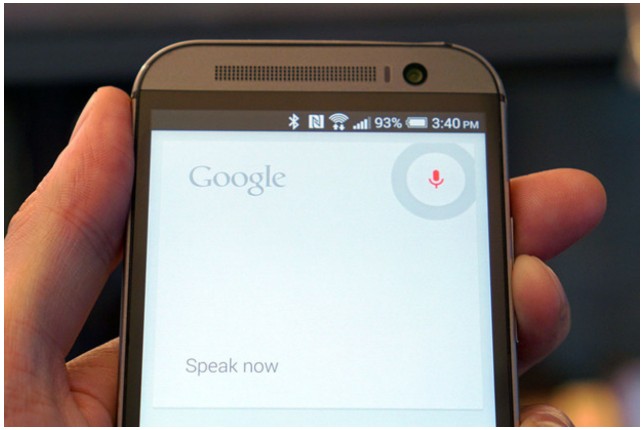
Speech Interfaces



Based on slides by R. Schwitter, E. Reiter

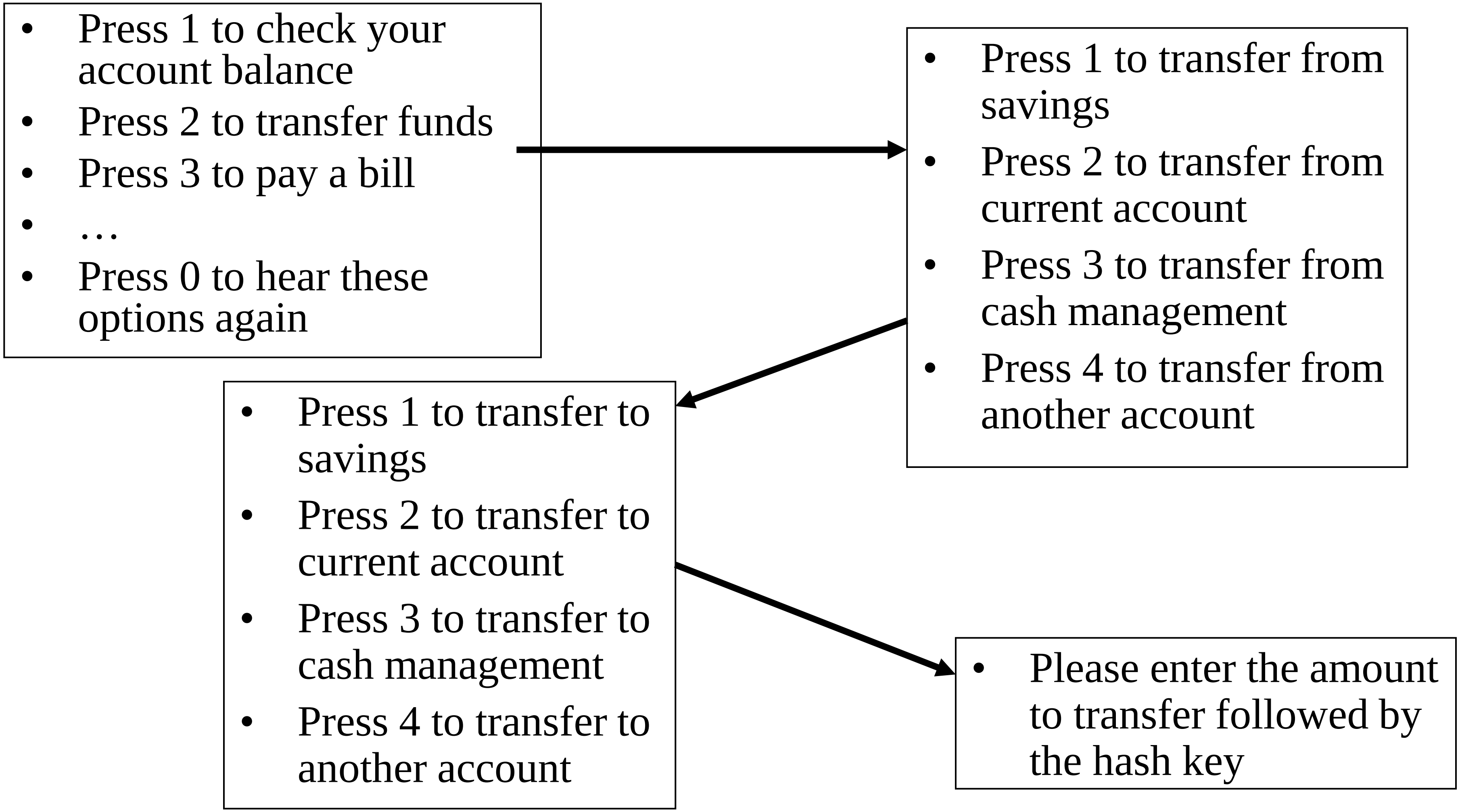
Benefits of Speech Input

- Makes data entry possible without keyboard
- Is excellent for hands/eyes busy situations
- Supports people with disabilities
- Is natural
- May be faster

Speech Input Technology

- Works to some degree, but some words not recognised correctly
- Works best if
  - Limited vocabulary (eg, name of movie)
  - Trained to individual voices
  - Good microphone (not telephone handset!!)
- Speech SDK for Windows
  - http://www.bing.com/dev/speech

Traditional Interactive Voice Response Systems



- Press 1 to check your account balance
- Press 2 to transfer funds
- Press 3 to pay a bill
- …
- Press 0 to hear these options again

- Press 1 to transfer from savings
- Press 2 to transfer from current account
- Press 3 to transfer from cash management
- Press 4 to transfer from another account

- Press 1 to transfer to savings
- Press 2 to transfer to current account
- Press 3 to transfer to cash management
- Press 4 to transfer to another account

- Please enter the amount to transfer followed by the hash key

Speech-Enabled Interaction

- “Transfer 500 pounds from savings to current account next Wednesday after 3pm”

~25 seconds via speech —as compared with two minutes via touch tone

If the computer understands what the user is saying…

Dialogue/Speech Interface

This is a conversational interface, BUT:

- Only speech interaction
  - – Computerised telephone helpline
    - Train information, weather information, …
- Very different kind of HCI!
  - – Is becoming more important
  - – Examples: Apple’s Siri, Google Now, Cortana

See video: http://www.redmondpie.com/microsoft-cortana-vs-apple-siri-vs-google-now-voice-assistants-compared-video/

Issues:

1. What did the user say/mean?
2. Natural interaction
3. What *can* the user say?
4. Prompts
5. Errors and feedback

Issue 1: What did the user say/mean?

- Recognition errors
- Ambiguity/missing information

Recognition Errors

- Causes for recognition errors:
  - user spoke at the wrong time,
  - sentence is not in the grammar,
  - user pause too long,
  - words sound alike,
  - over-emphasis,
  - user has a "cold".

Example

User: Hello, I want to fly to London next Thursday

System: What airport will you be flying from when you go to London, UK?

User: Aberdeen

System: What time on Thursday, 15 May, do you wish to depart from Aberdeen, Scotland?

User: mid-morning

System: BA 1305 leaves Aberdeen at 940 and arrives into London Heathrow at 1115. Should I book one seat for you on Thursday, 15 May?

Questions

- Did the user say “London”
  - If so, which London did he mean?
  - Does he want a particular airport in Lon?
- What does “next Thursday” mean
  - 4 May or 11 May?
- Other information needed
  - departure city
  - time
  - number of seats

Strategies

- Specifically ask about every unknown
  - users dislike
- Incorporate guess into next message
  - “What airport will you be flying from when you go to London, UK?”
  - Explicitly asks about departure airport
  - Implicitly verifies that destination is London, UK
    - User will hopefully tell us if this is incorrect!

Can guess using knowledge

- Statistics
  - London, UK is a more common destination than London, Ontario
- User knowledge
  - User lives in Aberdeen, so this is the likely departure airport
- General knowledge
  - Many people don’t care which London airport they fly into

Issue 2: Natural interaction

- Turn taking
- Use of pauses/silence
- Grounding

Turn-Taking

- Human-human dialog is characterized by turn-taking.
- Speaker A says something, then speaker B, then speaker A.
- How do speakers know when it is their turn?
- It seems that turn-taking is governed by rules.
- Rules apply at places where the structure of language allows a change of speaker to occur.
- The overlap of speech in human conversation is small (< 5%).

Silence

- Silence can occur after any turn in human-human dialog.
- Significant silence needs to be interpreted:

  A: Have you worked on your assignment? (1sec)
  A: Yes or no? (1.5 sec)
  A: Eh?
  B: No.
- Here silence is interpreted as a refusal to respond.
- Users are disturbed by pauses in dialog systems caused by a slow speech recogniser.

Grounding

- Speaker and hearer must constantly establish common ground.
- To achieve common ground, the hearer–must acknowledge the speaker’s utterance or–make it clear that there was a problem.
- For example in
  A: … returning on flight one one eight.
  B: Mm hmm
  the expression Mm hmm is a continuer, also called a backchannel.

Issue 3: What can the user say?

Example: Speech-enabled Calendar

- **Computer**: Currently, you have “lecture on speech interfaces." Later, at 2:00, you have “Seminar in room M2."
- **User**: What's on my calendar the Monday after Easter?
- **Computer**: On Monday, April 8, you have "Call John," you have "Barbara's birthday."
- **User**: What does Paul have?

Example: Speech-enabled Calendar

- **Computer**: On Monday, April 8, Paul has "Chris goes to London." From 10 o'clock to 10:30, "Meeting with Bob."
- **User**: What about the next day?
- **Computer**: On Tuesday, April 9, Paul has no appointments.
- **User**: What do I have tomorrow morning?
- **Computer**: Sorry?

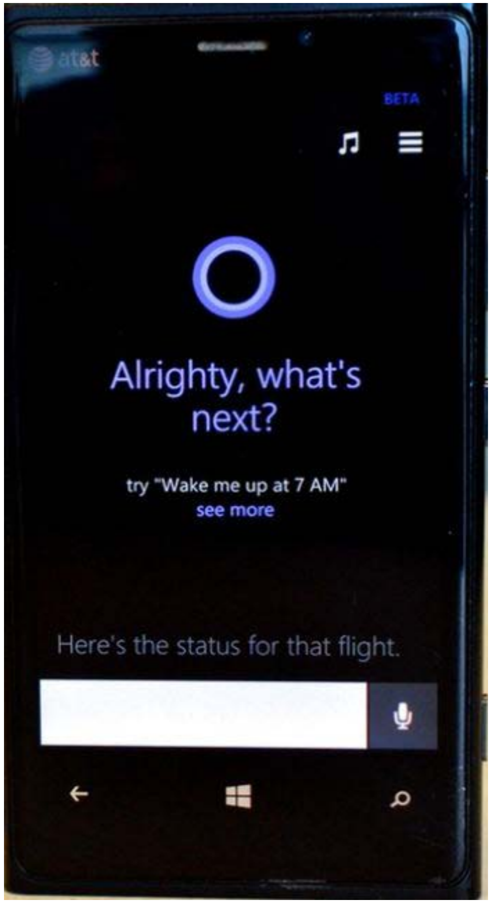
Example: Speech-enabled Calendar

- The expression `"tomorrow morning"` sounds reasonable but is not supported by this application.
- Once you allow users to talk about "morning", "afternoon", and "evening“ they expect to say things like "tomorrow morning", "early afternoon", "after 6 pm".
- Problem:
  - Users assume they can say things that are not supported.
  - Users don’t know all the things that are supported.

Constraints for Guidance

- Guiding users in speaking well-formed input should take into account:
  - coverage of speech recogniser (continuous speech/discrete words)
  - environment in which speech recogniser is used (over a telephone, or in conjunction with a gui)
  - user profile (who uses the speech recogniser).
- Main mechanism: prompts

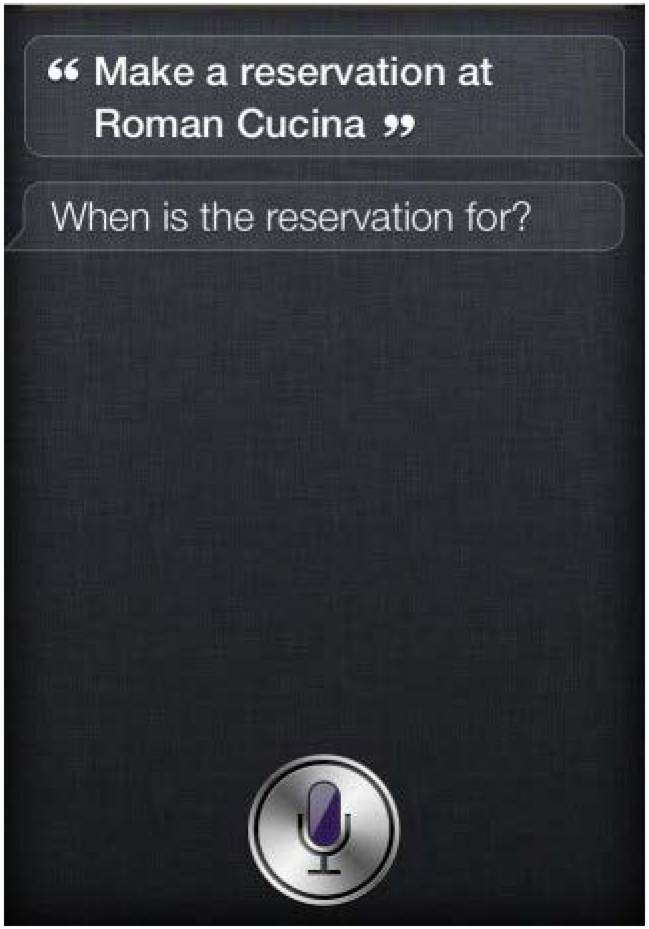
Issue 4: Prompts



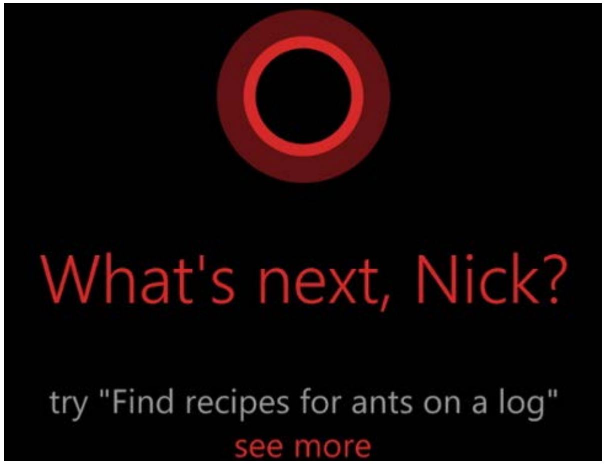
Alrighty, what's next?

try "Wake me up at 7 AM"

Here's the status for that flight.


“ Make a reservation at Roman Cucina ”

When is the reservation for?


What's next, Nick?

try "Find recipes for ants on a log"

Prompts

- Prompts are the turn-taking cues within spoken dialogs.
- Prompts have two purposes:
  - cause the user to speak,
  - convey to the user what may be spoken (optionally).
- Prompts fall along a continuum from implicit to explicit.

Implicit versus Explicit Prompts

- **Computer 1**: Welcome to ABC Bank. What would you like to do?
- **Computer 2**: Welcome to ABC Bank. You can check an account balance, transfer funds, or pay a bill. What would you like to do?
- **Computer 3**: Welcome to ABC Bank. You can check an account balance, transfer funds, or pay a bill. Say one of the following choices: check balance, transfer funds, or pay bills.

Explicit Prompts

- Designers constrained by a small vocabulary stick to explicit prompts.
- The most explicit prompts are called directive prompts.
- Directive prompts ask for the exact word or phrase.
  `Computer``: Please choose: leave a message, return my call, or talk to the secretary.`
- Listing options is most efficient, if the recogniser supports barge-in.
- While the prompt is playing, the user can interrupt and choose.

Example: Music Mall

- **Computer**: Please say the name of the artist you want.
- **User**:The Beatles.
- **Computer**:I have a lot of albums by the Beatles; please say the name of the album you want. For example, you could say "Magical Mystery Tour" or "Abbey Road". For a full list of albums I have by the Beatles, say "list".

Example: Music Mall

- **User**: List.
- **Computer**: Here are the albums I have by the Beatles. Say the one you want as soon as you hear it. "With the Beatles." "Beatles For Sale." "Revolver." "The White Album." "Please Please Me ....
- **User**: <interrupting> The White Album.

Example: Music Mall

- **Computer**: I think you said, "The White Album." Is that correct?
- **User**: Yes.
- **Computer**: Okay. To sample "The White Album"...

Example: Music Mall

- The user input is tightly constrained in the Music Mall application.
- It does not allow input such as:
  `**User**: Hey man, I would like to hear the Beatles, okay.`
- The constrained approach helps to minimize errors.
- Especially, if the application has to work for first-time callers.

Implicit Prompts

- Most people working on research-oriented applications are attempting to design some degree of natural interaction.
- In many of these systems the dialog with the user is a mixed-initiative interaction (i.e. user can take initiative as well).
- Grammars have to deal with a wide range of constructions.
- Cleverly designed prompts can encourage users to speak sentences that conform to more constrained grammars.
- People tend to mimic their conversational partners.

Example: Air Travel System

- **Computer**: Thanks for calling KuoniTravel Company. How can I help you today?
- **User**: I'd like to book a flight.
- **Computer**: Okay. What is your point of origin, and where are you going?
- **User**: I wanna fly from San Francisco, California to Boston, Massachusetts.
- **Computer**: Okay, travelling from San Francisco, California to Boston, Massachusetts. Is that correct?

Prompt Design

- Prompt design is at the heart of effective speech interface design.
- Prompts should follow a number of guidelines:
- Make prompts short (whenever possible).
- Precede prompts with instructions (if necessary).

Computer: Your plan requires that you select a PIN to use the system. The PIN must be between 5 and 9 digits in length. At the tone, please say your PIN.

Prompt Design

- Repeat only the prompt (if something goes wrong).
  Computer: Sorry, I didn’t understand that. Please say your PIN.
- Put key information immediately before expected user input.
  - If using barge-in, put the information at a phrase boundary.
  - If not using barge-in, put the information before the tone.

Prompts: Grammatical Forms

- Use active voice.
  Avoid: Your account number is requested.
  Use: Please enter your account number.

- Use second person.
  Avoid: The user should now say the number
  he wants.
  Use: Please say the number you want.

- Use present tense.
  Avoid: You will be asked for your
  ID number.
  Use: After the prompt, please say your
  ID number.

Prompts: Grammatical Forms

- Avoid subjunctive mood.

Avoid: Should the city name be incorrect, you may backup by saying "Cancel".

Use: If the city name is wrong, say "Cancel".

Yes-No Interrogative Prompts

- Use interrogative forms.

Avoid: `If this is correct, please say "yes" now.`

`For another transaction, say "yes".`

`If you want a quote on <fund name>, say "yes". Otherwise, say "no".`

Use: `Is this correct?`

`Do you want another transaction?`

`Did you say <fund name>?`

Yes-No Interrogative Prompts

- Include the verb on interrogative yes-no prompts.

  Avoid: `Correct?`

  Use: `Is this correct?`

- Reserve imperative form for recovery.

  Use: `For more quotations, say "yes" now.`

  `Please answer "yes" or "no".`

Yes-No Interrogative Prompts

- Avoid ambiguous questions.

  Avoid: `Are you travelling alone and will you need a rental car? Are the name and address correct?`

  Use: `Is this correct?`

- Avoid compound questions.

  Avoid: `Do you want to cancel or change your order?`

  Use: `To process your order, say one of the following: Confirm … Change … Cancel …`

Incremental and Expanded Prompts

- **Computer**: Welcome to ABC Bank. What would you like to do?
- **User**: <silence>
- **Computer**: You can check an account balance, transfer funds or pay a bill. What would you like to do?
- **User**: <silence>
- **Computer**: Say one of the following choices: check balance, transfer funds or pay bills.

Tapering

- Tapering is aimed at shortening the interaction for users as they gain experience with a system.
- For example.

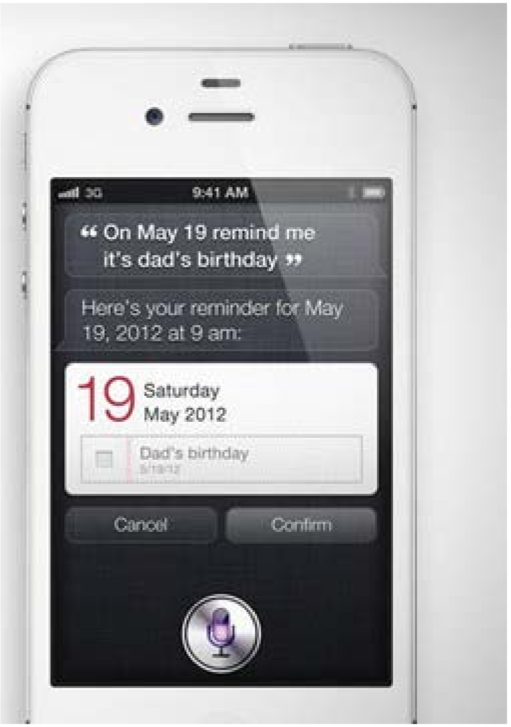
  Computer: Please say the first and last name of the person you want.

  If the user stays on the line to make a second request, the prompt is shortened to

  Computer: Say the name of the person you want.

Issue 5: Errors and feedback

- Errors
- Feedback
- Confirmation
- Help



“On May 19 remind me it's dad's birthday”

Here's your reminder for May 19, 2012 at 9 am:

Error Messages

- An inadequate error message gives no information or often simply repeats a command.

Computer: Say the departure date.
User: Tomorrow.
Computer: Say the departure date.
User: I want to travel tomorrow.
Computer: Say the departure date.

Error Messages

- Error messages should be specific.
- An adequate error message tells the user
  what is wrong, why it is wrong, how to correct it.

```
Computer: Say the departure date.
User:     Tomorrow.
Computer: I do not understand that
          date. Say the month, date,
          and year. For example, say
          October 16th, 2015.
User:     February 4th, 2016.
```

Feedback

- Avoid literal feedback.

  Avoid: Your entry was not seven digits.
  You said 12349670…

  Use: The PIN must be seven digits.
  Please repeat your PIN.

- Replace apology and blame with feedback.

  Avoid: Sorry, I don’t understand.
  You <did something wrong>.

  Use: Please repeat. Do you want help?

Relation between Feedback and Prompts

- Combine feedback with prompts.

```
User:     Call.
Computer: Name to call?
User:     James Martin
```

Feedback and recognition errors

- Recognition error:

  Computer: `Stock name?`
  User: `Texaco.`
  Computer: `Shares of PepsiCo to sell?`
  User: `… umh… No, that’s wrong …`

Confirmations

- You may have to use confirmation questions to assure that the computer has heard the right word or phrase.

```
Computer: What do you want to do next?
User:     I want to schedule an
          appointment with my manager.
Computer: Do you want to set up an
          appointment?
User:     Yes.
```

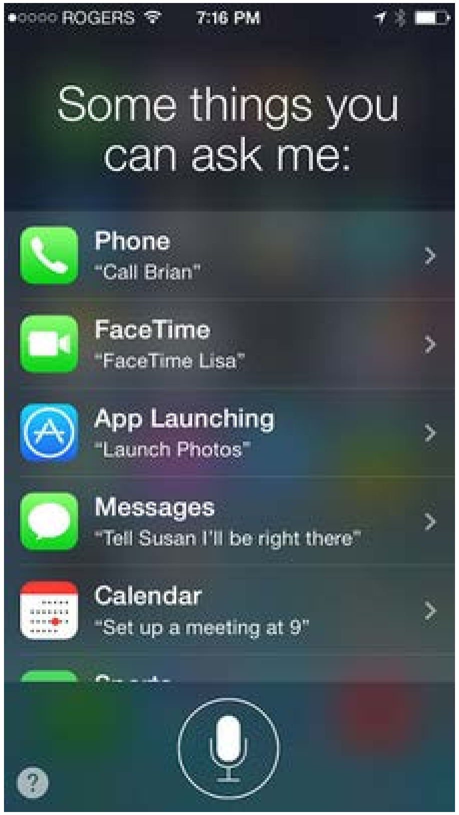
- You have to balance the cost of
  - making an error
  - with extra time to confirm a statement.

Help

- Empower the user with help availability.
- Let the user know that help is available.
- Let the user know how to get help.
- Once help is declared available, keep it available.
- Return to a logical starting point after help.
- Use examples for help.

Help Mode

- Differentiate between help and application mode.
- Let the user exit the help mode.



●○○○○ ROGERS 7:16 PM

Some things you can ask me:

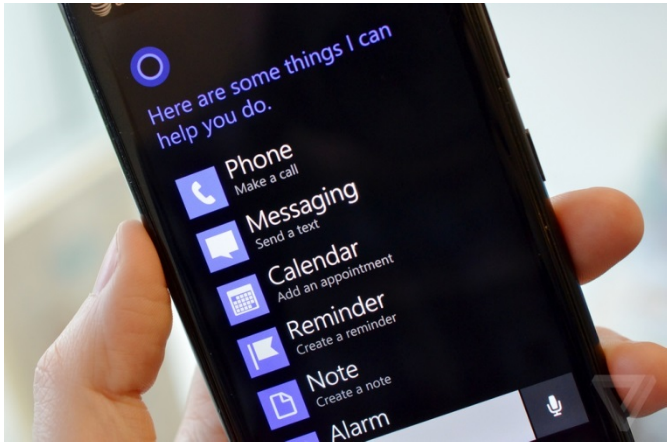
Phone
"Call Brian"

FaceTime
"FaceTime Lisa"

App Launching
"Launch Photos"

Messages
"Tell Susan I'll be right there"

Calendar
"Set up a meeting at 9"



Here are some things I can help you do.

Phone
Make a call

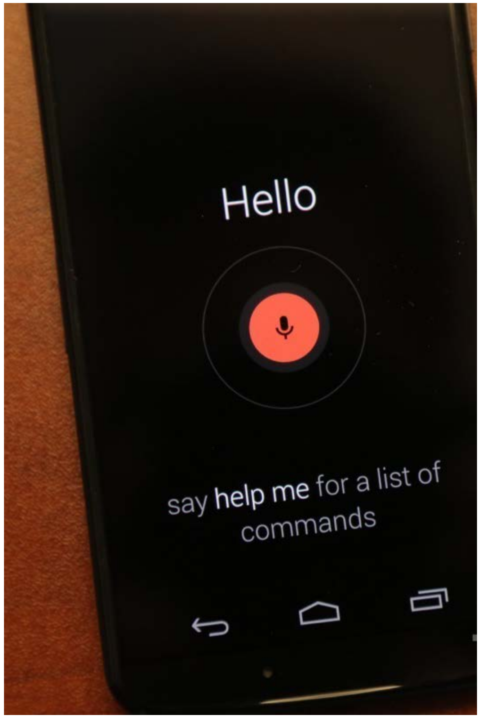
Messaging
Send a text

Calendar
Add an appointment

Reminder
Create a reminder

Note
Create a note

Alarm

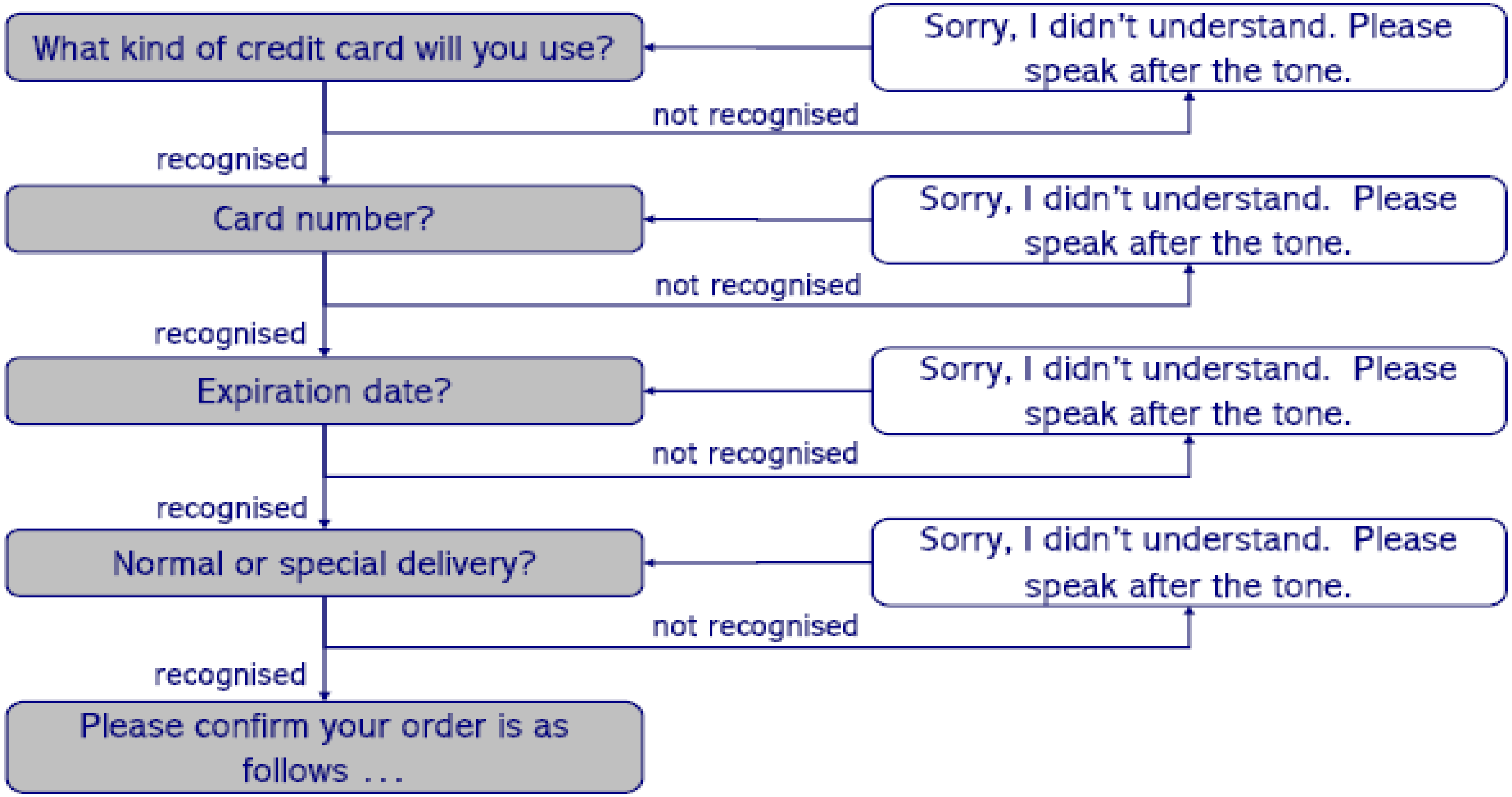


Hello

say **help me** for a list of commands

Designing a Speech Dialogue System

Example of Prototype Notation



What kind of credit card will you use?

Sorry, I didn't understand. Please speak after the tone.

not recognised

recognised

Card number?

Sorry, I didn't understand. Please speak after the tone.

not recognised

recognised

Expiration date?

Sorry, I didn't understand. Please speak after the tone.

not recognised

recognised

Normal or special delivery?

Sorry, I didn't understand. Please speak after the tone.

not recognised

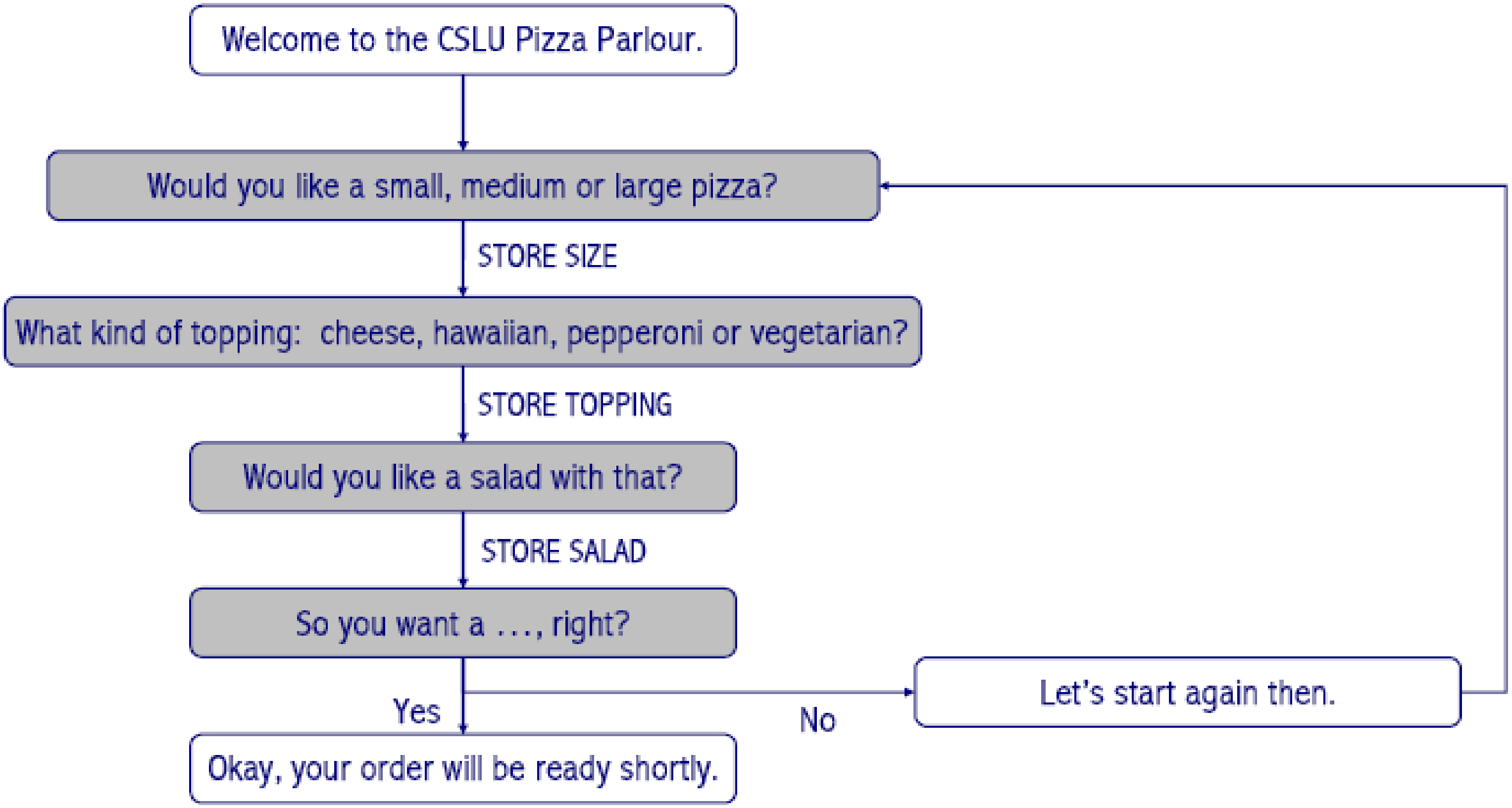
recognised

Please confirm your order is as follows . . .

Prototype Notation: Call Flow Diagram

- A call flow diagram shows the flow of the dialogue.
- Different types of notations exist.
- We will use the following conventions:
  - explicit prompts (or abbreviations) label nodes
  - system actions appear on arcs.
- Distinguish between recognition states and non-recognition states.
- Error recovery:
  - you can assume each state has a re-prompt capability built in
  - there’s no need to include it in your call flow diagram.

Example: Pizza Ordering



Welcome to the CSLU Pizza Parlour.

Would you like a small, medium or large pizza?

STORE SIZE

What kind of topping: cheese, hawaiian, pepperoni or vegetarian?

STORE TOPPING

Would you like a salad with that?

STORE SALAD

So you want a ..., right?

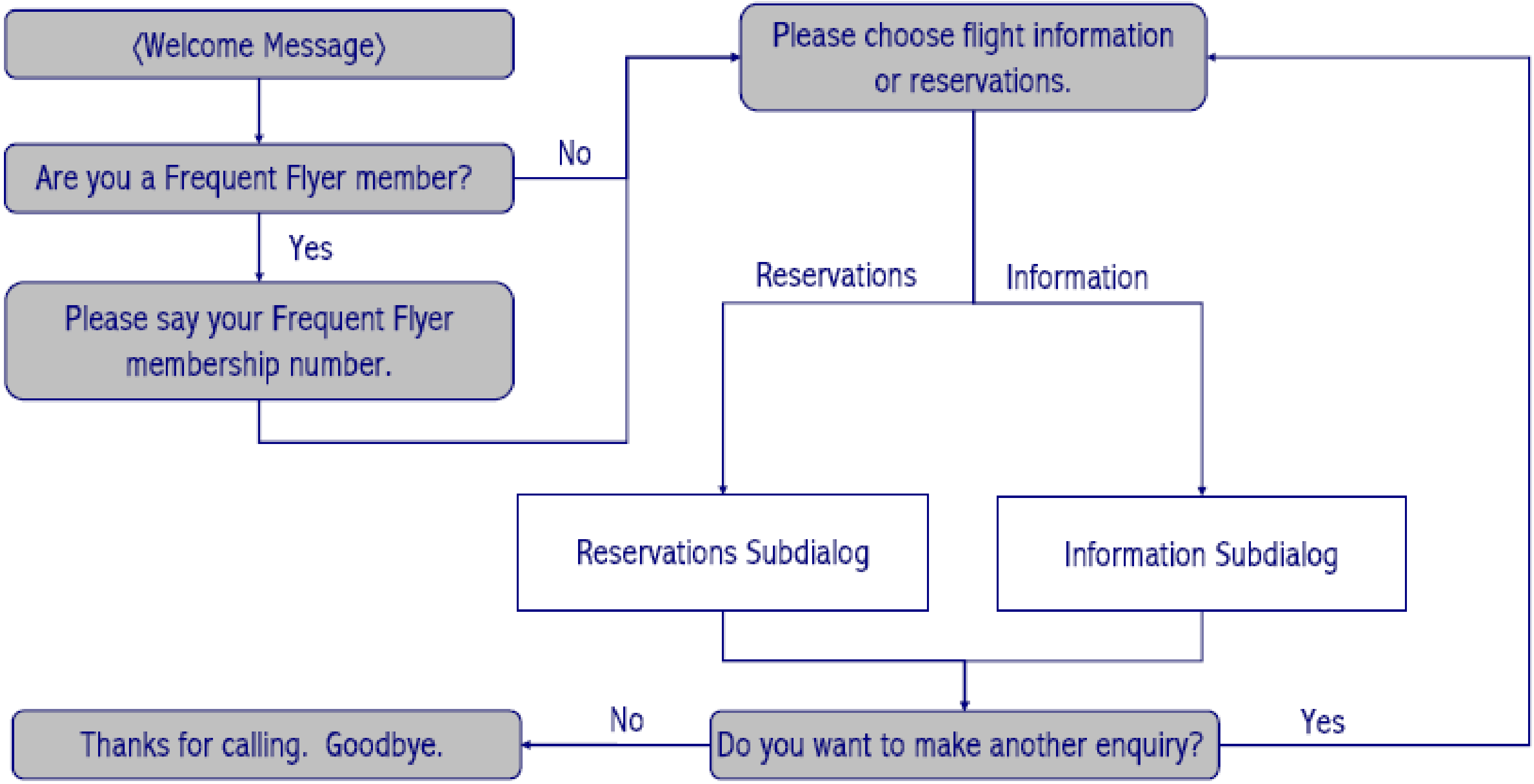
Yes

No

Let's start again then.

Okay, your order will be ready shortly.

Example: Flight bookings


⟨Welcome Message⟩
Are you a Frequent Flyer member?
No
Yes
Please say your Frequent Flyer membership number.
Please choose flight information or reservations.
Reservations
Information
Reservations Subdialog
Information Subdialog
Thanks for calling. Goodbye.
No
Do you want to make another enquiry?
Yes

Wizard of Oz Simulations

- A human experimenter (the Wizard) simulates an automated system.
- Uses the dialog specs (e.g., call flow diagram and prompts).
- Reads the appropriate prompt from the specs, waits for a response from the subject (or no response), checks the specs on how to proceed, and then speaks the next prompt.
- Very effective in uncovering problems with logic, navigation, awkward sequences of prompts, omissions, etc.
- Good to get a feeling for grammar coverage.

Some Limitations of Speech Interfaces

- Uncertainty about user input
- Only one thing at a time communicated
- No pointing
- Speech is transient,
  - say simple things, verify if understood

- Harder to program!
  - But more appropriate in many contexts?

Further Limitations of Speech Interfaces

- Speech interfaces and command-line interfaces have similar problems.
- The functionality of the application is hidden.
- The boundaries of what can and cannot be done are invisible.
- In speech-only environment it is not possible to– display menus,–show options,–highlight buttons.
- Other techniques must be used to guide users through an interaction.