1. Be adaptable
   - https://developer.amazon.com/docs/alexa-design/adaptable.html
2. Be personal
   - https://developer.amazon.com/docs/alexa-design/personal.html
3. Be available
   - https://developer.amazon.com/docs/alexa-design/available.html
4. Be relatable
   - https://developer.amazon.com/docs/alexa-design/relatable.html
5. Establish and maintain trust
   - https://developer.amazon.com/docs/alexa-design/trustbusters.html

Google's Guidelines for Assistant

- Conversation Design Guidelines – Getting Started
  - https://designguidelines.withgoogle.com/conversation/
- Style Guide
  - https://designguidelines.withgoogle.com/conversation/style-guide/language.html
- Error Handling
  - https://designguidelines.withgoogle.com/conversation/conversational-components/errors.html

References

- Voice First: The Future of Interaction?
  - https://www.nngroup.com/articles/voice-first/
- How to Design Voice User Interfaces
  - https://www.interaction-design.org/literature/article/how-to-design-voice-user-interfaces
- The Narrowing Rift: Voice UI and Conversational UI
  - https://medium.com/@muppetaphrodite/the-narrowing-rift-voice-ui-and-conversational-ui-7d5c95cf086c

License

- These slides are distributed under a Creative Commons license "**Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)**"

- **You are free to:**
  - **Share** — copy and redistribute the material in any medium or format
  - **Adapt** — remix, transform, and build upon the material
  - The licensor cannot revoke these freedoms as long as you follow the license terms.

- **Under the following terms:**
  - **Attribution** — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.
  - **NonCommercial** — You may not use the material for commercial purposes.
  - **ShareAlike** — If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.
  - **No additional restrictions** — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

- https://creativecommons.org/licenses/by-nc-sa/4.0/