

Dialog Styles: Pen & Gesture and Speech & Natural Language

John Stasko

Spring 2007

This material has been developed by Georgia Tech HCI faculty, and continues to evolve. Contributors include Gregory Abowd, Al Badre, Jim Foley, Elizabeth Mynatt, Jeff Pierce, Colin Potts, Chris Shaw, John Stasko, and Bruce Walker. Permission is granted to use with acknowledgement for non-profit purposes. Last revision: January 2007.

Agenda

- Pen & gesture
 - PDA overview
 - Pen input styles
 - Issues
- Speech & natural language
 - What is speech?
 - When to use speech
 - Speech output
 - Speech input
 - Designing the speech interaction



Dialog Design

- 1. Command language
- 2. WIMP
- 3. Direct manipulation
- 4. Pen, gesture
- 5. Speech, audio



How to use a PDA

DILBERT / SCOTT ADAMS, scottadams@aol.com



Personal Digital Asst. (PDA)



Apple Newton (1993)



Palm VII Palm IIIc



HP Jornada Handspring Visor



HP Jornada

PDAs

- Becoming more common and widely used
- Smaller display (160x160), (320x240)
- Few buttons, interact through pen
- Estimate: 14 million shipped by 2004
- Improvements
 - Wireless, color, more memory, better CPU, better OS
- Palmtop versus Handheld

Alas, No Shredder...

PEG-NX70V

CLiÉ™

WIRELESS expansion slot

310k pixel built-in CAMERA

MPEG4 VIDEO recording/playback

voice RECORDER

PRODUCT TOUR

screen image simulated

IT'S A COMBINATION PDA, PHONE, PAGER, DIGITAL CAMERA, FAX, E-MAIL, LAP-TOP AND SHREDDER.



6750-Spr '07



7

Input

- Pen is dominant form
- Main techniques
 - Free-form ink
 - Soft keyboards (tapping)
 - Numeric keyboard => text
 - Stroke recognition – strokes not in shape of characters
 - Hand printing/writing recognition
- Sometimes can connect keyboard



Free-form Ink

- Ink is the data, take as is
- Human is responsible for understanding and interpretation
- Like a sketch pad



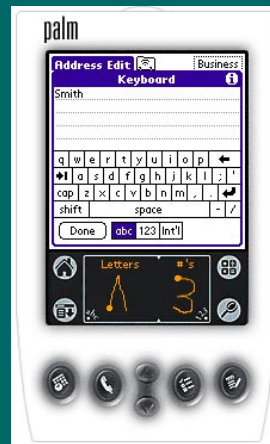
Example

- Digital Ink - CMU
 - video, CHI '98
- Flatland - Xerox PARC
 - video, CHI '99



Soft Keyboards

- Common on PDAs and mobile devices
- Many varieties
 - Tapping interface
 - Stroking interface



6750-Spr '07



11

Tapping Interface

- Presents a small diagram of keyboard
- You click on buttons/keys with pen
- QWERTY vs. alphabetical
 - Tradeoffs?
 - Alternatives?



6750-Spr '07



12

Tegic Communications-T9

- Tapping interface that uses phone pad
- Press out letters of your word, it matches the most likely word, then gives optional choices
- Used in mobile phones
- www.tegic.com/t9



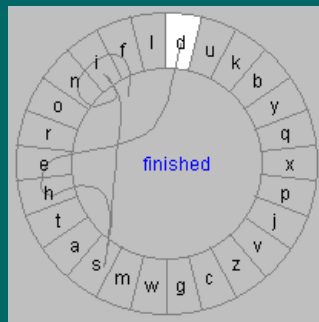
6750-Spr '07



13

Cirrin

- Developed by Jen Mankoff (GT->CMU)
- Word-level unistroke technique



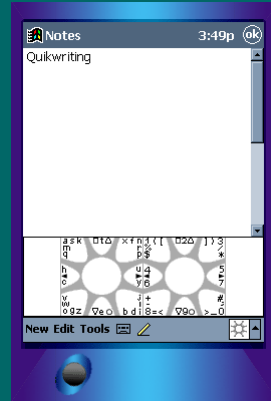
6750-Spr '07



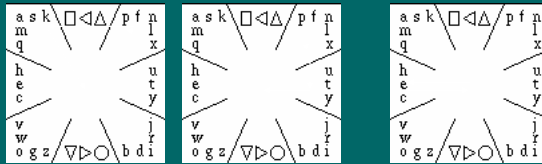
14

Stroke Recognition - Quikwriting

- Developed by Ken Perlin
- UIST '98 paper



Quikwriting Example



p

l

e

Said to be as fast as graffiti, but have to learn more

<http://mrl.nyu.edu/projects/quikwriting/>



Recognition Systems

- Recognizing letters and numbers
- Special symbols



Handwriting Recognition

- Lots of systems (commercial too)
- English, kanji, etc.
- Not perfect, but people aren't either!
 - People - 96% handprinted single characters
 - Computer - >97% is really good
- OCR (Optical Character Recognition)



Recognition Issues

- Off-line vs. On-line
 - Off-line: After all writing is done, speed not an issue, only quality
 - On-line: Must respond in real-time but have richer set of features such as acceleration, velocity, pressure
- Bitmapped vs. Vectorized
 - Bitmapped: Usually off-line, like OCR
 - Vectorized: On-line, uses angle, direction, speed, pressure, acceleration, etc.



More Issues

- Boxed vs. Free-Form input
 - Sometimes encounter boxes on forms
- Printed vs. Cursive
 - Cursive is much more difficult
- Letters vs. Words
 - Cursive is easier to do words



More Issues

- Using context & words can help
 - Usually requires existence of a dictionary
 - Check to see if word exists
 - Consider 1/I/l
- Training - Many systems improve a lot with training data



Special Alphabets

- Graffiti - Unistroke alphabet on Palm PDA
 - Experience?



- Other alphabets or purposes
 - Gestures for commands



Pen Gesture Commands



- Might mean delete

Define a series of (hopefully) simple drawing gestures that mean different commands in a system



Pen Use Modes

- Often, want a mix of free-form drawing and special commands
- How does user switch modes?
 - Might use visible mode switch on screen
 - Might have pen action buttons/switches



Error Correction

- Having to correct errors can slow input tremendously
- Strategies
 - Erase and try again
 - When uncertain system shows list of best guesses
 - ...



Interesting Applications

- Signature verification
- Note-taking
 - Academic course
 - Corporate meeting
- Sketching systems
 - Designers' aids

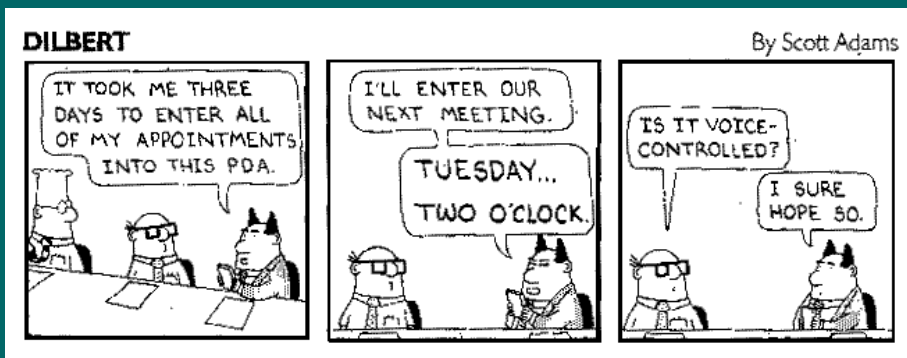


Dialog Design

- 1. Command language
- 2. WIMP
- 3. Direct manipulation
- 4. Pen, gesture
- 5. Speech, audio ←



A Voice Interface



When to Use Speech

- Hands busy
- Mobility required
- Eyes occupied
- Conditions preclude use of keyboard
 - Vibration, cold, water, hygiene, public use
- Visual impairment
- Physical limitation



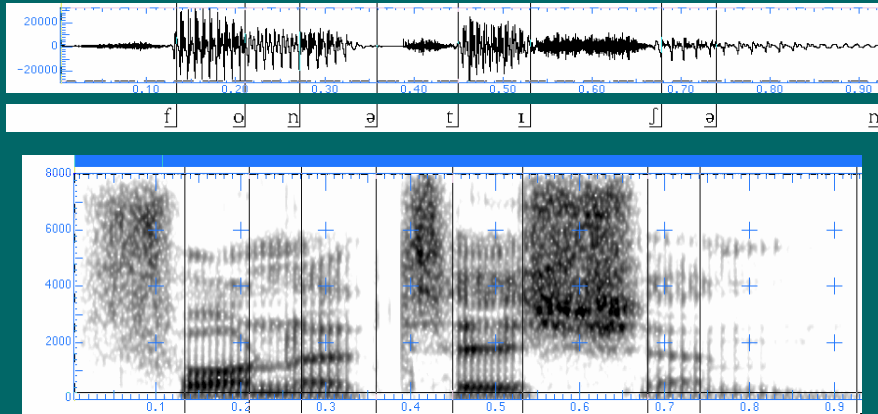
Speech

- What is speech?
 - Vibrations of vocal cords creates sound “ahh”
 - Mouth, throat, tongue, lips shape sound
- English speech
 - 40 phonemes; 24 consonants, 16 vowels
- Sounds transmit “language”



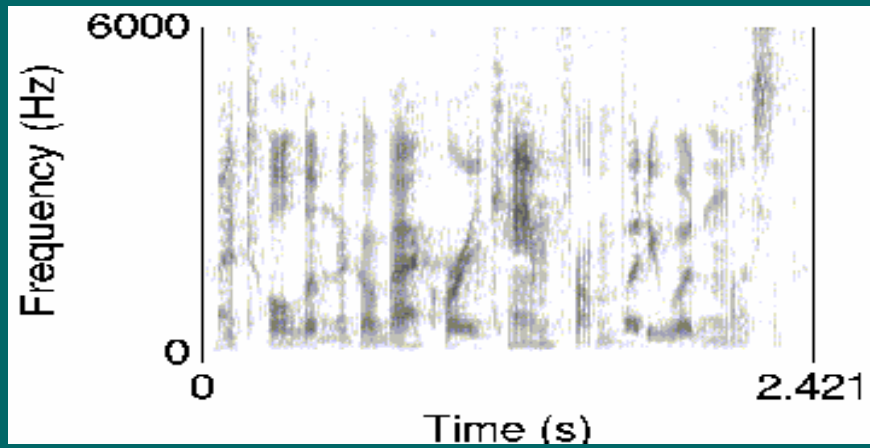
Waveform & Spectrogram

- Speech does not equal written language



Parsing Sentences

"I told him to go back where he came from, but he wouldn't listen."



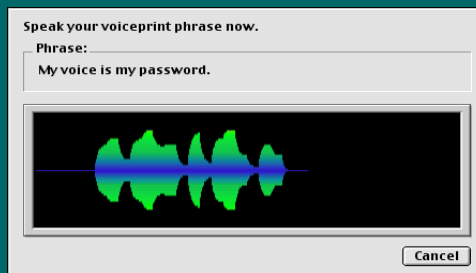
Speech Input

- Speaker recognition
- Speech recognition
- Natural language understanding



Speaker Recognition

- Tell which person it is (voice print)



- Could also be important for monitoring meetings, determining speaker



Speech Recognition

- Primarily identifying words (not meaning)
- Improving all the time
- Commercial systems:
 - IBM ViaVoice, Naturally Speaking, ...



Recognition Dimensions

Did you
vs.
Didja

- Discrete vs. Continuous speech
- Discrete
 - Say one word at a time
- Continuous
 - Sayallthewordsruntogether
 - Computer calculates where one word ends and the next starts - much harder than discrete



Recognition Dimensions

- Speaker dependent/independent
 - Parametric patterns are sensitive to speaker
 - With training (dependent) can get better
- Speaker-independent
 - Are mostly discrete word-oriented
 - Must work with male, female & accented voices
 - Typically used with phone-based systems
 - Banking, Airline reservations
 - Keys to success
 - Limited set of choices at each step
 - “Would you like to make domestic or international reservations?”
 - “Speak your frequent flyer number”
 - Frequent feedback and error-correction opportunities
 - “Did you say 434568432?”



Recognition Dimensions

- Speaker dependent systems require initial training
 - User reads text (several pages) known to system
 - Continues to get better after initial training
 - Partly by learning from mistakes/corrections
 - Partly by training user :)
- Vocabulary
 - Some have 50,000+ words



Recognition Systems

- Typical system has 5 components:
 - Speech capture device - Analog -> digital converter
 - Digital Signal Processor - Gets word boundaries, scales, filters, cuts out extra stuff
 - Preprocessed signal storage - Processed speech buffered for recognition algorithm
 - Reference speech patterns - Stored templates or generative speech models for comparisons
 - Pattern matching algorithm - Goodness of fit from templates/model to user's speech
 - Make heavy use of probabilities and large finite state machines



Errors

- Systems make four types of errors:
 - Substitution - one for another
 - Rejection - detected, but not recognized
 - Insertion - added
 - Deletion - not detected
- Which is more common, dangerous?
- MUST HAVE means for user recovery from system errors!



Natural Language Understanding

- Giving *meaning* to the words
- Input might be spoken or typed
- Holy grail of Artificial Intelligence problems



Eliza - a "Faked" Understanding

- **Person** : Men are all alike.
ELIZA : In what way?
Person : They're always bugging us about something or other.
ELIZA : Can you think of a specific example?
Person : Well, my boyfriend made me come here.
ELIZA : Your boyfriend made you come here?
Person : He says I'm depressed much of the time.
ELIZA : I am sorry to hear that you are depressed.
Person : It's true. I am unhappy.
ELIZA : Do you think coming here will help you not to be unhappy?
Person : I need some help, that much seems certain.
ELIZA : What would it mean to you if you got some help?
Person : Perhaps I could learn to get along with my mother.
ELIZA : Tell me more about your family
- <http://www-ai.ijs.si/eliza/eliza.html>
- Weizenbaum, J., "ELIZA -- A computer program for the study of natural language communication between man and machine", *Communications of the ACM* 9(1):36-45, 1966



NL Factors/Terms

- Syntactic
 - Grammar or structure
- Prosodic
 - Inflection, stress, pitch, timing
- Pragmatic
 - Situated context of utterance, location, time
- Semantic
 - Meaning of words



SR/NLU Advantages

- Easy to learn and remember
- Less transfer problems
- Powerful
- Fast, efficient (not always)
- Little screen real estate
- Enormous potential



SR/NLU Disadvantages

- Doesn't work good enough yet
- Assumes knowledge of problem domain
 - Not prompted, like menus
- Requires confirmation/clarification
- Requires typing skill (if keyboard)
- Enhancements are invisible
- Unrealistic expectations



Speech Output

- Male or female voice?
 - Technical issues (freq. response of phone)
 - User preference (depends on the application)
- Rate of speech
 - Technically up to 550 wpm!
 - Depends on listener (blind: 150-300 wpm)
- Synthesized or Pre-recorded?
 - Synthesized: Better coverage, flexibility
 - Recorded: Better quality, acceptance



Speech Output

- Synthesis
 - Quality depends on software (\$\$)
 - Influence of vocabulary and phrase choices
- Recorded segments
 - Store tones, then put them together
 - The transitions are difficult (e.g., numbers)
- Numbers
 - Record three versions (rise, flat, fall)
 - Logic to determine which version to play



Designing the Interaction

- Constrain vocabulary
 - Limit valid commands
 - Structure questions wisely (Yes/No)
 - Manage the interaction
 - Examples from the airline systems?
- Slow speech rate, but concise phrases
- Design for failsafe error recovery
- Process preview & progress indicator



Speech Tools/Toolkits

Talking Clock

- Java Speech SDK
 - FreeTTS 1.1.1 <http://freetts.sourceforge.net/docs/index.php>
 - *"For 3/4 or 75% of his time, Dr. Walker practices for \$90 a visit on Dr. Dr., next to King Philip X of St. Lameer St. in Nashua NH."*
- Cepstral TTS (probably the best, right now)
- Microsoft Speech SDK
- IBM JavaBeans for speech
- Visual/Real Basic speech SDK
- OS capabilities (speech recognition and synthesis built in to OS) (TextEdit)
- VoiceXML



Notes to Remember

- A natural language interface need not be speech
 - Pen and typing are also natural
- A speech interface need not use natural language (might be more command language-like)
- Wizard of Oz evaluations are particularly useful in this area



HW 3

- Speech interfaces
 - Try out two airline reservation systems that use speech
 - Brief evaluation per assignment
 - A short one
 - Due Tuesday



Upcoming

- Predictive Models
- Cognitive Models

