CPE/CSC 486: Human-Computer Interaction

Franz J. Kurfess

Computer Science Department
California Polytechnic State University
San Luis Obispo, CA, U.S.A.

CAL POLY

Course Overview

- Introduction
- Cognitive Foundations
- Input-Output Devices
- Interaction Spaces
- Interaction Styles
- Interaction with Mobile Devices
- Speech-Based Interaction
- User Assistance
- Natural User Interfaces
- Case Studies
- Project Presentations

Logistics

- **Term Project**
  - quick review of mid-quarter project displays Thu, May 3
  - opening ceremony ("ribbon cutting") on Thu, June 1
    - guests
    - student presentations
- **Research Activity**
  - status update on Thursday
- **Talk by Dr. Leilei Chu**
  - 10:00 am, 14-302

Course Overview

- Introduction
- Interacting with Devices
- Interaction Styles
- UI Elements
- UI Design Guidelines
- UI Development Tools
- User Assistance
- Interaction through Speech
- Interaction with Mobile Devices
- Project Presentations

Chapter Overview
Speech User Interfaces

- Motivation
- Objectives
- Speech Technologies
- Speech Recognition
- Speech Applications
- Speech User Interface Design
- Natural Language
- Important Concepts and Terms
- Chapter Summary

Speech Recognition

- motivation
- terminology
- principles
- discrete vs. continuous speech recognition
- speaker-dependent vs. speaker-independent recognition
- vocabulary
- limitations

[Mustillo]

Motivation

- **speaking is the most natural method of communicating between people**
- **the aim of speech recognition is to extend this communication capability to interaction with machines/ computers**
  - "Speech is the ultimate, ubiquitous interface." Judith Markowitz, J. Markowitz Consultants, 1996.
  - "Speech is the interface of the future in the PC industry." Bill Gates, Microsoft, 1998.
  - "Speech technology is the next big thing in computing." BusinessWeek, February 23, 1998.
  - "Speech is not just the future of Windows, but the future of computing itself." Bill Gates, BusinessWeek, February 23, 1998.

[Mustillo]

Terminology

- **speech recognition (SR)**
  - the ability to identify what is said
- **speaker recognition**
  - the ability to identify who said it
  - also referred to as speaker identification, voice recognition
- **speech recognition system**
  - produces a sequence of words from speech input
- **speech understanding system**
  - tries to interpret the speaker's intention
  - also sometimes referred to as Spoken Dialog System

[Mustillo]

Terminology (cont.)

- **talk-through (barge-in)**
  - allows users to respond (interrupt) during a prompt
- **word spotting**
  - recognizer feature that permits the recognition of a vocabulary item even though it is preceded and/or followed by a spoken word, phrase, or nonsense sound
    - example: “I’d like to make a collect call, please.”
- **decoy**
  - word, phrase or sound used for rejection purposes
    - natural decoys - hesitation "ah", user confusion "What?", "Hello", ...
    - artificial decoys - unvoiced phonemes used to identify "clunks" (phone hang-ups) and background noises.

[Mustillo]

SR Principles

- process of converting acoustic wave patterns of speech into words
- true whether speech recognition is done by a machine or by a human
- seemingly effortless for humans
- significantly more difficult for machines
- the essential goal of speech recognition technology is to make machines (i.e., computers) recognize spoken words, and treat them as input

[Mustillo]

Speech Recognizer

Input speech

Feature extraction: Extract salient characteristics of user's speech

Channel equalization and noise reduction

End-point detection: Obtain start and end of user's speech

Acoustic Models of Phonemes

Recognition: Score list of candidates

Vocabulary

Confidence measurement:
In or out vocabulary
Correct or incorrect choice

Recognized word or rejection decision

[Mustillo]

Discrete Speech Recognition

- **requires the user to pause briefly between words**
  - typically > 250 ms of silence must separate each word
  - common technology today
  - example:
    - entering a phone number using Isolated-Digit Recognition (IDR)
    - "7" (pause), "6" (pause), "5" (pause), "7" (pause), "7" (pause), "4" (pause), "3" (pause)

[Mustillo]

Connected Speech Recognition

- **isolated word recognition without a clear pause**
- **each utterance (word/digit) must be stressed in order to be recognized**
- **Connected-Digit Recognition (CDR)**
  - e.g., 765-7743
  - becoming common technology for interactions between users and computers via phone
    - banking
    - reservations

[Mustillo]

Continuous Speech Recognition

- **most natural for humans**
  - users can speak normally without pausing between words
- **these speech systems can extract information from concatenated strings of words**
- **continuous-digit recognition**
  - e.g., “I’d like to dial 765-[77][43].”
- **some companies have deployed this technology commercially**
  - often domain-specific, limited vocabulary

[Mustillo]

Speaker-Dependent Recognition (SDR)

- system stores samples (templates) of the user's voice in a database, and then compares the speaker's voice to the stored templates
- also known as Speaker-Trained Recognition
  - recognizes the speech patterns of only those who have trained the system
- can accurately recognize 98%-99% of the words spoken by the person who trained it
  - training is also known as enrollment
- only the person who trained the system should use it
- examples: dictation systems, voice-activated dialing

[Mustillo]

Speaker-independent Recognition (SIR)

- capable of recognizing a fixed set of words spoken by a wide range of speakers
- more flexible than STR systems because they respond to particular words (phonemes) rather than the voice of a particular speaker
- more prone to error
  - the complexity of the system increases with the number of words the system is expected to recognized
  - many of samples need to be collected for each vocabulary word to tune the speech models

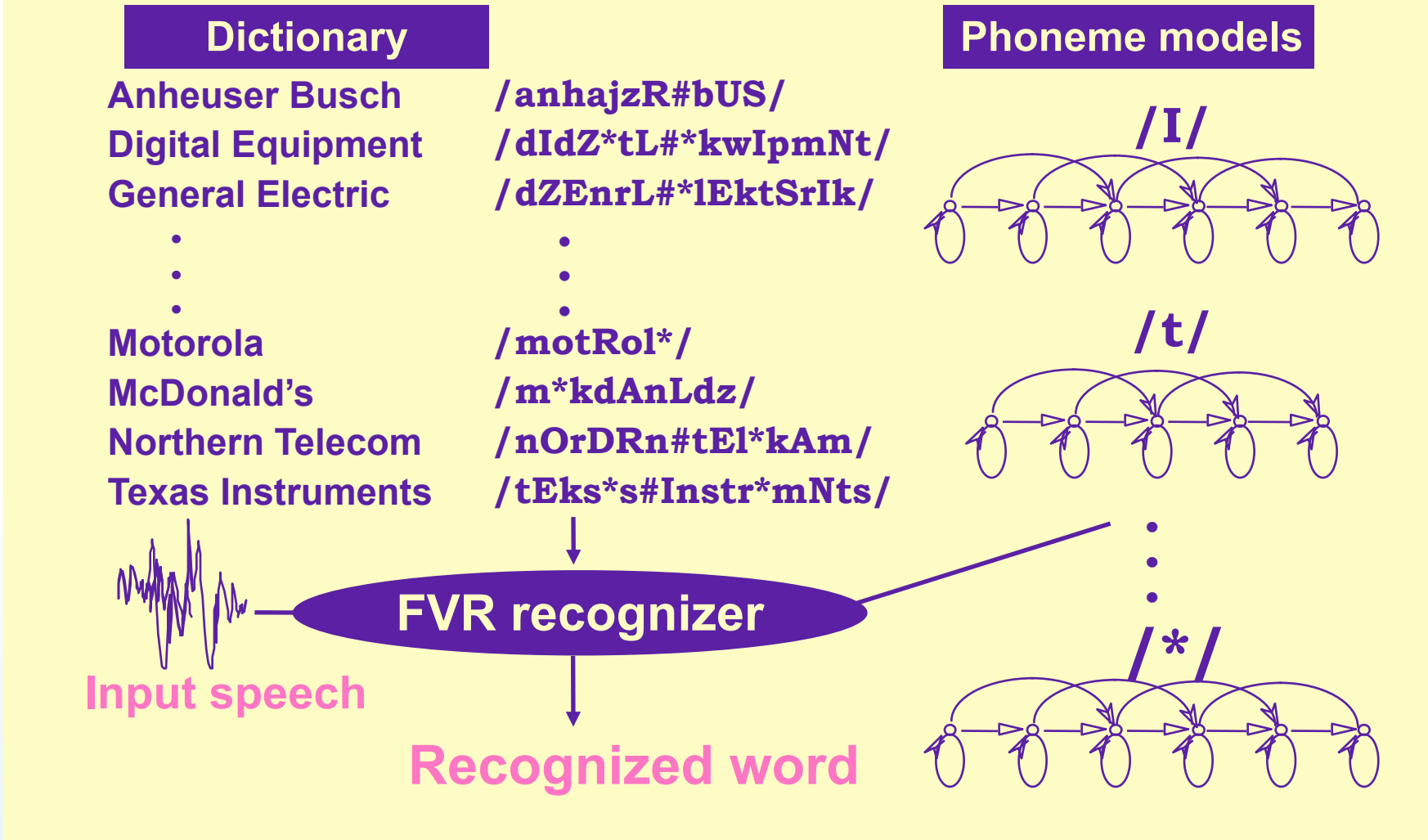
[Mustillo]

Phonemes

- **smallest segments of sound that can be distinguished by their contrast within words**
  - 40 phonemes for English: 24 consonants and 16 vowels
    - example: consonants - /b/ bat or slab, d/ dad or lad, /g/ gun or lag, ... vowels - /i/ eat, /I/ it, /e/ ate, /E/ den, ...
  - in French, there are 36 phonemes: 17 consonants and 19 vowels
    - example: /tC/ tu, /g!/ parking, /e/ chez, /e!/ pain, ...

[Mustillo]

Example SIR


Dictionary
Anheuser Busch /anhajzR#bUS/
Digital Equipment /dIdZ*tL#*kwIpmNt/
General Electric /dZEnrL#*lEktSrIk/
Motorola /motRol*/
McDonald's /m*kdAnLdz/
Northern Telecom /nOrDRn#tEl*kAm/
Texas Instruments /tEks*s#Instr*mNts/
Phoneme models
/I/
/t/
/*/
FVR recognizer
Input speech
Recognized word

Differences SDR-SIR

- **dictionary composition:**
  - dictionary entries in SDR are determined by the user, and the vocabulary is dynamic
  - best performance is obtained for the person who trained a given dictionary entry
  - dictionary entries in SIR are speaker independent, and are more static
  - training of dictionary entries:
    - for SDR, training of entries is done on-line by the user
    - for SIR, training is done off-line by the system using a large amount of data

[Mustillo]

SR Performance Factors

- **physical characteristics**
- **geographic diversity of the speaker**
  - regional dialects, pronunciations
- **age distribution of speakers**
- **ethnic and gender mix**
- **speed of speaking**
- **uneven stress on words**
  - some words are emphasized
- **stress on the speaker**

CAL POLY

[Mustillo]

SR Performance Factors (cont.)

- **phonetic**
  - "a" in "pay" is recognized as different from the "a" in "pain" because it is surrounded by different phonemes
- **co-articulation**
  - the effect of different words running together
  - "Did you" can become "dija"
- **poor articulation**
  - people often mispronounce words
- **loudness**
- **background noise**

[Mustillo]

SR Performance Factors (cont.)

- **phonemic confusability**
  - words that sound the same but mean different things
    Example: “blue” and “blew”, “two days” and “today’s”, “cents” and “sense”, etc.
- **delay**
  - local vs. long distance
- **quality of input/output**
  - wired vs. wireless

[Mustillo]

Vocabulary

- **small vocabulary**
  - 100 words or less
- **medium vocabulary**
  - ~ 100 to ~ 1,000 words
- **large vocabulary**
  - currently 1,000 words or more
  - ideally, this should be unlimited
    - very challenging
    - may include proper names, words from other languages

[Mustillo]

Vocabulary

- **SIR systems generally support limited vocabularies of up to 100 words**
  - Many are designed to recognize only the digits 0 to 9, plus words like “yes”, “no”, and “oh”
- **some SIR systems support much larger vocabularies**
  - Nortel’s Flexible Vocabulary Recognition (FVR) technology
- **constraints for vocabulary size in SIR systems**
  - amount of computation required to search through a vocabulary list
  - probability of including words that are acoustically similar
  - need to account for variation among speakers

[Mustillo]

Usage of Speech Recognition

- **user knows what to say**
  - person's name, city name, etc.
  - habitable vocabulary
- **user's eyes and hands are busy**
  - driving, dictating while performing a task
- **user is visually impaired or physically challenged**
  - voice control of a wheelchair
- **touch-tone (i.e. dial pad) entry is clumsy to use**
  - airline reservations
- **user needs to input or retrieve information infrequently**
  - not recommended for taking dictation or operating a computer

[Mustillo]

Usage of SR (cont.)

- **suitable usage of SR**
  - vocabulary size is small
  - usage is localized
  - large number of speech samples have been gathered
    - in the case of SIR/FVR
  - dialog is constrained
- **background noise is minimized or controlled**
  - more difficult with cellular telephone environments
  - high-quality microphone helps

[Mustillo]

Speech Applications

- command and control
- data entry
- dictation
- telecommunications

[Mustillo]

Command and Control

- control of machinery on shop floors

[Mustillo]

Data Entry

- order entry
- appointments

[Mustillo]

Dictation

- **examples**
  - Dragon Systems
    - true continuos speech, up 160 words/minutes
    - very high accuracy (95-99%)
    - can be used with word processing applications lik Microsoft Office
    - large vocabulary (42K words)
    - different price points
  - IBM ViaVoice
    - Continuous speech software for editing and formatting text documents

[Mustillo]

Telecommunications

- Seat Reservations (United Airlines/SpeechWorks)
- Yellow Pages (Tele-Direct/Philips; BellSouth/SpeechWorks)
- Auto Attendant (Parlance, PureSpeech)
- Automated Mortgage Broker (Unisys)
- Directory Assistance (Bell Canada/Nortel)
  - ADAS+ (411)
- Stock Broker (Charles Schwab/Nuance; E*Trade/ SpeechWorks)
- Banking/Financial Services (SpeechWorks)
  - simple transactions
- Voice-Activated Dialing (Brite “VoiceSelect,” Intellivoice “EasyDial”)
- Google Voice iPhone app

[Mustillo]

Newer Applications

- **voice-based Web browsing**
  - Conversá/Microsoft Explorer
- **intelligent voice assistant (Personal Agent)**
  - Apple Siri
  - Wildfire, Portico, ...

[Mustillo]

SR Demos

- http://www.intellivoice.com
- http://www.speechworks.com
- http://www.nuance.com

[Mustillo]

Human Factors and Speech

- speech characteristics
- variability
- auditory lists
- confirmation strategies
- user assistance

[Mustillo]

Speech Characteristics

- **speech is slow**
  - listening is much slower than reading visually
  - typical speaking rates are in the range of 175 to 225 words per minute
  - people can easily read 350-500 words per minute
  - has implications for text-to-speech (TTS) synthesis and playback
- **speech is serial**
  - a voice stream conveys only one word at a time
  - unless recorded, it disappears immediately
- **speech is public**
  - it is spoken (articulated), and can be perceived by anybody within hearing distance
  - the range is quite limited, unless amplified or transmitted

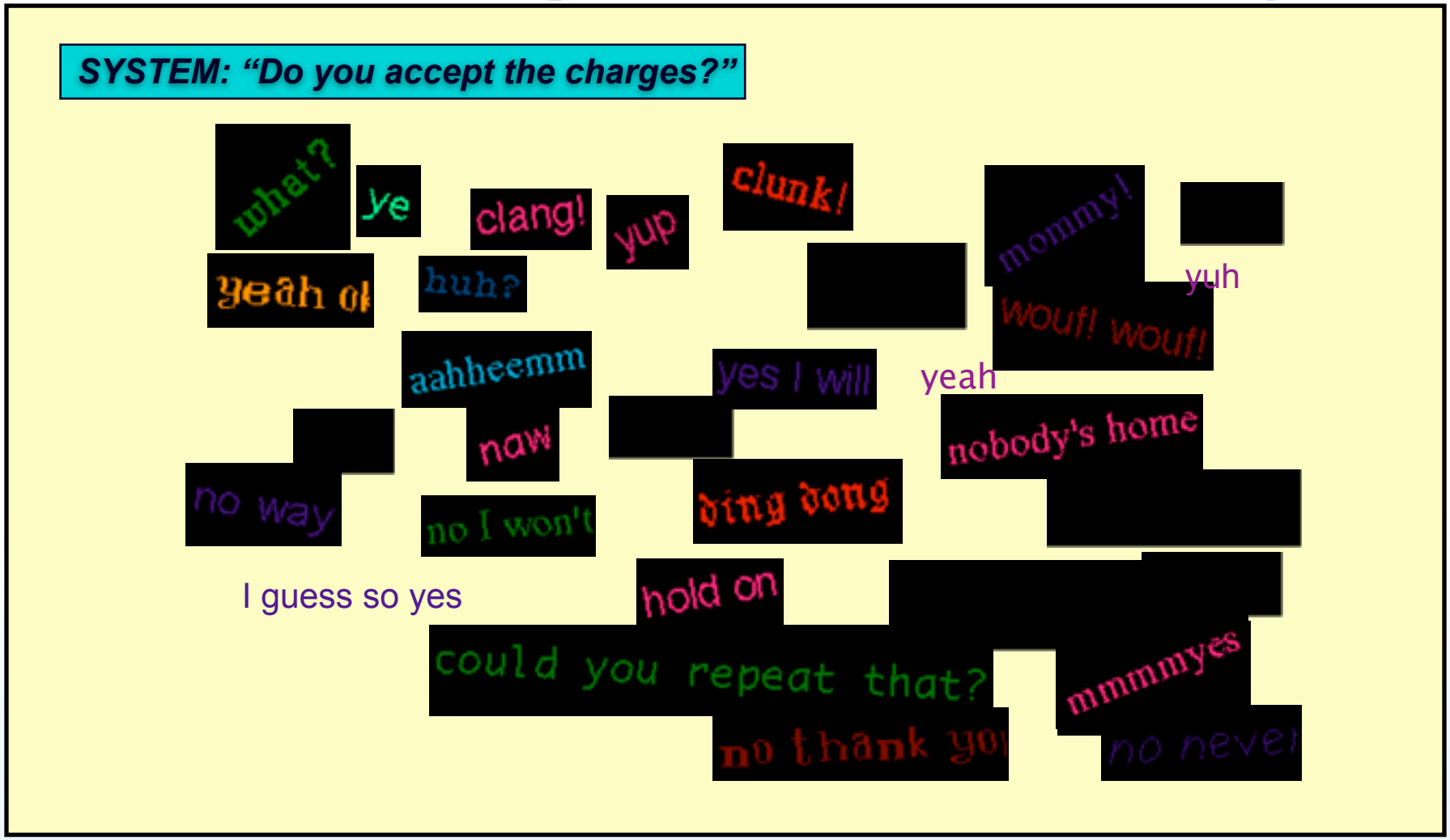
[Mustillo]

Speech Characteristics

- **speech is temporary**
  - acoustic phenomenon consisting of variations in air pressure over time
  - once spoken, speech is gone
  - opposite of GUIs, with dialog boxes that persist until the user clicks on a mouse button
- **recorded speech needs to be stored**
  - the greater the storage, the more time will be required to access and retrieve the desired speech segment
- **speech retrieval is challenging**
  - unless converted to text, it is difficult to search recorded speech

[Mustillo]

User Response Variability



[Mustillo]

Interpretation

- **users are sensitive to the wording of prompts**
  - "You have a collect call from Christine Jones. Will you accept the charges?" "Yeah, I will."
  - "You have a collect call from Christine Jones. Do you accept the charges?" "Yeah, I do."
- **users find hidden ambiguities**
  - "For what name?" "My name is Joe."
  - "For what listing?" "Pizza-Pizza"

[Mustillo]

Auditory Lists

- **specify the options available to the user**
- **variations:**
  - detailed prompt
  - list prompt
  - series of short prompts
  - questions and answers
  - query and enumeration
- **Detailed Prompt**
  - Present one long prompt, listing the items with a short description of each item that can be selected

[Mustillo]

Example: Conference Room Reservation

- **"After the beep, choose one of the following options:**
  - To make a conference room reservation or to reach a specific Admirals Club, say "Admirals Club"
  - For general enrollment and pricing information, say "General Information"
  - To speak with an Admirals Club Customer Service representative, say "Customer Service"
  - For detailed instructions, say "Instructions"
  - <beep>

Auditory Lists - Evaluation

- **Pros:**
  - descriptions help users make a selection
  - implicitly constrain the vocabulary
- **Cons:**
  - lists with multiple items can be challenging
    - require the user to remember the different options (recall)
  - potential mismatch between the user's mental model of the task and the one implied in the structure of the auditory lists
    - "voice response hell"
  - without talk-through, users have to wait until the entire prompt is played before being able to make a selection
  - May invite talk-through since users don't know the end of the prompt

Detailed Prompt

- **present one long prompt, listing the items with a short description of each item that can be selected**
  - example: “After the beep, choose one of the following options:
    - To make a conference room reservation or to reach a specific Admirals Club, say “Admirals Club”
    - For general enrollment and pricing information, say “General Information”
    - To speak with an Admirals Club Customer Service representative, say “Customer Service”
    - For detailed instructions, say “Instructions”” <beep>

[Mustillo]

Detailed Prompt (cont.)

- **pros:**
  - descriptions help users make a selection
- **cons:**
  - without talk-through, users have to wait until the entire prompt is played before being able to make a selection
  - may invite talk-through since users don't know the end of the prompt

[Mustillo]

List Prompt

- **present a simple list without any description of the items that can be selected**
  - example: “Say “General Information”, “Customer Service”, or a specific conference room or Admirals Club city location. For detailed instructions, say “Instructions”.”
  - pros:
    - quick
    - direct
  - cons:
    - users have to know what to say
    - list categories and words must be encompassing and unambiguous

[Mustillo]

Series of Short Prompts

- **present a series of short prompts with or without item descriptions**
  - example: “Choose one of the following options:
    - To make a conference room reservation or to reach a specific Admirals Club, say “Admirals Club” <-
  - For general enrollment and pricing information, say “General Information” <-
  - For detailed instructions, say “Instructions”” <-
  - pros:
    - easy to understand
  - cons:
    - may invite talk-through
    - users may not know when to speak unless they are cued

[Mustillo]

Questions and Answers

- **present a series of short questions, and move users to different decision tree branches based on the answers**
  - example: “Answer the following questions with a yes or no:
    - Do you wish to make a conference room reservation or call an Admiral’s Club location? <-
      - Do you wish to hear general enrollment and pricing information? <-
    - Do you want detailed instructions on how to use this system?” <-
  - pros:
    - easy to understand, accurate
    - requires only Yes/No recognition
  - cons:
    - slow, tedious

[Mustillo]

Query + Simple Enumeration

- **query the user, and then explicitly list the set of choices available**
  - example: “What would you like to request? <-
    - Say one of the following: “General Information”, “Customer Service”, “Admirals Club Locations”, or “Instructions””
  - pros:
    - explicit
    - direct
    - accurate
  - cons:
    - users have to know what to say
    - list categories and words must be encompassing and unambiguous

[Mustillo]

Confirmation Strategies

- explicit confirmation
- implicit confirmation

[Mustillo]

Explicit Confirmation

- **confirmation that an uttered request has been recognized**
  - <Name X>. Is this correct? or, Did you say <Name X>?
- **usage**
  - when the application requires it
    - or when the customer demands it
  - when executing destructive sequences
    - e.g., remove, delete
  - when critical information is being passed
    - e.g., credit card information

[Mustillo]

Explicit Confirmation (cont.)

- **benefits**
  - guarantee that the user does not get receive the wrong information, or get transferred to the wrong place
  - give users a clear way out of a bad situation, and a way to undo their last interaction
  - since users are not forced to hang up following a mis-recognition, they can try again
  - clear, unambiguous, and leave the user in control
  - responses to explicit confirmations are easily interpreted
- **drawbacks**
  - very slow and awkward
    - requires responses and user feedback with each interaction

[Mustillo]

Implicit Confirmation

- application tells the user what it is about to do, pauses, and then proceeds to perform the requested action
  - e.g., User: “<Name X>” System: “Calling <Name X>”
- faster and more natural than explicit confirmation
- more prone to error
  - particularly if recognition accuracy is poor
- users frequently hang up after a misrecognition
- from a human factors perspective, implicit confirmations violate some of the basic axioms of interface design
  - there is no obvious way for the user to exit the immediate situation,
  - there is no obvious way to undo or redo the last interaction
  - the system seems to make a decision for the user

[Mustillo]

User Assistance

- **menu structure and list management**
  - how should menus be structured (i.e., flat, hierarchical)?
  - how should auditory lists be managed in a SUI?
- **acknowledgment**
  - implicit or explicit confirmation
  - what/where are the cost/benefit tradeoffs?
- **beeps/tones**
  - to beep or not to beep?
  - What kind? Is there room for beeps/tones in a SUI?

[Mustillo]

User Assistance (cont.)

- **clarification, explanation, and correction sub-dialogs**
  - what is the best way to handle errors and different levels of usage experience?
- **help**
  - when to provide it, how much to provide, what form to provide it in?
- **context**
  - using accumulated context to interpret the current interaction
- **intent**
  - e.g., “Do you know the time?”

[Mustillo]

Speech User Interface Design (SUI)

- GUI vs. SUI
- SUI principles
- anatomy of SUIs
- types of messages
- SUI design guidelines

[Mustillo]

Speech vs. Vision

- **designing speech user interfaces (SUIs) is different, and in some ways, more challenging than designing graphical user interfaces (GUIs)**
- **speech**
  - slow, sequential, time-sensitive, and unidirectional
  - speech channel is narrow and two-dimensional
  - speech provides alternate means of providing cues
    - prosodic features, shifting focus of discourse, etc.
- **vision**
  - fast, parallel, bi-directional, and three-dimensional
  - visual channel is wide
  - immediate visual feedback is always present

[Mustillo]

GUI Design

- well-defined set of objects
  - e.g., buttons, scroll bars, pop-up, pull-down menus, icons, operations - click, double click, drag, iconify, etc.
- hierarchical composition of objects
  - e.g., placing them together to form windows, forms
- clearly understood goals
  - customizable to the user's needs
  - lead to consistent behavior
- well accepted and widely available guidelines
- well accepted methods of evaluation
- tools for fast prototyping
  - e.g., MOTIF, UIM/X, etc.
- standards that make portability feasible
  - e.g., X-Windows, client-server model

[Mustillo]

SUI Design

- standards are just starting to emerge
- conferences and workshops devoted exclusively to SUI design are slowly becoming more available
- people are starting to get interested in SUIs as core SR technologies mature and prices come down
- customers are starting to demand SR solutions
- guidelines are sparse, and expertise is localized in a few labs and companies
- development tools and speech toolkits are emerging

[Mustillo]

SUI Principles

- **context**
  - users should be fully aware of the task context
  - they should able to formulate an utterance that falls within the current expectation of the system
  - the context should match the users' mental model
- **possibilities**
  - users should know what the available options are, or should be able to ask for them
    - "Computer, what can I say at this point? What are my options?"
- **orientation**
  - users should be aware of where they are, or should be able to query the system
    - "Computer, where am I?"

[Mustillo]

SUI Principles (cont.)

- **navigation**
  - users should be aware of how to move from one place or state to another
    - can be relative to the current place (next, previous), or absolute (main menu, exit)
- **control**
  - users should have control over the system
    - e.g., talk-through, length of prompts, nature of feedback
- **customization**
  - users should be able to customize the system
    - e.g., shortcuts, macros, when and where/ whether error messages are played

CAL POLY

[Mustillo]

SUI Components 1

- **every SUI has a beginning, middle, and an end**
- **greeting message**
  - entry point into the system,
    - identifies the service, and may provide basic information about the scope of the service, as well as some preliminary guidance to its use
  - usually not interactive, but sometimes involves enrollment
- **main body**
  - series of structured prompts and messages
    - guide the user in a stepwise and logical fashion to perform the desired task
      - e.g., make a selection from an auditory list
  - may convey system information
    - but may also require user input

[Mustillo]

SUI Components 2

- **confirmation**
  - users require adequate feedback
    - where they are in the dialog, or what to do in case of an error
    - error messages and prompts, error recovery prompts, and confirmation prompts
- **instructions/help**
  - general as well as context-sensitive help
    - required whenever the user is having difficulty in using the system
  - state the basic capabilities and limits of the system
- **exit message**
  - relates success or failure of the task/query
  - should be polite, may encourage future use
  - not necessary if the caller is transferred to a human operator

[Mustillo]

Types of Messages

- **greeting messages**
  - e.g., “Welcome to…”
- **error messages**
  - identify a system or user error
    - who, what, when, and where of the error
    - the steps to fix the situation
      - e.g., “The system did not understand your response. Please repeat.”
- **completion messages**
  - feedback that a step has completed successfully
    - including what happened and its implications
      - e.g., “Your are now being connected. Please hold.”
- **working messages**
  - inform the user that work is in progress
    - provide a time estimate to completion
    - e.g., “The person you wish to speak with is on the phone. Do you wish to wait? Yes or No?”)

[Mustillo]

SUI Design Guidelines

- **avoid short words and letters of the alphabet**
  - longer utterances are more discriminable and easier to learn to pronounce consistently
- **maximize phonetic distance/discriminability**
  - words with similar sub-parts (e.g., repair/despair) are easily confused
- **avoid numbers, letters, and words that can be easily confused**
  - b,c,d,e,g,p,t,v, z
  - A, 8, H, J, K
  - THIS, HIS, LIST, IS
- **use words that users are familiar with**
  - users are able to pronounce familiar words more consistently than less familiar or unfamiliar words
- **do not use different words to mean the same thing**
- **keep prompts and messages brief and clear**
  - longer prompts and messages tend to be wordy, and require more storage space
    - System: "Do you want services or sales?"
    - User: "Sales"

[Mustillo]

SUI Design Guidelines (cont.)

- **ask questions that correspond to familiar user vocabularies**
  - System: "Please say a company name"
  - User: "Sears"
- **make use of intonation cues**
  - system: "Pour service en français, dites français. For service in English, say English."
  - User: "Français."
- **keep lists in auditory short-term memory limitations**
- **allow for synonyms in prompts**
  - it is natural for people to use a variety of ways to say the same thing
- **provide simple error correction procedures**
- **provide clear and constructive error messages**
- **play error messages as soon as possible after the occurrence of an invalid user input or system error**

[Mustillo]

SUI Design Guidelines (cont.)

- **phrase error messages politely**
  - they should not place fault on the user, or use patronizing language
- **error messages should provide information as to what error has been detected, where the error occurred, and how the user can correct the error**
- **provide prompts rather than error messages in response to missing parameters**
- **keep listeners aware of what is going on**
  - e.g. “Your call is being transferred to <Department X>. Please hold.”
- **provide users with sufficient but brief feedback**
- **use progressive assistance to provide granulated levels of help**
- **establish a common ground between the user and the system**
  - to engage the user in the interaction, the system should let the user know at each step of the interaction that it is recognizing what the user is saying
  - at the same time, the system should confirm what it is recognizing

[Mustillo]

SUI Design Guidelines (cont.)

- **good example of effective error handling (time outs) and disambiguation (AITech auto attendant system)**
- **System: ¨Thank you for calling AITech. What can I do for you?"**
  - User: Silence
  - System: ¨Sorry. I did not hear you. Please tell me who you would like to speak with."
  - User: ¨Well. I'd sure like to talk to Joanne, if she's around. Is she in today?"
  - System: "Sorry, I did not understand. Please just say the name of person you want to speak with."
  - User: "Joanne."
  - System: "Got it. We have more than one Joanne here. Which one do you want?"
  - User: "Umm... Joanne..uh.. Smith."
  - System: "Was that Joanne Smith?"
  - User: "Yes."
  - System: Thanks. Please hold while I check to see if she is available."

[Mustillo]

SUI Design Guidelines (cont.)

- use implicit confirmation to verify commands that involve simple presentation of data
- use explicit confirmation to verify commands that may alter data or trigger future events
- integrate non-speech audio where it supplements user feedback
- ask yes/no questions to get yes/no answers
- give users the ability to interrupt messages or prompts
- give users a way to exit the application
- design for both experienced and novice users
- novice users require auditory menus expert users who are expected to make frequent use of a system, prefer dialogs without prompts
- design according to the users' level of understanding
  - protect novices from complexity, and make things simple for them make complex things possible for expert users

[Mustillo]

SUI Design Guidelines (cont.)

- **structure instructional prompts to present the goal first and the action last - GOAL --> ACTION**
  - e.g. To do function X, say Y, etc.
  - format is preferred because it follows the logical course of cognitive processing, while minimizing user memory load in other words, listeners do not have to remember the command word or key word while they listen to the prompt
- **place variable information first**
  - e.g. “Three messages are in your mailbox.” vs. “Your mailbox contains three messages.”
  - permits more frequent or expert users to extract the critical information right away, and then perform an action based on a specific goal
- **place key information at the end of prompts**
  - e.g. “Is the next digit three?” vs. “Is three the next digit?”
- **provide immediate access to help at any time during a dialog**
- **use affirmative rather than negative wording**
  - e.g. “Say X,” instead of “Do not say Y”
  - affirmative statements are easier to understand
    - tell the user what to do rather than what to avoid
- **use an active rather than a passive voice**
  - e.g. “Say X,” rather than “The service can be reached by saying X”
- **be consistent in grammatical construction**
  - even minor inconsistencies can distract a listener

[Mustillo]

SUI Design Considerations

- **voice behind the prompts**
  - callers pay a lot of attention to the voice
  - they like to hear a clear and pleasant voice
  - the voice can be either male or female, depending on the application and customer requirements
  - voices can be mixed to distinguish different decision tree branches, but be careful with using this strategy
  - male and female voices can be used to distinguish or emphasize critical dialog similar to using color or italics to emphasis a word
- **order of options**
  - menu items should be ordered in a list on the basis of a logical structure
  - if the list has no structure, then items should be ordered according to a ranking of their expected frequency of use
    - determined by a task flow analysis
- **talk-through (barge-in)**
  - use of talk-through affects SUI design

[Mustillo]

Conversational User Interfaces

- natural dialog
- principles
- examples

[Mustillo]

Natural Dialog

- support an interactive dialog between the user and a software application
- more natural than using just speech recognition
- open new channels for communication
  - communication is fundamentally social
  - can enhance approachability
- enhancement to rather than a replacement for current speech recognition

[Mustillo]

Principles

- **research:**
  - interactive speech interface applications
    - MailCall - M. Marx (MIT)
    - NewsTalk - J. Herman (MIT)
    - SpeechActs - N. Yankelovich (Sun)
- **commercial:**
  - first-generation personal agents
  - telecommunications - Wildfire, Webley, General Magic's Portico
  - desktop agents
    - Open Sesame! - Desktop automation
    - Microsoft Bob - Household management
    - Microsoft Office 97 - Active user assistance
  - social metaphors - Peedy the Parrot, animated characters

[Mustillo]

Example: SpeechActs

- **SpeechActs (Sun Microsystems)**
  - Conversational speech system that consists of several over-the-phone applications:
    - access to email
    - access to stock quotes
    - calendar management
    - currency conversion
  - System composition:
    - audio server
    - natural language processor
    - discourse manager
    - text-to-speech manager

SpeechActs:	Welcome to SpeechActs. Would you like mail, calendar …
User:	I'd like mail please.
SpeechActs:	You have 47 old messages and 78 new messages; 14 are high priority. The first new message is from Paul Martin regarding "Sailing this weekend".
User:	Skip that one.
SpeechActs:	Message 2 is from Nicole Yankelovich regarding "Change in itinerary"
User:	Let me hear it.
SpeechActs:	"Ric, can you meet me in Paris on the first Friday after New Year's instead of in Brussels next week?"
User:	Switch to Calendar.
SpeechActs:	Switching to Calendar. Today you have "Lunch with visitors".
User:	What do I have the Friday after New Year's?
SpeechActs:	On Friday, January 5th, you have no appointments.
User:	How about the day before that?
SpeechActs:	On Thursday, January 4th, you have "Remember to call florist".
User:	What does Sandy have on Wednesday?

[Mustillo]

Example: Integrated Messaging

- example: next-generation integrated messaging
  - AGENT: “Good morning, Pardo. While you were away, you received 3 new calls, and have 2 unheard messages.”
  - User: “Who are the messages from?”
  - AGENT: “There’s a voice mail message from your boss about the meeting tomorrow afternoon....’
  - User: “Let me hear it.¨
  - AGENT: “Pardo, the meeting with Radio-Canada has been moved to Wednesday afternoon at 3:00 p.m. in the large conference room. Hope you can make it.”
  - User: “Send Mark an e-mail.”
  - AGENT: “OK. Go ahead.”
  - User: “Mark. No problem. I'll be there.”
  - User: “Play the next message.”
  - AGENT: “....”

[Mustillo]

Principles - Conversational Interfaces

- **principles and guidelines that apply to SUIs apply equally well to the design of conversational UIs**
- **in addition, social cues play an important role in conversational UIs**
  - tone of voice, praise, personality, adaptiveness
- **conversational UIs employ natural dialog techniques:**
  - anaphora - use of a term whose interpretation depends on other elements of the language context
    - e.g. “I left him a message saying that you had stepped out of the office.¨
  - ellipsis - omitted linguistic components that can be recovered from the surrounding context
    - e.g. ¨Do you have a check for $50? Yes, I do. Is the check made out to you. Yes, it is.
  - deixis - use of a term whose interpretation depends on a mapping to the context
    - e.g. “It’s cold in here.”
- **conversational UIs establish a “common ground” between the user and the system**

CAL POLY

[Mustillo]

Natural Language

- NL basics
- language understanding
- complexities of natural language
- recent developments

[Mustillo]

NL Basics

- **natural language is very simple for humans to use, but extraordinarily difficult for machines**
- **words can have more than one meaning**
- **pronouns can refer to many things**
- **what people say is not always what they mean consider the sentence - "The astronomer saw the star."**
  - does "star" in this sentence refer to a celestial body or a famous person?
  - without additional context, it is impossible to decide
- **consider another sentence**
  - "Can you tell me how many widgets were sold during the month of November?"
  - What is the real answer? Yes, or, the number of widgets sold?
- **people constantly perform such re-interpretations of language without thinking about it, but this is very difficult for machines**

[Mustillo]

Language Understanding

- from a systems perspective, understanding natural language requires knowledge about:
- how sentences are constructed grammatically
- how to draw appropriate inferences about the sentences
- how to explain the reasoning behind the sentences

[Mustillo]

Complexities of Natural Language

- **one of the biggest problems in natural language is that it is ambiguous ambiguity may occur at many levels:**
- **lexical ambiguity occurs when words have multiple meanings**
  - example: “The astronomer married a star.”
- **semantic ambiguity occurs when sentences can have multiple interpretations**
  - example: “John saw the boy in the park with a telescope.”
    - Meaning 1: John was looking at the boy through a telescope.
    - Meaning 2: The boy had a telescope with him.
    - Meaning 3: The park had a telescope in it.
- **pragmatic ambiguity occurs when out-of-context statements can lead to wild interpretations**
  - example: “I saw the Grand Canyon flying to New York.”

[Mustillo]

Chapter Summary

- **spoken language as an alternative user interaction method changes many aspects of user interface design**
- **natural language is rich and complex**
  - full of ambiguities, inconsistencies, and incomplete/irregular expressions
- **humans use natural language with little effort**
  - machines (computers) have a considerably more difficult time with it
- **progress continues to be made in the areas of speech technologies and natural language processing**
  - the dream of completely natural, spoken communication with a computer (like HAL or Star Trek) still remains largely unrealized
- **some speech technologies are not mature enough for wide-spread use**
  - continuous, speaker-independent recognition
- **in limited domains and for specific tasks, spoken language is already being used**
  - seat reservation, directory assistance, yellow pages