

Lecture 11-1: Telecommunications

- Telecommunications
 - IVR Design and Speech Recognition
 - Unified Messaging
 - Videotelephony
- Anthropomorphism
 - Studies: Shneiderman, Nass, Boyce
 - Natural Language Understanding

Telecommunications

IVR Design

- Interactive Voice Response (IVR)
 - Telephone user interfaces using touch-tone keypad or command-based speech recognition
 - Applications:
 - Call routing
 - Automated customer service (banking, pay bill, get information, etc.)
 - Voice mail

Bond & Camack's IVR Design Principles

- Bond & Camack's Guidelines
 - Limit the menu tree to 3 levels and menu options for 4 or fewer
 - Eliminate padding, repetition and conversational pleasantries
 - Efficient dialog
 - Organize menus by importance, followed by frequency of use
 - Provide “escape hatches”
 - Always provide a key (“0”) to go to an attendant (“operator”)
 - Avoid language that might be taken as patronizing
- Usability testing
 - Focus on performance, usability
- Focus groups
 - User preferences
 - Usefulness of the application

Some Observations on Bond & Camack

- Any system that replaces a live attendant will be disliked
- A system which provides ability to do new things has the potential of being liked
 - Banking functions over the phone previously unavailable
 - 24-hour service availability
 - Voice messaging, call interception
- It *is quite* possible to do think-aloud studies with IVRs
- It is *not* always possible to provide an operator
 - Business requirements may implement an IVR solely for the purpose of reducing or eliminating use of live agents

Blanchard & Lewis: Other Guidelines

- Action-response order
 - To get your messages, press 1
 - Consistent with human conversational structure (“given-new” information)
 - Testing shows faster responses and fewer errors (Englebeck & Roberts, 1990)
- Interrupt-ability
 - Dial-through
 - Dial-ahead
- All user actions should have feedback
 - Inform users of wait after three seconds
- More ...

Depth vs. Breadth of Menus (1)

- Revisited for IVR
 - Kent Norman’s review of computer-based menus concluded that breadth was better than depth (although both extremes proved to be bad)
 - IVR menus are different
 - limited to 10 keys (two-digit menu responses are awkward)
 - auditory medium imposes more constraints on human memory
 - navigation required to revisit missed options, etc.
- Accepted IVR practice
 - Gould et al. IBM research (1980s)
 - “Speech Filing System” -- early version of voice mail
 - Iterative testing showed that keeping options down to 3 was best
 - Accepted practice and guidelines recommend four options
 - Thus, trade-off favors depth, not breadth

Depth vs. Breadth of Menus (2)

- Virzi & Huitema (1997) HFES Conference Paper
 - Experience showed that splitting top-level options across menus makes system harder, not easier to use
 - Experiment compared broad versus deep version of menu systems
 - Eight option menu given as one menu or split into 2 4-option menus
 - Results
 - Splitting menus slowed selection time for options placed on second menu
 - Conclusion: Breadth is better than depth
 - Aligned with results of visual menu research?
- What might the problems of this study be?
 - What is the cause of the extra time, is it simply navigation?
 - Virzi & Huitema did not logically group submenus
 - Might results be different if grouping were meaningful
 - might faster times be possible or not?

Alternate Approaches to IVR Menus

- Resnick & Virzi (1995) (see Blanchard & Lewis reading)
 - Example: “Rejection Menu Dialog”
 - Twenty-questions approach

System: If you'd like to get messages, please wait, otherwise, press #.

User: #

System: Send a message. Wait to select or press # to reject.

User: #

System: Change your greeting. Wait to select or press # to reject.

Voice Technologies

- Dependent speech recognition
 - Requires individual user training of words
- Independent speech recognition
 - Trained by data collection during development, user does no training
- Hybrid systems
 - Recognition works out of the box with no training, but user training improves performance
 - Recognition models updated during everyday use of system
- Speaker verification
- Synthesized speech
 - Concatenated speech
 - Text-to-speech (TTS)
- Natural Language Understanding

Voice Recognition

- Some applications
 - Input to personal computer
 - Dragon Dictate, IBM Voice ViaVoice, others
 - Telephone-based IVR
 - Hands-free environments
 - Make cell phone calls in car, On-Star controls

What Can User Interface Design Do for ASR Systems?

- Bank transaction example
- Industry stories abound:
 - American Express (Network World 7/90) –
 - Only 3 out of 3,000 callers spoke dollar amounts as required by service
 - 30 sec explanation of rounding amount reduced to 2 sec
 - Now Amex trials all new voice processing technologies before deploying them
 - AT&T Voice Response Operator Service – iterative user testing of prompt designs lowered transaction times by half a second, working out to savings of millions of dollars for AT&T and customers

***Customers respond positively to ASR ... but ...
they expect it to be accurate, efficient, and satisfying***

What Can User Interface Design Do for ASR Systems?

- Fit the task to the user
- Shape human behavior to naturally fit system limitations
 - Accommodate Inaccuracies
 - Humans conversation contains many sophisticated strategies for error detection, correction, and recovery which ASR systems will lack for a long time
 - Must prevent user frustration from error correction (repeated loops of “please repeat”) — *sphexishness*
 - Error feedback and recovery – quick, clear, polite – with backup strategies
 - Fast, efficient transactions and response times
 - Must accommodate both novices and experts
 - Optimal response times exist around natural conversation response time
 - Increase naturalness
 - Conversational prompts and natural responses (words and phrases)
 - But they must discourage extraneous speech or outstrip the vocabulary

Example: Voice Response Operator Service

Old: “Welcome to AT&T. Please say collect, calling card, third number, person-to-person, or operator, now.”

New: “Welcome to AT&T. What type of call would you like to make? (PAUSE) Please say collect, calling card, third number, person-to-person, or operator, now.”

- ***Simulated and real laboratory and field testing***
- ***Takes advantage of barge-in technology***
- ***Pause allows short prompts for experienced users with menu provided for new users***
- ***Word spotting technology mitigates problems with open-ended prompt***
- ***AT&T and customers realized savings in terms of reduced network holding time (during non-billable time) and reduced usage of operator assistance***

Redesigning Voice Mail User Interfaces

- Learn from customer feedback
- Minimize keypresses for frequent operations
- Prompt for three options per menu: Hide complexity
- Use yes/no dialogs where one option is overwhelmingly chosen
- Use prompts that are conversational and personal agent-like

Redesign Voice Mail: Example 1

Minimize the time it takes to do frequent operations

Example: Getting your messages

Old system

Welcome to Audix. For help at any time press *
H.

Please enter your extension followed by the
pound sign. *62819#*

Please enter your password, followed by the
pound sign. *123456#*

One new message. To record messages press
1, to get messages press 2, to administer
personal greetings, press 3. *2*

New message. Call answering message from
Harry Blanchard. Received 8:39 am Friday
October 1st. Ten seconds. Extension
62819. To listen press zero to delete press
* D to skip press the pound key. *0*

(Message plays.)

Redesigned system

AT&T Mailbox. Good
morning.

PIN please. *1234#*

You have one new message.
First message.
(Message plays.)

Redesign Voice Mail: Example 2

Make use of yes/no dialogs when one choice is common

Example: Sending a message

Old system

Record message.

Enter recipient's telephone number

“To send this message, press the # key, to make private, press 1, to make priority, press 2, to schedule delivery, press 3, to file a copy, press 4. ...”

Redesigned system

Enter recipient.

Record message

“Do you want to send it now?
... Press 1 for yes or 2 for no, please.”

Design for Novices and Experts

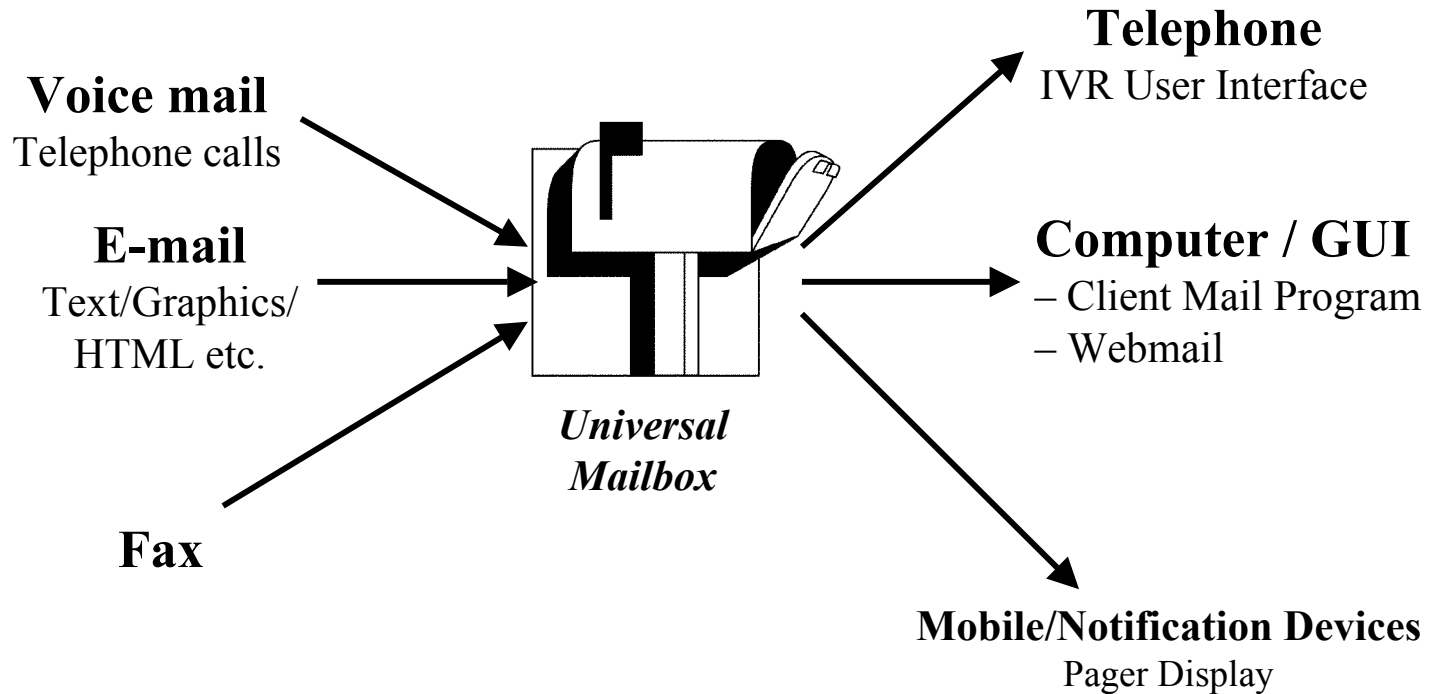
- Three choice menus
- Delayed prompting of more options
- Type ahead active at all times

Example

If you want your messages now, press 1
To send a message, press 2
To change your greeting, press 2
If you need something else, press 0.

*Other options active: 4, 5, 6, 8, 9, * 8, *7, *9
– prompted after 0, time-out, and errors.*

Unified Messaging



Natural Language Systems

- Differ from ASR systems in that users can respond in a free form manner
 - No specific vocabulary
 - No ‘out of vocabulary’ responses
 - Allows system to automate many more activities with ease of use
 - Contrast to menus or lists of keywords
 - Limited number of activities can be offered on a menu
 - Or must have deep, hard to use, hierarchy
- Data from user responses is used to ‘train’ system
 - Machine categorization
- Two distinct user interface techniques:
 - Open-ended prompting
 - Unrestricted responses

Open Ended Prompting

- Does not specify exact responses
 - “How may I help you?”
 - “What can I help you with?”
 - Contrast with
 - “Please say get messages, send a message, ...”
 - “To get your message, press or say one”
- Does not have to be used with natural language technology
 - Build grammars for large majority of response forms
 - Inexact and may reject many reasonable responses
 - Prompt-pause-options format
 - Brems et al. (1995) *Human Factors*
 - Open ended prompt – pause – prompt for specific responses
 - Technology is ASR, restricted response
 - Allows experts to use quicker (voice response operator service)

Open Ended Prompting Advantages

- Advantage – Encourages conversational interaction
 - Natural language system built for free-form response
 - Encourages natural, free form responses
 - User satisfaction does increase with more natural dialogs
 - But – it is a mistake to try to fool user
 - “Welcome. I am an automated assistant. Please speak naturally to me”
 - Boyce (1999)

Open Ended Prompting & “Priming” (1)

- Sheeder & Balogh (2003) *Intl. J. of Speech Tech.*
- Does it help to “prime” user how to respond?
- Place examples of responding in initial query
 - Following query (like prompt-pause-options)
 - Preceding query (actually within query)
- Specific versus “natural” priming
 - Specific keywords
 - Example phrases of natural questions

- Why??
- How natural is this?
- Do people need to be told how to talk naturally?

Open Ended Prompting & “Priming” (2)

- Keyword / Following
 - Welcome to Clarion Wireless Customer Service. How can I help you? < PAUSE > You can ask me about things like “minutes used,” “automatic payments,” and “calling plans”. So, what can I help you with?
- Natural / Following
 - Welcome to Clarion Wireless Customer Service. How can I help you? < PAUSE > You can ask me about things like “how many minutes have I used?,” and “I’d like to set up automatic payments.” So, what can I help you with?
- Natural / Preceding
 - Welcome to Clarion Wireless Customer Service. You can ask me things like “how many minutes have I used?,” and “I’d like to set up automatic payments.” So, how can I help you with your account?

Open Ended Prompting & “Priming” (3)

- Instructed people to ask about
 - Activate Service
 - Copy of Bill
 - Change Plan
- Results
 - Natural / Preceding resulted in improved performance

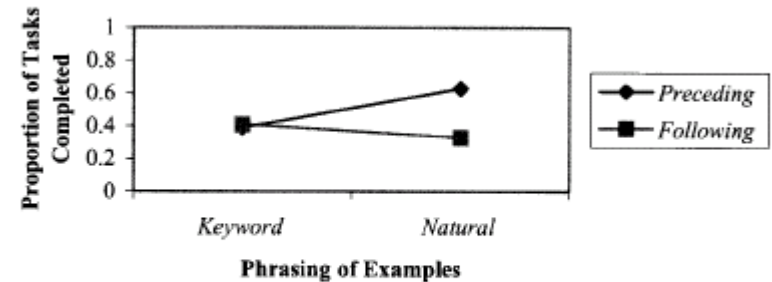


Figure 1. Correct task completion as a function of the phrasing and placement of examples in initial prompts.

- What might explain a ‘need’ for priming responses?
- Why would users not speak naturally
- What problems might there be with this study?
- Could you imagine different results in the field?

Videotelephony

- Videotelephones
 - Picturephone: advanced technology for the 50s/60s, spectacular failure
 - Videophones: 80s/90s, MCI, AT&T, others: also failures
- Video Conferencing
 - Video rooms
 - Video roll-about devices (PictureTel)
 - Thriving business
- Desktop video and CSCW (Computer Supported Cooperative Work)
 - Groupware systems with camera connections
 - Make video calls on computer
 - Internet video calling
- Video walls
 - Continuously open connection in hallway or common room
 - Link between remote sites of same lab or company
 - Encourage casual use of video
 - Device for cohesion among different groups

Core Human Factors Issues for Videotelephony

Components of Usability

- *Familiarity* — the interface should use concepts and procedures similar to ones users know
- *Transparency* — underlying technology should be invisible to the user
- *Simplicity* — operations should be self-evident and easy to remember
- *Predictability* — the interface should be consistent with user expectations
- *Consistency* — similar products and services should work similarly ... "driveability"
- *Attractiveness* — the interface should look good and be fun to use
- *Adaptation of Design to Functionality* — different interface styles are needed as complexity increases

Core Human Factors Issues for Videotelephony

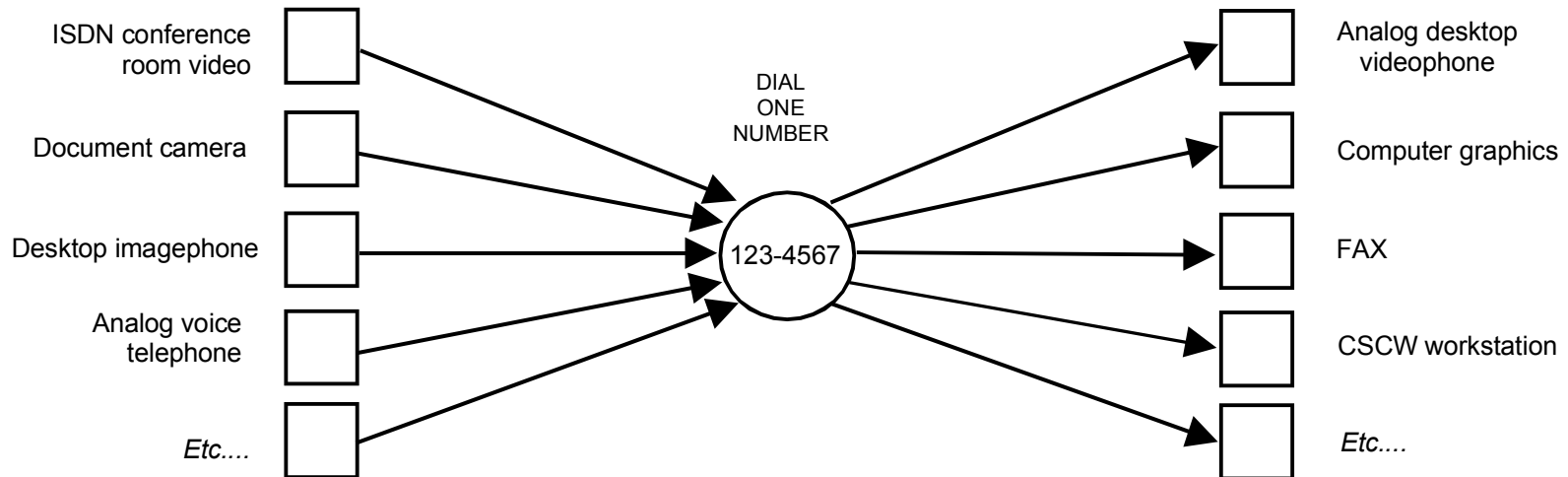
Call Setup and Dialing

Making a video call should be as simple as making a voice call.

- no special prefixes to identify the call as a video call
- only one number to dial, not multiple numbers for voice and video
- no two-stage dialing (dialing an access number then a second number)

Dialing in a multimedia network:

interworking between heterogeneous devices and media should be transparent to the user.



Core Human Factors Issues for Videotelephony

Summary: Usability Issues

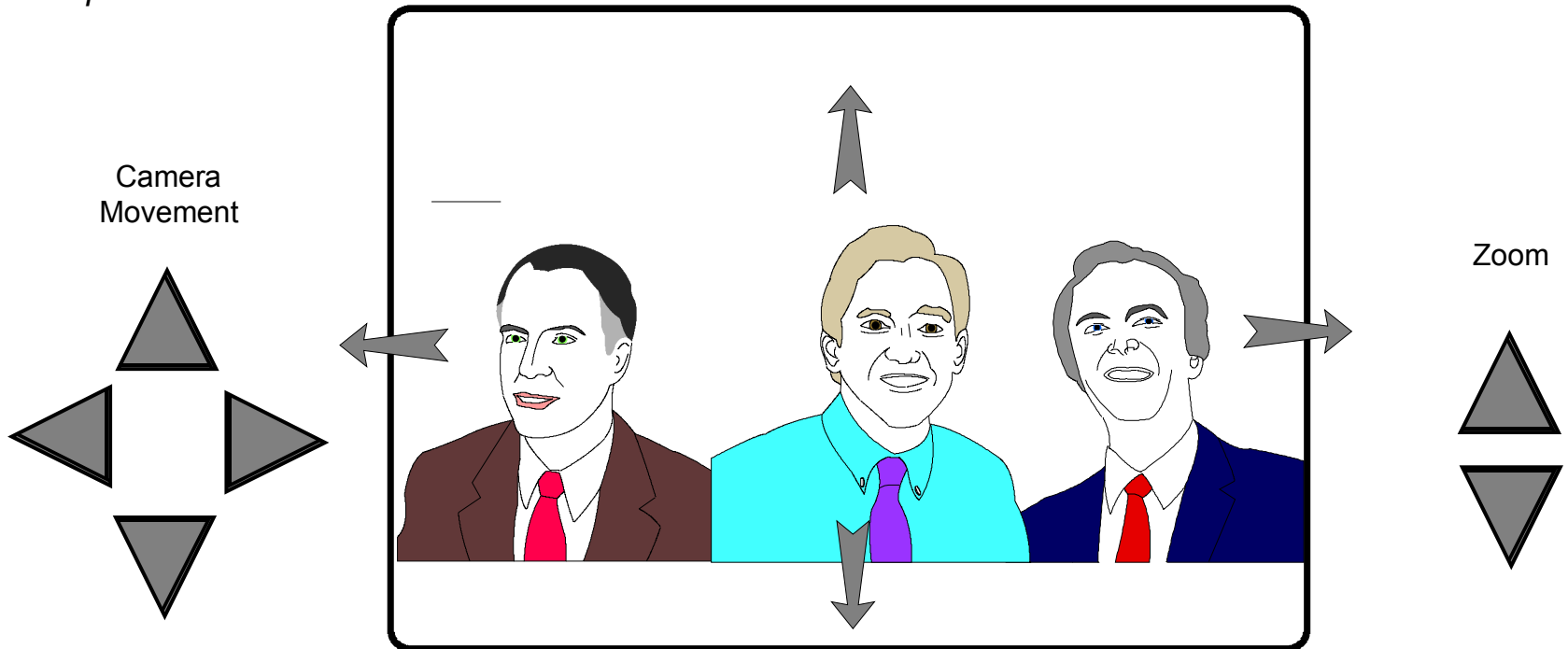
<i>Video Call Setup</i>	<ul style="list-style-type: none">• Simple dialing• Intelligent network identifies communication mode (video, voice, graphics) and terminal equipment
<i>Spontaneity of Communication</i>	<ul style="list-style-type: none">• Prevent call blocking• Portable terminals
<i>Privacy</i>	<ul style="list-style-type: none">• Calls begin voice-only• Users have complete control over transmission and reception of video
<i>Self View</i>	<ul style="list-style-type: none">• Provide intuitive controls to view transmitted and/or local views• Mirrored for personal self view, non-mirrored for document views• No surprises: prevent being seen unintentionally while in self view
<i>Camera Control</i>	<ul style="list-style-type: none">• Controls to allow natural scanning of other party
<i>Delay and Lip Sync</i>	<ul style="list-style-type: none">• Minimize delays• Preserve lip sync
<i>Physical Environment</i>	<ul style="list-style-type: none">• Prevent misattribution of environmental problems (e.g. bad lighting) to video hardware

It is desirable to share and standardize solutions to many of these user interface issues, before many products and services with incompatible interfaces are available.

Core Human Factors Issues for Videotelephony

Remote Camera Control

Which way does the picture move?



ARROW
CONTROLS

Control operation should take the perspective of the user looking at the other party and not the camera's perspective.

Core Human Factors Issues for Videotelephony

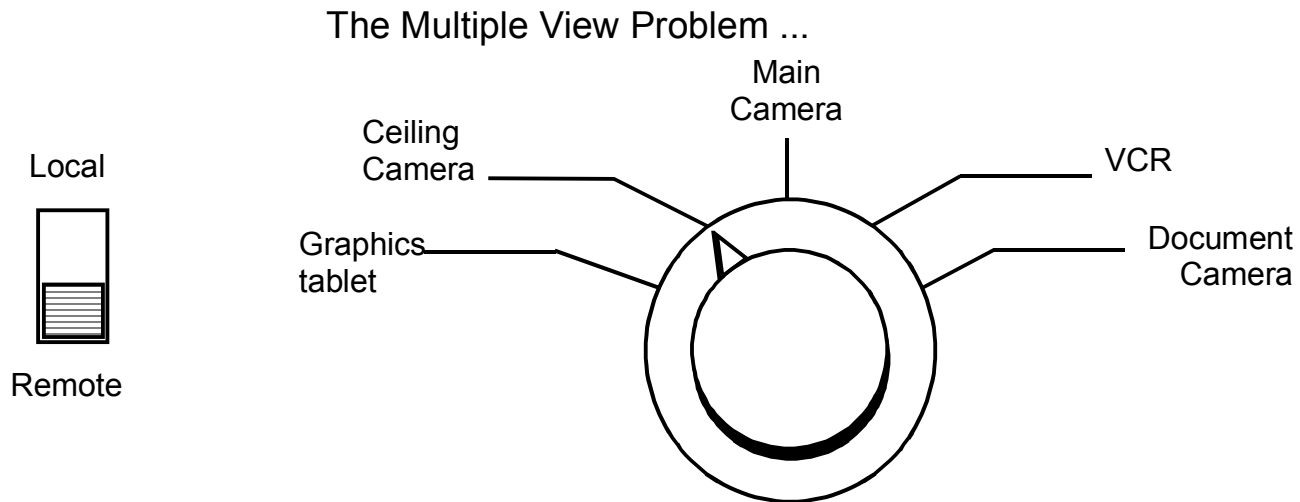
Privacy and Control of Video Modes

- Customers ...
- want no possibility of being seen when they don't want to be seen
 - want control over what is displayed on their own video screen
- Design Decisions
- all calls on a videotelephone begin as voice-only calls
 - full two-way video is established by an action by both parties, *e.g. both must press "video" button*
 - during a video call, both parties have control over whether they can be seen by the other party
- Exceptions ?
- group video / videoconferencing
 - computer-based video / cooperative work systems

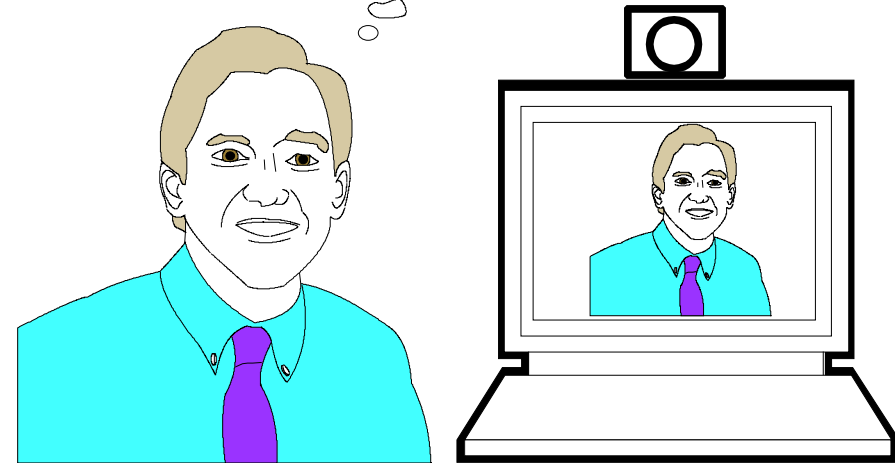
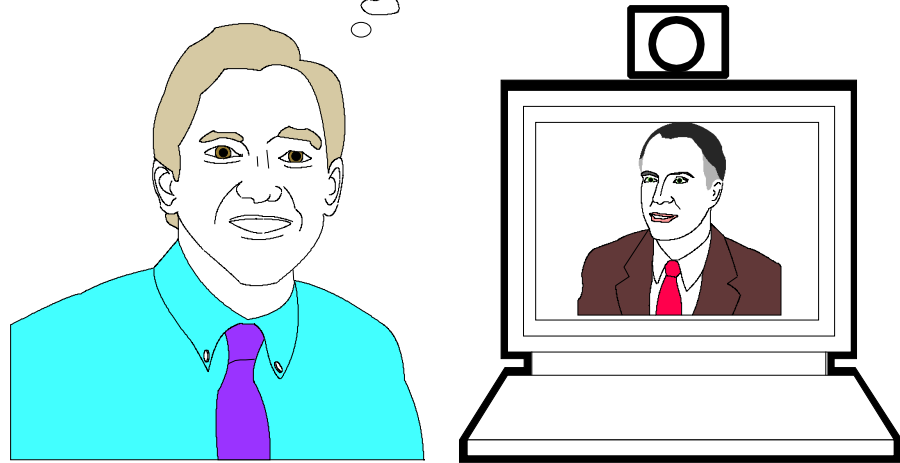
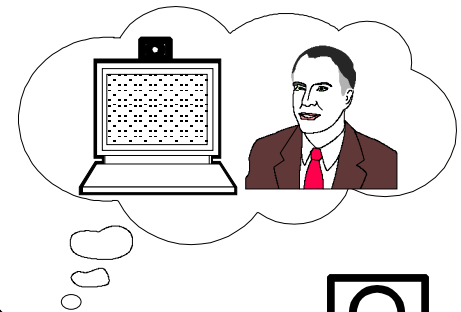
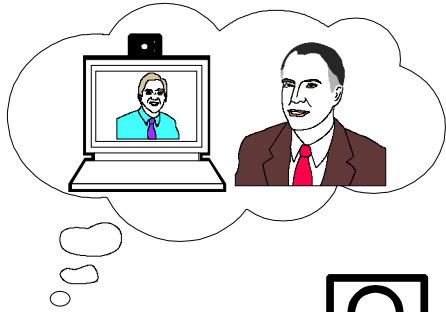
Core Human Factors Issues for Videotelephony

Self View

-
- Local View* View of the signal from a local video source.
- Self View* The local view from the camera pointed at the user, a "mirror".
- Monitor* View of the video signal which is being transmitted.
- Preview* Video of a local video source which is not being transmitted.
-



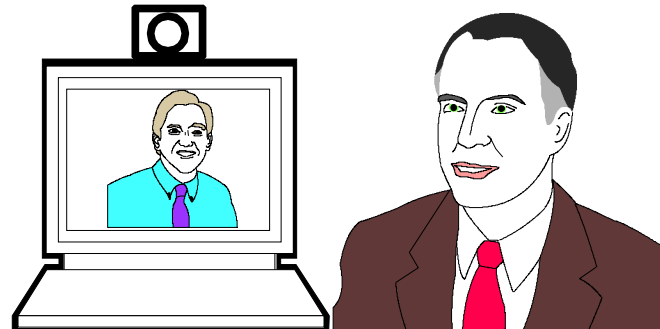
How do you design controls so that people can choose what to see, know what they're seeing, and know what they're sending?



In normal two-way communication, who sees whom is clear.

When choosing self-view, users have the impression they cannot be seen.

What the other party really sees →



... even when the system is designed so the other party can still see them

Anthropromorphism and Natural Language

Anthropomorphism

- Anthropomorphism
 - Designing the user interface of a computer system to look and/or act like a human being
 - May be applied to desktop computers or telephone systems
- Objections to anthropomorphism
 - Shneiderman (1992) “Designing the User Interface”
 - Common guideline for many years to not make systems act, talk like, or refer to themselves as humans
 - Creates unrealistic expectations which computer system cannot maintain, users then become disappointed or confused with system
 - Older research supported this
 - However, this research was on single line, text based dialogs with computer systems which are much less capable than current ones
 - Negative reactions as too cutesy or talkative, wastes time

Current Examples: Failures

- More recent examples of computer-based anthropomorphism
 - Microsoft BOB
 - Abject failure from the moment of first trade magazine reviews
 - Microsoft Office Assistant
 - Longer lived, but the target of many user complaints
 - “How do I turn this off?”
 - Finally cancelled with next release of Microsoft Office
 - See <http://www.microsoft.com/office/>
- What is the basis for the failures?
 - Too cute (badly done design?)
 - Technology not capable of fulfilling human-like capabilities
 - Or, perhaps, technology used where it is not ready
 - Office assistant incapable of accurately identifying what kind of text user is entering
 - Yet the designers insist on interrupting the user at work with suggestions
 - Is it just bad design? Can anthropomorphism be applied with good user interface design?

Current Examples: Success

- Wildfire Assistant: A media success
 - Telephone based call management and messaging system
 - “Virtual, voice-activated personal assistant”
 - Prompting very frankly anthropomorphic
 - Uses personal pronouns (“I”)
 - Refers to itself as a person
 - Displays deliberated engineered personality
 - Young, efficient, and smart, female, executive assistant
- Took the industry by storm in its introduction in the mid 90s
 - Very positive reaction from news media and keen interest from telecommunications executives and community
 - Surprisingly little interest in user interface community outside the telecommunications industry
 - Company is *no longer* business
 - <http://www.wildfire.com/>

Wildfire

From company literature

What is Wildfire®?

Wildfire is a pioneer in personal assistant technology that humanizes communications™ through a simple, intuitive voice interface that gets to know you. “She” provides a single way to manage all of your mobile communications using simply your voice. She makes calls on your behalf, manages your contacts, takes messages, routes calls, screens incoming calls and even takes and sends faxes, all through intuitive voice commands given directly to Wildfire over your phone.

What differentiates Wildfire from other voice-activated technology?

Unlike other systems that have appended speech onto a touchtone interface, Wildfire’s unique interface was designed from the ground up for use with speech recognition. Most importantly, Wildfire has a personality. It’s not just what “she” says but what “she” does, how “she” does it and how “she” interacts with the subscriber. “She” develops relationships with her subscribers, learning from them and for them, like an actual person. And because the recognition is done with software rather than expensive, special-purpose hardware, Wildfire is significantly more cost-effective than any other voice-activated solution available.

Wildfire Example Dialog



Wildfire

Here I am !

Where were we?

You have a new
message from
Bill Jones.

What's it say?

Message Plays

WILDFIRE®

Give him a call

At the number
he left?

Yes

Dialing

Why Might Things Have Changed?

- Why might users be ready for anthropomorphism?
 - Younger users experienced with computer systems
 - Familiar with games and other anthropomorphic software
 - Familiar with the capabilities and limitations of computer systems, and thus less likely to be misled by anthropomorphism
 - Bored with formal and humorless user interfaces
- Why might designers be ready?
 - Well designed, and restrained, anthropomorphism might not engender the supposed drawbacks
 - More recent research does not support older, anti-anthropomorphic studies
- Natural Language Understanding Technology
 - Technology ready for use in limited knowledge domains
 - By its nature, requires conversational responses from user to work
 - Thus, it only seems logical for computer to elicit these responses by using natural, conversational dialog

Anthropomorphism: Recent Research (1)

- Boyce (1999, 2000)
 - Natural language understanding systems for routing operator assisted call and for routing calls to customer care subsystems
 - System dialog based upon
 - Research testing traditional telephone prompting versus causal, conversational dialog
 - Modeling of actual conversational behavior of human operators
- Anthropomorphism study
 - Tested causal versus formal prompts and use of first person pronouns
 - Casual, “I”
 - Casual, no “I”
 - Formal, “I”
 - Formal, no “I”
 - Note that prompting was no where near as ‘extreme’ as Wildfire

Anthropomorphism: Recent Research (2)

- Results
 - Users quite significantly more satisfied with first person prompts
 - Casualness did not make much of a difference, but had higher avg. rating
 - Ratings of satisfaction with speed of the user interface correlated with rated satisfaction (liking) of the user interface not with the actual time it took to perform task

- Clifford Nass
 - Nass & Lee (2000) reading (cited in NYT article)
 - Reeves & Nass (1996) book “The Media Equation”
 - Users treat computers as if they were real people
 - Replications of social psychology experiments substituting computer user interface for a person in person-to-person interactions, subjects show same effects

Anthropomorphism: Recent Research (3)

- Nass & Lee (2000)
 - Text-to-speech system engineered with introvert or extrovert “personality”
 - Users reactions to system matched known social phenomenon:
 - liked system that matched their own personality
 - People automatically apply social rules to computers
 - even though they are aware of what they are doing
 - even when the computer is imperfectly anthropomorphic
- Recent Nass research
 - See Nass web site preprints
 - <http://www.stanford.edu/~nass/>
 - <http://www.stanford.edu/~nass/comm369/index.html>
 - New York Times article