Expressive Human and Command languages

- We now discuss the rapidly growing speech interfaces that includes both speech recognition and speech production .
- We also discuss human language technologies including translation and educational applications.
- Finally, we review the traditional, yet expressive, command language interfaces

Speech technologies

- Store and replay (museum guides)
- Dictation (document preparation, web search)
- Close captioning, transcription
- Transactions over the phone
- Personal “assistant” (common tasks on mobile devices)
- Hands-free interaction with a device
- Adaptive technology for users with disabilities
- Translation
- Alerts
- Speaker identification

Speech Recognition

- Speech recognition has made significant progress in recent year and is now being used in a number of well targeted knowledge domains such as airline information, lost luggage, medical-record data entry, and personal digital assistants
- . Driven by the difficulty of typing while using mobile devices (phones or touch tablets), spoken input has gained acceptability.
- More users learn to use spoken commands such as “Where is the closest coffee shop?” or “Tell John I will be late.”

- Early applications of speech recognition were mostly limited to discrete-word recognition (with extensive training for the system to learn a particular user’s voice), the major breakthrough in the past decade has been the improvement of continuous-speech recognition algorithms and the availability of very large repositories of voice data on the web, which can be analyzed to train algorithms.
- The significant advance that made speech recognition possible on mobile devices is the ability to process the spoken input remotely and quickly enough for rapid interaction.
- Reduced training (or its elimination with speaker-independent systems) has greatly expanded the scope of commercial applications.
- Quiet environments, head-mounted high quality microphones, and careful choice of vocabularies improve recognition rates in all cases.

Speech recognition and production: Opportunities and obstacles

- ***Opportunities***
- When users have physical impairments
- When the speaker's hands are busy
- When mobility is required
- When the speaker's eyes are occupied
- When harsh or cramped conditions preclude use of a
- keyboard
- When application domain vocabulary and tasks are limited
- When the user is unable to read or write (e.g., children)

Obstacles to speech recognition

- Interference from noisy environments and poor-quality microphones
- Commands need to be learned and remembered
- Recognition may be challenged by strong accents or unusual vocabulary
- Talking is not always acceptable (e.g., in shared office, during meetings)
- Error correction can be time-consuming
- Increased cognitive load compared with typing or pointing
- Math or programming difficult without extreme customization

- **Obstacles to speech production**
- Slow pace of speech output when compared with visual
- displays
- Ephemeral nature of speech
- Not socially acceptable in public spaces (also privacy
- issues)
- Difficulty in scanning/searching spoken messages
- use can be significant for users who take the time to learn and remember what can be accomplished with spoken commands, but

general users of office or personal computers are not rushing yet to adopt speech input and output device

Speech Recognition applications

- The goal of speech recognition is primarily to *produce text based on spoken input* , the most straightforward application being *dictation*. Dictation systems have now reached recognition rates that are acceptable in many situations (e.g., Google Docs' Voice Typing).
- They allow users to compose a document or speak search terms such as "movie theater in college park" and then correct mistakes with the keyboard instead of typing all the text.
- It can be a big time-saver with mobile devices, but keyboards, function keys, and pointing devices with direct manipulation often remain more rapid, depending on the quality of the recognition and the context of use (mobile or not), user's typing abilities, vocabulary complexity, nonnative speaker, and so on.

- The other large category of speech recognition use is to allow users to *speak commands* that the user interface is trained to recognize effectively.
- This includes completing transactions over the phone, interacting with a device when direct manipulation is not convenient or possible, and using specialized voice services or "assistants."
- Dictation without using a keyboard will also require the use of commands to correct errors, start a new paragraph, or request the possibility to spell a name

- **Specialized voice services** or *“personal assistants”* like Siri, Google Cortana, and Hound have become the more visible use of speech recognition.
- Because mobile interaction makes the use of keyboards impractical, speech becomes attractive to allow users tospeak commands that execute the most common tasks performed on those devices, such as finding a location of interest, setting reminders, calendaring, communicating with others, or launching apps.

- ***Initiation***

- The first step in using spoken interaction is for users to indicate that they wish to start the spoken interaction.
- In phone systems, a welcome prompt is sufficient to get started,
- On the screen, a start button is needed (usually in the shape of a microphone), or an option
- is available to use a voice command to turn on the listening (e.g., “Hey Siri” or “Wake up”).
- This spoken command has to be very carefully chosen so that it is not misrecognized, but false positives will inevitably occur, causing frustration and possible chaos if further commands are recognized without users noticing it.
- The initiation may be done for each command, or a separate spoken command may be needed to stop the recognition process

Knowing what to say

- Next, users need to know what can be said and reliably recognized.
- Learnability is one of the main issues of human language technologies that attempt to mimic natural language.
- In IVR phone systems, spoken prompts guide users and invite them to press keys or speak one of the proposed menu choices.
- Because they are typically used by novices or intermittent users, the possible transactions remain
- simple and the dialogue entirely directed (e.g., users are instructed to please say "account balance," "bill pay," or "fund transfer").
- Some IVR systems use more open-ended prompts (e.g., "What service do you need?") and rely on a series of dialogues to clarify and confirm choices.
- The use of speech recognition allows users to shortcut through menu trees, which can be successful when users know the names of what they seek, such as a city, person, or stock name.

- ***Recognition errors***
- Common errors occur when the vocabulary includes similar terms ("dime/time" or
- "Houston/Austin").
- Challenges include dealing with regional or foreign accents and background noise.

- ***Correcting errors***
- Correcting errors can be very taxing, especially when users do not have access to a keyboard or pointing device so all corrections have to be done using speech, possibly compounding errors with new ones
- ***Mapping to possible actions***

- ***Feedback and dialogues***

- During dictation or transcription, the recognized text is shown in the document being composed or in a dictation buffer, usually after a short delay (one to two seconds).
- Users can continue speaking or start correcting errors with the keyboard or by speaking navigation or editing commands.
- After correction, the text can also be transferred to a search box, the body of an e-mail message, or a field in a form.