

Next-Generation Spoken Language Dialog Systems

Robert Dale

rdale@ics.mq.edu.au
www.clt.mq.edu.au

The Aims of This Talk

- To establish where the technology is today
- To look at what's coming out of the research labs in the next few years
- To see what voice applications of the future will look like

Outline

- Spoken Language Dialog Systems: What They Are and Why They're Important
- Advances Enabling Better Quality Applications
- Advances Going Beyond Telephony
- Advances Providing Broader Application Possibilities
- How Does the Future Look?

The Bigger Picture

- Language Technologies:
 - Technologies for working with the spoken voice
 - Technologies for working with documents
 - Search engines and information retrieval
 - Text summarisation and information extraction
 - Machine translation
 - Question answering
 - Automated email handling and forms handling
- Find out more at www.clt.mq.edu.au

The Evolving Interface

- 1960s through 1980s: command line interfaces
- 1980s and 1990s: graphical user interfaces
- The new millenium: the voice user interface

Spoken Language Technologies

- The Early Days
- Hollywood's Vision
- Real Systems Today
- What's in the Labs Today

What is a Spoken Language Dialog System?

- An SLDS is a computer system that you can talk to in order to carry out some task
- SLDSs are typically of two kinds:
 - Transaction-based systems allow you to undertake some transaction, such as buying or selling stocks, or reserving a seat on a plane
 - Information-provision systems provide information in response to a query, such as a request for timetable information or weather information

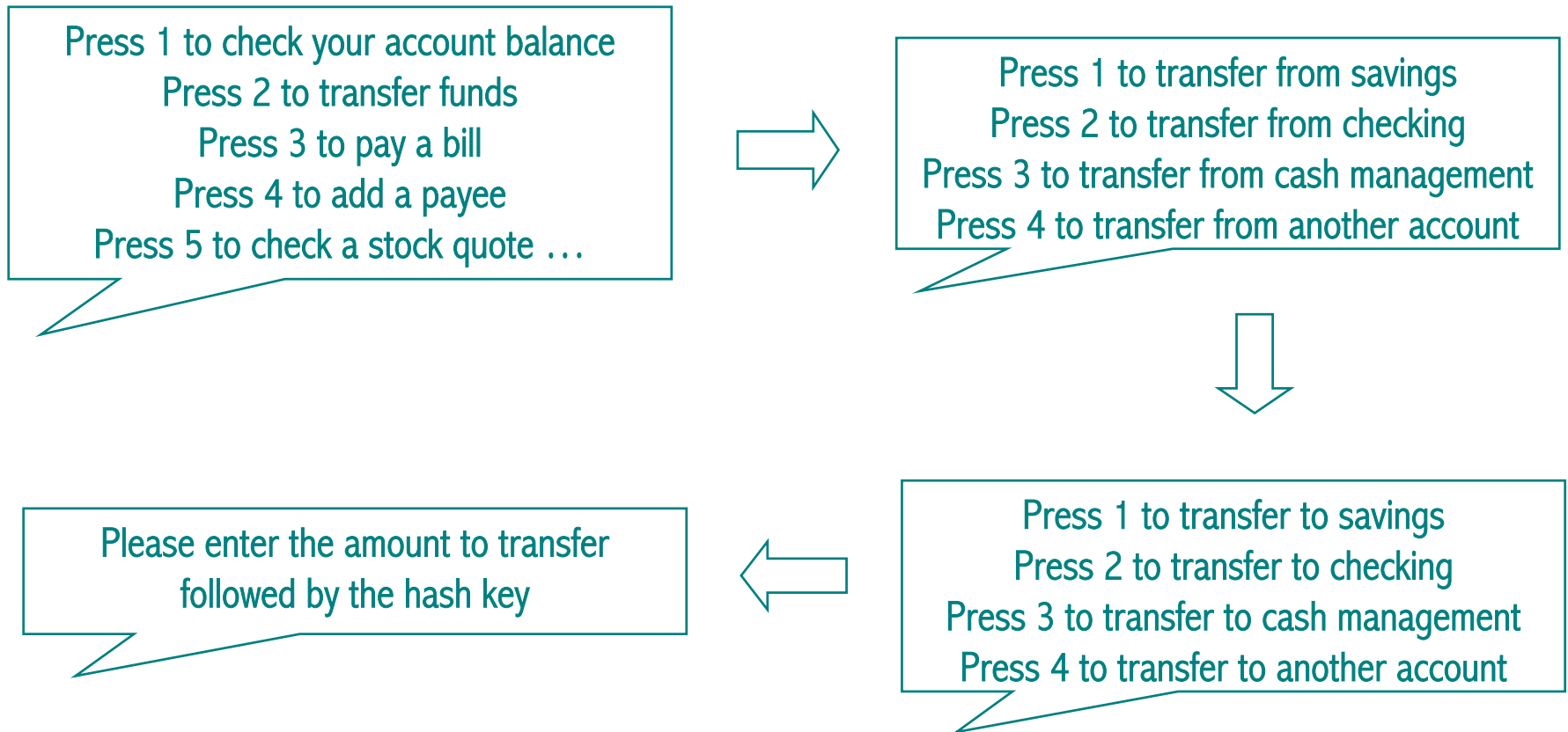
Two Uses of Speech Recognition Technology

- On the desktop:
 - Speaker-dependent
 - Large vocabulary (tens of thousands of words)
 - Dictation tasks
- Telephony-based:
 - Speaker-independent
 - Relatively small vocab (hundreds of words)
 - Interactive tasks


Uses of Telephony-Based SLDSs

- Remote banking
- Travel reservations
- Information enquiry
- Stock transactions
- Auto-attendants
- Directory assistance
- Taxi bookings
- Pizza ordering

Traditional Interactive Voice Response Systems



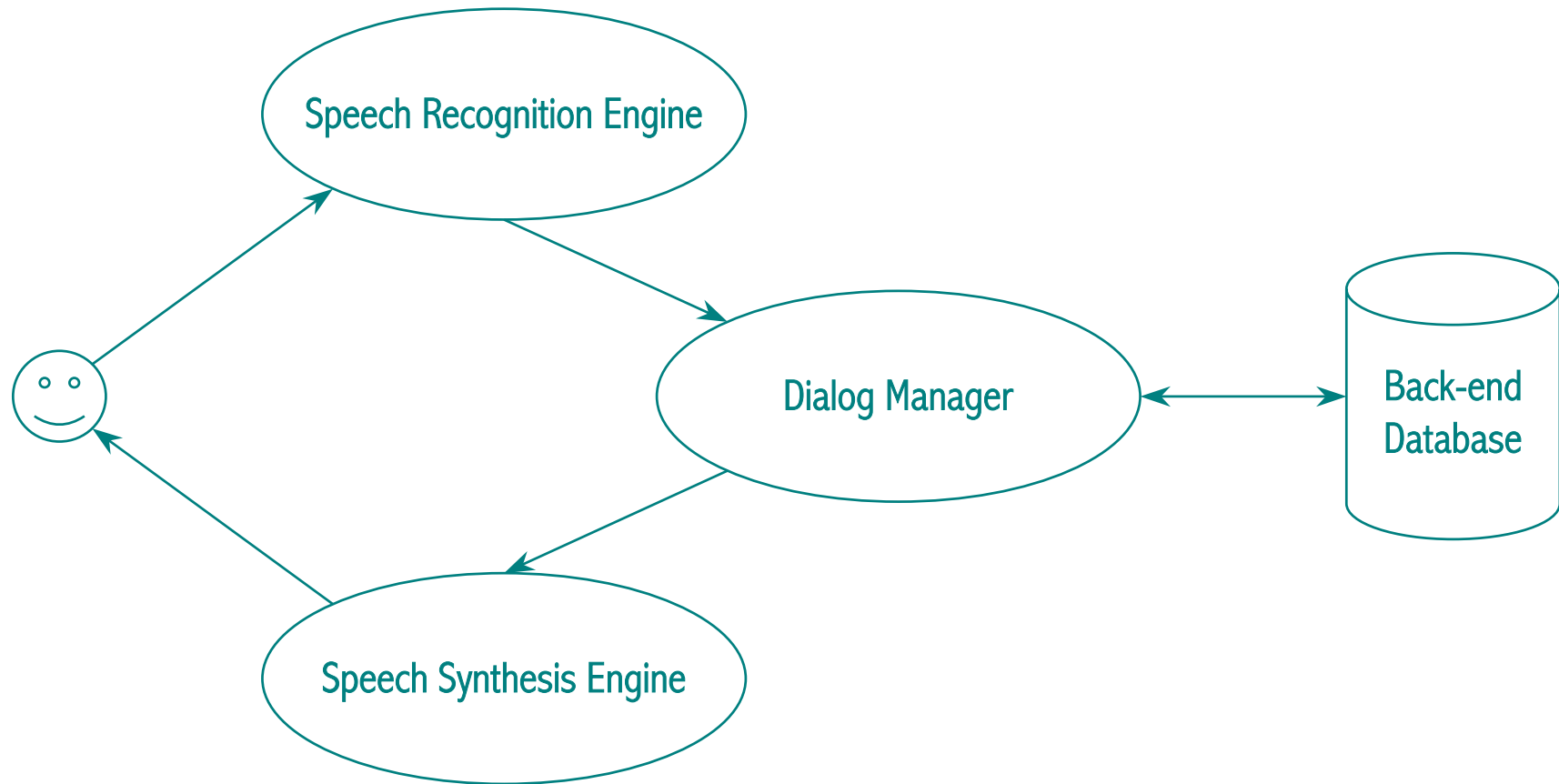
Speech-Enabled Interaction



Transfer 500 dollars from
savings to checking next
Wednesday after 3pm

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

The Architecture of a Spoken Language Dialog System



Areas of Future Technical Development

- Techniques for processing the acoustic signal, not only to improve the accuracy of speech recognition but also, for example, for speaker identification;
- Techniques for managing the interaction with the user, usually with the aim of making dialogs more natural; and
- Techniques for improving the quality of synthesized speech

How These Developments Will Impact Applications and Users

- Some advances will result in better quality applications
- Some advances will take us beyond telephony
- Some advances will broaden the range of possible applications
- So, ten things to watch for ...

Outline

- Spoken Language Dialog Systems: What They Are and Why They're Important
- Advances Enabling Better Quality Applications
- Advances Going Beyond Telephony
- Advances Providing Broader Application Possibilities
- How Does the Future Look?

Advance #1: Better Synthesised Output

- Current systems:
 - for short texts of a few sentences, current TTS systems are of high quality
 - less acceptable for larger chunks of information
- In 3-5 years:
 - synthesis driven by the purpose of the utterance
 - better pausing, phrasing and prosody; expression of emotion
 - ‘concept-to-speech’ rather than text-to-speech

Advance #2: Better Tools

- Currently:
 - Languages like VoiceXML provide abstractions that it make it easy to build simple dialog systems ...
 - ... but they also make it easy to build bad dialog systems
- In 3-5 years:
 - Next generation tools will incorporate templates and wizards that support best practice

Advance #3: Speaker Identification

- Current systems:
 - Speakers identified by caller line identification and database profiles
- In 5 years:
 - Identify speaker from his or her speech; use tailored recognition, web-accessible profiles
 - Handle multi-speaker contexts – determine who is speaking when

Advance #4: Real Natural Language Processing

- Current systems:
 - rely on hand-coded or statistically-derived correspondences between user utterances and semantic slot fills
 - very brittle, not very reusable
- In 5-10 years:
 - based on broader linguistic principles
 - larger more powerful components
 - more robust, more reusable

Outline

- Spoken Language Dialog Systems: What They Are and Why They're Important
- Advances Enabling Better Quality Applications
- Advances Going Beyond Telephony
- Advances Providing Broader Application Possibilities
- How Does the Future Look?

Advance #5: Multimodal Apps

- Almost here: integration of speech, text and graphics
 - Route directions via graphics and voice: ‘Turn here.’
- Soon: integration of touch
 - Point to maps on your PDA screen: ‘How do I get there?’
 - Point to emails: ‘Read me these.’
- In 5-10 years:
 - Recognition of facial expressions?

Advance #6: Embedded Apps

- Already feasible to embed speech recognition and speech synthesis in other devices
 - Cars, PDAs, tablet PCs
- Next:
 - fridges, microwaves, shoes
 - smart spaces
- Key issues
 - Back-end intelligence needed to populate the interface
 - How does the shoe know you're not talking to the fridge?

Outline

- Spoken Language Dialog Systems: What They Are and Why They're Important
- Advances Enabling Better Quality Applications
- Advances Going Beyond Telephony
- Advances Providing Broader Application Possibilities
- How Does the Future Look?

Advance #7: Smart Responses

- Current systems:
 - use ‘canned text’ or simple templates to create responses
- In 5 years:
 - automatic summarisation of complex data sources using Natural Language Generation
 - systems will reason about and plan their responses

Advance #8: Multilingual Systems

- In 3-5 years:
 - Systems which work out the language of the speaker and adapt accordingly
- In 5-10 years:
 - Systems which work reliably with multiple languages, eg to translate from one into the other: book a hotel room in Tokyo without knowing any Japanese

Advance #9: Large Vocabulary SR

- Currently you can have any two of:
 - speaker independence
 - large vocabulary
 - high accuracy
- In 5-10 years:
 - unlimited vocabulary, speaker-independent continuous dictation
 - makes possible sophisticated apps, eg audio data mining, email dictation

Advance #10: Conversational Interfaces?

- Not in our lifetimes
- Science Fiction has raised the consumer's expectations to unrealistic levels: a recalibration may be needed
- Radical approaches: eg CMU's Universal Speech Interface, European standards for voice-interface vocabulary

Outline

- Spoken Language Dialog Systems: What They Are and Why They're Important
- Advances Enabling Better Quality Applications
- Advances Going Beyond Telephony
- Advances Providing Broader Application Possibilities
- How Does the Future Look?

How Does the Future Look?

- SLDSs are a viable technology now
- SLDSs of the future will be:
 - better constructed, more robust, more pleasant to listen to, smarter, multimodal, multilingual
- ... but they won't be like 2001's HAL

Concluding Moral

- Badly chosen applications generate bad press
- Well-chosen applications are quietly successful
- Choose your application well!