Design of Interactive Multimodal Media Systems

Kellogg Booth (organizer)
Sidney Fels
Brian Fisher
Karon MacLean
Ronald Rensink

Updates for course notes at
www.icics.ubc.ca/hci/multimodal/

SIGGRAPH 2002 THE UNIVERSITY OF BRITISH COLUMBIA MAGIC

Abstract

The course is intended for people involved in the design of interactive media and applications for emerging computer graphics display technologies. It will provide essential background on multimodal perception and explain how to integrate these constraints into the design process.

Current HCI methods are optimized for situations where interaction is perceptually straightforward, but computer graphics increasingly is multimodal. Immersive displays provide a large visual field of dynamic high-resolution information, but require designers to possess knowledge of visual parsing. Haptic (touch) techniques offer benefits of tangibility, but require knowledge of haptic and multimodal ability. Sound is both an independent channel (e.g. system status, speech, music, and background) and an integrated part of a multichannel event (e.g. a collision), but requires knowledge of sound perception.

Participants will learn the theory and practice of multimodal interaction design in a multidisciplinary setting.We demonstrate how traditional HCI methodologies augmented with theories from cognitive science address challenges posed by multimodal interaction using vision, haptics, and sound in conventional and immersive computer graphics environments. Topics include the cognitive science of intersensory processing (vision, hearing, haptics) in scene understanding and interaction, including attention, change blindness, haptics, ventriloquism, and space constancy; enhanced iterative design (Schön reflective practitioner) for integration of visual display design, haptic devices, and sonified and integrated visual/auditory environments including virtual environments and community/performance spaces.

Presenters

Dr. Kellogg Booth, Professor and Director, UBC Computer Science and the Media and Graphics Interdisciplinary Centre, B.S. Mathematics, Caltech; M.A. and Ph.D. Computer Science, University of California, Berkeley.

Dr. Sidney Fels, Assistant Professor, Dept. of Electrical & Computer Engineering, UBC Ph.D. and M.Sc., Computer Science, University of Toronto; B.A.Sc., Electrical Engineering, University of Waterloo.

Dr. Brian Fisher, Associate Director, UBC Media and Graphics Interdisciplinary Centre (MAGIC), Ph.D. Experimental Psychology, Univ. of California at at Santa Cruz, B.A. Biology Hiram College.

Dr. Karon Maclean, Assistant Professor, UBC Department of Computer Science, Ph.D. and M.Sc., Mechanical Engineering, Massachusetts Institute of Technology; B.Sc., Biological Sciences & Mechanical Engineering, Stanford University

Dr. Ronald Rensink, Assistant Professor, Computer Science and Psychology. Ph.D. University of British Columbia, M.Sc . University of British Columbia, B.Sc. University of Waterloo.

Morning Schedule

Welcome and Course Overview: (Booth)

- Multiple perspectives on design
- Adaptation of the design processes to include new theory-based approaches

Intersensory Interactions: (Fisher)

- Integrating cogsci theory with design
- Information integration models
- Dynamics of multimodal processing
- Multimodal events and attentional limitations
- Sensory fusion: Visual capture and visual speech
- Incorporating haptics

Attentional & Nonattentional Processes in Vision (Rensink)

- Relevance to visual display design
- Change blindness: Failure to see unattended stimuli
- Attentional processes - formation of coherent structure
- Nonattentional processes - early vision; scene structure
- Visual perception as a dynamic, "just in time" system
- Lessons for visual display design

Afternoon Schedule

Physical Interaction Design (Maclean)

- Physical interaction design: for haptics and multimodal interfaces:
- Human haptic sensing and motor performance
- Tenets of physical and multimodal interaction design
- How force feedback works
- Rendering haptic and multimodal models
- Areas of basic research
- Computational issues

Closing the Loop in Virtual Environments & Ubiquitous Computing (Fels)

- Novel human interface technologies:
- Overview of human cognitive and physical performance characteristics
- Input technologies: tracking, gloves, various sensors for measuring human activity
- Display technologies: large and small visual displays, haptics, speech, music and non-vocal audio, olfactory and others
- User-centered and non-user-centered approaches design considerations for VEs:
- Designing for intimacy and embodiment
- Survey of research and systems

Intersensory Interactions

Brian Fisher

SIGGRAPH 2002

THE UNIVERSITY OF BRITISH COLUMBIA

MAGIC

Intersensory Interactions

- Intro
- Integrating cogsci theory with design
- Cognitive Architecture: Modularity
  - Information hiding-- conflict resolution
  - Cognitive impenetrability
  - Performance differences between modules
  - Recalibration
- Cognitive Architecture: Spatial indexes
  - Multimodal cue matching within modules
- Intro to haptics

Human-centred design at (and around) UBC

- Institutes and Centres:
  - MAGIC: Media And Graphics Interdisciplinary Centre
  - ICICS: Inst. for Computing, Information, & Cognitive Systems
  - NewMIC: NewMedia Innovation Centre
- Departments and programs:
  - Cognitive Systems, Commerce, Computer Science, Engineering, Psychology, Interdisciplinary Studies
- Collaborations:
  - SFU GRUV, Banff New Media Inst., National Film Board, Nissan, HRL, Boeing, IBM, Sun,ThoughtShare etc.

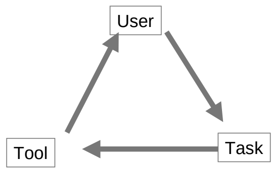
I'm not a ThoughtShare employee, rather I work at UBC MAGIC. MAGIC collaborates with a number of private sector and academic institutions on development projects.

The research I will describe here was done at Simon Fraser university in a group headed by John Dill. I participated in this project as an SFU employee and later at MAGIC

Tool/User/Task Model

- "Classic" HCI
  - Early Cogsci-- GOFSAI
  - Conscious thought- Learning, Memory, Reasoning
  - Sequences of operations
- Task, Protocol & GOMS Analyses
- Built for command-line, menus, workplace systems

Limits of Ergonomic models



What if I am doing this for fun?
What if I want new insights?
What if I want to communicate with someone?

New displays create rich sensory user experience

- WIMP interface
  - – Metaphorical tool icons on desktop
  - – Direct manipulation
- Visualization
  - – Visual analogs of information
  - – Spatial Instruments
- Embodied interaction in immersive environments

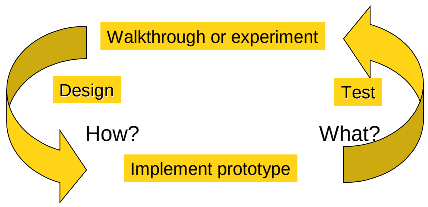
RSEs and UbiComp

- Interconnected devices, distributed applications
- Ubiquitous computing requires HCI to move into the world
- Applications in entertainment, cognition, communication
- Broad user population with acessability, culture, language issues
- Implementation challenges
- Institutional challenges

Opportunities

- Environments: Affordances for exploration
  - Spatial cognition, human space constancy theory
- Support for creative & logical thinking
  - Problem solving, embodied cognition models
- Media-based communication & collaboration
  - Metacognition, distributed cognition
- Experience (Kansai) engineering: Moving beyond usability

Iterative development cycle



Spiral with increasing detail and test specificity

Effective interface design for RSEs

The interaction between display characteristics and the information processing characteristics of the user's perceptual, motor, and cognitive processes will largely determine interface performance

- Cognitive Architecture—The structure of the mind
- Input Representations and Schemata—Structures for information
- Intentions, Goals, and Plans—Semantics of processing

Need for new methodologies

- Low level processes are inaccessible to intuition--Theory based approach is needed
- Intuitions about cognitive processing, e.g. what people think,their goals and plans (“folk Psychology) are reasonably accurate
- Intuitions about processing mechanisms, e.g. how people see, hear, and remember are very inaccurate
- “Metacognitive gap” of technology design

Example of metacognitive gap

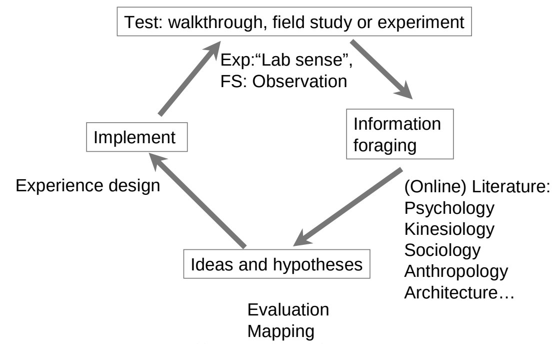
- The “computer metaphor” of mind suggests that a single processor may operate upon a variety of sensory inputs in a variety of tasks.
- Conversely, evidence from functional neurophysiology and sensory psychophysics leads to the conclusion that divisions of processing exist.
- Modules are defined by restriction in the flow of information and control.
- Their processing characteristics are often counterintuitive

Problems in applying theory to HCI

- Convincing designers that there is something to understand—the “metacognitive gap” of folk Psychology
- Convincing Cognitive Scientists to answer relevant questions—Complex data displays and multiple tasks and the Psychophysics reductionist approach
- Integrating research and design—Finding a common language

Fisher: Intersensory Interactions

Reflective HCI Practice (Schön)



Fisher: Intersensory Interactions

Fields of interest: Experimental Psychology:

- Founded~ 100 years ago by Wilhelm Wundt
- Areas of study
  - Psychophysics—Vision, hearing, tactile senses
  - Attention—Endogenous, exogenous, sustained
  - Learning and memory
  - Goal is often information processing algorithms
  - Discussed in Module 2

Useful things from Psychology

- Basic Perception
  - Visual & auditory acuity & discrimination
  - Colour perception
  - Salient display changes, change blindness
- Learning and Memory
  - Primacy, recency
  - Skill acquisition
  - STM limits
  - State-dependent learning
- Decision making
  - Base-rate neglect
  - End effects, biases

Fields of interest: Kinesiology & related disciplines

- Science of human movement
  - Neuroscience
  - Mechanics
  - Anthropometry
- Examples: Fitts' law, GOMS/Keystroke
- Goal is often perceptually guided behaviour
- Discussed in Module 3

Fields of interest: Cognitive Science

- Cogsci Society founded 22 years ago
- Combines Experimental Psychology, AI, Philosophy and Neurophysiology
- 3 levels of analysis
  - Semantics: Intentions, Goals, and Meanings
  - Syntax: Information processing
  - Implementation: Neural processing
- Goal is often Cognitive Architecture

Modularity of processing

"Horizontal" divisions produce stages:

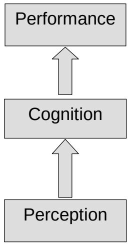
- Response
- Decision
- Perception

Horizontal modularity

Cognitive impenetrability (Pylyshyn, 1984) refers to the inability of observers to use semantic information (such as what the person believes or intends to do) to influence the operation of the input stage.

CA: Horizontal Modularity

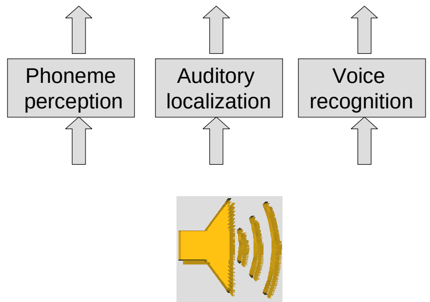
- Model Human Processor (MHP)
- Serial stages of processing
- Information flow is Bottom-up



Vertical modularity

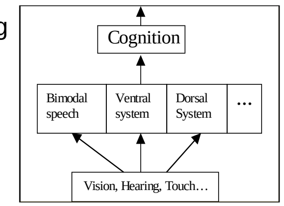
Information encapsulation (Fodor, 1983) refers to structural barriers within the cognitive architecture that prevents internal data stores from being shared between modules in the same stage of processing.

CA: Vertical modularity



One input, many processes

- Intuitions about thinking
  - Fails at low levels
- Cognitive Architecture
  - Multiple brain areas
  - Interconnected
  - Informationally encapsulated
  - Multimodal inputs, parsed from scene and fused



Advantages of modular processing

- Download task to module, reduce cognitive limits
- Fast, effortless information processing
- Information integration between cues and sensory modalities
- Disadvantages?

Information encapsulation

Disadvantage: Modules can't accept information from other modules.

- The same stimuli might give rise to a single multimodal construct in one task, and two unimodal events for another.
- Ventriloquism meets the McGurk effect.
- Vary location of visual and auditory phonemes in a simple teleconferencing-style video display
- Vary information carried by using synthetic speech stimuli (5 levels).

Modularity issues for RSEs

- Stress within a module is not accessable
- Complex stimuli may be processed differently in different modules
- Tasks may access different modules with different performance characeristics
- VEs may introduce discrepencies that impact different modules (and tasks) differently

Stress within a module may be undetected:

- VDT Stress Syndrome example
  - Sampling" raster during saccades overloads saccadic suppression
  - Space constancy perspective allowed us to:
    - Find a better task and measure
    - Make substantive recommendations (>250 Hz)
  - Predict where problems will be greatest (i.e. when larger saccades are made in a high-contrast display)

Example: VDT stress syndrome

Anecdotal reports are backed up by studies:

- Pupillary tremor
- Erratic eye movements (regressive saccades)
- Overall slower (~20%) reading

Theory of space constancy in active vision

- Eye movements create confusing image shifts
- Maintaining space constancy requires
  - Efference copy
  - Passive blur
  - Lateral masking
  - Saccadic suppression

Study Example

- “Sampling” raster during saccades reduces intra-saccadic blur, and may overload saccadic suppression
- Space constancy perspective allowed us to:
  - Isolate the important factors in a complex situation
  - Find a more sensitive task and measure (suppression thresholds for flickering stimuli)
  - Make substantive recommendations (>250 Hz)
  - Predict where problems will be greatest (i.e. when larger saccades are made in a high-contrast display.

Use of mathematical modeling tools allow us to address

- Sensory input from a number of channels simultaneously
- How stimuli from multiple channels are matched and partitioned into mental representations
- How information from multiple senses is integrated to give rise to trans-modal mental events

Modularity of reading

- Name the colour of the text
- Respond as quickly as possible
- Measure response time
- 2 trials

Dog
Cat
Fish
Bird
Cow
Horse
Pig
Cow
Fish

Green
Red
Orange
Red
Blue
Blue
Orange
Green
Red

Cognitive Impenetrability & Modularity

- Stroop effect
  - Reading is data-driven module
  - Competition for response
- Other Modularity phenomena
  - Modularity of perception for action
  - Modularity of visual/auditory integration
  - Modularity of eye movement control
  - Modularity for Models of Minds

SREs are processed differently in different modules

- When a module uses multiple sources of information, it must solve the feature assignment problem.
  - Different modules should have access to a different set of matching cues.
  - When visual and auditory stimuli do NOT fit together, illusory conjunctions can occur:
  - Phoneme perception: The McGurk effect
  - Auditory localization: The ventriloquist effect

Tasks may use modules with different performance characteristics.

- 2 visual systems—“ventral stream” for recognition and “dorsal stream” for action.
- Where vs how
- Different impact of illusions
- Lesion data

Recalibration by pairing (Epstein, 1975)

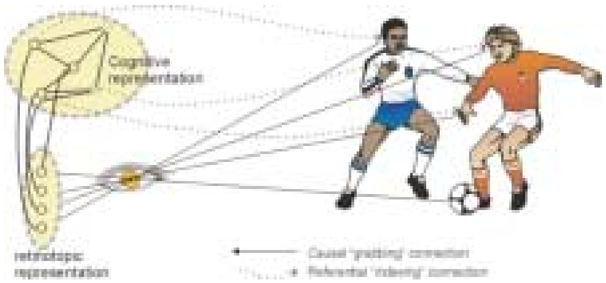
- Sensory modalities calibrate each other: haptics, vision, sound
- Observed actions calibrate visual space (space constancy)
- Vision calibrates hearing for the location of a multimodal event
- Sound calibrates vision for the time of a multimodal event

Effective interface design for RSEs

The interaction between display characteristics and the information processing characteristics of the user's perceptual, motor, and cognitive processes will largely determine interface performance

- Cognitive Architecture—The structure of the mind
- Input Representations and Schemata—Structures for information
- Intentions, Goals, and Plans—Semantics of processing

Indexical cognition

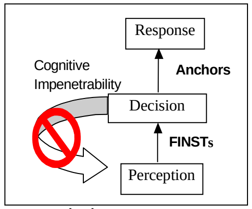


Mental representations of space

- Cognitive architecture perspective requires that links be established between lower level perceptual qualities and cognitive symbols—i.e. a pointer, called a FINST.
- FINSTing allows us to interact with perceptual objects and events without the need for mental images per se.
- symbolic representation + pointers makes different predictions than intuitive picture-in-the-head
- Coping with spatial transformations in complex data spaces:

“FINSTs... make thoughts true”

- Perception
  - “Hotlink” tokens
  - Drawn to salient events
  - Object-centred, “sticky”
  - Visual routines
  - Finite number ~ 4
- Cognition
  - Maintain object history
  - Implicit memory of object associations
  - Sparse cognitive representation
  - Just-in-time delivery of information
  - Atom of intentionality



In recent years there has been a great deal of interest in Indexical, situated, embodied, deictic cognition: Minimal mental model. Leave as much of the representation in the world as possible, retrieve information as needed

Pylyshyn’s FINST hypothesis describes a minimal mechanism that can support this high level of interaction of perception and cognition. At the perceptual level, there is evidence for a small number (~4) of attentional tokens that index perceptual primitives as originating from a given object or event in the world. When information about aspects of an object are needed, they are recalled by reference to the token.

Among other things, FINSTs enable you to track a subset of identical moving targets, subitize a small number of items and perform simple visual routines such a collinearity quickly. They are drawn by new display items and provide potentially parallel access to a small number of them.

At the cognitive level, FINSTs provides the underlying atom of semantics-- the token that enables you to believe something about a specific object or event

Philosopher Jerry Fodor says FINSTs are the things that "make thoughts true”

More about FINSTs

- Role of FINSTs: Link mind and world
  - Visual routines: (collinear, inside, subitizing)
  - History of an object
  - Object-centred, “sticky”
  - Drawn to salient changes-- onsets, luminance increments, oddballs
  - Finite number ~ 4-7
  - FINSTs + ANCHORs for motor behaviour

Multiple object tracking demo

QuickTime™ and a
Animation decompressor
are needed to see this picture.

Inseparability of Mind & World

- Embodied cognition-- mind/body
- Situated cognition-- mind/world
- Distributed cognition-- mind/mind
- Ecological theories (Vygotski, Luria, Bateson, Gibson) can be linked to sensory phenomena and inform interaction design

Implications for interface design

- Interface designs improve performance by “downloading” information processing operations to input modules.
- Interaction of display characteristics with capabilities and characteristics of the functional architecture will determine performance.
- Distortions in location, timing, and category- relevant information may lead to the formation of conflicting representations in different modules.
- Errors and conflicts within a module can create errors and increase cognitive load. (CRT flicker example)
- Cognitive impenetrability of modules makes it difficult for operators to determine the reasons for their poor performance.

Mental representation of space

- Cognitive architecture perspective requires that links be established between lower level perceptual qualities and cognitive symbols—i.e. a pointer, called a FINST.
- FINSTing allows us to interact with perceptual objects and events without the need for mental images per se.
- Symbolic representation + pointers makes different predictions than intuitive picture-in-the-head
- Coping with spatial transformations in complex data spaces

Ventriloquism meets the McGurk effect.

- Vary location of visual and auditory phonemes in a simple teleconferencing-style video display
- Vary information carried by using synthetic speech stimuli (5 levels).
- Subjects report sound location and syllable heard,
- Analyses included testing a variety of mathematical models of information integration by fitting free parameters with STEPIT.

Use of mathematical modeling tools allow us to address

- Sensory input from a number of channels simultaneously
- How stimuli from multiple channels are matched and partitioned into mental representations
- How information from multiple senses is integrated to give rise to trans-modal mental events

Results:

- Visual capture of auditory source location, resulting in a shifting of unimodal auditory location estimation (ventriloquism after-effect).
- No effect of location difference on phoneme perception as measured by statistical or modeling tests.
- No correlation between errors in the two tasks (i.e. subjects could not selectively attend to the auditory phoneme on trials when visual capture failed).
- Overall, modularity of phoneme perception is supported.

Multimodal perception and motor performance

Functional Neurophysiology of Perception for action.

- 2 visual systems—"ventral stream" of vision for pattern recognition and "dorsal stream" for motor performance.
- Stein (1993) proposed that the posterior parietal terminus of the dorsal stream serves as a multimodal motor control system.
- As described by Stein and others, the dorsal stream may be a module that cuts across both sensory channels and stages of processing.
- The disassociation of perception for event classification ("what") and for action ("how") has been tested with visual, but not auditory stimuli.
- To test this, we asked subjects to categorize phonemes and either point to the auditory source or tell us where it is.

2 visual systems lesion evidence

lesion	performance deficits	spared abilities
V1 (blindsight)	detection and identification	pointing
Ventrolateral occipital (DF)	identification, shape recognition, object orientation	object manipulation (orientation matching, grip scaling)
Posterior parietal (RV)	object manipulation (orientation matching, grip scaling)	identification, shape recognition, object orientation

2 visual system illusions

stimuli	deficits	spared abilities
Tichner circles	size report	grip scaling
displacement during saccade	detection of displacement, location report	pointing
Moving or off-centre frame	induced motion, location report	pointing
Sound with displaced visual distractor	pointing	apparent location of sound

Different modules access different matching cues.

- Location
- Synchrony
- Category fit
- When visual and auditory stimuli do NOT fit together, illusory conjunctions can occur:
  - Phoneme perception: The McGurk effect (McGurk, 1964)
    - describes the pre-attentive integration of visual information of speakers' lip movements in observers' report of the auditory phoneme.
  - Auditory localization: The ventriloquist effect (Jack and Thurlow, 1973)
    - describes the strong influence of a visual distractor on auditory source localization

Information encapsulation

- Since modules are cognitively impenetrable and encapsulated they can neither accept information from cognition (expectancies) or other modules.
- If different modules use different matching cues, the same stimuli might give rise to a single multimodal construct in one task, and two unimodal events for another.
- Differences in processing would be expected to occur at the sensory stage, producing d' effects.

- Current research interests:
- Mental representations of space
- Cognitive architecture perspective requires that links be established between lower level perceptual qualities and cognitive symbols—i.e. a pointer, called a FINST.
- FINSTing allows us to interact with perceptual objects and events without the need for mental images per se.
- Symbolic representation + pointers makes different predictions than intuitive picture-in-the-head
- Coping with spatial transformations in complex data spaces:
- Interaction of dynamic data displays and attention

Ventriloquism meets the McGurk effect.

- Vary location of visual and auditory phonemes in a simple teleconferencing-style video display
- Vary information carried by using synthetic speech stimuli (5 levels).
- Subjects report sound location and syllable heard,
- Analyses included testing a variety of mathematical models of information integration by fitting free parameters with STEPIT.

Use of mathematical modeling tools allow us to address:

- Sensory input from a number of channels simultaneously
- How stimuli from multiple channels are matched and partitioned into mental representations
- How information from multiple senses is integrated to give rise to trans- modal mental events
- How errors in visual display geometry, sound location, synchrony of visual and auditory stimuli, and poor resolution affect these operations

Results:

- Visual capture of auditory source location, resulting in a shifting of unimodal auditory location estimation (ventriloquism after-effect).
- No effect of location difference on phoneme perception as measured by statistical or modeling tests.
- No correlation between errors in the two tasks (i.e. subjects could not selectively attend to the auditory phoneme on trials when visual capture failed).
- Overall, modularity of phoneme perception is supported.

Functional Neurophysiology of Perception for action.

- 2 visual systems—“ventral stream” of vision for pattern recognition and “dorsal stream” for motor performance.
- Stein (1993) proposed that the posterior parietal terminus of the dorsal stream serves as a multimodal motor control system.
- As described by Stein and others, the dorsal stream may be a module that cuts across both sensory channels and stages of processing.
- The disassociation of perception for event classification (“what”) and for action (“how”) has been tested with visual, but not auditory stimuli.
- To test this, we asked subjects to categorize phonemes and either point to the auditory source or tell us where it is.

Results.

- Higher level of visual capture in pointing (dorsal stream) than
- location estimation (ventral stream).
- Neither pointing nor location estimation was affected by differences between visual and auditory phonemes.

Conclusions

- Modularity of phoneme perception.
- Modularity of perception for action.
- Dorsal stream is multi-modal.
- Visual capture after-effect is not limited to a single location pair

Implications for interface design

- High-realism interface designs improve performance by “downloading” information processing operations to input modules.
- Interaction of display characteristics with capabilities and characteristics of the functional architecture will determine performance.
- Distortions in location, timing, and category-relevant information may lead to the formation of conflicting representations in different modules.
- Errors and conflicts within a module can create errors and increase cognitive load. (CRT flicker example)
- Cognitive impenetrability of modules makes it difficult for operators to determine the reasons for their poor performance.

VII

O thin men of Haddam,
Why do you imagine golden birds?
Do you not see how the blackbird
Walks around the feet
Of the women about you?

VIII

I know noble accents
And lucid, inescapable rhythms;
But I know, too,
That the blackbird is involved
In what I know.

IX

When the blackbird flew out of sight,
It marked the edge
Of one of many circles.

X

At the sight of blackbirds
Flying in a green light,
Even the bawds of euphony
Would cry out sharply.

XI

He rode over Connecticut
In a glass coach.
Once, a fear pierced him,
In that he mistook
The shadow of his equipage
For blackbirds.

XII

The river is moving.
The blackbird must be flying.

XIII

It was evening all afternoon.
It was snowing
And it was going to snow.
The blackbird sat
In the cedar-limbs.

"Civilization advances by extending the number of important operations which we can perform without thinking about them."

Alfred North Whitehead

Thirteen Ways of Looking at a Blackbird
Wallace Stevens

I
Among twenty snowy mountains,
The only moving thing
Was the eye of a blackbird.

II
I was of three minds,
Like a tree
In which there are three blackbirds.

III
The blackbird whirled in the autumn winds.
It was a small part of the pantomime.

IV
A man and a woman
Are one.
A man and a woman and a blackbird
Are one.

V
I do not know which to prefer,
The beauty of inflections
Or the beauty of innuendoes,
The blackbird whistling
Or just after.

VI
Icicles filled the long window
With barbaric glass.
The shadow of the blackbird
Crossed it, to and for.
The mood
Traced in the shadow
An indecipherable cause.

Alternative approaches

- SOAR (Newell): models learning and cognitive processing
- ACT (Anderson): theory
  - ACT-R implementation
  - ACT-R/PM adds perception/motor performance

Embodied cognition models (Ballard 97)

- GOMS/keystroke task breakdown: ~1/3 second
- Time range of orienting movements of the body
- *Embodiment level*: physical constraints determine the nature of cognitive operations.
- Sequentiality and timing of body movements determine computational processes
- Pointing movements are used to bind objects in the world to cognitive programs.
- Working memory costs are key

Research recap

- GOMS and related approaches
  - Physical systems determine cognitive
- Embodied cognition models
  - Interplay between processing and information availability
- FINSTs and perceptual bottleneck
  - 4 FINSTs, sticky, onsets, history
- Modular/parallel processes

References for Brian Fisher's presentation

Also papers at www.cs.ubc.ca/~fisher

Ballard, D.H., Hayhoe, M.M., & Pook, P.K. Deictic codes for the embodiment of cognition. Behavioural and Brain Sciences (in press).

Bertelson, P. and Radeau, M. (1976). Ventriloquism, sensory interaction, and response bias: Remarks on the paper by Choe, Welch, Gilford, and Juola. Perception and Psychophysics 19(6), 531-535.

Bertelson, P. and Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. Perception and Psychophysics 29(6), 578- 585.

Bridgeman, B.(1981). Cognitive factors in subjective stabilization of the visual world. Acta Psychologica 48, 111-121.

Bridgeman, B. (1992). Conscious vs. unconscious processes: The case of vision. Theory and Psychology 2(1), 73-88.

Bridgeman, B. Kirch, M., and Sperling, A. (1981). Segregation of cognitive and motor aspects of visual function using induced motion. Perception and Psychophysics, 29, 336-342.

Bridgeman, B., Hendry, D., and Stark, L. (1975). Failure to detect displacement of the visual world during saccadic eye movements. Vision Research, 15, 719,722.

Bridgeman, B., Lewis, S., Heit, G., and Nagle, M. (1979). Relation between visual and motor-oriented systems of visual position perception. Journal of Experimental Psychology, 5, 692-700.

Bridgeman, B. van der Heijden, A. H. C. and Velichkovski, B. M. (1991). Visual stability and saccadic eye movements. Behavioural and Brain Sciences

Choe, C. S., Welch, R. B., Gilford, R. M., and Juola, J. F.(1975). The "ventriloquist effect": Visual dominance or response bias? Perception and Psychophysics 18(1), 55-60.

Engleken, E. J. and Stevens, K. W. (1989). Saccadic eye movements in response to visual, auditory, and bimodal stimuli. Aviation, Space, and Environmental Medicine 60. 762-768.

Epstein, W. (1975) Recalibration by pairing: A process of perceptual learning.Perception, 4(1): 59-72.

Fisher, B.D. (1992a). Integration of visual and auditory information in the perception of speech events. (Doctoral Dissertation, University of California at Santa Cruz, 1991). Dissertation Abstracts International, 52, 3324B.

Fisher, B.D. (1992b, June). Cognitive architecture and bimodal integration. Paper presented at the 2nd Annual Meeting of the Canadian Society for Brain Behaviour and Cognitive Science.

Fisher, B.D. and Pylyshyn, Z.W. (1994) The cognitive architecture of bimodal event perception: A commentary and addendum to Radeau. Current Psychology of Cognition 13:1. pp. 92-96.

Fodor, J. A. (1983). The modularity of mind : an essay on faculty psychology Cambridge, Mass.: MIT Press.

Geilen, S. C. A. M., Schmidt, R. A., and van den Heuvel, P. J. M. (1983). On the nature of intersensory facilitation of reaction time. Perception and Psychophysics 34(2), 161-168.

Goodale, M. A. & Milner A. D. (1992). Separate visual pathways for perception and action. Trends in Neurosciences 15(1), 20-25.

Jack, C. E. and Thurlow, W. R. (1973). Effects of degree of visual association and angle of displacement on the "ventriloquism" effect. Perceptual and Motor Skills. 37, 967-969.

Massaro, D. M. (1987). Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry. Hillsdale, NJ.: Lawrence Ehrlbaum Associates.

Mateeff, S. Hohnsbein, J. and Noack, T. (1985). Dynamic visual capture: Apparent auditory motion induced by a moving visual target. Perception 14, 721-727

Pylyshyn , Z. W.(1989). The role of location indexes in spatial perception: A sketch of the FINST spatial index model. Cognition (32), 65-97.

Pylyshyn, Z.W. The role of cognitive architecture in theories of cognition. In K. VanLehn (Ed.), Architectures for Intelligence. Hillsdale: Lawrence Erlbaum Associates Inc. 1991.

Pylyshyn, Z.W. "Here" and "There" in the visual field. In Z.W. Pylyshyn (Ed.), Computational Processes in Human Vision: An Interdisciplinary Perspective. Norwood, N.J.: Ablex Publishing. 1988.

Pylyshyn, Z.W., Elcock, E.W., Marmor, M. and Sander, P. A system for perceptual-motor based reasoning. (Tech. Rep. no. 42) London Ontario: Dept. of Computer Science, University of Western Ontario.1978.

Pylyshyn, Z., Burkell, J. Fisher, B. Sears, C. Schmidt, W. Trick, L.Multiple parallel access in visual attention. Canadian Journal of Experimental Psychology 48:2, 1993, pp.260-283

Radeau,M. and Bertelson, P. (1974) The after-effects of ventriloquism. Quarterly Journal of Experimental Psychology 26, 63-71.

Radeau,M. and Bertelson, P. (1977). Adaptation to auditory-visual discordance and ventriloquism in semirealistic situations.Perception and Psychophysics 22(2), 137-146.

Radeau, M. and Bertelson, P.(1976) The effect of a textured visual field on modality dominance in a ventriloquism situation. Perception and Psychophysics 20(4), 227,235

Rhodes, G. (1987). Auditory attention and the representation of spatial information. Perception and Psychophysics 42(1), 1,14.

Stein, B. (1992). Posterior parietal cortex and egocentric space. Behavioral and Brain Sciences, 15, 691-700.

Treisman, A. and DeShepper, B. Object tokens, attention, and visual memory. In T. Inui and J. McClelland (Eds.) Attention and Performance XVI: Information Integration in Perception and Communication, Cambridge, MA: MIT Press, 199. pp.15-46.

Thurlow, W. R. and Jack, C. E. (1973). Certain determinants of the "ventriloquist effect". Perceptual and Motor Skills, 36, 1171-1184.

Trevarthen, C. B. (1968). Two mechanisms of vision in primates. Psychologische Forschung 31, 299-337

Ullman, S.Visual routines. Cognition, 18, 1984 pp. 97-159.

Ungerleider, L. G. & Mishkin, M. (1982) Two cortical visual systems. in D. J. Ingle, M. A. Goodale, and R. J. W. Mansfield (Eds.), Analysis of Visual Behavior Cambridge: M.I.T. Press.

Warren, W. H., Kim, E. E., and Husney, R.(1987). The way the ball bounces: visual and auditory perception of elasticity and control of the bounce pass. Perception 16, 309-336.

Warren, D. H. Welch, R. B., and McCarthy, T. J. (1981). The role of visual-auditory "compellingness" in the ventriloquism effect: Implications for transitivity among the senses. Perception and Psychophysics 30(6), 557-564.

Attentional and Nonattentional Processes in Human Vision

Ronald A. Rensink
Dept's of Computer Science and Psychology
University of British Columbia

SIGGRAPH 2002 · THE UNIVERSITY OF BRITISH COLUMBIA · MAGIC

These notes contain a superset of the slides presented in my talk for the course "Design of Interactive Multimodal Media Systems" (SIGGRAPH 2002). These slides are extracted and modified from several earlier talks on scene perception, and my presentation "Internal vs External Information in Visual Perception", given at Smart Graphics 2002.

For more information:

Mail: Ronald A. Rensink
Department of Computer Science
University of British Columbia
2366 Main Mall
Vancouver, BC V6T 1Z4
Canada

Email: rensink@cs.ubc.ca

http://www.cs.ubc.ca/~rensink/flicker/ .

Overview

- 1. Change Blindness
- 2. Implications for Human Vision
  - Attentional processes
  - Virtual representation
  - Triadic Architecture; Nonattentional processes
- 3. Implications for Visual Displays
  - Optimized information pickup
  - Invisible transitions
  - Coercive graphics

This course begins with a striking demonstration of a serious limitation on the human visual system, namely, the inability to see large changes that are made at the same time as a disturbance elsewhere in the display. Observers remain "blind" to these, even when the changes are large, anticipated, and repeatedly made. (See Rensink et al., 1997; Rensink, 2000a)

This phenomenon of "change blindness" runs counter to several intuitions about the operation of the human visual system. Among other things, it destroys the idea that vision operates by building up a static, detailed "picture" somewhere in our heads. Instead, it is a much more dynamic "just-in-time" system based on the interplay of attentional processes and nonattentional processes. More precisely, our visual experience is based on coherent representations that are formed whenever attention is allocated to them, and which dissolve as soon as attention is withdrawn.

In light of this, it is suggested that interface design should rely not only on knowledge of the attentional system (which provides visual experience), but also of the nonattentional systems that guide it. Magicians have exploited attentional control for centuries to make objects seem to appear and disappear. A few proposals are put forward here on how we may begin to create a new generation of graphics systems that effectively put a similar kind of magic into visual displays and interfaces.

1. Change Blindness

- 1.1 How Do People See Scenes?

Rensink: Attentional and Nonattentional Processes in Human Vision 3

Our impression as observers is that we see a coherent, detailed image of the world in front of us (at least the part of it in our visual field).

How is this done?

Visual buffer: accumulates information

Rensink: Attentional and Nonattentional Processes in Human Vision 4

High-resolution pickup of information [as well as good color perception] only occurs in the central few degrees of the visual field; the rest (peripheral vision) has very poor spatial resolution. How does this become the basis of the highly-detailed "picture" we experience?

Until recently, it was widely believed that the high-resolution contents of each each fixation were added together—more precisely, the contents were accumulated in a "visual buffer" that maintained the highest-resolution information it received. After several eye movements, the buffer would contain a high-resolution "picture" of the world in front of the observer, and this picture was then the basis of all subsequent perceptual processing.

Rensink: Attentional and Nonattentional Processes in Human Vision 5

The final result—a complete, highly-detailed picture somewhere inside the head.

Question:

How is this accumulation implemented?

Suggestions:

- shift retinal contents (Trehub, 1991)
- data fusion (e.g. Clark & Yuille, 1990)

All attempts to create complete visual descriptions have failed

Rensink: Attentional and Nonattentional Processes in Human Vision 6

Unfortunately, there were several problems with this "solution":

1. There never was found a part of the brain that could convincingly serve as the neural correlate of the internal "picture".
2. There were considerable problems even at a more abstract level. For example, it was found to be extremely difficult to take into account any shifts of the head or body during perception. We can see even when walking around; there are no artifacts of the kind that you might expect with this model.
3. And even if this were somehow solved, it was found to be extremely expensive computationally (both in time and space) to combine the individual "snapshots" in any meaningful way. All "quick and dirty" solutions were found to introduce serious artifacts.

After years and years of trying without solving these problems, people began to wonder…

Looking Again at the Basic Assumptions…

Question:

~~How is detailed visual information accumulated?~~

Question:

Is detailed visual information accumulated?

Back to basics. Perhaps the error was in the initial assumption made about vision. Perhaps our intuitions concerning the accumulation of information were wrong. After all, it initially seemed obvious that the sun went around the earth. Maybe something similar is happening here.

So—how do we know if we are really accumulating information?

Because we accumulate detailed information, it's always easy to see changes...

One simple argument in favor of accumulation is that if there is a change anywhere in our visual field, we will always see it. If we didn't have a complete picture in our heads, wouldn't that mean that we would often miss sudden changes (e.g a cat suddenly running in front of us)?

But is this always true?

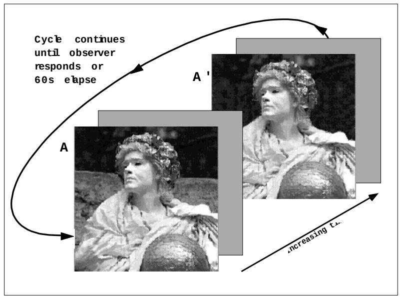
Make change during brief blank interval between original and changed images

But what about if viewing conditions are made slightly different?

What if a brief blank were inserted between the presentation of the original and modified display? Would our ability to detect change be unaltered? If the blank field is brief enough (e.g. 100 ms), there shouldn't be much decay of the visual buffer, and so it should still be easy to see change.

So, another test would be to see what happens under these conditions.

flicker paradigm (Rensink et al., 1997)



Demo

Consider now the situation where, brief blanks are inserted between original image (statue with background wall) and modified image (statue with wall removed). Each image is typically presented for 200-300 ms (a time typicaly of an eye fixation); blank intervals are typically between 80-250 ms. Let this sequence be cycled so that it continues indefinitely. The result is a sequence that looks like an image with a slight amount of flicker; thus the name "flicker paradigm" (Rensink et al., 1997).

Observers are then shown this sequence,told that there is a change between the two images, and then asked to press a button when they see the change.

The question is: will the ability to see change be any different than in "normal" viewing?

This is known as change blindness

also occurs for changes simultaneous with:

- image flicker
- saccades
- eyeblinks
- "splats" not on change
- movie cuts
- real-world interruptions
- gradual change

Proposal: All of these have the same cause

As it turns out, changes are extremely difficult to notice under these conditions. Observers can often go for 20-30 seconds without seeing them, even though the changes can be quite large and very easy to see once noticed.

This inability to see change under these conditions is called “change blindness” (Rensink et al., 1997).

Interestingly, change blindness also occurs for changes made contingent upon other events, eg. moving the eye, blinking the eye, or even when there are visual disturbances elsewhere in the image (disturbances that don’t even cover the image being changed). For a review of the various methods that have been used, see (Rensink, 2002).

One possibility is that these are all due to various specialized mechanisms (e.g. particular to eye movements or eye blinks or visual flicker). However, it seems rather odd that all of these mechanisms would have the same kind of “weak spot”. A more likely possibility is that there is something common to all of these, something that is central to the way that we see.

Proposal: Attention is needed to perceive change in an object.

Under normal circumstances, a change creates a motion transient, which draws attention.

- When change is made same time as other event, transients interfere with drawing of attention, causing change to become “invisible”.

One hypothesis (Rensink et al, 1997) is that attention is needed to see change. The reason we normally see change is that the motion transient that accompanies it draws our attention to the location of the change, allowing us to see it. If this mechanism fails for some reason (e.g. there are transients all around the visual field, so that it is no longer informative), then our attention is no longer automatically drawn to the change. We must then instead “hunt around” with our attention to find the item that is changing. Until we successfully “latch on” to the correct item, we will remain blind to the change.

Implication:

Change blindness shows that
we only integrate what we attend to

And we can't attend to very much…

→ **Never build a complete representation of a scene**.
Our impression that we do is just an illusion.

Athough this explanation solves one problem, it creates several others. For instance, it is known that humans can attend only a few items at a time. If we can't attend to much, then according to this view we can't see much. But that seems to fly in the face of our impression that we <u>can</u> see everything just fine.

Is our impression of seeing everything coherently and in detail just an illusion?

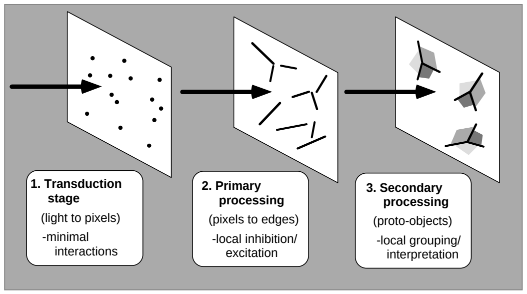
2. Implications for Human Vision

- How Do People See Scenes?

Taking a step back again, let's start with the initial issue of how humans see. Vision is very difficult and involves many complex processes, of which only a small number are understood well. Let's try to see if there's a way to relate these processes to what these new results are telling us.

2.1 Visual Attention

Initial stages of visual processing



Best place to start is at the beginning, i.e., what happens when light first enters the eyes. This is quite well known; it involves processes that occur rapidly (within 200 ms) and in parallel across the visual scene; these processes also appear not to rely on visual attention. At least three stages may be distinguished:

1. Transduction stage. Here, photoreceptors in the eye convert photons to neural signals on a point-by-point basis.
2. Primary processing. Local operations combine the receptor signals into simple edge fragments. These processes are fast, but the fragments formed have no great spatial extent, and no great complexity; they are “image-based”, being simple descriptions of the image, and nothing more.
3. Secondary processing. More sophisticated operations combine edges, luminance information, and some degree of high-level knowledge to form structures (“proto-objects”) that describe the world rather than the incoming image. For example, estimates of surface reflectance (rather than image intensity) can be made at this level. The processing involved in the creation of these proto-objects is not complex; it relies on “quick and dirty” heuristics to get an approximation of the properties at each point of the scene. (See e.g., Rensink & Enns, 1998).

Proto-objects

- bits and pieces of real-world properties
  - e.g., surface slant, true surface color
  - limited extent in space
- obtained rapidly and in parallel
  - ”quick and dirty” interpretation
- obtained **without attention** (preattentive) (e.g., Rensink & Enns, 1998)

Proto-objects are the last frontier of what is well understood about vision at the present time. Indeed, much about proto-objects is still poorly understood.

From various experiments, it appears that proto-objects are fairly complex fragments, but fragments nonetheless—they have a limited extent in space. (By some accounts, the limit is about 2-4 degrees of visual angle.) Interestingly, there appears to be no involvement of attention in the formation of these structures—they are created rapidly and in parallel across much of the visual field, without much apparent effort on the part of the observer.

2.1.1 Coherence theory

1

Without attention, proto-objects are volatile, i.e., have limited coherence in space and time. Thus, they are replaced by new stimuli.

We now need to explain how attention may enter into visual perception. According to coherence theory (Rensink, 2000b), attention influences the proto-objects that have been created. Coherence theory describes this in three different parts, depending on whether attention has yet been allocated.

Part 1 states that before attention is given to them, proto-object are not only limited in their spatial extent, but also in their temporal extent. They are volatile, lasting only a few hundred milliseconds. Proto-objects need to be continually regenerated, staying in existence as long as light continues to enter the eyes, and decaying rapidly after the eyes are closed. They can also be "knocked out" by new proto-objects that appear at their (retinal) location. Thus, if an unattended item is changed, the new representation simply replaces the old.

Thus, any change in an unattended image simply gives rise to a new set of proto-objects. No sense of object continuity has been established; in the absence of attention there is no real memory that could support this (Rensink, 2000b).

Coherence theory— (cont'd):

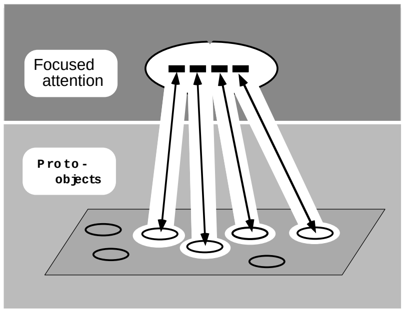
2

Focused attention acts as a metaphorical hand

- "grabs" selected proto-objects and makes them coherent. As such, they maintain an identity, and thus can change

Part 2 states that attention acts by using feedback links to stabilize a small number of selected proto-objects. Once stabilized, these form part of a larger, more extended complex that corresponds to an (attended) objects. In this view, then, attention creates structures with a larger spatial and temporal extent. Given that items can now be perceived as extending through time, any change to a stimulus that is attended will now be perceived as a change of a (persisting) stable structure.

[Note that in this view, attention acts somewhat like a hand, "grabbing" a small number of items from the constantly-regenerating flux of proto-objects, and stabilizing them into a structure with a greater degree of coherence.]



Feedback from higher-level areas allows the stabilization of a small number of proto-objects.

Coherence theory— (cont'd):

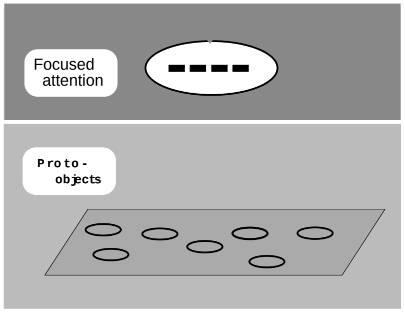
3

Once attention is released, objects "dissolve" back into proto-objects

- There is no buildup of information after attention is withdrawn from items (see also Wolfe, 1999)

Part 3 concerns the fate of items once attention has been withdrawn. Rather than having attention "weld" items into relatively permanent representations, it is posited that the attentional complexes quickly destabilize after attention is withdrawn. Thus, there is no problem of what to do with built-up structure, since there never is much.

[Note that the "hand" metaphor continues to be appropriate: the items that have been "held" are simply let go, reverting to their original status.]



Rensink: Attentional and Nonattentional Processes in Human Vision 21

After attention has been withdrawn from a location (i.e., the feedback links have been broken), the previously-stabilized items reverted back to their original status as volatile proto-objects.

2.1.2 Exploring Attentional Mechanisms

Can use experimental techniques and theories to explore this **attentional** system

Example: Use them to explore:

- **capacity** — how many items are "held" at a time?
- **speed** — how fast are attended objects formed?
- **coding** — what are the "primitives" of attention?
- **guidance** — what attracts visual attention?

Rensink: Attentional and Nonattentional Processes in Human Vision 22

Given the sheer size of the effects encountered change blindness, and given the existence of a framework that might help us make sense of all this, the possibility arises to use these techniques and this framework to explore the nature of visual attention.

Attentional Capacity

Approach: Visual Search for Change

- use images that change back and forth, like the scene examples

- **but** images that are **much simpler**
  - can control the number of items, the type of change, etc.

To illustrate the potential of this approach, examine the issue of attentional capacity, ie. how many items attention can "hold" at a time.

Here, the idea is to use displays similar to the original flicker paradigm, but with much simpler structure. Although this is less natural than viewing real-world scenes, it allows greater control over what the observer is doing. Note the tradeoffs:

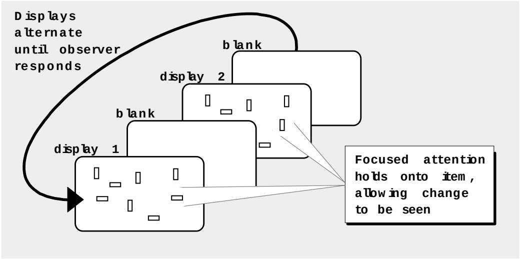
1) Images of real-world scenes: Good for discovering what kinds of strategies an observer would use in real life; bad for determining the characteristics of the mechanisms themselves (get lower bounds only).

2) Images of simple items: Good for discovering the limits of the mechanisms involved; typically bad for determining the strategies used in many real-life tasks. (Although can be useful to study tasks centered around simple visual displays.)

Visual Search for Change *(Rensink, 2000c)*

-on half the trials, one of the items changes (target)

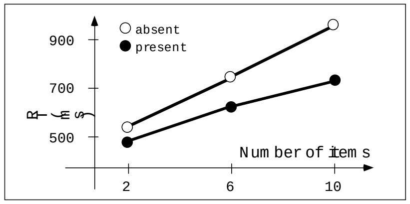
-observer must report if change present or absent



Example of a simplified display. Here, the observer sees a set of flickering items. On half the trials, there is an items changing one of its properties. (Here, the property is orientation.) If coherence theory is correct, to see the change, observers must send their attention to each item in turn, until the target item (ie. the change) is seen.

Measure: Reaction time (RT) vs. set size

RT often a linear function of number of items



search slope = Δ (reaction time) / Δ (# of items)

target-absent slope is 2x target-present slope
-> **serial, self-terminating search**

These kinds of studies typically describe the performance of the observer in terms of the speed of attention. This can be determined by using displays containing different numbers of items, and then measuring the reaction time as a function of number of items.

<u>To determine attentional capacity</u>:

find search speeds for **various display times**
the longer the display, the more items are held

loading will eventually **saturate**
asymptotic value of hold = attentional capacity

Given the speed, the attentional "hold", i.e., the number of items being "held" across each display alternation follows in a straightforward manner.

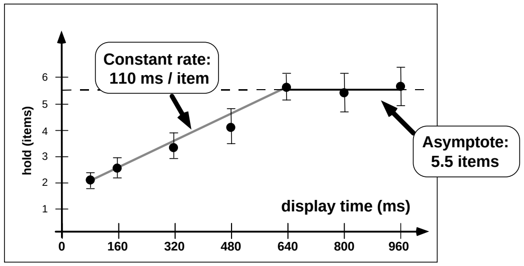
Imagine for example that attention could only hold on to 1 item. If so, then the observer could only check one item at each alternation. Thus, search speed would be the same as alternation rate.

If, however, the observer could hold on to 2 items, then search would go twice as fast. More generally, if the alternation rate is known (it always is) and the search speed has been measured, it is possible to determine the attentional hold. (For details, see Rensink, 2000c)

Since it is possible that more items can be picked up if attention is given more time to act on a display, a critical factor is display time. If the attentional hold becomes independent of display duration, this indicates that it has saturated; the value of the hold at this "saturated" state is the <u>attentional capacity</u>.

Results: Search for **orientation change**

(task: look for horizontal items changing to vertical)



As seen from the graph, hold increases until about 600ms, after which it no longer increases. The asymptotic value is 5.5 items, which is then taken as an estimate of attentional capacity.

Results:

Attention loads up over time
loading rate = 8 items/sec

Attention has a capacity of 5 items
similar to other estimates of attentional capacity

Demonstrates that visual detail is not built up
otherwise, capacity estimate would be unlimited

In summary, this example has illustrated how these techniques can be used to map out various characteristics of visual attention.

In this regard, it is worth pointing out the estimate of attentional capacity (5.5 items) is similar to that obtained by other means. This helps assure us that the "attention" being studied here is the same as the "attention" studied in more classical approaches. (See also Rensink 2000c for more tests of this.)

These experiments also provide a nice demonstration that there really is a severe limit to how many items can be held in memory at any one time.

2.2 Virtual Representation (Rensink, 2000a)

If we only represent a few objects at a time, why do we feel we see all objects at once?

Observation:

- Although objects appear to be present simultaneously, do not all need to be **represented** simultaneously

All that is needed is that the properties of the objects can be **accessed when requested.**

Rensink: Attentional and Nonattentional Processes in Human Vision 29

Back to the main question of this section: Why do we have an impression of seeing everything?

The proposal here is that this impression is an illusion: we don't really represent everything in our field of view at the same time—we just represent what we need when we need it.

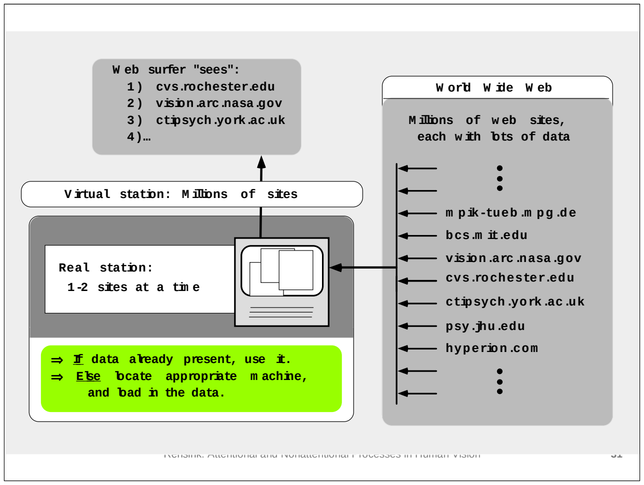
*This is **virtual representation***

If co-ordination is successful, it will appear to higher levels as if representation is "real", i.e., as if all items present simultaneously.

*In such a case, the sparse nature of the object representation is completely **transparent** to higher-level processes.*

Rensink: Attentional and Nonattentional Processes in Human Vision 30

Technically, this strategy is known as virtual representation. If coordination is such that objects can be created whenever needed (and dissolve immediately thereafter), there will be no functional difference between these two forms of representation, at least from the point of view of higher-level processes.



Virtual representation is used in many systems. For example, most computers have virtual memory, in which a limited amount of primary memory (RAM) is made to appear much larger to (higher-level) operations by a coordinated use of secondary memory (disk).

Another example, somewhat simpler and perhaps more illuminating, is the use of virtual representation to make it appear that a work station not only has immense memory, but also contains the information from all machines on the World Wide Web. Although a workstation can contain only a small fraction of this information at any time, it can (usually) obtain any information that is requested, making it appear that it was in the machine all along. Although response times to requests are currently slow enough that it becomes clear that the workstation is accessing other sources, as responses to requests become faster there will likely be an increasing tendency on the part of unsophisticated users to believe that their workstation contains all the information all the time.

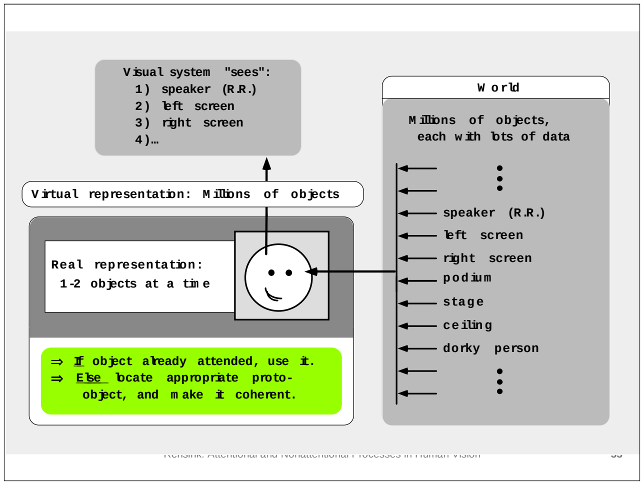
Can this work for the visual system? **Yes!!**

- can always **obtain information from the world**
  - **use the world itself as an external memory**

    (Brooks, 1991)

- to build a coherent representation of an object, focus eyes and attention on appropriate location whenever that object is needed

This strategy should also work for visual perception—although virtual representation is applicable to only a small subset of tasks, perception has the right kind of task structure.



In this view, there are a number of interesting similarities between visual perception and workstation access to networks such as the World Wide Web. For the moment, though, it is enough to point out that virtual representation could account for the apparent contradiction between our limited abilities to represent objects and our impression of seeing everything simultaneously.

[Interestingly, given that our visual systems do use this strategy, it follows that the first use of virtual representation (i.e., that in biological systems) was actually hundreds of millions of years ago…]

Note 1:

- attended representations do not contain a complete description of an item at any instant in time

  (e.g., in airplane example, don't see engine change, even though it is part of attended object [i.e. airplane])

- attention traverses the object hierarchy, holding onto a few details at any time

An important point: Although attending to an item creates a stable representation, this representation does not necessarily contain all the properties of what intuitively appears to be a single objects. Rather, attention is able to hold on to just a few parts and properties.

The apparent detail of an attended object is due to the fact that attention can move up or down a structural hierarchy of an object whenever required. Thus, the impression we have that we see all parts of an object in detail simultaneously is again an illusion, similar to the illusion that we see all objects simultaneously.

Note 2:

- although world is an **external memory**, it is not **an external representation** (as proposed by e.g., Brooks, 1991)
- representations are still needed at early levels for various purposes, e.g.
  - compensating for object occlusion
  - linking together elements in the image that are related in the scene

An important confusion to avoid: Even though the world is constantly supplying the information (via light), there is still a need to do more with it. The world cannot be an "external representation", except in a very minimal sense.

Note 3:

- using the world as an external memory means **perception is not carried out in isolation** in the perceiver
- rather, the perceiver and environment form a **partnership**.
- environment not only an **external memory**; can also be an **external processor**
  - situated cognition (see Clark, 1997).

Note that this view of perception involves a fundamental switch in how observers relate to their surroundings. The older view of perception tacitly had a "lone observer" take in information and then create a representation (visual buffer) which could be maintained largely independent of the environment.

In the view here, such a separation between observer and environment is much less marked—the observer is continually relying on the environment, which for the most part can be trusted to supply the needed information at all times.

Developments in the field of situation cognition (see e.g., Clark, 1997) take this view even further, with the environment acting not only as an external memory, but as an intimate part of many perceptual and cognitive operations (e.g. road signs, calculators, etc.) In a very real sense, then, perception and cognition are extending beyond the physical body of the "lone observer". And this opens up the possibility of extremely natural and effective modes of human-machine interaction.

2.3 Triadic Architecture (Rensink, 2000a)

Question:

How might a virtual representation be implemented in the human visual system?

Need to do this in a way that is compatible with what is known about the visual system.

Rensink: Attentional and Nonattentional Processes in Human Vision 37

Before doing anything more, a reality check: Even if the nature of scene perception is such that virtual representation can be used, is this strategy really used in humans? Is such an approach consistent with what we know about the mechanisms of human vision?

2.3.1. Proposal : Triadic Architecture

- **Nonattentional** extraction of aspects of scene (I):

Gist: abstract meaning of scene (farm, harbor, etc.)

obtained within 150 ms (Biederman, 1981)

obtained without attention (Oliva & Schyns, 1997)

Rensink: Attentional and Nonattentional Processes in Human Vision 38

It is argued here that this is indeed the case. In fact, an architecture will be sketched showing how a virtual representation might be implemented.

Curiously, this architecture is based on an area of perception that traditionally has seemed somewhat baffling: the ability of observers to obtain considerable amounts of information about their surroundings without the involvement of attention. (For references, see Rensink 2000a.)

The first of these is the perception of gist (the abstract category of the scene being viewed). It has been found that observers can classify the type of scene within 100-150 milliseconds, and do so without attention.

[These finding have been difficult to reconcile with the traditional idea of scenes being built up via an attentional “welding” of structures, which should have taken at least one or two seconds.]

Rensink: Attentional and Nonattentional Processes in Human Vision 39

Example of getting the gist of a scene from a brief exposure.

Example of getting the gist of a scene from a brief exposure.

Rensink: Attentional and Nonattentional Processes in Human Vision 41

Example of getting the gist of a scene from a brief exposure.

2.3.1. Proposal : Triadic Architecture

- **Nonattentional** extraction of aspects of scene (I):

- **Gist**: abstract meaning of scene (farm, harbor, etc.)

obtained within 200 ms (Biederman, 1981)

obtained without attention (Oliva & Schyns, 1997)

Possibly derived via statistics of low-level structures (e.g. Swain & Ballard, 1991)

Rensink: Attentional and Nonattentional Processes in Human Vision 42

Recent results in computer vision point to the possibility of scene classification based on simple statistics (e.g. distribution of colors or frequencies in an image). Thus, the existence of such a system (i.e. one not based on coherent structures) is becoming more plausible.

- **Nonattentional** extraction of aspects of scene (II):

  Layout: arrangement of items in the scene.

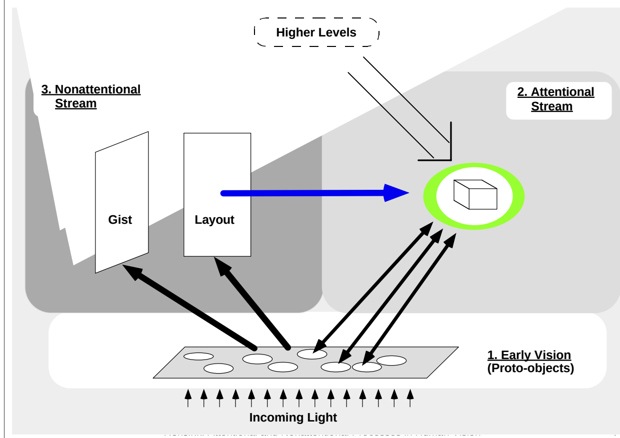
  –nonvolatile (Simons, 1996)

  –can be learned without attention (Chun & Jiang, 1998)

There are other aspects of scene perception that may not require attention.

For example, the layout of items might be picked up (or at least remembered) without the active involvement of attention.

Note that it may be possible to build up layout information to some extent. However, such information is likely to be relatively sparse: a small number of properties about a small number of objects. As such, it is a far cry from the dense accumulation of all visual detail.



All this can be put together in a triadic architecture containing all the components needed for a virtual representation of a scene (Rensink, 2000b).

1. The early vision system creates a constantly-regenerating set of proto-objects. These structures are limited in space and time, but to a considerable degree describe scene-rather than image-based properties.
2. The attentional system can select a number of proto-objects and form them into a structure with a large amount of spatio-temporal coherence.
3. The nonattentional system has two (and possibly more) streams that operate independently of attention to determine various aspects of scene structure.

Operation of this system is relatively straightforward. The early vision system serves as a front end to minimize the influence of extraneous factors such as lighting and object orientation. The nonattentional system then extracts gist (which helps to determine which items to attend) and layout (which helps to determine where these objects likely are). As such, the nonattentional system can help guide the attentional system, so that attention can be sent to the appropriate item, creating the appropriate object at the appropriate time.

2.3.2 Nonattentional Streams

Triadic architecture implies an important role for nonattentional streams in vision

These streams are not primarily concerned with explicit perception of change
(this is done via attentional mechanisms)

-> Mapped out via implicit detection of change?

This architecture emphasizes the importance of nonattentional processes. As opposed to "attento-centric" views that regard all aspects of perception as ultimately being mediated by attention, the view put forward here treats attentional and nonattentional processes as separate processing streams.

According to this view, there is a possibility that nonattentional processes may be able to detect change. However, since attentional processes are needed for the conscious (or *explicit*) perception of change, nonattentional processes would likely support completely nonconscious (or *implicit*) perception.

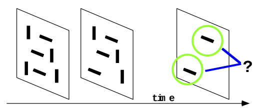
a) Implicit Detection of change: Visuomotor

- Bridgeman et al. (1975) — **oculomotor response**
  - target moves while observer saccades to it
  - eye makes corrective saccade, even though observers have no explicit perception of change
- Goodale et al. (1986) — **manual pointing**
  - target moves while observer saccades to it
  - hand corrects its trajectory while reaching to target, even though observers have no explicit perception of change

Such nonconscious (or *implicit*) detection has been found. Several studies have shown that various motor systems appear to respond to change that the observer does not consciously see.

b) Implicit Detection of change: Perceptual

- Fernandez-Duque & Thornton (2000)
  - observers view 2-display sequence; each display is a simple array of rectangles
  - observers tested on two items: the item changed, and the item diagonally across from it



  - If observer did not notice change, asked to **guess which item changed**.

Rensink: Attentional and Nonattentional Processes in Human Vision 47

Implicit perception of change doesn't only involve motor systems. It can show up in other ways as well.

Results

- Observers could guess **better than chance** (55-63%) even though change was not consciously noticed
  - (a form of blindsight in normal observers)
  - → **involvement of limited-capacity system**
- No attentional priming at location of unnoticed change
  → **involvement of purely nonattentional system**

Rensink: Attentional and Nonattentional Processes in Human Vision 48

These results provide further evidence that:

1. Conscious perception of change occurs via the attentional system.
2. Nonconscious perception of change occurs via the nonattentional system.

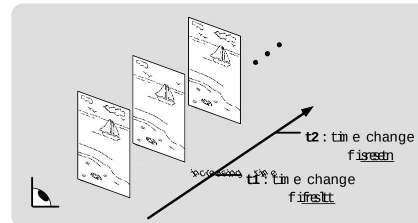
c) Visual Awareness without Visual Experience

Origin - reports by some observers that they "**sensed**" the change long before they saw (= **visually experienced**) it.

A final interesting effect. Here, there is no "visual experience" of change occurring; observers simply have a "feeling" that something is happening.

[Note: This was discovered by observers (in the original flicker experiments) spontaneously asking whether they should hit the button when they "saw" the change, or when they "felt" it.]

- Rensink (2000a)
  - observers view flicker sequence (natural images)
  - asked to hit button (t1) when change was **felt**
  - then hit button (t2) when change was **seen**



The experimental setup here was exactly like the "standard" setup, except that observers were asked to press the response button twice.

Later analysis showed that the "seen" responses were no different from the "seen" responses in the "standard" setup, indicating that asking for two button presses created no interference in the task.

Results

- 1/2 of observers had no feeling of change without visual experience of it
- 1/3 of observers could feel a change before seeing it
  - (t2-t1) > 1 second on 20% of trials
  - average duration = 3.7 seconds
- not a result of guessing:
  - accuracy on catch trials is good (82%)

Curiously, whereas a large fraction(1/3) of observers could get this feeling on occasion, another large fraction (1/2) never did. Later testing failed to find any differences in attentional abilities between the two groups.

Mindsight: Conscious (mental) awareness without an accompanying visual experience

Different than seeing with visual experience
- different sensitivities to types of change

Mindsight due to **a nonattentional system (alert?)**
- basis of the belief in a “sixth sense”???

It still remains to explore this phenomenon more thoroughly—even its general nature is still unknown. However, it does open up the interesting possibility that visual displays of the right type can create not only visual experience, but “gut feelings” (or the “sixth sense”) as well.

3. Implications for Visual Displays

- 3.1 Optimized Information Pickup
  - max. amount of information that can be picked up (consciously) is only 4-5 items
    - only a small amount for each item
  - -what (and how much) is selected depends on the viewer & the task involved
    - different people can literally see the same world very differently

The view of visual processing set out here has several implications for the design of visual displays and visual interfaces. Just a few are described here; hopefully they will give an indication of the possibilities.

The first general implication is that visual designers need to be aware that only a small amount of information is ever consciously picked up by the attentional system at any time. Not only are at most 4-5 items accessed, but only a small part of the information in these items is accessed, too.

This result itself has a number of important consequences. To begin with, since vision is very dynamic, the representations formed are not all-encompassing, "all-purpose" structures. Rather, they depend critically on the abilities and strategies of the individual viewer (these vary with age, culture, etc.), and on the task at hand. As such, each individual experiences the world in a very different way.

- 3.1.1. Displays
  - rendering is often computationally expensive;
    - need to decide what to render
    - eye movements used to find important parts of objects, events (O'Sullivan et al., 2002)

Can use flicker paradigm to find which parts and properties of objects are most important. (= most easily seen to change)

→ can use to provide descriptions of objects using small number of vertices

In terms of visual displays, these limitations can be used in a positive way. A viewer cannot represent an object in all detail (at least ,simultaneously). Given that rendering of objects (and particularly movements of multiple objects) cannot always been done in an completely accurate way within a reasonable amount of time, it is possible to create graphics that will still look reasonably life-like to a viewer, provided that they do not scrutinize it very hard. (But scrutiny is unlikely in many situations, such as interactive animation.)

Moreover, the techniques described earlier can be adapted to determine which aspects of an object are typically picked up during casual viewing, and which are not. For example, if something in an object is changed but this change is not noticed (e.g. changing the engine of an aircraft), then it is likely that the "basic" representation of the airplane that the viewer has does not contain a description of the engine.

- in <u>active graphics</u> (e.g gaze-contingent rendering), important to know correct level of detail to give to non-foveated areas

- coherence theory: nonattended (= foveated) areas have relatively simple descriptions
    - distributions of proto-object properties

Flicker paradigm can determine the properties represented in non-foveated areas
(= changes in average properties of groups)

According to this view of visual processing, it also becomes possible to leave unattended (or non-foveated) areas in a relatively undetailed state; all that is required is that the properties picked up by the nonattentional systems (e.g.,distributions of colors and orientations; overall layout) be approximately correct.

Recent experiments (e.g. Ariely, 2001) show that when viewers see a group of items (as opposed to a single one), they are only sensitive to average properties of the group. This appears to be an interesting line of research to follow.

- 3.1.2. Interfaces

- given limited amount of information that can be attended, important to use items that correspond to attentional “chunks”.

Can use flicker paradigm to find the basic units of visual attention.
(= Properties which have capacity of 4-5)

The techniques and theoretical framework described here can also be applied to provide guidelines for visual interface design. For example, given that attention can only pick up 4-5 items, it becomes important to determine what those “items” are. Optimal information pickup is possible when the items in the display correspond to attentional “chunks”.

- given that there is little visual memory,
  -> information conveyed via changes must not rely on memory, but must use other techniques (Nowell et al., 2001)

Another interesting application of change blindness research is the realization that if changes themselves are to be used to convey information, then precautions should be taken to ensure that the display really does allow the user to see the change (Nowell et al., 2001).

- 3.2 Invisible Transitions
  - make sudden transitions effectively invisible (e.g. those due to change in level of detail)
  - can be done by making transition:
    - during eye blink or saccade
    - during occlusion by object passing by
    - as a slow fade (blend)
  - if transitions cannot be avoided, then make them simultaneously (minimizes effects)

A more direct application of change blindness is to use it to render invisible any (unattended) transitions in the image.

- 3.2.1. Displays
  - can create smooth active displays
    - may not need gaze-contingent rendering
  - can create "supernatural" effects
    - transitions such that events could not occur in the real world. E.g.,
      - sudden (dis)appearances
      - sudden change to objects, regions

Note: Magicians have been doing this for years…

In displays, such transitions can arise for example via the "popping" that can occur when there is a change in the level of detail, or when an active display changes in response to an eye movement or mouse event. One application is to use change blindness to create smoother displays that (since irrelevant transitions are reduced) seem more realistic.

Another—perhaps more interesting—possibility is to make transitions that do not correspond to anything in the world (e.g. a sudden change when an item is occluded). Magicians have been doing these kinds of manipulations for years; this approach would effectively put a similar kind of "magic" into visual displays.

- 3.2.2. Interfaces
  - fewer visible transitions -> less change blindness
  - especially important in animated displays, where it is important to detect change

→ Only one moving information source at a time (any more would create change blindness)

→ Methodologies to evaluate robustness of interface to change blindness?

The work on change blindness shows that irrelevant transitions in the display can greatly disturb the ability of the visual system to access information in the external world; moreover, observers are largely unaware that such a disturbance is even happening. As such, interface designers much attempt to minimize such events as much as possible.

What about when the transition itself is another source of information? (e.g. two movies playing side-by-side). Since only one item can be seen to change at a time (Rensink, 2001, 2002), the second source of information is effectively a distractor, and will increase the likelihood of change blindness. Thus, only have only moving source of information at a time.

Another interesting possibility (so far undeveloped) is to use the experimental techniques described here to evaluate the robustness of an interface to change blindness.

- 3.3 Coercive Graphics

-display can control user's attention; effectively "highjacks" virtual representation

- attentional control via
  - high-level interest (cf. movies)
  - low-level salience
  - small set of cues (e.g., pointing finger)
    - such cues are what magicians use to control what audience "sees"

Given that our representation of the world is a dynamic one that operates via the careful coordination of attention, the possibility arises that external displays can "take control" of attentional allocation, effectively making the user see whatever part of the display that the designer wants them to see.

- 3.3.1. Displays

- coercive displays could induce a viewer to attend to a given location at a given time

- could enable invisible transitions, even if the effect is transitory

It is unlikely that attentional "hijacking" could be maintained for long. However, even a transient effect would be useful: for example, if attention could be briefly sent to a particular part of the display at a particular time, it might be possible to make invisible transitions elsewhere. If this were practical (and the success of magicians indicates that it might be), this would enable the creation of invisible transitions without any need for eye tracking, or any other monitoring of the observer.

- 3.3.2. Interfaces

- coercive displays could send attention of user to appropriate location at appropriate time

-> **Soft warning**: user automatically "sees" what they should see (e.g. incoming mail)

- no need for hard warning (e.g. beep); attention is controlled in more natural way

An important application of coercion is to make the viewer notice new events that may have occurred. Given that an interface user is often deeply engaged in a task, they may not notice when a new event (such as arrival of email) has occurred. Current systems will grab the attention of the user by a "hard" warning—a noticeable alert. However, it may also be possible to direct the attention of the user in a less disruptive way by the use of "soft" warnings, which would take effect when the user was in an appropriate state (e.g., just finished reading a particularly interesting section of text). In such a situation, the user would simply notice that the event had occurred; the announcement would have appeared as if by magic.

Recap

- 1. Change Blindness
- 2. Implications for Human Vision
  - Attentional processes
  - Virtual representation
  - Triadic Architecture; Nonattentional processes
- 3. Implications for Visual Displays
  - Optimized information pickup
  - Invisible transitions
  - Coercive graphics

Wrap-up.

A striking phenomena—change blindness—was introduced, in which observers can be blind to large changes, even when these changes are large and repeatedly made. To account for this counterintuitive phenomenon, it was argued that perception does not form a detailed, all-purpose "picture" of the world, but rather, is based on a much sparser and more dynamic "just-in-time" system in which attention plays a key role. In this view, visual perception is an inherently interactive process, relying on the environment as an external memory, and even an external processor. Finally, some suggestions are made as to how visual displays can take advantage of this new perspective. Among other things, a few approaches are sketched showing how it may be possible to put magic into visual displays.

References

Ariely, D. (2001). Seeing sets: Representation by statistical properties. Psychological Science, 12 (2), 157- 162.

Biederman, I. (1981). "On the semantics of a glance at a scene". In M. Kubovy and J.R. Pomerantz (Eds.), Perceptual Organization (pp. 213-253). Hillsdale, NJ: Erlbaum.

Bridgeman, B., Hendry, D., & Stark, L. (1975). Failure to detect displacement of the visual world during saccadic eye movements. Vision Research, **15**, 719-722.

Brooks, RA (1991). Intelligence without representation. Artificial Intelligence, **47**, 139-159.

Chun, M.M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. Cognitive Psychology, **36**, 28-71.

Clark, A. (1997). Being There: Putting Brain, Body, and World Together Again. Cambridge, MA: MIT Press.

Clark JJ, Yuille AL (1990). Data Fusion for Sensory Information Processing Systems. Boston : Kluwer Academic.

Fernandez-Duque, D., & Thornton, I.M. (2000). Change detection without awareness: Do explicit reports underestimate the representation of change in the visual system. Visual Cognition, **7**, 323-344.

Goodale, M.A., Pelisson, D., & Prablanc, C. (1986). Large adjustments in visually guided reaching do not depend on vision of the hand or perception of target displacement. Nature, **320**, 748-750.

Nowell LT, Hetzler EG, Tanasse T. (2001). Change Blindness in Information Visualization: A Case Study. Proceedings of the IEEE Symposium on Information Visualization 2001 (INFOVIS'01), 15-22.

O'Sullivan C, Dingliana J, Howlett S. (2002). Eye Movements and Interactive Graphics. In The Mind's Eyes: Cognitive and Applied Aspects of Eye Movement Research. Hyönä, J. Radach, R. and Deubel, H. (Eds.) Elsevier Science, Oxford. (To appear, 2002).

Oliva, A. & Schyns, P. (1997). Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. Cognitive Psychology, **34**, 72-107.

Rensink RA (2002). Change Detection. Annual Review of Psychology,53:245-277.

Rensink RA (2001). Change Blindness: Implications for the Nature of Attention. In MR Jenkin and LR Harris (eds.), Vision and Attention (pp. 169-188). New York: Springer.

Rensink RA (2000a). Seeing, Sensing, and Scrutinizing. Vision Research,40: 1469-1487.

Rensink RA (2000b). The Dynamic Representation of Scenes. Visual Cognition,7:17-42.

Rensink RA (2000c). Visual Search for Change: A Probe into the Nature of Attentional Processing. Visual Cognition,7:345-376.

Rensink RA, Enns JT (1998). Early Completion of Occluded Objects. Vision Research, 38:2489-2505.

Rensink RA, O'Regan JK, Clark JJ (1997). To See or Not to See: The Need for Attention to Perceive Changes in Scenes. Psychological Science, 8:368-373.

Simons, D.J. (1996). In sight, out of mind: When object representations fail. Psychological Science, **7**, 301-305.

Swain, M.J., & Ballard, D.H. (1991). Color indexing. International Journal of Computer Vision, **7**, 11-32.

Trehub, A. (1991). The Cognitive Brain (pp. 55-78). Cambridge, MA: MIT Press.

Wolfe, J.M. (1999). Inattentional amnesia. In V. Coltheart (Ed.), Fleeting Memories. (pp. 71-94). Cambridge, MA: MIT Press.

Selected Papers (pdf files available at http://www.cs.ubc.ca/~rensink/publications/).

Rensink RA (2002). Change Detection. Annual Review of Psychology,53:245-277.

Rensink RA (2000a). Seeing, Sensing, and Scrutinizing. Vision Research,40: 1469-1487.

physical interaction design:
haptic and multimodal interfaces

Karon MacLean
Dept. of Computer Science
University of British Columbia

SIGGRAPH 2002 | THE UNIVERSITY OF BRITISH COLUMBIA | MAGIC

These notes contain a material based on the UBC Computer Science graduate course, **"Physical Interface Design & Evaluation"** (more information at http://www.cs.ubc.ca/~cs554/physical/). Some material was previously presented in an invited haptics overview course at UIST 2001.

For more information:

Mail: Karon E. MacLean
Department of Computer Science
University of British Columbia
2366 Main Mall
Vancouver, BC V6T 1Z4
Canada

Email: maclean@cs.ubc.ca

http://www.cs.ubc.ca/~maclean

outline

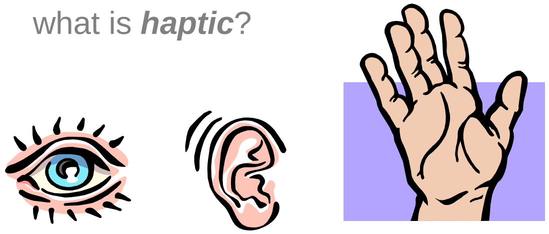
- definitions & orientation
- historical context of force feedback
- tenets of physical and multimodal interaction design
- human haptic sensing and motor performance
- how force feedback works
- rendering haptic and multimodal models
- areas of basic research
- getting going

definitions & orientation

MacLean: Physical Interaction Design 3

We’ll start with some background on the sense of touch, and the kind of research and applications that currently focus on the touch sense.

what is ***haptic***?



from Greek *haptesthai* : **to touch**

MacLean: Physical Interaction Design 4

“Haptic” refers to touching, as “visual” refers to seeing and “auditory” to hearing.

types of human haptic sensing

cutaneous / tactile:

- **heat, pressure, vibration, slip, pain**
- sensation arising from stimulus to the skin

kinesthesia / proprioception:

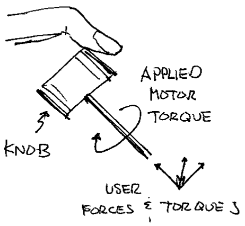
- **limb position, motion, force**
- end organs located in muscles, tendons, and joints
- stimulated by bodily movements

MacLean: Physical Interaction Design 5

The haptic sense has two main components: taction and kinesthesia. Tactile sensors in your skin allow you to feel surface textures and qualities. Your kinesthetic sense, on the other hand, provides your sense of body forces and motions. If you held a brick in your hand, you would use kinesthesia to tell how heavy it is. If you closed your eyes and someone else moved your arms around, kinesthesia would tell you where they were without your seeing them.

"Haptic interfaces" are generally directed at either the tactile or the kinesthetic sense, with the latter being much more common. These are also called "force feedback" interfaces.

what is haptic force feedback?



a small personal robot:

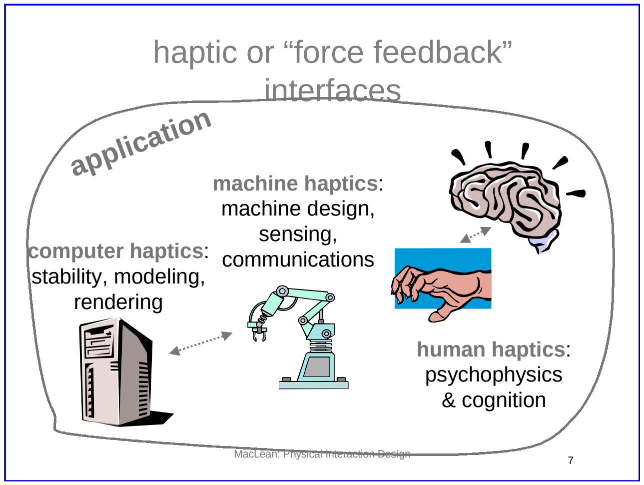
- applies computer-controlled forces to user's hand
- represents a virtual environment
- acts as both an input and output device: user feels & controls at same time.

MacLean: Physical Interaction Design 6

Mechanically, a haptic force feedback interface is in effect a small robot: an actuator applies force or motion through a linkage of varying complexity. The sketch here illustrates a very simple, one degree-of-freedom (DOF) haptic interface, with a configuration like a knob. When the user turns the knob, he feels a torque applied by the actuator resisting or aiding his motion.

Typically, a computer models a "virtual environment", and sends force commands to the actuator that correspond to the current state of this model. At the same time, the computer receives position data from the device which inform it how the user is responding to the environment. This position data is used to update the model, closing the feedback loop.

The haptic interface is thus both a control and display channel at the same time, in a way no sense other than touch can support. This leads to powerful possibilities, but also presents design challenges.



The craft of designing, building and developing applications for haptic interfaces is intensely interdisciplinary. Some of the areas that feed it include:

Human haptics: psychological study of the sensory and cognitive capacities relating to the touch sense, and the interaction of touch with the other senses. This knowledge is crucial to effective interface design; and in turn, haptic interface designers are a source of interesting new research problems for psychologists.

Machine haptics: the province of mechanical and electrical engineers, this consists of the design of the robotic machine itself, including its kinematic configuration, electronics and sensing, and communications to the computer controller. New devices are being invented all the time, both for particular applications and as mechanical design concepts looking for applications.

Computer haptics: computed algorithms and models run on a computer host to control the mechatronic haptic display. This includes creation of virtual environments for particular applications, general rendering techniques and control issues such as sustaining robot stability in the face of a changing virtual and manual environment.

Interaction design: puts the above together to accomplish a particular interaction task.

some current & future application areas:

- desktop GUI augmentation
- medical robotics
- physical rehabilitation
- entertainment
- embedded consumer electronics
- telerobotics and virtual environments
- training and education
- CAD tools
- creative & expressive tools

desktop "HUI"

one approach:

- everyone has a haptic mouse (force or tactile feedback)
- render GUI is easier with edges, textures, etc

more ambitious:

- redesign GUI's from ground up with haptic feedback in mind

a bit further out still:

it's not a desk any more…

The first and perhaps most obvious use for a "HUI" (Haptic User Interface) is as an enhancement to the GUI. The enhanced mouse has been a frequently cited haptic interface goal for many years. There are multiple approaches that could be taken here – the first has already been frequently attempted with mixed results, the others are much more ambitious.

1. Simply add haptic feedback to existing GUI's. Use the haptic feedback to solve common problems, like sliding off menu bars, clicking on small buttons and finding locations in word processor text. It has often been suggested that such an approach can be used to make GUI's more accessible to users with visual and/or motor disabilities, since the graphical terrain can be felt; and reduced motor skills might be required to manipulate the mouse.

2. Start from scratch, and redesign the interface into a graphic+haptic one – a GHUI! I don't believe anyone's tried this, and there are obvious barriers to acceptance – the QUERTY problem. But it's what might be required to do it right.

3. Finally, throw away the desktop metaphor entirely. Haptic interfaces will enable new kinds of interactions which are no longer tied to sitting down in front of a graphical monitor. While more of a jump, this may be easier than 2. since there's no existing standard to battle.

medical applications

- training & simulation, for:
  - diagnosis
  - tissue palpation
  - measures… swelling, bone fracture, pulse, lumps
- Minimally Invasive Surgery (MIS)
  - training through simulation
  - the real thing: "fly by wire" to improve interaction
- telesurgery
  demonstrated on a human in France in October 2001

Medical applications are the most popular (in a research sense) to date. The bulk of haptic research applications and many commercial endeavors are developing different aspects of a number of surgical training and operational aids, from mechanical device design to enhancing communications bandwidth for remote telepresence and studying acceptance of these tools among medical professionals.

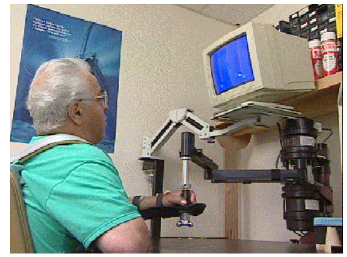
Intuitive Surgical, Inc.
Mountain View, CA

Intuitive Surgical has been working on one of the most elaborate laparoscopic surgical aids for several years.

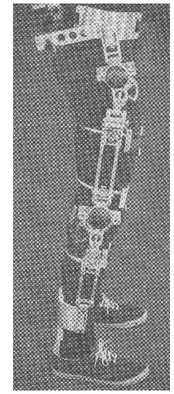
Other companies include Computer Motion (Santa Barbara, CA) and divisions of Immersion Corp. (San Jose, CA).

rehabilitation

Goldfarb & Durfee: controllable brake aids paraplegics in walking



Krebs, Hogan et al: retraining stroke patients while measuring their progress.



Some of the earliest force feedback applications were in the rehabilitation area. MIT's Biomechanics Lab developed many such precursor projects in the 80's and 90's, intended to restore function or retrain individuals with different types of disabilities.

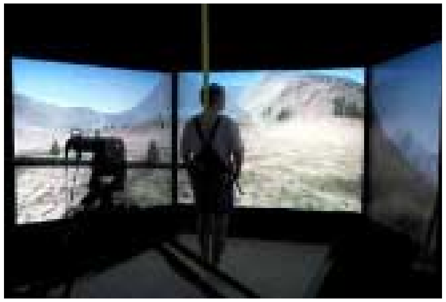
The planar manipulator (left) is used to retrain stroke patients, replacing the tedious job of a human therapist of guiding the individual's limbs through repetitive motions while precisely monitoring the resistance or assistance he's able to provide as he progresses. The controllable brake (right) is part of a system that helps paraplegics walk: the individual's leg muscles are electrically stimulated to provide motive forces, which with current technology is crude and poorly controlled. The brake orthotic is activated sequentially through the gait cycle to provide the cyclic control required for smooth walking. It effectively dissipates some of the force generated by the individual's muscles, to produce the desired activation levels.

entertainment: greatest cost pressure

- virtual reality arcades (body-sized systems)
- home-based gaming systems:
  - vibration feedback
  - force feedback joysticks and mice
- coming soon (?):
  - model-based force feedback in synthesizer keys
  - more expressive & creative apps as quality goes up and cost goes down

Consumer products provide the greatest potential market, but they can't cost very much. The first entries have been (in addition to pager buzzers, which are tactile rather than force feedback) gaming joysticks and actuated electronic piano keyboards. As we improve inexpensive actuator and sensor designs, there will be more possible applications.

military applications

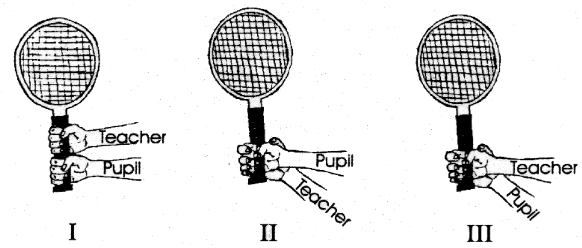


combat simulators: e.g. Hollerbach, Univ. of Utah projection with treadmill display

Some of the most elaborate applications are in service of military training and field devices. An example is the combat simulator designed by Hollerbach's team at the University of Utah with Sarcos Research, where a variable-orientation (inclination and horizontal direction) treadmill is combined with a projected visual display and a support applied to the user's torso to simulate loads of gravity and inertial accelerations.

education

- includes medical simulators
- the “virtual teacher”: three paradigms
  (Gillespie, O’Modhrain et al 1998)



Educational applications are a rich area of development. There have been numerous variations on the “virtual teacher” idea, teaching a student motor skills relating to, for example, calligraphy or playing a musical instrument. The three approaches suggested above vary in the degree of control applied by the teacher vs. the student; a novice student will require more guidance from the teacher, while an advanced student needs much less.

historical context

a brief
historical review of force feedback:

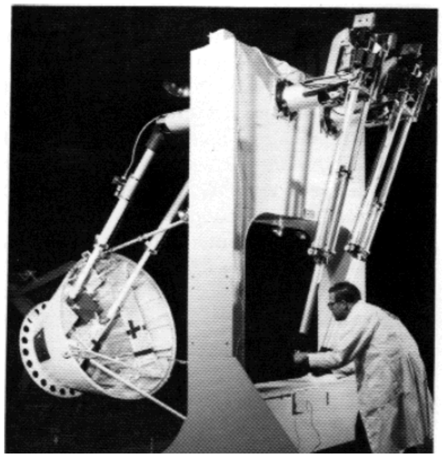
roots in **robotics** and **teleoperation**

MacLean: Physical Interaction Design 17

From the machine side, force feedback interfaces have derived from robotics and teleoperation applications. Just for fun, we'll look at some of these old applications.

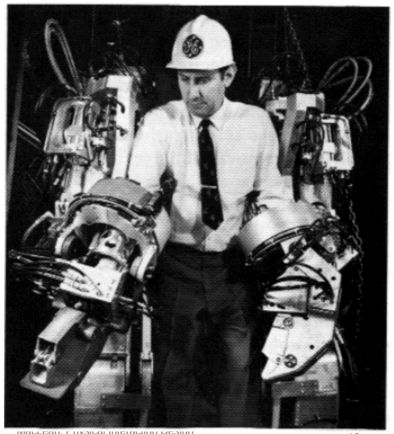
60's

General Electric robotics program



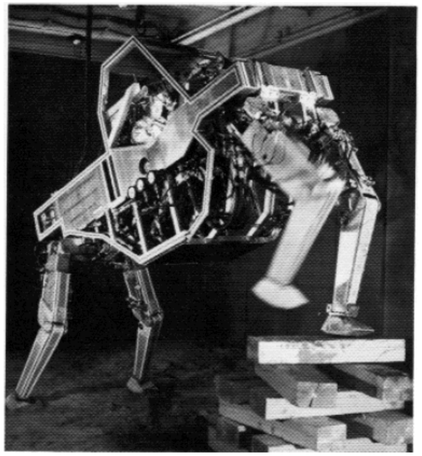
A very early teleoperated system: the user moves handles on the right, while a much stronger "slave" actuators follow these "master" commands on the right. The user's handles aren't actuated; if the slave bumped into something, he wouldn't be able to "feel" it, but must rely on visual inspection of the slave to see how he's doing. Such non-force-feedback manipulation is analogous to manipulating an object with a hand and arm that's numb.

powered telerobot: “man magnifier”



19

Here, we have powered “exoskeleton”. The user wears a powered shell and provides motion commands; these are amplified by the shell so the user can pick up and maneuver heavier objects than he’d otherwise be able to.



the elephant:

4-legged walking machine

20

No force feedback here, but it’s certainly an impressive control problem. This machine is quite large – the size of a car.

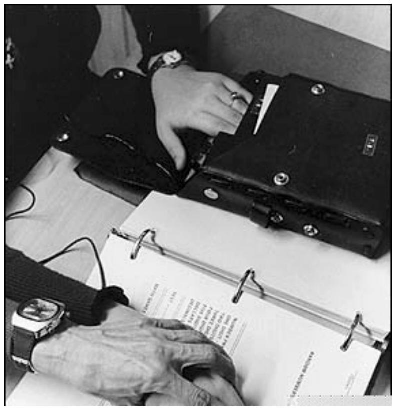
70's

driven by rehabilitation research

Optacon:

Braille reading device,

first tactile display

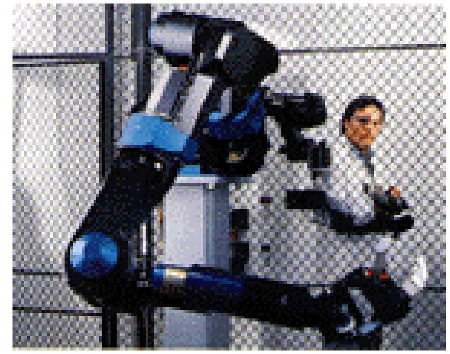


Unlike most of the other examples given in this course, the Optacon is a tactile rather than force feedback display: it acts on the surface of the skin rather than the joints and muscles. An aid for the blind, it dynamically displayed Braille images to one of the user's hand as she moved the other across a page. The Braille display consisted of an array of pins. This device, developed in the 70's, only recently went out of production in response to modern blind assistive reading technology.

80's: "force feedback"

improved teleoperation performance

Sarcos Arm



Several research labs began adding more sophisticated forms of "force feedback" to teleoperated systems in the 80's. This picture shows an arm and hand developed by Sarcos Research, a hydraulically powered slave arm controlled by a "master" exoskeleton worn by the user and designed to be used in remote undersea environments where a human operator couldn't go. Because of limited visibility, it was particularly important in such an application for the user to be able to feel what the slave arm was feeling – that is, to "de-numb" the master's arm. Here, environmental disturbance forces encountered by the slave are "fed back" to the master. If the slave encountered a wall, the master would have feel the wall's resistance as well and not be able to push through it. If the slave picks up an object, the user can gauge proper grip force via the force feedback through the master's fingers.

early 90’s

- 1990: Minsky’s “virtual sandpaper” - caused a “sensation”
- 1992: first “conference” on haptic feedback (session at a large mechanical engineering conference – mainly about device design)
- 1994: PHANToM invented - for sale in **1995**

In the early 90’s, the field of haptic force feedback began to develop an identity of its own. There were research projects specifically oriented at providing force or tactile feedback, rather than teleoperation; and there were enough research practitioners to support a session at an engineering conference specifically on the subject. The first commercial device was for sale in 1995 – a profound influence on the field, since now it became possible to develop haptic applications without being a robotics engineer.

This gave rise to a new parallel focus on on application development and computer haptics, in addition to machine development alone.

Here’s the Phantom (sold by Sensable Devices Inc.) today. It has 3 actuated translational degrees of freedom as well as 3 sensed angles at the end effector. It is still the most popular commercial research device today, used most commonly in conjunction with 3D graphical models.

late 90's:

- sensory and cognitive psychophysics studies
  undertaken explicitly to improve haptic displays
- 1997: first haptic game joystick for sale
- 1999: Immersion.com went public
- today, proliferation of research displays
- integration into new environments – e.g. automotive

In the last four years, the field has gone almost mainstream: it is possible that when you tell someone you work on haptic interfaces, they'll have an idea what "haptic" means (and not think it's some kind of disease). A second company (Immersion) went into business selling devices, aiming more at the consumer market, and went public in 1999 (just before the tech bust). Today, there are many companies selling haptic displays of different configurations and cost, some general purpose devices like the Phantom and some oriented at particular application areas, such as surgical simulation. Many researchers still design and build their own devices, and many more purchase devices to develop applications and algorithms.

Consumer adoption of haptic technology remains at the primitive level, mainly in games joysticks and pager buzzers. However, as various kinds of information displays become more visually overloaded, there has been a commercial push to offload information display and manipulation to the haptic sense, for example in automobile cockpit controls. These augmented interfaces are in their infancy, with prototype releases by BMW and Nissan in 2001; but can be expected to become more sophisticated out of sheer necessity.

tenets of physical interaction design

touchable interfaces:

what's so special about touch?

what kind of interactions is it good for?

That brings us up to the present. We'll talk now about the actual design of touchable interfaces, beginning with a discussion of the big deal about the sense of touch.

For a more thorough treatment of this topic, see MacLean (2000), "Designing with Haptic Feedback".

special qualities of touch

bidirectionality: encompasses intention, manipulation, gesture and perception

social loading: intentional, socially invasive and committing

gesture and expression: convey functional and emotional signals through touching

multi-parametered: force, pressure, moisture, temperature, texture…

resolution and associability: precise control & discrimination, poor absolute resolution

MacLean: Physical Interaction Design 27

There are several attributes of the touch sense – from the perspective the human - which are unique to it, and make it both potentially powerful, and quite tricky to design for.

motivations for touching

we touch intending to -

- do a task
- probe an object
- poke to elicit a reaction
- fidget to relieve tension
- communicate a message
- verify that an action is completed
- enjoy aesthetic pleasure or comfort
- connect physically or emotionally to living things.

MacLean: Physical Interaction Design 28

It's important to understand why your user might choose – or be persuaded to – touch an interface. Basically, it will be either to accomplish some task or satisfy some other kind of urge, aesthetic or otherwise.

inhibitions to touching

we avoid a touch through perception that it would be

and then, there are the “haptically challenged.”

MacLean: Physical Interaction Design 29

However, an touchable interface will fail miserably if the user doesn’t want to touch it. Keep in mind all the reasons someone might be reluctant to expose their body to this intimate experience, and try to deal with them up front. A force feedback interface, which will for some time be novel to most new users, must meet the additional challenge of fear: if the inexperienced user every sees it do something unexpected, he may be afraid that it will hurt or startle him.

information available from touching

assessments of an object’s dynamic and material properties.

verification of engagement and completion

continuous monitoring of ongoing activity and gradual doneness.

building mental models for invisible parts of a system.

judgments of other people.

MacLean: Physical Interaction Design 30

These are some of the things we get out of everyday touching. If an interface is to benefit from having force feedback as a component, it should provide at least one of these functions.

mediums of tangibility

old-fashioned tools and textures: specific, modified, distinctive, worn in significant ways.

synthesized haptic feedback: active, passive & parasitic force feedback; “multihaptics”.

mediating haptic interfaces: abstract relation between user and environment (Snibbe, MacLean et al 2001)

haptic language: shares attributes with visual gesture; lexicon & grammar of touching

MacLean: Physical Interaction Design 31

What forms can haptic feedback take, in an interface sense?

In traditional (non-active) interfaces such as hand tools, forces and textures provide a wealth of information. Tools are generally designed for specific tasks, with handle and business end suggesting that function. Patterns of wear on old tools give signs of how (and how much) it’s been used. Heft and balance make a tool work better, and well designed handles allow the user to apply large forces or delicate movements appropriately.

Synthesized (active) haptic feedback is what we’ve been talking about: virtual or remote environments are rendered haptically to allow the user to interact, learn or manipulate a distant environment.

Mediating interfaces interpose a rendered physical model in between the user and some application. This kind of environment assists the interaction rather than being its point – in the same way that a pair of scissors assists you to cut a piece of paper, rather than explore and experience the textural properties of the paper directly.

Finally, “haptic language” represents a further level of abstraction in active haptic interfaces: when haptic feedback is used to communicate information to the user, what do these feels mean? In the simples sense, “haptic icons” or hapticons can be used in the same way as visual or auditory icons.

when active touching helps: potential benefits

reconfigurability:

- represent environment changes haptically (e.g. # of knob detents)

dealing with complexity:

- offer cues to user options
- differentiate buttons
- merge discrete steps into fluid continuous control gesture
- transport electronic tool use away from desktop

MacLean: Physical Interaction Design 32

Adding haptic feedback to an interface will generally make it more expensive, so we need to be realistic about the extra value it provides. Some of these are purely functional…

when active touching helps:
potential benefits

comfort and aesthetics:
- pleasant tactility
- satisfying motion & dynamics
- ergonomics
- bidirectional environment coupling
- muscle memory
- personalization

affect and communication:
- adds social context and presence to mediated user-user or user-machine connections.

… while other functions are more subjective or emotional.

abstraction in user-target mediation

abstraction

direct manipulation: environment signal or model is rendered literally

container manipulation: simple hierarchical layer of abstraction. raw signal / model is rendered in “bins”

annotation: haptic marks on signal or model indicate interesting content

mediating dynamic system: user manipulates virtual dynamic system to observe / change environment

Within the realm of active force feedback, there is a range of abstraction inherent in the rendered virtual environment.

haptic feedback in discrete and continuous control

buttons for discrete control and information:

- distinction / ID of objects
- impose edges on continuous input
- notify of failure / confirm operation success
- reflex-rate user reactions

handles for continuous control & monitoring:

- expressive input
- lo-res, lo-attention monitoring
- teaching, training & guiding

(MacLean et al, 2000)

The distinction between "discrete" and "continuous" control is an important one, in understanding the potential value of manual control. The analogy to "buttons" and "handles" in our more familiar manual world is helpful for distinguishing the types of tasks each sort of function is good for.

Active haptic feedback can provide particular kinds of augmentation in **discrete "button" applications**, where the task is to trigger something automatic. It may make it easier to locate the button, distinguish or even preview its function, and let the user know something happened. If the event must be activated quickly, the haptic feedback can help by providing the necessary information through the same channel that will be used to activate.

In continuous, "handle" applications, the user's job is to maintain control or contact over an operation – for example, drawing or playing a musical instrument (requiring expressive interaction); low-attention monitoring, where different levels of status can be conveyed without the user making them the principal focus of attention; learning a new manual task.

challenges for haptic interaction design

continuous / discrete manual control: many tasks require both; transitions are interesting

displaying interaction potential: one handle / many functions best when tasks use same rules

embedding haptic interfaces: can be customized to a specific task context (okay when cheap & simple).

tight sensory coupling for perceived control: attainable through wisely designed software architecture

Haptic interaction design isn't easy, in part because we're new to it and conventions / technology for solving some of its challenges haven't become widespread yet.

•When a task ideally requires both continuous and discrete control, how do you make transitions without resorting to a modal design? (see MacLean, Snibbe et al 2000 and Snibbe, MacLean et al 2001 for some ideas)

•One of the powers of haptic feedback is that one handle can do a lot of things –its feel and function are programmable. However, the downside is that if you use a generic physical handle, it is harder to make it apparent what the interface can do, or even how to use it.

•When we leave the desktop and embed inexpensive physical interfaces throughout our everyday environment – e.g. in cars, buildings, public spaces, clothing and portable devices – we can address the generic-handle problem by making each interface serve a specific purpose with a specific, custom handle. However, the process of creating the cheap devices and embedding them effectively is an engineering work in progress.

•High quality physical interactions must be tightly coupled among the different sensory modalities to give the user a sense of tight control: when you make a movement, you must perceive the result – be it visual, auditory or haptic – to happen simultaneously. This can require careful design of software architecture and communication networks. (See MacLean & Snibbe, 1999, "An Architecture for Haptic Control of Media")

human haptics

research leading to the characterization of human sensory, motor and cognitive capabilities

➡ insights for device design & applications

- types of haptic sensing
- taction
- kinesthesia
- sensorimotor control

MacLean: Physical Interaction Design 37

Our next topic is **Human Haptics**, the study of human sensory, motor and cognitive capabilities with respect to the sense of touch. We'll go over a bit of background on the haptic sense and motor control, keeping in mind its impact on interface design.

types of haptic sensing

cutaneous / tactile:

sensation arising from stimulus to the skin

kinesthesia / proprioception:

a sense mediated by end organs located in muscles, tendons, and joints

stimulated by bodily movements.

MacLean: Physical Interaction Design 38

As we mentioned earlier, there are two components to the haptic sense; the majority of contemporary active-haptic applications are addressed at the kinesthetic sense, but in real-world interactions, taction is equally important and we hope that as technology improves, tactile interfaces will become more sophisticated.

tactile sensory receptors

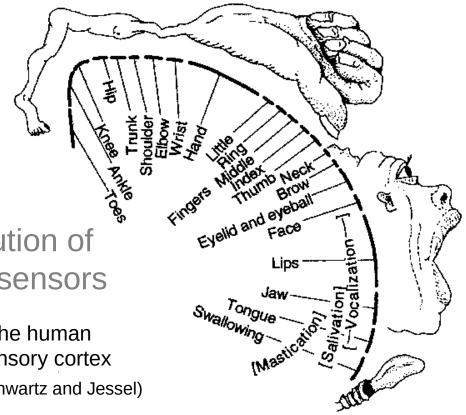
thermoreceptors
- change in skin temperature

mechanoreceptors
- pressure, vibration, slip

nocioreceptors
- pain

There are three main functional categories of tactile receptors, all located in various layers of the skin. We won't go into the fascinating details of how they work mechanically and neurologically; but many perception textbooks will go into this (e.g. Rock 1984, Goldstein 1999).

distribution of tactile sensors



mapping the human somatosensory cortex

(Kandel, Schwartz and Jessel)

To the interface designer, the distribution (in particular, the density) of tactile sensors is more relevant. The higher the density, the more detail the user will be able to feel when a tactile stimulus is applied to that region of the skin.

sensorial adaptation

receptors have different rates of adaptation to stimuli:

Slowly Adapting (SA):

- respond throughout stimulus
- e.g. joint angle information from skin stretch

Rapidly Adapting (RA):

- respond at start/end of stimulus;
- optimized to “block out” extraneous signals, e.g. wearing gloves

Another aspect that interface designers need to know about is the rate of adaptation of the different receptors have. “Adaptation” is the phenomenon of a sensor “getting used to” a stimulus and after a while, failing to pass it on up the neural pathways towards the brain. Adaptation is why after a while, you don’t notice a loud steady noise. Some sensors adapt quickly, and others more slowly.

spatiotemporal resolution

spatial limen (resolution):

- depends on size of receptor field;
- resolution reduced by crosstalk & overlap

successiveness limen:

- 5 msec to perceive as separate
- 20 msec to determine order
- …but much more for cortex to process.

masking:

- stimuli interfere, either spatially or temporally
- limits the maximum information transmission rate

Finally, it’s important to understand a few details about tactile sensor resolution, both spatial and temporal. A “limen” is a specific receptor’s resolution. Spatially, these are determined by the size of the receptor’s field; if there is a high density of receptors, then the resulting overlap and “crosstalk” reduces effective resolution – when a point stimulus is applied in overlapping fields, the perceptual resolution becomes the size of the union of the two receptor fields.

Successiveness limen relate to temporal resolution – how closely spaced a series of stimuli to be for a person to distinguish them as separate. This determines, for example, the lowest and highest frequency of vibrations we can distinguish.

kinesthesia

- perception of limb position, motion, force
- some cutaneous information is used, especially in hairy skin
- main information from mechanoreceptors in muscles

MacLean: Physical Interaction Design 43

Kinesthesia is what we're using when we get information from most force-feedback displays. Its receptors are mostly embedded in our muscle fibers and joints; sometimes skin stretching also gives kinesthetic cues.

muscle mechanoreceptors

2 types:

Golgi tendon organs:

measure force via localized tension

located serially between muscles & tendons;

muscle spindles

located in parallel between individual muscle fibers

excited by active & passive stretching

MacLean: Physical Interaction Design 44

Again, we'll skip neural mechanism for our purposes here, but look it up in Rock or Goldstein or any other good perception text – very cool stuff.

sensorimotor control

exploratory tasks:

dominated by sensing, with limb under force control

manipulation tasks:

motor dominant - use both position & force receptors

key components of control (differ widely by body region):

- maximum & sustained force exertion
- force tracking resolution
- torque & compliance resolution

MacLean: Physical Interaction Design 45

Sensorimotor control guides our physical motion in coordination with our touch sense. There is a different balance of "position" and "force" control when we're exploring an environment – e.g. lightly touching a surface – versus manipulation, where we might be going through a pre-programmed sequence of movements and relying only subconsciously on the touch sense.

A few specifications of our sensorimotor capacity are relevant to physical interface design, to be expanded on below.

maximum & sustained force exertion

finger contact forces: depend on

grasping geometry, strength (indirectly on gender, age)

power grasps:

high stability and force (200-400 N maximum)

precision grasps:

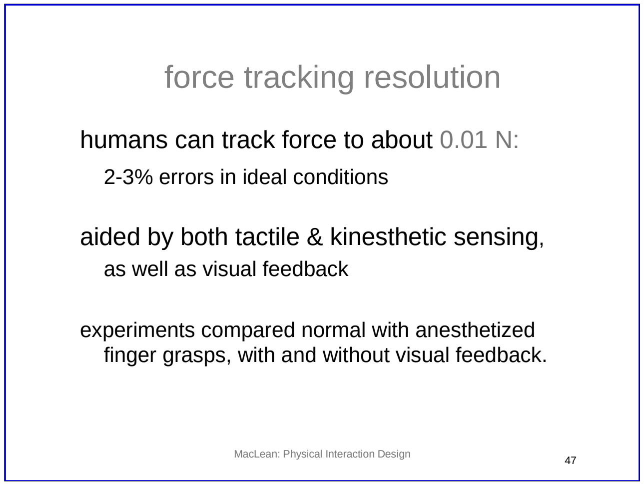
less force but higher dexterity

fatigue: compromises ability to control grip force

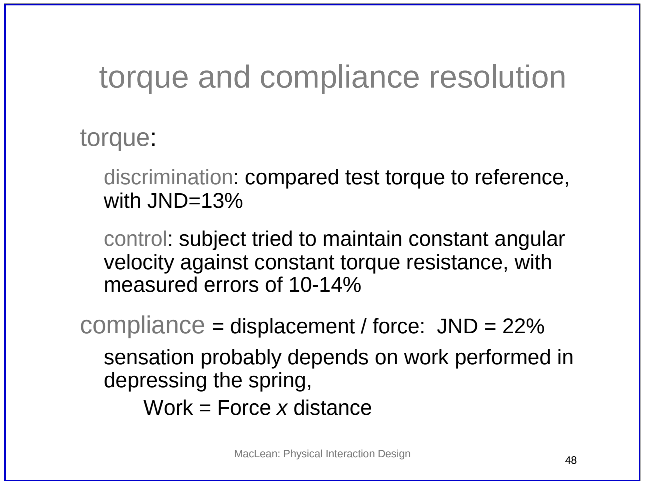
MacLean: Physical Interaction Design 46

As anyone knows intuitively, the amount of force you can generate depends on the way you hold something – the grip brings into play different muscle groups, which in turn differ in their force generation / sensory resolution capabilities.

When you get tired, not only can you generate less force – you aren't as good at controlling it.



Many psychophysical studies have been done to determine such human capabilities as force tracking resolution. As an example of how these studies are done, to determine the involvement of different kinds of sensory receptors in tracking resolution, the hand might be anesthetized either skin only, or entire hand, or not at all.



"JND" means "Just Noticeable Difference", and is a common measure of sensory resolution. A JND of X% implies an exponential resolution curve: at low torque levels, we can sense values relatively close together, but as the absolute torque or force level increases, absolute sensory resolution decreases accordingly.

Interestingly, it wasn't initially clear if people discriminated torque/force directly, or compliance, or something else. Some experiments by Srinivasan's group at MIT helped clear this up: it looks like what we're really paying attention to is the work done in compressing a spring (e.g. Wu et all, 1999).

sensing & control bandwidth
upper limits

variation in frequency limit = function of receptor type:

sensing (kinesthetic): 20-30 Hz

sensing (tactile): 10-10,000 Hz

control bandwidth: 5-10 Hz

Frequency is an important sensory parameter. We can sense tactile frequencies much higher than kinesthetic; this makes some sense, since skin can be vibrated much more quickly than a heavy limb, even a fingertip.

Our control bandwidth is how fast we can move our own limbs or digits – this frequency is much lower than the rate of motion we can perceive.

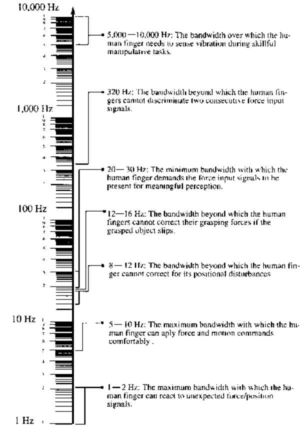
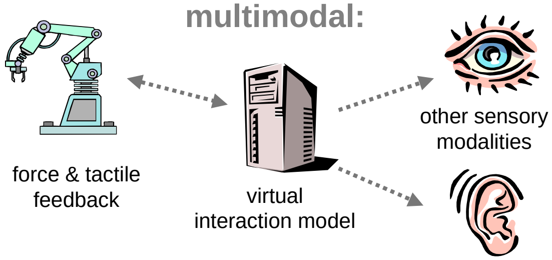


Figure 2.14 Human finger sensing and control bandwidth. Adapted from Shimoga [1992]. © Éditions Hermès. Reprinted by permission.

bandwidth detail

For reference, here’s some more detail on what happens at different bandwidths.

physical interfaces are usually **multimodal:**



→ psychophysicists have discovered a few things about how stimuli to multiple senses **influence** one another

Force display technology works by using mechanical actuators to apply forces to the user. By simulating the physics of the user's virtual world, we can compute these forces in real-time, and then send them to the actuators so that the user feels them/

some distinctions

these terms sometimes used interchangeably in literature:

- **crossmodal, intermodal**: one modality subconsciously influences perception in another modality
- multimodal: an event is perceived and integrated by multiple senses
- supramodal: phenomenon that applies to all senses
- intramodal: all in one sense

Here, we'll focus on crossmodal effects.

early examples of crossmodal effects

- Bartholmus (1669) noted that partially deaf people seem to hear better in light than in dark.
- …1929: corroborated; found (1950's) to be dependent on other factors
  – e.g. light hue, auditory pitch, relative stimuli onsets.
- Urbantschitsch, 1888, found that thresholds for touch and pressure are lowered by exposure to weak sounds but raised when the accessory auditory stimulation is intensified.
- Johnson (1920) found that tactual discrimination is slightly better (2%) in light than in dark.

As you might guess from your own personal experience, there's been curiosity about crossmodal sensory interactions for a long time. These are a few representative studies from the past.

the McGurk effect

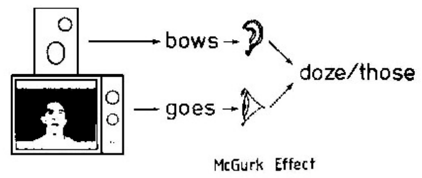


Fig. 1 The McGurk effect is an example of a cross-modality illusion. If one hears the word "bows", but sees "goes", one perceives a synthesis of the two inputs: either "doze" or "those". Reprinted from Dodd and Campbell (1987)

McGurk & MacDonald 1976

The McGurk effect is between vision and audition, and shows one way that input to one sense can modify the interpretation made from another sense – and the type of study that elucidates this sort of thing.

Crossmodal effects involving the haptic sense haven't been much studied until much more recently.

as interface designers, why do we care?

"multimedia" interfaces deliver input in many sensory modalities → need to understand how these are processed perceptually:

- design rules: control net percept produced in user
- ecological verity: respect perceptual latency thresholds for perceived synchrony
- avoid overkill: find most efficient path to desired result
- exploit illusions: work-around hardware limitations through clever compensation

important factors in crossmodal effects

stimulus intensity:

- moderate accessory stimulus facilitates primary stimulus
- intense accessory stimuli inhibits primary stimulus

habituation: initial effects of accessory stimulus differ from later effects

enhancement:

- neurons sum + transform input from different senses
  … effect most dramatic when unimodal stimuli are weak.

masking:

- occurs when stimuli are further apart spatially:
  → degradation in overt orientation behavior

sharing of attentional resources by different modalities

implication:

- if attentional resources are shared & tactile information is presented visually (e.g. texture, weight)

  this may reduce visual resources for processing purely visual tasks.

- resource-sharing hypothesis supported by study of linked attentional spatial shifts
  (Spence & Driver, 1997)
- area of ongoing research

MacLean: Physical Interaction Design 57

As the need for interfaces making extremely efficient use of limited perceptual resources, sharing of attention becomes something we need to understand better.

There's been quite a bit of study about attentional distribution within vision; less with audition, and virtually none with touch. Even less studied is attention as shared among senses. If, for example, we plan to offload the visual sense by delivering information haptically, we better know whether this transfer of work will actually unload total attention required – or make the situation even worse. A group at UBC is working on this problem right now.

multimodal temporal thresholds for synthesis

temporal thresholds:

delays between onset of stimuli in different modalities

e.g strike a drum, but hear the sound early or late:

at what delay do events fail to fuse?

- poorly studied
- many variables
- highly context-dependent

MacLean: Physical Interaction Design 58

There are different thresholds to consider:

•For perceptual synchrony: "the same event created both stimuli"

•For presence/absence of "covert" attending & orienting: subconscious connections between stimuli in different modalities

•For performance degradation: task-based

And different threshold factors to consider:

•relative stimulus intensities

•stimulus duration relative to onset delay

•training / accommodation

•consistency in delays - e.g. playing a pipe organ

•task complexity - whether task is identification, orientation, or complex manipulation

sensory substitution

- synesthesia
- substituting:
  - vision or audition for touch
  - touch for audition
  - touch for vision

many examples (most in rehabilitation research and product development)

(discussion beyond scope of this course)

MacLean: Physical Interaction Design 59

Synaesthesia is a condition present in some individuals (1/25,000) who involuntarily make crossmodal associations; e.g. hear colors, taste shapes, or experience other sensory blendings.

categories of crossmodal effects

- detection and identification
- attention and orientation
- dominance and interpretation

MacLean: Physical Interaction Design 60

These are the principle ways in which crossmodal sensory interactions can be manipulated to improve an interface – or might degrade it if not controlled.

detection and identification

- cross-modal stimulation can lower detection threshold
- appears important to co-locate visual /auditory and somatosensory stimuli
- tactile stimuli can reinforce & clarify marginal stimuli from other modalities (and vice versa)
- potentially increase sense of personal presence or relation to a situation
  e.g. seeing image of your hands can reinforce or manipulate the proprioceptive sense of what your hands are doing
- detection of light tactile stimulation facilitated by vision of own hand (or image, or dummy) (Tipper 98)

attention and orientation

- orienting spatial attention in one modality results in complementary shifts of attention in other modalities
- further:
  - orienting in one modality → enhanced perceptual processing of subsequent stimuli in another modality
  - stimuli should be co-located in space & time
- but if information conflicts
  (e.g. relates to separate phenomena)
- then sources should be separated.

dominance and interpretation

when one of two conflicting stimuli dominates,
the dominating sense is said to "capture" the other.

e.g. stiffness experiment (DiFranco 1997)
vision "captures" touch in stiffness perception

vision / audition usually dominate haptic perception
with key exceptions.

most studies compare vision and touch; audition/touch less understood

MacLean: Physical Interaction Design 63

A few examples of exceptions to visual / auditory dominance of haptics:

•**Texture / surface property judgments**: tactile and visual perception are equally weighted as information sources (e.g. Jones and O'Neil, 1985). Subject produces weighted, arithmetic mean of multiple judgments → compromise rather than capture.

•**Judgments of size**: no single modality dominates. Both modalities seem to be perceived: assessment depends on task and attentional requirements. E.g. if you feel and see discordant versions of an object's size, then you report size by matching it with another object; you may match it with the seen size if matching visually, but with the felt size if matching haptically. It is possibly the same story with shape, orientation and location.

why bother with haptic feedback in multimodal interfaces?

if tactile is dominated by visual / auditory stimuli,
what's the point?

- reinforce dominant modality
- clarify ambiguity in dominant modality
- some parameters not available in dominant modality

  e.g. appropriate grasp force in a minimally invasive surgical interface
- continuous control

MacLean: Physical Interaction Design 64

In the face of conflicting information, the haptic sense is often dominated by other senses. This doesn't make it pointless – far from it. Its information may serve to reinforce or disambiguate other information when it reproduces it; and in other situations, the information might be delivered *only* through the touch sense.

Finally, as we've seen earlier, touch is closely aligned with motor control and this makes it uniquely suitable for some kinds of tasks, particularly those involving continuous control.

force feedback

Now, we get into the heart of what force feedback is and how it works.

force feedback: how does it work?

a few basic principles of **robot control**

- open and closed loop control
- requirements of a control system
- control actions (P, I, D)

To understand its function, you need to know just a little bit about robot control. We'll go over some introductory principles very quickly.

open and closed loop control

the basic problem:

make output = desired

control system design means:

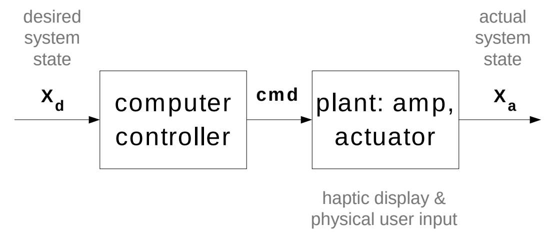
- defining input and output
- techniques for applying corrective action
- techniques for stability & performance

Really what you need to know about a feedback control system is that you're trying to bring the system output in line with some desired input. There are many ways of doing that, some quite simple (we'll talk about those here) and going on to very complex techniques for difficult-to-control systems.

First, you have to understand what your output is (what you're trying to achieve with your control), and what your input is (the parameter which the system can affect and use to bring about a change in the output). You need to understand at least roughly the relation between the system input and output (this is a system model). Then, you need a means of enforcing the relation – usually with either a computer algorithm or an electronic circuit. Finally, you need to make sure the system behaves itself nicely – when you give it a change in desired state, the system should go to that state as smoothly and quickly as possible. Much of control theory is in the interest of achieving this smooth, fast response for difficult cases.

Haptic control systems can be difficult cases, mainly because the human user is part of the system and it is quite hard to predict what the user is going to do. However, the simple techniques presented here are adequate for many situations.

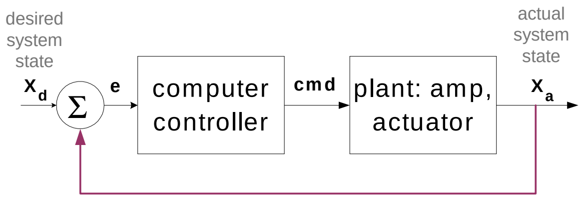
open loop system



The "plant" (a term derived from the first electronic control systems, which were literally chemical process plants) for a haptic display system is the **robotic device** (typically powered by amplified current driven through a DC motor and mechanical linkage) interacting with the user, who also supplies physical input by imposing force or position onto the device when in contact with it. "U" and "X" are the conventional representations for desired system state and the measured state: for example, desired and measured end effector position. "CMD" is the magnitude of, for example, current sent to the actuator which the controller has determined will achieve the desired system state. That is, the job of the computer controller is to translate the desired signal into a **control action**.

In an "open loop system", there is no direct comparison of the output (e.g. measured position) and input (desired position). To be effective, the controller must have a good model of the "plant" so it can estimate the correct control action. If it is wrong, it will never know.

closed loop system



A “closed loop system”, by comparison, “feeds back” the measured state to be compared with the desired state. The resulting error signal, “e”, is input to the computer controller, and the control action is derived from the error rather than the desired value.

Thus, the action of the closed-loop controller is to drive the actual system state towards the desired state, much as a mechanical spring tries to hold two objects together at a specified gap.

human-in-the-loop: closed or open?

in a haptic display system,

- what parameters are controlled?
- what are parameters are observable?
- which ones must we infer?

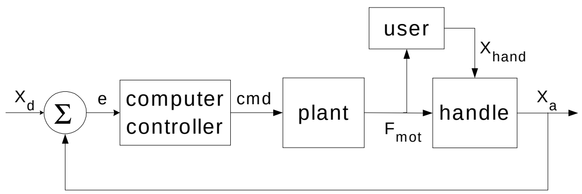
some loops are closed, others are open.

e.g. in most force feedback systems, we close the robotic loop, but have little direct knowledge of user’s state

An “observable” parameter is one which can be measured directly – e.g., you might measure position with a digital encoder or an analog potentiometer. If you can measure position, you can also “observe” velocity, by differentiating position. But some parameters you might not be able to observe even that directly, but estimate using a dynamic model of the whole system and the parameters which you *can* observe.

With contemporary systems, the latter is usually true of most user state parameters. For example, we might have a pretty good guess of their hand position based on reading the device position, *if* we are confident the user is holding the device. But we don’t know where the user is looking, or even the position of the wrist, elbow or any other part of the body.

“closed loop” system with human-in-loop



Here, we show schematically the user added into the loop. The user holding onto the device end effector experiences the same force as the handle; and applies a position to the handle. The device is shown here as an impedance (determining force) and the user as an admittance (determining position in response to the felt force). When this is true, Xhand = Xa.

multirate systems

- fastest: haptic loop (.5-10 Khz)
- fast: auditory, visual refresh (30-100 Hz)
- [usually] slower: virtual environment “state”
- slowest: modes, large user changes

One of the things that makes multimodal interface control systems challenging is the variety of dynamic scales involved.

Haptic servo rates are typically 1 KHz, and involve small amounts of data for point interactions; graphics displays usually update around 30 Hz and may involve massive amounts of data. Thus one is a fast serial loop, the other a relatively slow parallel loop. If these occur on the same machine, tricks may be required to ensure that neither process is starved.

Other aspects of the system probably change more slowly still – for example, the regime or state of the virtual environment (is the probe in contact with the wall or not?); or the mental / emotional / physical state of the user. A person may react differently to the system depending on state of alertness, strength or training.

something to think about: "adaptive" control

control / feedback parameters change as the controller's understanding of the system changes

for haptic feedback, this could mean "as the user's [wishes / experience / fatigue / temper] changes".

… user modeling is a hot area

requirements of a control system

- stability
- responsiveness
- error reduction & steady state accuracy
- disturbance rejection

These are the most important things a control system has to accomplish. The control actions below will accomplish these for simple systems.

system state

e.g.: $\mathbf{X} = \{x, dx/dt, d^2x/dt^2\}$

- defines the current state of the system
- can be forces, positions or anything else.
- because of typical haptic display control methods, they typically employ kinematic state (position, velocity, acceleration).

Estimating system state:

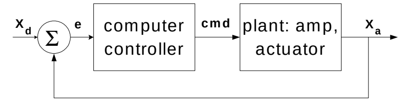
Don't confuse the measured and fed-back signal with the "real" state.

Some estimates are better than others, though. Some of the reasons for imperfect estimates are:

- •sensor problems - e.g. nonlinearity, offset, drift
- •transmission noise, lag, delay
- •discretization

There are many techniques to deal with imperfect signal estimates, beyond the scope of this course.

control actions



function applied to error signal:

$$cmd = f(\text{gain } K, e = x_a - x_d)$$

proportional

derivative

integral

often used in combination:

PD, PI, PID controllers depending on response required

How do you tune a PID controller? Iterate!

- •start with P: increase until unstable
- •add D until stable
- •if SS error is a problem, add I
- •(add D until stable)
- •add more P, adjust D, etc.

→ The result is a compromise between responsiveness, stability. The setpoint depends on the context in which the device will be used: *how* stable does it need to be?

proportional

- $cmd = K_p\ e$
- effect: big signal when large error, small signal when small error
- results in steady-state error
  need a nonzero error to generate a control action
- essentially a spring centered at the desired position (F = K x):
  → high gain makes a stiffer spring & increases tendency to oscillate

A proportional control is “static”: it acts like a spring connecting the actual system state to the commanded value . A high gain makes the system feel like a stiff spring, a low gain like a soft spring.

You can see that this won’t work very well when there’s a large initial error. The resulting command will be very large, and – especially for a high gain system – can cause overshoot. Then there’s a large command in the other direction, and so on. This results in oscillation.

If you try to fix this by reducing gain, the system might be stable but it won’t respond very fast.

derivative

- $u = K_d\ d(e)/dt$ or $u = K_d\ d(x_m)/dt$
- effect: slows control action down and damps oscillations; increases system stability
- can also make response more sluggish
- based on 1st derivative: either noise or filtering phase lag can actually introduce instability again

A derivative element acts sort of like a brake; it tends to increase the system stability. Whenever the rate of change of either the desired signal or the error (difference between command and actual) is large, the command is reduced accordingly. It doesn’t improve system responsiveness, but it helps with smoothness.

integral:

- cmd = $K_i \int(e)dt$
- effect: integrates error over time
  control signal builds up and eventually becomes strong enough to bring error to zero

→ reduces steady-state error and improves system responsiveness

- unpleasant side-effects:
  – integrator windup
  – decreased stability

Finally, the integral term helps to reduce “steady state error”. Adding a lot of integral gain can reduce system stability, and needs to be balanced with a corresponding increase in derivative gain.

types of force feedback displays

- configurations:
  grounded / ungrounded
- actuation
- sensing

grounded interfaces:

solidly connected to the “worlc

- similar to a robot
- compute kinematics to locate “user” relative to “world”
  1. determine endpoint position
  2. derive velocities
  3. calculate desired endpoint force and send to motors
- sometimes need to model & compute device dynamics to compensate for dynamics in device

ungrounded haptic interfaces

body-mounted

- joint angles are measured relative to mount point
- pros: follow *you* around the world
- cons:
  - forces push against body, not world
  - donning/doffing
  - user usually bears full weight of device
  - one-size-doesn’t-fit-all
  - power / tethering
- many other ungrounded *tactile* displays
  these are just force feedback

example of an **ungrounded** mechanism:

Virtual Technologies' (Immersion now) Cybergrasp

MacLean: Physical Interaction Design 83

haptic actuators

force / torque

motion or motion arrest

vibration or impulse

temperature

pressure - *inflate/deflate/vibrate*

touch – *make/break physical contact*

MacLean: Physical Interaction Design 84

An "actuator" is any device that transmits some sort of power. Haptic displays are designed to many different types, each operating on a different parameter.

common actuator technologies
for haptic displays

- electric motors – DC brush most common
- brakes - stable!
- pneumatic – no messy fluids + fast;
  springy + require compressed air source
- hydraulic – strong, fast, valves nonlinear.
  large systems: safety issues + nasty liquids
- SMA – shape memory alloy. tiny, fragile, tricky cooling requirements but can be densely packed.
- piezo – small displacement, strong / fast

These are some of the technologies used for haptic interface actuators, to create the effects listed above.

sensing:
what is there to sense?

pick your body part:

contact

position, velocity, acceleration

applied force

pressure (squeeze, press)

type of grasp

temperature

Likewise there are many different user parameters that a physical interface might need to know, and many technologies used to sense them.

computer haptics: rendering force feedback

MacLean: Physical Interaction Design 87

Today, rendering is where most practitioners will enter the haptic field: you'll buy or borrow a display, and program it to do something. We'll talk a little about rendering works.

basic procedure for haptic rendering

1. read device position
2. determine which regime we're in …
   changes triggered by, e.g.
   - penetrating a virtual object
   - user- or application-triggered events
3. calculate display force based on mode & position
4. send corresponding force command to motors
5. inform other parts of the system of changes

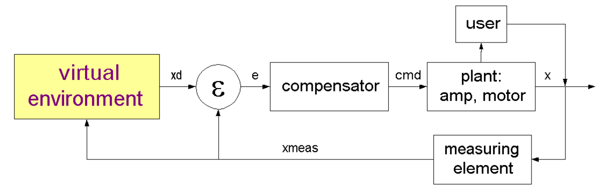
MacLean: Physical Interaction Design 88

A rendering algorithm will generally include a code loop that looks like this, at minimum. The parts which may be most unfamiliar to a graphics designer could be (a) the strict temporal requirements – the loop must execute in about 1-2 milliseconds – and (b) the process of reading and writing to the computer's I/O ports, to get sensor data and output actuator commands. Depending on the computer and software architecture, the latter might happen in many different ways.

simple rendering:
rigid bodies & linear building blocks

- rendering an attached rigid body
- blitzkrieg on Physical Modeling
- rendering a wall
- tricks
- two-body dynamic systems
- rendering textures

where we are in the system...



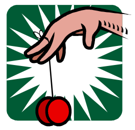
Compare this diagram with the one for Closed Loop Control earlier. The main difference is the presence of the Virtual Environment block at the left. The VE takes as input the measured system state, and supplies a desired state. This is now what's used to create the error signal.

The compensator is the "controller", which tries to drive the error signal to zero. That's where the PID controller goes. But when we talk about haptic rendering, we're usually referring to what's in the VE box.

direct manipulation of a rigid body

rigid body

dynamic system



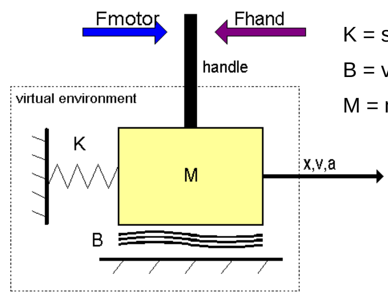
- rigid body: the VE pretends the user is holding directly onto a fictional *rigid body* through the handle.
- dynamic system: the VE pretends the user has a more complex relationship with an *independent dynamic system* (also fictional)

MacLean: Physical Interaction Design 91

One of the simplest virtual environments to simulate is a rigid body; whereas a dynamic system can be arbitrarily complex. The main difference between the two is the number of inertial objects able to move relative to one another, and how their motion is coupled to the user.

background: $F_{on\_handle} = F_{hand} + F_{motor}$

F = ma $F_{motor} = [K(x_o - x_m) + Bv + Ma]_{VE}$



K = spring force

B = viscosity or damping

M = mass or inertia

MacLean: Physical Interaction Design 92

When haptically displayed, this **rigid body system** feels like you're holding something heavy – like a brick attached to a spring – and pushing it back and forth. There's a compliance in the system, but only one moving object with inertia, and it is directly coupled (no relative motion) to the position of your hand.

rendering a virtual mass-spring-damper

- F_{mot} is applied to user's hand (effort)
- user adjusts position in response (flow)

→ x changes and is re-measured

→ F_{mot} (x, v, a) is recalculated.

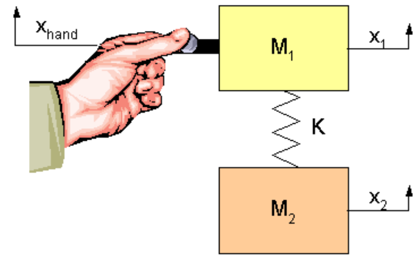
- first you must differentiate position to get velocity, acceleration (noisy)
- or, you can measure acceleration from an accelerometer for a smoother estimate.

MacLean: Physical Interaction Design 93

How do you render it? This is perhaps conceptually the simplest, although it doesn't necessarily work the best.

next: 2-body dynamics

a virtual yo-yo



MacLean: Physical Interaction Design 94

Now, here's the way it's more commonly done. In a "two-body" dynamic system, there are multiple masses moving relative to one another. The user will feel the motion of M2 indirectly, in the same way you can feel the position of a yo-yo through the connecting string.

We won't go very far into this here, but the system modeling part is covered in any engineering dynamics textbook (e.g. Ogata). In the diagram here, both M1 and M2 are virtual, as is the spring K. They are part of a computational model updated each servo cycle, with the new position of M1 and M2 based on measured xHand.

In this system, the position of M1 is constrained to xHand; and very often, M1 << M2 and can be neglected altogether. Now, you don't need to know x1's acceleration

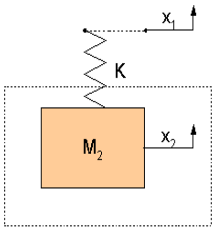
steps to rendering a 2-body dynamic system:

1. simulate virtual forces imposed on a virtual mass, created by the compression of a virtual spring displaced by user's hand motion
2. compute acceleration of virtual mass
3. integrate twice to get its virtual position
4. use the new virtual position to compute the new force in the spring
5. apply this back to the hand

The main idea in modeling this VE is to look at all the imaginary forces acting on the imaginary M2, compute its kinematic state as a result, then go back and compute the effect of this imaginary object's imaginary motion on M1 (essentially the user's hand).

draw a box around m2 and look at forces

(a "free-body diagram"):



$$\sum external\ forces = body\ forces$$

$$0 = M_2\ddot{x}_2 + K(x_2 - x_1)$$

$$\ddot{x}_2 = \frac{K(x_2 - x_1)}{M_2}$$

$$and, \quad x_1 = x_h = x_{measured}$$

$$\dot{x}_2 = \int \ddot{x}_2 dt$$

$$x_2 = \int \dot{x}_2 dt$$

Here's the equations. We've isolated M2 and shown the forces that go through the box: here it's just the spring force imposed by M1 (F=K(x2-x1). Body forces for this system consist only of gravity (F=ma).

(Notation: xdotdot=acceleration; xdot = velocity)

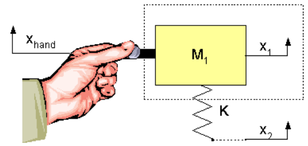
working back to the hand…

So that's how we simulate the virtual mass's position (M_2) based on the hand's position.

But, what we really want to know is,

What force should we apply to the hand?

free-body diagram on M_1:



$$\sum external\ forces\ = body\ forces$$

$$F_{hand} = M_1\ddot{x}_1 + K(x_2 - x_1)$$

$$let's\ say\ M_1\ is\ very\ small \rightarrow neglect$$

$$F_{hand} = K(x_2 - x_1)$$

$$at\ least,\ this\ is\ what\ we\ want\ the\ user\ to\ feel.$$

$$\rightarrow F_{motor} = K(x_2 - x_1)$$

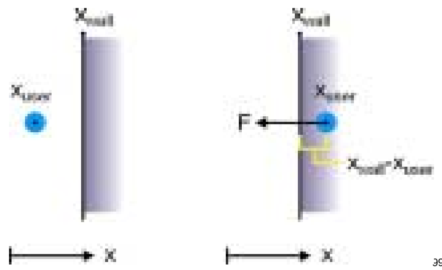
For the final step, we go back to M1 and compute the effect of M2 on it. Use the same procedure as for M2, with the free body diagram, and note that there's the same spring force except in the opposite direction.

In the end, we get Fmotor = K(x2-x1). This should look familiar! It is just like a proportional controller. And in fact, that is exactly what is going on here in the end: we are simulating a mass in motion to generate the desired position, and driving the hand's position to equal it.

rendering a wall

If $x_{user} > x_{wall}, F = K(x_{wall} - x_{user})$

stiffness $K > 0$



99

A wall is a very common thing to want to render; it's not quite like either of the other systems. For one thing, it's "nonlinear", involving two different modes (contact and non-contact) and a discontinuity between them.

In non-contact mode, your hand is "free" (although holding the haptic display), until it encounters the virtual object. It feels like you tapped something through the device handle. This is fundamentally different from the other rigid body simulation described earlier: in that one, you were holding on to the rigid body, here you bump into it. What makes them them similar in terms of rendering is that in both cases, only one "inertia" is modeled.

tricks:
1-way damping to increase passivity

- a pure spring force for a wall may seem "active" (jittery)

$\rightarrow$ add a dissipative term, where *B* is the damping coefficient

- only damp when going *into* the wall

$$F = \begin{cases} K\Delta x + B\dot{x} & \text{for } \dot{x} > 0 \\ K\Delta x & \text{for } \dot{x} < 0 \end{cases}$$

MacLean: Physical Interaction Design 100

Rendering a wall is actually one of the hardest tests of a haptic display system even though it's conceptually so simple. This is because you're transitioning from a regime with very low (in fact, zero) stiffness to a regime with very high stiffness – usually the highest the system can manage, with nothing in between. This has two problems.

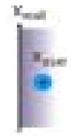
1. It is what we call a "stiff system" and is computationally unstable – also physically unstable! Rendering can lead to instability
2. For all that, it might not even feel very "hard". This is because when you first enter the wall, the force you're feeling isn't very large: K(xwall-xuser) is close to zero when you're near the surface. With a large K, Fwall rises quickly but that still doesn't give you a nice crisp feel of contact.

This slide shows a trick that helps to solve Problem 2. When you're entering the wall, you apply damping – that is, a term proportional to velocity. This means you have a large Fwall right at entry if you're moving with any speed. You have to turn off this term when you're backing out of the wall, or else it will feel sticky.

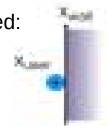
tricks: visual / aural augmentation

- visual: never show the point penetrating the surface, even if it is
- aural: play a crisp contact sound on contact

actual:



displayed:



psychophysical studies have shown that this makes the surface appear stiffer/harder

Here’s another trick to make the wall seem harder: a very strong illusion involving multimodal interaction. Adding either visual or aural reinforcement of the wall’s entry make the wall seem much harder than if you relied on haptic feedback alone.

This is nice, because if you have a visual / auditory simulation anyway, then now you can get by with a cheaper haptic display (it’s expensive to build a haptic display capable of very high stiffness).

rendering textures

- typical haptic devices display shape very well, but don’t feel “realistic”:
  everything is smooth and slightly spongy
- rendering material surface properties increase model realism
- method depends on
  – surface model
  – complexity of the surface property

Textures are a whole different kind of rendering. They may or may not be based on a realistic physical model.

what kinds and what methods?

make these feel different:

- plastic
- steel
- glass
- rubber
- foam
- fur ???

using:

- hardness
- height maps
- spatial and/or temporal functions
- dynamic methods: e.g. damping
- friction: many models

We'll go over methods to render each of the texture qualities in the right column…

realistic or imaginary? (the usual dimension)

you can try to reproduce surface properties recognizable from the real world - "high fidelity":

- e.g. surgical simulations

or you can try something new:

- e.g. computer music handles

another familiar dimension: to base physically or not?

physically based models:

- at least begin with a physically accurate model for the haptic sensation
  - e.g. compute a wave equation for a vibrating violin string
  - then (if you want) tweak the parameters wildly to make something that couldn't exist in the real world!

or, just go for an interesting equation or shape

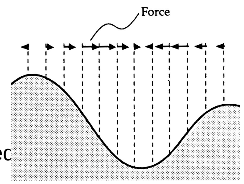
which may or may not try to achieve perceptual fidelity with something real.

hardness

this just means stiffness…

- $F = Kx$
- an important perceptual component of any surface rendering
- augmented by one-way damping and/or visual/auditory reinforcement

height maps: bumps



Minsky, 1990

basic idea:

- as a user moves "up" a bump, motion is opposed (energy hills and wells)
- spring force ∝ to "height" of bump

force vs. position gradients:

- force gradient: user's hand never moves out of level plane
- position: actually move up and down. requires at least 2df display.
- feel very similar

MacLean: Physical Interaction Design 107

how are bumps calculated? spatial or temporal functions

evaluate a linear or nonlinear equation:

- e.g. sinusoid (wave.exe example)

$$Fdes = A\cos(\omega x)$$

- or, any other shape.

spatial: $Fdes = \frac{f(x)}{dx}$

temporal: $Fdes = \frac{f(t)}{dt}$

a deterministic method:

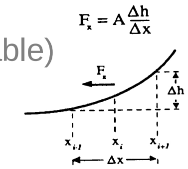
get same value every time for same input

or both: $Fdes = \frac{f(x,t)}{dxdt}$

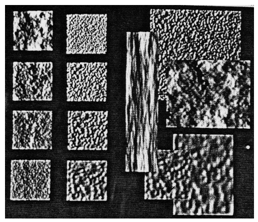
MacLean: Physical Interaction Design 108

or,

empirical (i.e. lookup table)



- hand code a lot of values
- or, use a textured picture:
  - dark is low, light is high
  - then compute local gradient:



(Minsky, 1990)

MacLean: Physical Interaction Design 109

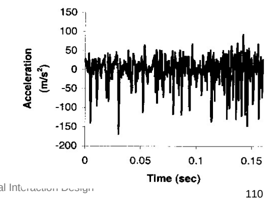
Margaret Minsky's "Virtual Sandpaper" system was one of the first to use the height map method, based on textured pictures of different materials.

stochastic models

- evaluate a randomly valued function with some structure
- derive stochastic parameters (amplitude, band limits) by analyzing actual force data

nondeterministic:

- random result at each point.



MacLean: Physical Interaction Design 110

dynamic methods:
e.g. simple damping

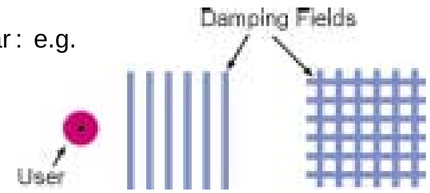
if $\vec{p}_{user}$ < inside a damping area

$F_{des} = B\vec{v}_{user}$

$(\vec{v}_{user} = \dot{\vec{p}}_{user})$

or, nonlinear: e.g.

$F_{des} = Bv^2$

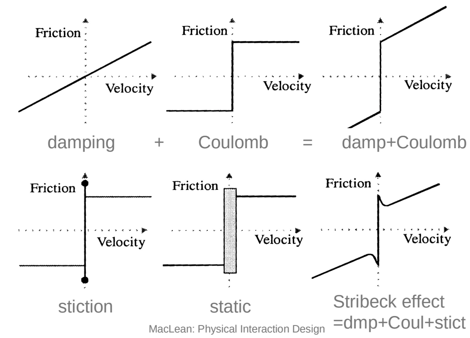


To achieve a viscous texture, compute some relation to velocity. These fields can be unidirectional (feels like you’re moving through a river with a current in one direction) or bidirectional (resists motion no matter which direction you go).

displaying friction
hard to render:

- non-linear
- discontinuities at low velocities

(Richard, Cutkosky & MacLean, 1999)



There are many ways to model friction, most of which have been tried with haptic feedback. These include the bristle model, the Dahl and Karnopp models. (Richard 1999) provides more information on these.

areas of basic research

- hardware
- rendering
- uni- and multimodal cognition & psychophysics
- networking

MacLean: Physical Interaction Design 113

That's it for techniques. We'll finish up by reviewing the hot current research areas and problems. As pointed out in the beginning, these follow the lines of machine, computer and human haptics.

research areas: hardware

- current constraints
  (actuators need more help than sensors):
  - cost
  - size
  - weight for performance
  - robustness
  - controllability
  - bandwidth
- → need to develop novel methods for force and pressure display and sensing.
- tactile interfaces especially demanding…

MacLean: Physical Interaction Design 114

research areas: rendering
how to make it feel like something

unimodal haptic rendering:

- techniques for fast, smooth, stable display of models
  - 2- and 3-D
  - dynamic / static
  - literal / abstract
  - many researchers

multimodal rendering research examples:

- physically based haptic/auditory acoustic modeling
  Pai et al, UBC
- “fast” perception: display to automatic, reflexive sensory system.
  Rensink & MacLean, UBC

research areas: communication
what can feels mean?

- lowest level: design and learn people’s ability to distinguish & associate complex “hapticons”
- highest level: people and machines communicate with a two-way haptic language
- in between: a lot of work!

research areas: adaptive interfaces

- e.g. multimodal application feedback that adjusts itself to the user's context:
  - emotional state (upset, excited, tired?)
  - attention
  - learning level
- input: biometric sensing

the future (?)

- improved environment modeling
  physically based & empirical
- tactile feedback mechanisms
- novel force actuation & sensing
- immersive virtual reality systems
- integrated multimodal feedback
- haptic language
- *ubiquitous haptics*

getting going

- who does it?
- where is the haptics community?
- what's for sale?

who can do it?

haptic feedback design is pretty multidisciplinary:

application interface design (conceptualizing it)
- application immersion
- interface creation & integration
- concept prototyping

human side (perceiving and interpreting it)
- perceptual psychology
- cognition
- user experimentation & analysis
- biomechanics & kinesiology

machine side (making it happen)
- multisensory display design & control
- realtime software architecture design
- rendering algorithms
- physical system modeling

different approaches to multidisciplinary, application-driven design

going it alone: you have a choice of

- depth: psychophysics, mechanism design, rendering
- breadth: integration

another way: collaborative research
among academics & within companies

MacLean: Physical Interaction Design 121

I want to learn more: where is the haptics community?

conferences & publications

- Haptics Symposium ('92-present)
  http://www.VR2002.org/
- haptics-e: The Electronic Journal of Haptics Research
  http://www.haptics-e.org/
- some selected readings
  http://www.cs.ubc.ca/~cs532/references/readings.html

community websites

- the Haptics Community Webpage: http://haptic.mech.nwu.edu/
- Haptics-L: the electronic mailing list for the int'l haptics community
  http://www.roblesdelatorre.com/gabriel/hapticsl
- the Eurohaptics Community: http://www.eurohaptics.org/

MacLean: Physical Interaction Design 122

what's for sale?

I have a great idea for a haptic feedback application, but I don't want to build my own haptic display.

what can I buy? here's a start:

- Sensable Technologies: http://www.sensable.com/
  6-degree of freedom desktop displays ($8K USD and up)
- Immersion Inc.: http://www.immersion.com/
  entertainment & medical devices; embedded automotive apps
- Reachin Technologies http://www.reachin.se/
  integrated haptics/graphics VR systems (based on Sensable)
- ACT Labs: http://www.act-labs.com/
  high-end force feedback steering wheel ($170 USD)

References:
Physical Interaction Design

For links to many of the publications listed here, see
http://www.cs.ubc.ca/spin/publications/index.html

[1] Burdea, G., Force and Touch Feedback for Virtual Reality: John Wiley & Sons, 1996.

[2] DiFranco, D. E., Beauregard, G. L., and Srinivasan, M. A., “The effect of auditory cues on the haptic perception of stiffness in virtual environments,” in Proc. of the 6th Ann. Symp. on Haptic Interfaces for Virtual Environments and Teleoperator Systems, ASME/IMECE, DSC-Vol. 61, pp. 17-22, 1997.

[3] Gillespie, B. and Cutkosky, M., “Stable user-specific haptic rendering of the virtual wall,” in Proc. of Proceedings of the ASME Dynamic Systems and Control Divisio, DSC-Vol. 58, pp. 397-406, 1996.

[4] Gillespie, B., O'Modhrain, S., P. Tang, C. P., and Zaretsky, D., “The Virtual Teacher,” in Proc. of Proceedings of the ASME Dynamic Systems and Control Division, DSC-Vol. 64, pp. 171-178, 1998.

[5] Goldfarb, M. and Durfee, W., “Design of a controlled-brake orthosis for FES-aided gait,” IEEE Trans. Rehab. Eng., vol. 4:1, pp. 13-24, 1996.

[6] Goldstein, E. B., Sensation and Perception, 5 ed. Belmont, CA: Wadsworth Pub. Co, 1999.

[7] Heller, M. A., Calcaterra, J., A., Green, S. L., and Brown, L., “Intersensory conflict between vision and touch: the response modality dominates when precise, attention-riveting judgments are required,” Perception & Psychophysics, vol. 61, pp. 1384-1398, 1999.

[8] Hershberger, W. A. and Misceo, G. F., “Touch dominates haptic estimates of discordant visual-haptic size.,” Perception & Psychophysics, vol. 58, pp. 1124-1132, 1996.

[9] Hollerbach, J. M. and Jacobsen, S. C., “Haptic interfaces for teleoperation and virtual environments,” in Proc. of First Workshop on Simulation and Interaction in Virtual Environments, Iowa City, 1995.

[10] Hollerbach, J. M. and Johnson, D. E., “Virtual Environment Rendering,” in to appear in Human and Machine Haptics: MIT Press, 2000.

[11] Immersion Corporation, The Immersion I-Feel Mouse & I-Force Game Controllers. San Jose, CA, 2000.

[12] Jacobsen, S. C., Smith, F. M., Backman, D. K., and Iversen, E. K., “High performance, high dexterity, force reflective teloperator II,” in Proc. of ANS Topical Meeting on Robotics and Remote Systems, Albuquerque, NM, 1991.

[13] Kandel, E. R., Schwartz, J. H., and Jessell, T. M., “Principles of neural science,”,, 4th ed. New York: McGraw-Hill, 2000.

[14] Klatzky, R. L., Lederman, S. J., and Matula, D. E., “Haptic exploration in the presence of vision,” Journal of Experimental Psychology: Human Perception and Performance, vol. 19, pp. 726-743, 1993.

[15] Kontarinis, D. A. and Howe, R. D., “Tactile display of vibratory information in teleoperation and virtual environments,” Presence, vol. 4, pp. 387-402, 1995.

[16] Krebs, H. I., Hogan, N., Aisen, M. L., and Volpe, B. T., “Robot-Aided Neuro-Rehabilitation,” IEEE - Transactions on Rehabilitation Engineering, vol. 6:1, pp. 75-87, 1998.

[17] MacLean, K. E., “Designing with Haptic Feedback,” in Proc. of IEEE Robotics and Automation (ICRA'2000), San Francisco, CA, 2000.

[18] MacLean, K. E., Shaver, M. J., and Pai, D. K., “Handheld Haptics: A USB Media Controller with Force Sensing,” in Proc. of the IEEE VR2002 10th Symp. on Haptic Interfaces for Virtual Environment and Teleoperator Systems (HAPTICS 2002), Orlando, FL, 2002.

[19] MacLean, K. E. and Snibbe, S. S., “An Architecture for Haptic Control of Media,” in Proc. of the 8th Ann. Symp. on Haptic Interfaces for Virtual Environment and Teleoperator Systems, ASME / IMECE, Nashville, TN, DSC-5B-3, 1999.

[20] MacLean, K. E., Snibbe, S. S., and Levin, G., “Tagged Handles: Merging Discrete and Continuous Control,” in Proc. of ACM Conference on Human Factors in Computing Systems (CHI '2000), The Hague, Netherlands, 2000.

[21] Marks, L. E., “Multimodal perception,” in Perceptual Coding, vol. 8, Handbook of Perception, E. C. Carterette and M. P. Friedman, Eds.: Academic Press, 1978.

[22] Martino, G. and Marks, L. E., “Cross-modal interaction between vision and touch: the role of synesthetic correspondence.,” Perception, vol. 29, pp. 745-754, 2000.

[23] Massie, T. H. and Salisbury, J. K., “The PHANToM haptic interface: a device for probing virtual objects,” in Proc. of the Third Ann. Symp. on Haptic Interfaces for Virtual Environment and Teleoperator Systems, ASME/IMECE, 1994.

[24] McDonnell, P. M. and Abraham, W., “Observations on a new illusion demonstrating touch dominance of vision,” Perceptual and motor skills, vol. 46, pp. 1240-1242, 1978.

[25] McGurk, H. and MacDonald, J., “Hearing lips and seeing voices,” Nature, vol. 264, pp. 746-748, 1976.

[26] Minsky, M., Ouh-young, M., et al., “Feeling and seeing: issues in force display,” in Proc. of ACM Symposium on Interactive 3D Graphics, Snowbird, UT, 24, pp. 235-242, 1990.

[27] Morganbesser, H. B. and Srinivasan, M. A., “Force shading for shape perception in haptic virtual environments,” in Proc. of the 5th Ann. Symp. on Haptic Interfaces for Virtual Environment and Teleoperator Systems, ASME/IMECE, Atlanta, GA, DSC:58, 1996.

[28] Ogata, K., Modern Control Engineering. Englewood Cliffs, N.J.: Prentice-Hall, 1970.

[29] Richard, C., Cutkosky, M., and MacLean, K., “Friction Identification for Haptic Display,” in Proc. of the 8th Ann. Symp. on Haptic Interfaces for Virtual Environment and Teleoperator Systems, ASME/IMECE, Nashville, TN, 1999.

[30] Rock, I., Perception. New York: Scientific American Library, 1984.

[31] Sensable Technologies, PHANTOM Desktop 3D Touch System. Cambridge, MA, 1998.

[32] Snibbe, S. S., MacLean, K. E., et al., “Haptic Metaphors for Digital Media,” in Proc. of ACM Symp. on User Interface Software & Technology (UIST 2001), Orlando, FL, 2001.

[33] Spence, C. J. and Driver, J., “Cross-modal links in attention between audition, vision and touch: implications for interface design.,” International journal of Cognitive Ergonomics, vol. 1, pp. 351-373, 1997.

[34] Srinivasan, M. A., Basdogan, C., and Ho, C. H., “Haptic Interactions in Virtual Worlds: Progress and Prospects,” in Proc. of Proceedings of the International Conference on Smart Materials, Structures, and Systems, Bangalore, India, 1999.

[35] Srinivasan, M. A., Beauregard, G. L., and Brock, D., “The Impact of Visual Information on the Haptic Perception of Stiffness in Virtual Environments,” in Proc. of the 5th Ann. Symp. on Haptic Interfaces for Virtual Environment and Teleoperator Systems, IMECE, Atlanta, GA, DSC:58, 1996.

[36] Tan, H. Z., Pang, X. D., and Durlach, N. I., “Manual Resolution of Length, Force, and Compliance,” in Proc. of the 1st Ann. Symp. on Haptic Interfaces for Virtual Environment and Teleoperator Systems, ASME/IMECE, 42, pp. 13-18, 1992.

[37] Tan, H. Z., Srinivasan, M. A., Eberman, B., and Cheng, B., “Human Factors for the Design of Force-Reflecting Haptic Interfaces,” in Proc. of the 3rd Ann. Symp. on Haptic Interfaces for Virtual Environment and Teleoperator Systems, ASME/IMECE, Chicago, IL, DSC:55-1, 1994.

[38] Tipper, S. P., Lloyd, D., et al., “Vision influences tactile perception without proprioceptive orienting,” NeuroReport, vol. 9, pp. 1741-1744, 1998.

[39] van den Doel, K., Kry, P. G., and Pai, D. K., “FoleyAutomatic: Physically-based Sound Effects for Interactive Simulation and Animation,” in Proc. of SIGGRAPH 2001, 2001.

[40] Wu, W.-C., Basdogan, C., and Srinivasan, M. A., “Visual, Haptic and Bimodal Perception of Size and Stiffness in Virtual Environments,” in Proc. of the 8th Ann. Symp. on Haptic Interfaces for Virtual Environment and Teleoperator Systems, ASME/IMECE, Nashville, TN, DSC:67, pp. 19-26, 1999.

[41] Zilles, C. B. and Salisbury, J. K., “A Constraint-based God-Object Method for Haptic Display,” in Proc. of IEEE Conference on Intelligent Robots and Systems (IROS '95), 1995.

Multimodal Design & Technologies

Sidney Fels

SIGGRAPH 2002

THE UNIVERSITY OF BRITISH COLUMBIA

MAGIC

These notes contain a superset of the slides that are presented at Siggraph 2002. The set of slides are extracted and modified from my course on Human Interface Technologies (HIT) at the University of British Columbia (http://www.ece.ubc.ca/~elec596).

Overview

- Introduction
- Human I/O: A look at the body and mind
  - human information processing
- Bringing Modalities Together
  - Intimate and Embodied interfaces
- Summary

This section of the course starts by looking at the human information processing system. I begin with the most understood systems, vision and audition, to the least, gustation and olfaction. It is through understanding the varieties of the human mechanisms that we appreciate the way to integrate them into a multimodal, interactive graphics system. The human body and mind works in complementary ways so that information received from various modalities combines together and compensates or supports missing information or contradictory information. It is for these reasons that multimodal interfaces are so appealing. We can create illusions through one channel to affect another or we can support performance by providing multiple paths of information facilitating attentional mechanisms.

Once the basics of human processing are understood we move on to issues for bringing the modalities together. I introduce a design framework based on intimacy and embodiment of the interface.

Introduction

- Virtual Environments
  - large, medium, and small scale
  - include people
- Communicating human experience
  - information, emotion, environment
  - people to people
  - people to machine
- Engagement of Body and Mind

Introduction

- Techniques needed for:
  - sensing, encoding, transmitting, storing, indexing, retrieving, compressing, recognizing and synthesizing
- Human body has many I/O channels
- Integrate Cognitive, Physical and Emotional aspects of interaction
- Interface should disappear

Human Information Processing

- Input
    - Visual channel
    - Auditory channel
    - Position and Motion Sensing Channel
    - Somatic Channel
    - Taste and Smell Channels
- Output
    - Intentional
        - neuromuscular, movable, verbal
    - Non-intentional
        - GSR, Heart Rate, Brain, Muscle, other

One way to decompose the human system is in terms of input and output. Some channels have both. Likewise, within the output channels some are volitional (intentional) and some are not (biopotentias). Each are opportunities to create new, multimodal interfaces that take advantage of the ability of each channel.

Readings: #1, #2, #3, #5, #6

Human Information Processing

- Decisions
    - Tracking
    - Memory
    - Learning
    - Indviduals vs. Groups

The mind also has unique features that can be exploited in interface design.

Driving Trends in Interface Technologies

- Virtual Reality, Immersive Environments, AR
- Ubiquitous computing/Intelligent Environments
- Wearable Computing, Tangible Bits,
- Games, Arts, Interactive Theatre, Interactive Art
- WWW, Agents, Collaborative work

Lately, many research areas are driving multimodal research top down.

Readings: #4, #7, #8, #11, #20, #27, #28

Visual Channel

- Senses electromagnetic radiation (wavelength = 0.3-0.7 microns)
  - Cones: fovea (600um), high resolution, independent
  - Rods: periphery, low resolution, integrated
- 2 eyes for binocular vision
- 100,000 fixation points (100deg circular)
- Types of eye movement (six muscles):
  - compensatory (must have target)
  - pursuit (must have target)
  - Tremor, flick and drift
  - saccadic (jump from one fixation to another)

Let’s start with the visual channel. It is the most understood and is directly relevant to graphics.

Visual Channel

- Rods and Cones
  - Fovea is all cones (6 million)
  - Periphery is mostly rods (125 million)
  - interleaved
- Rods activate neurons in groups
  - higher sensitivity less resolution
- Cones are more one-to-one
  - lower sensitivity more resolution

The eyes are amazing!

Visual Channel: Movement

- 8-10Hz gives sensation of motion.
- 5 ways to make a light move
- Familiarity helps interpret movement
- Movement implies life
- Movement links images (strongly)

This is how to make movies.

Visual Channel

- Seeing depth:
  - Binocular vision:
    - Disparity can be used to determine distance
    - Frontal plane horoptors (Helmholtz)
    - non-euclidean space
  - Other cues:
    - overlap, relative size, relative height, atmospheric perspective, texture gradients, parallel line convergence, motion parallax, accommodation and convergence
- Seeing size:
  - size constancy

But what if we want 3D? Much research in this area along with products for seeing 3D. What is interesting is that there are many sources of information that the brain uses to determine depth. Each of these can be used in a multimodal interface and linked with a separate channel. Likewise, it is smart to integrate more than one cue. This is why stereo and motion parallax together makes sense.

Visual Channel

- Colour perception is very complicated - refer to readings
- Adapts to light conditions
- lots of illusions to play with size and distance
- other interesting things:
  - retina is reflective
  - eye blink does not affect perception
  - pupil is normally black and circular
  - attention and gaze direction are correlated
  - people wear glasses and/or contacts

I'm not covering colour since it is such a large area on its own.

One thing to remember, illusions are your friends. When you encounter an illusion as a designer, you should be thinking, "how can I take advantage of this?" You want to think this since illusions are windows into how the brain works and the assumptions it makes about the real world. The right illusion coupled with the appropriate cues from the other sensory channels can give a powerful sense of immersion.

Visual Displays

- Main Issues
  - field-of-view
  - resolution
  - update rates, animation
  - stereopsis
  - perspective
- Main Technologies
  - CRTs, LCDs, projectors
  - HMDs, CAVEs, Cubby, FishTank VR, VRDs

Of course, we need displays for our eyes so we can see something.

Visual Displays: Issues

- Field of view vs resolution
  - f.o.v. = total angular deviation that can be seen
    - horizontal 180º (no eye movement), 270º (with eye movement)
    - vertical 120º
  - need about 8000x8000 pixels for full resolution
  - tradeoff f.o.v with resolution using lenses
  - low resolution ->
    - lack of sharpness
    - can see individual pixels
- 60-100º for immersion

Visual Displays: Issues

- Update rate
  - Eye detects flicker below 50-60Hz
    - critical fusion frequency
    - changes with age
    - rods more sensitive than cones
    - function of illumination
  - Animation
    - less than 10-15Hz will result in non-cts motion
- typical video update rates are 50Hz (PAL) or 60Hz (NTSC)
  - 4:3 aspect ratio, HDTV 16:9
- interlaced vs. non-interlaced (progressive)

Visual Displays: Issues

- Stereopsis
  - present two different perspectives for each eye
  - Inter Pupilar Distance (IPD)
    - focal point can change IPD
- Perspective
  - "walk around" effect
  - multiple people should see from different perspectives

Visual Displays: Technologies

- Cathode ray tube (CRT)
- Liquid Crystal Display (LCD)
- Head Mounted Displays (HMD)
- Projectors
  - CRTS
  - LCDs
- Virtual Retinal Display (VRD)

Fels: Design of Interactive Multimodal Graphics

These are some of the main visual display technologies out there.

Visual Displays: CRT

- Scanning electron beam strikes phosphors
  - blue-green phosphors use zinc sulphide
    - longer persistence
  - red phosphors do not
- video streams are interlaced
  - 525 vertical lines per *frame*; odd/even (60Hz - NTSC)
  - 625 lines/frame (50Hz - PAL)
- CG data is non-interlaced (progressive scan)
  - various resolutions (EGA, VGA, SVGA, XVGA)
    - I.e. 1280x1024 pixels for SVGA
    - also include number of bits/pixel
  - various update rates 60-120Hz

Fels: Design of Interactive Multimodal Graphics

Visual Displays: Stereo Tech.

- Stereo systems
  - shutter glasses/ “flicker glasses”
    - actively alternate LCD panel to switch between right and left eyes
    - crosstalk is a problem
      - green phosphors are main culprit
    - Crystal Eyes
  - passive
    - polarized left and right images
      - use glasses
    - red/green glasses
      - crosstalk is a problem not to mention colour problems
      - check out history of cost/quality tradeoff

Visual Displays: Stereo Tech.

- Autostereo Systems
  - No glasses needed
  - limited viewing position
    - left/right - use tracking to move active area
    - back/front difficult
  - Fair amount of research but few products
    - Stereoscopic Displays and Virtual Reality Systems VI Proceedings of the SPIE (1999)
      - 9 papers on autostereoscopes

This is a growing, hot area. Some is coming to siggraph: SIGGRAPH 2000: Perlin’s group’s autostereoscop (http://www.mrl.nyu.edu/projects/autostereo/), SIGGRAPH2001 (SynthaGram Monitor: http://www.inition.co.uk/inition/product_stereovis_stereographics_synthagram.htm) to name just two.

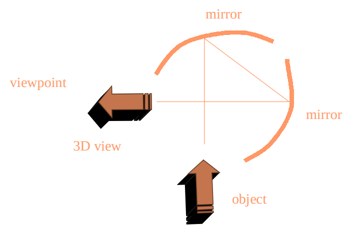
Visual Displays: Autostereoscope

- Parallax systems
  - thin barrier
  - backlit LCD (Dimension Systems Inc.)
- Lenticular Lenses
  - 3D postcards
  - Philips and others
- Holographic Optical Elements (HOE)
  - RealityVision, Trayner and Orr
- Dual Concave Mirrors
  - Dimensional Media
- holography, lasers...

Visual Displays: Lenticular Lens

This is like the 3D postcard and the 3D toy in the Cracker Jack box.

Visual Displays: Concave Mirrors



There is a toy that uses this effect. You put a quarter at the bottom and it looks like it is floating in the air ready for you to grab.

Visual Displays: LCD

- LCD = liquid crystal displays
  - two polarizers,
  - two pieces of glass,
  - some form of switching element or electrode to define pixels, and
  - driver Integrated Circuits (ICs) to address the rows and columns of pixels.
  - TN (Twisted Nematic)/ TNFE (Twisted Nematic Field Effect)
- LCD need separate light source
  - can make projectors easily
- Active vs. Passive
  - active uses diode + extra capacitor to isolate charge
  - active switching times much faster

Visual Displays: LCD

```
Cross Section of a Simple LC Display viewer

/////////////////////////////////////  Polarizer
_____________________________________  glass
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~  Liquid
                                       Crystal
_____________________________________  glass
\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\  Polarizer
        backlight
```


Visual Displays: LCD vs CRT

- CRTs generally have more
  - brightness
  - contrast
- LCDs have
  - smaller footprint and weight
  - less power
- CRT currently cheaper than similar LCD

Visual Displays: Virtual Retinal Display (VRD)

- scan light directly onto retina
    - no need for screens
- Motivated from scanning laser ophthalmoscope (SLO)
    - technology to get picture of retina
- Work done at HITL at U. of Washington
    - Now done by Microvision

This is very exciting work.

Readings: #23, #26

Visual Displays: VRD

- Currently use VGA video source (640x480)
- Currently using lasers for desktop version
    - argon for blue and green
    - laser diode for red
        - used for luggable version
- Control and drive circuits
    - direct modulation of laser diode
    - indirect modulation of argon source
        - acoustic-optical modulator (AOM)

Visual Displays: VRD

- Scanned onto retina using:
  - mechanical resonant scanner (MRS) for horizontal
  - galvanometer for vertical
- scan width for each pixel
  - 40nsec on retina (retinal)
  - no persistence
- scan loops instead of flyback
  - 60Hz interlaced
- final scanned beams exit through a lens
- user puts eye at exit pupil of VRD to see image
- Total: 307,200 spots of non-persistent lights

Visual Displays: VRD

- images are
  - perceived without flicker,
  - perceived to have vibrant color,
  - able to be seen both in occluded or augmented viewing modes
- extremely small exit pupil
  - large depth of focus

Visual Displays: VRD

- large colour range
- theoretical resolution limits set by eye
- f.o.v: 50º horizontally by 40º vertically.
- Luminance should be safe
  - 60-300nW for perceived equivalent brightness
  - 3-4 times less power than CRT
- better contrast ratio than CRT
- better depth of focus
  - like a pinhole camera
- Low power consumption (if using laser diodes)
- Theoretically very cheap

Visual Displays: VRD

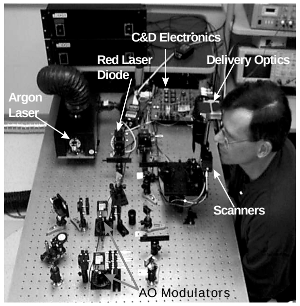
- Applications (theoretical)
  - replace CRTs and LCDs
  - low vision aid
    - exit pupil is very small (0.6x0.7mm)
    - eye doesn't need to focus
    - only uses small area of lens
      - corneal problems can be helped
    - can place image in places other than fovea
    - have some successful results

Visual Displays: VRD

- Issues unresolved:
  - not sympathetic to head movement
    - lose the image
  - Needs an argon laser
    - large and costly
    - Red, Green and Blue laser diodes are coming
  - Safety issues still not clear
    - coherent light vs. non-coherent
  - better resolution and larger f.o.v.
  - portable version
    - eye glasses

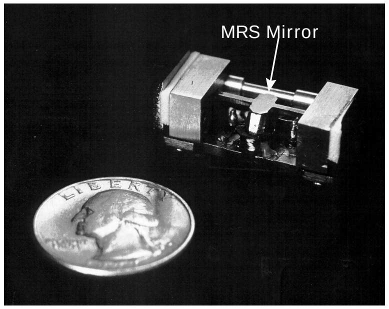
These problems should be considered opportunities for the HCI researcher. Head tracking and eye tracking are important not only for the VRD but for many multimodal interfaces.

Visual Displays: VRD

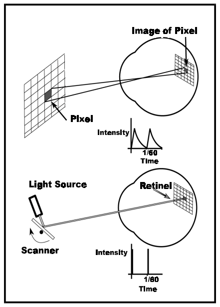


From HIT lab,
U. Of Washington

Visual Displays: VRD



Visual Displays: VRD



Visual Displays: Applications

- Fish-tank VR
- Cubby
- CAVE
- HMDs and VR
- Other systems:
  - ImmersaDesk

Visual Display: Fish Tank VR

- Stereo image of 3D scene on a workstation
  - binocular vision
- perspective projection coupled to head position
  - movement parallax cues
- Advantages:
  - resolution
    - 2 minutes of arc/pixel
    - simulate depth of field (working area is generally known)
    - stable w.r.t. eye movements
      - off axis eye movement effects are small (since monitor is far away)
    - virtual workspace does not exclude real workspace

See reference #24.

Visual Display: Fish Tank VR

- Which is most important for 3D effects
  - Arther, Booth and Ware, 1992
  - stereopsis or head coupling?
- Two experiments
  - experiment 1: Ss decide which is best 3D appearance
  - experiment 2: tracing tree paths
- Experiment 1:
  - 89% felt HC binocular (no stereo) is better than stereo only
  - Overall: HC binocular (no stereo) preferred 82% of time
    - possibly due to crosstalk as well

Notice that this is not exactly multi-modal, but it does illustrate that more information (that is consistent) is better and will be taken advantage of by the brain.

Visual Display: Fish Tank VR

- Experiment 2
  - HC + stereo best for understanding complex scenes
  - Practical implications
    - blood pathways
    - software structures
- Unresolved Issues:
  - when head moves what is best projection?
  - How to do non-invasive head tracking?
  - Large volumes?
  - Multiple viewers?

Visual Display: Fish Tank VR

- Applications:
  - VR workbench
  - scientific visualization
  - medical data visualization
  - interactive artwork
  - entertainment

Visual Display: Cubby

- What if we take Fish-tank VR and use 3 orthogonal screens?
  - Get CUBBY(J.P. Djajadiningrat, Royal College of Art, London)
- Uses 3 projectors arranged in a corner
- Uses headtracked monocular perspective
  - Delft Virtual Window System (DVWS)

This is a mini-CAVE. (see reference #9)

Visual Display: Cubby

- Non-immersive display
- DVWS ensures that manipulation space and virtual space can be unified
  - various experiments with stylus and virtual tip
    - compare with Ishii (1994)
    - other systems: Schmandt (1983), Kameyama (1993), Responsive Workbench (Kruger & Frohlich, 1994) and ImmersaDesk (Czernuszenko et al, 1997)
- head tracking technology is important
  - same point noticed in Fish-Tank VR

Visual Display: Cubby

- Advantages:
  - large movement space
  - excellent 3D effect using monocular motion parallax
  - compact
  - manipulation space and virtual space can be unified
- Disadvantages:
  - made for monocular viewing
  - non-immersive
  - small work space

Visual Display: CAVE

- What if we make the screens wall sized?
  - CAVE (CAVE Automatic Virtual Environment)
- Use 4 projection surfaces
  - left, front, right and bottom
- 3 rear projectors and one front projector
  - 1280x1024 pixels @ 120 fields/sec
  - special projector tubes for green
  - special synchronization for frames

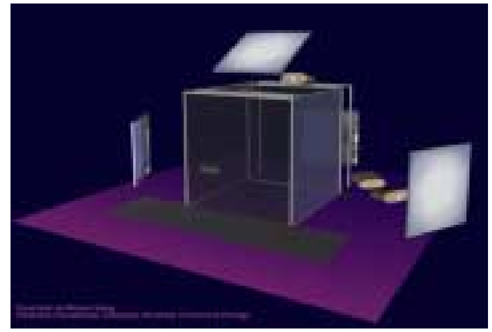
Fels: Design of Interactive Multimodal Graphics

Check out http://archive.ncsa.uiuc.edu/VR/cavernus/users.html to see how many of these type of environments are around. These are very compelling environments - what can multimodal scenarios do to make it better?

One major problem that still has to be sorted out is that there is usually a discrepancy between how much you move physically in a CAVE and how much the graphics scenery moves.

Readings: #25, #27

Cave Illustration: U. of Illinois



Fels: Design of Interactive Multimodal Graphics

Visual Display: CAVE

- Room sized CUBBY
- fit many people
  - whose perspective should be used?
- Fills complete field of view so very immersive
- Uses magnetic or sonic tracker for heads
- Uses stereo flicker glasses for stereopsis
- screens are calibrated to get rid of seam effects
- Uses window projection for two views
  - most systems use two camera perspective views

Visual Display: CAVE

- Disadvantages:
  - cost
  - size
  - interference
  - multiple view points
- Applications:
  - scientific visualization
  - art
  - other 3D visualizations

Visual Display: CAVE

- Other related systems:
  - spherical screens
    - VisionDome from eLumens)
      - spherical distortion needed
  - cheaper rhombic dodecahedron structure (Siggraph’97)
    - Garnet Vision - Hiroo Iwata

Visual Display: Head Mounted Displays (HMDs)

- Put one screen on each eye
- Typical for VR applications
- Very immersive
- Trades off field-of-view with resolution
  - various manufacturers make different tradeoffs
- Most use LEEP optical system
  - Large Expanse Extra Perspective (LEEP)

I know these seem a little old-fashioned now, but they are still an interesting way to get immersive effects and overcome many difficulties with other techniques. The are also reasonably well suited for multimodal interfaces. As for the technology, patience, patience, patience.

Visual Display: (HMDs)

- Advantages:
  - cost, size
  - immersive
  - cool
- Disadvantages
  - resolution
  - comfort
  - rotational error
  - motion sickness

Visual Displays: Applications

- CAD environments
- Interactive Art
  - Iamascope
  - World Skin
- Games
- scientific visualization
  - universe demo at U. of Illinois
  - chemical/biological modelling

Visual Displays: Products

- HMDs
  - over 60 different models
  - 24*768*3=2,359,296 pixel, Kaiser ProViewXL
  - 800*600*3=1,440,000 pixel, Sony Glasstron LDI-D100BE 640*480*3=921,600 pixe,l i-glasses Protec
  - 263*230*3=181,470 pixel(180k), VFX-1, i-glasses, CyberMaxx, Scuba, EyeTrek, Sony, LDI-D50BE
  - BOOM (Binocular Omni-Orientation Monitor) - Fakespace
- Shutter Glasses
  - over 100 models from <$100 to > $1000
  - CrystalEyes from StereoGraphics is probably the current winner
  - come with drivers, wired or not, dubious quality out there

These are constantly changing, the list here is just to indicate that it is an active area for product development. It also suggests that there is a market for these products.

Visual Displays: Products

- Passive Stereoscopes
  - Nuvision shutter screen add on or built in
  - Stereographics Z-Screen polarization-shutter-screen add-on
  - VRex Cyberbook - polarized laptop display
- Autostereoscopes
  - DTI - Dimension Technologies Inc. (parallax barrier)
  - Philips 3D-LCD (lenticular lenses)
  - Richmond Holographic Studios Ltd. (RHS) (holographic optical elements - HOEs)
  - Sanyo 3D Screen (image splitter with head tracking)
  - VISUREAL Displaysysteme GmbH (Holotron)
  - RealityVision (HOE)
  - Technische Uni Dresden - D4D (image splitting with head tracking)

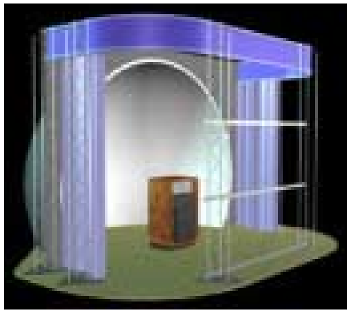
Visual Displays: Products

- Autostereoscopes ctd.
  - 3D EXPERIENCE LtdSpexfree 3D Monitor (looks like image splitting)
  - Dimensional Media Associates: high definition volumetric display (HDVD) - two concave mirrors + audio control optics
  - CRL (Kakeya et. Al.) use Fresnel lens plus comp. Cont. polarizing filters.
  - NYU, Perlin et al., dynamic parallax barrier using LCDs

Visual Displays: Products

- Cave - Fakespace
- ImmersaDesk - Fakespace
- Other
  - Cyberscope Virtual Reality Hood: (mirrors and divide screen in two)
  - n-Vision Virtual Binoculars handheld stereo display
  - visionDome - hemisphere projection and screen (like Imax)

VisionDome Illustration



Visual Displays: Summary

- Visual displays that are:
  - Low-cost,
  - immersive,
  - multi-person,
  - stereo,
  - head coupled, and
  - small

  do not exist.
- Opportunity for research

Auditory Channel

- Senses mechanical vibration of air molecules
  - sound travels about 350m/s (1260km/hr)
- Human range is 16Hz to 20,000Hz
- Pressure waveform causes hair cells to move (23,500 cells)
- Perceived loudness is approximately logarithmic
- Perceived sound is highly dependent upon environment
  - without reverberation unusual effects are noticed

Fels: Design of Interactive Multimodal Graphics

The auditory channel (your ears) can provide functions that complement the ears. It is really good at things the eyes are not good at. Read on...

This course will only cover the basics of audio as an interface modality. The readings cover many aspects of audio.

Readings: #29 - #50, #52, #53-#58, #62

Auditory Channel

- Ears can localize sound using:
  - Intensity difference between ears
    - better for high frequency (>1000Hz)
  - Time difference between ears
    - better for low frequency (< 1000Hz)
    - difference must be > 0.03sec (0.65s for complete localization)
  - Frequency profiles
    - ear shape (pinnae) acts as directional filters

Fels: Design of Interactive Multimodal Graphics

Audio: Benefits

- eyes free
- rapid detection
- alerting
- backgrounding
- parallel listening
- acute temporal resolution
- affective response
- auditory gestalt formation
  - trend spotting

With all this, what can't they do?

Audio Technologies: Audio Displays

- Disadvantages
  - low resolution
  - limited spatial resolution
  - lack of absolute values
  - lack of orthogonality
    - audio parameters not perceived independently
  - annoyance
  - interference with speech
  - not bound by line of sight
  - absence of persistence
  - no printout
  - user limitations

But these are alot of what the eyes are good at.

Audio: types of sound

- speech vs. non-speech
- musical vs. non-musical
- real vs. synthetic
  - – water drops vs. computer beeps

Fels: Design of Interactive Multimodal Graphics

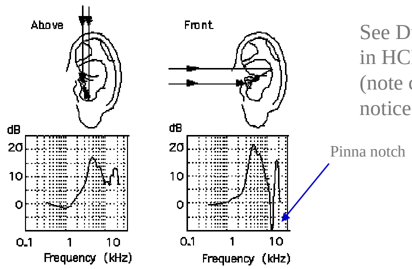
Audio: Spatialization

- Duplex Theory (azimuth)
  - – Interaural Time Difference (ITD)
    - ITD = a (θ + sin(θ)) $-\pi/2 <= \theta <= +\pi/2$
      θ is azimuth, ITD is up to 0.7msec
    - 10-20 degree accuracy
    - ears can detect change in position in ideal setting of 1 degree
      - – change in ITD of 10 microseconds (CD sample 22.7 microseconds)
  - – Interaural Intensity Difference (IID)
    - head shadow effect
    - high frequency attenuation up to 20dB

Fels: Design of Interactive Multimodal Graphics

Audio: Spatialization

- Elevation - monoaural and binaural
  - pinna is like acoustic antenna or directional filter
    - resonance, interference and directional sensitivity



See Duda, 1996/7 in HCI Course (#3) (note copyright notice in notes)

Richard Duda has been active in sound spatialization and HRTF work. I recommend reading his work on the area (#3).

The figures indicated for the spatialization section (pg 65-78) of this course come from http://www-engr.sjsu.edu/~knapp/HCIROD3D/3D_home.htm. Copyright 1996-7 by Richard O. Duda.

Audio: Spatialization

- Difficulty in reproducing directional sounds makes realism difficult in recordings
    - concert halls, stages etc.
- If pinna convolution could be simulated we could synthesize sound location
  - also a function of head and shoulder reflections
  - need Head Related Transfer Function (HRTF)
    - every person is different

Audio: Spatialization

- Range
  - poorest estimation
- Cues:
  - Loudness
    - power drops with range (square law)
    - need info. about sound source
  - Motion parallax
    - head movement changes azimuth
      - close source -> large change
      - far source -> small change
  - Excess interaural intensity difference (IID)
    - head shadow severe for close objects
    - single ear sounds (I.e. mosquitos) are threatening
  - Ratio of direct to reverberant sound
    - add reverb to help with range cues

Audio: Spatialization

- Reverberation issues:
  - 30-50msec delay is perceived as echo
  - precedence effect (Law of First Wavefront)
    - first arrival used for localization
    - check out Franssen effect
  - low-frequency info useless in highly reverberant spaces

You can try out a demo of the Franssen effect at:
http://www.parmly.luc.edu/parmly/franssen.html.

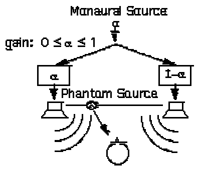
Audio Technologies: 3D audio

- Simple techniques for 3D audio
  - stereo (two-channel)
  - surround (I.e. Imax)
  - binaural recordings

Fels: Design of Interactive Multimodal Graphics

Audio Technologies: Stereo

- Two channel

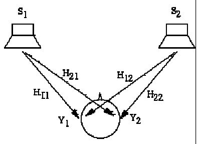


- adjusting gains will create phantom source
- cross-talk cancelled stereo can move phantom source off line
- can also use precedence effect (I.e. add delay)

Fels: Design of Interactive Multimodal Graphics

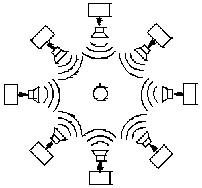
Audio Technologies: 3D audio

- Deliver using headphones or loudspeakers
  - Headphones
    - people don't like wearing them
    - possibly bad frequency response (like pinna filter)
      - problems with elevation
    - sounds sound close
    - low frequencies are not felt
  - Speakers (crosstalk is problem)
    - can use cancellation technique



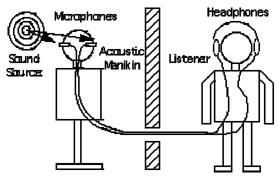
Audio Spatialization: Multichannel

- Can implement with:
  - lots of little speakers (High freq.)
  - one large speaker (Low freq.)
  - Franssen effect
- Dolby Pro Logic Surround Sound



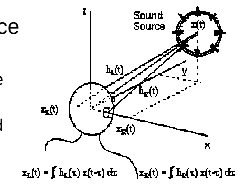
Audio Technologies: Binaural

- Binaural recordings
  - good if manikin has same
    - head size
    - pinna shape
    - body size
- Problems:
  - require use of headphones
  - They are not interactive, but must be prerecorded
  - If the listener moves, so do the sounds
  - Sources that are directly in front usually seem to be much too close
  - Because pinna shapes differ from person to person, elevation effects are not reliable
- Improvements made with HRTFs



Audio Technologies: Binaural

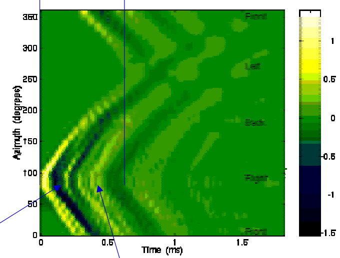
- Impulse response from source to ear drum:
  - head related impulse response (HRIR)
  - freq. Response is head-related transfer function (HRTF)
- With HRTF you only need monaural source
- HRIR measurements:
  - Knowles Electronics Manikin for Auditory Research (KEMAR)



Audio Technologies: HRIR

- Right ear measurement
- source in horizontal plane

about 0.7 s



Pinna reflections

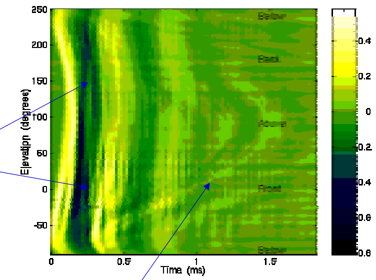
Shoulder reflection

Fels: Design of Interactive Multimodal Graphics

Audio Technologies: HRIR

- Source: median plane
- right ear

asymmetric

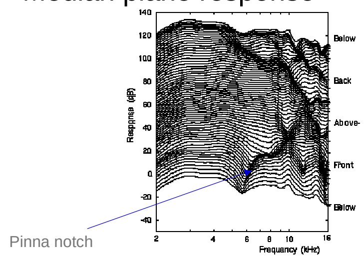


Floor echo

Fels: Design of Interactive Multimodal Graphics

Audio Technologies: HRIR

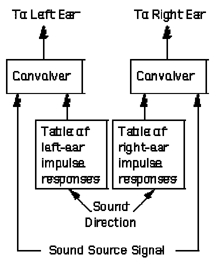
- Median plane response



Fels: Design of Interactive Multimodal Graphics

Audio: Convolvotron

- Developed by Crystal River Engineering
  - technology used in Aureal - vortex sound cards



Fels: Design of Interactive Multimodal Graphics

Audio: Getting HRTF

- Where to get HRTF?
  - standard HRTF
    - poor elevation
    - no standard available yet
  - set of standard HRTF's
    - improve by tailoring
  - individualized HRTF
    - takes time but produces good results
  - model HRTF
    - parameters adjusted for individuals

Audio: Model HRTF

- Three basic techniques
  - Rational function or pole/zero models
    - system identification
    - coefficients complicated functions of azimuth and elevation
  - Series expansions (principal components analysis)
    - high run-time computation costs
  - Structural models
    - based on physical model parameterization
    - maybe ray tracing literature could help here?
    - http://www-engr.sjsu.edu/~knapp/HCIROD3D/3D_sys2/models.htm

Audio Technologies: 3D Audio

- Still lots of work to achieve:
  - positional accuracy
  - relational accuracy (two people agree where sound is)
  - ambience
    - modeling room characteristics

Auditory Channel: Summary

- Least sensitive on median plane (due to symmetry)
- Perception of multiple pure tones complicated
- Ears are well adapted for speech
- Best 3D sound with
  - personalized HRTFs
  - difficult to get

There isn't time to discuss speech and music. Please refer to the reading for many papers that provide surveys as well as some of the key technical papers in these areas.

Position and Motion Sensing

- Inner ear has mechanisms for attitude
  - vestibular sensing system
  - like a biological gyroscope
  - 19,000 nerve fibres
  - six orthogonal semicircular canals
  - head movement and eye movement coordinated instantaneously

I mention this here since these mechanisms are part of the ears. It is not audio.

Position and Motion Sensing

- Body has *proprioceptors*
  - embedded in muscles, joints and tendons
  - provide kinesthetic sensation for position information
  - important for balance

Position and Motion Sensing

- Semicircular canals are sensitive to angular acceleration
- Acceleration has interesting effects:
  - visual-g illusion: light, instruments hard to read
  - autokinetic illusion: fixed light appears to move randomly
  - oculogyral illusion: after large acc. spins sensation of rotating the other direction

Position and Motion Sensing

  - oculogravic illusion: feeling of moving and/or objects are moving
    - works in both directions (acceleration = tilt)
- Non-visual illusion
- Audiogyral illusion
  - sounds are misinterpreted

Position and Motion Sensing

- Vertigo: loss of up direction
  - sensation of climbing while turning
  - sensation of diving while recovering from turn
  - sensation of opposite tilt while skidding
  - coriolis phenomenon - head movement out of plane of rotation
  - sensation of reversed rotation
- Motion sickness
  - problem in VR

Somatic

- Sense of touch
- sensations of:
  - heat (temp), cold (temp), touch (pressure), pain (various)
- 7 distinctive receptors
- one cold and one warm receptors
  - more cold than warm
  - over 45 deg can activate some cold sensors
  - sensitive to changes in temperature

This is critical for haptic interfaces.

Readings: I put gesture based interfaces in here as well even though they may not have any force feedback. They do require proprioception though for people to know which gestures they are making.

#12, #13, #15, #19, #22, #57, #58, #59, #60, #61, #71, #73, #74, #74, #77, #78

Somatic

- Tactile Sensing
  - rate is very important
    - light touch quickly applied produces sensation
  - Hair acts as lever
  - same as proprioceptors
  - negative adaptation occurs
    - high pass filter effect
  - 20 Hz is maximum for separability
    - above 20Hz it is like audio signal

Somatic

- Pain sensing
  - mechanical, chemical, thermal or electrical sensitive
- Some effort to communicate using somatic channel
  - vision to somatic (Paul Bach-y-Rita et al, 1969)
  - encode symbols with vibration
- critical feedback channel for manual tasks
- considerable work with touch and force feedback (haptic feedback)

Olfactory

- Olfactory cells for:
  - different theories: chemical, infrared absorption,
  - different perceptual mappings:
    - small prism
    - four odours: fragrant, acrid, burnt and caprylic
  - Acuity is great - 10,000 times more sensitive than taste
  - negative adaptation occurs

Smell interfaces are fast becoming a hot (and smelly) topic.

Taste

- "Chemical" sense
- taste buds for
  - sensations of sour, salty, bitter and sweet
  - receptor issues unresolved
  - extremely complex and poorly understood

Summary of Input Channels

- Usually combination of senses active
- We also can sense:
  - time (protensity)
  - probability
  - intensity
- Break-off phenomenon

Volitional Output: Neuromuscular

- Motor control associated with cerebral cortex
- volitional and non-volitional
  - can see in facial expression
- muscles contract when stimulated by nerves

Volitional Output: Movable Controls

- Affordances
  - keyboards, touch pads, phone dials, etc.
- verbal control/non-verbal control
- tongue movement
- breath control
- facial control
- gait
- Hand motion - see table:

Biopotential: Galvanic Skin Response

- Surface conductance of skin changes
- Related to mental activity
- 1000s ohm with up to 25% variation

GSR coupled with heart rate is used in polygraph testing. Both of these are good indicators of arousal levels. However, whether the correlation of these are indicators of lying is yet to be determined.

Biopotential: Heart Response

- Resting range around 72 pulses/sec
  - varies from 45 to 90 normally
- change related to mental state
- measure electrical change during beating
  - up to 2V can occur
  - electrocardiogram (EKG)
  - signal processing of EKG is correlated with stress for H.I. (Rowe, 1998)

Biopotential: Brain Response

- Brain produces electrical activity under various conditions
- electroencephalogram (EEG)
- difficult to interpret what is going on

Biopotential: Muscle Response

- Nerves electrically stimulate muscles
  - electromyogram (EMG)
  - order of 25 microvolts, 400Hz is highest energy
  - should measure between 0 and 1000Hz
- rest at 3-4 pulses, thinking about moving or moving will increase this
- Reliable measure of fatigue cost (Inman et al.)
- Gradient may be useful

Human Memory and Learning

- Significant constraint for learning and working effectively
- Magic number seven, Plus or Minus Two (Miller, 1956)
  - Chunking is key
- Capacity: 43 billion bits to 1.5 million bits?
- Short term memory and long term memory
  - different models for how memory is structured
- Brain: 10^{12} neurons and 10^{15} connections
  - connections change strength as we learn

Summary of Human I/O

- Multitude of input/output channels
  - all active at once
- I/O mechanisms usually depend upon
  - cognitive context
  - emotional contexts
- All these channels available to assist people
  - complement each other
- Multimodal looks at:
  - integration
  - substitution
  - complement

Driving Trends in Multimodal Research

- Tangible Bits
- Wearable Computing
- Ubiquitous Computing, Pervasive Computing, Intelligent Environments
- Art, Music and Entertainment
- World Wide Web
- VR/AR
- Information Appliances (see Invisible Computer)
  - Don Norman's design of everyday things

Virtual Reality/Environments

- Real-time, interactive graphics with 3D models + display technology that gives user immersion in the model world with direct manipulation
- interactive information visualization
- Devices used:
  - 3D graphics, trackers, gloves, HMDs and more
- Ivan Sutherland (1965), Jaron Lanier, Myron Krueger and lots more...

VR/VE

- Applications:
  - entertainment
  - vehicle simulation
    - airplanes, cars, expensive machinery
  - physical data visualization
    - planet surfaces
    - NMR data
  - information visualization
    - chemical models
    - mathematical relationships

VR/VE: Research areas

- Visual displays
  - field of view, resolution
- Audition (speech and non-speech, input and output)
- Haptics (forcefeed back and tactile feedback)
- Tracking (still)
- Emotion
- Motion sickness
- Software tools and models
- Evaluation

Fels: Design of Interactive Multimodal Graphics

VR/VE

- Depends on:
  - high speed computing
  - high speed rendering
  - low latency
  - good engineering design

Fels: Design of Interactive Multimodal Graphics

Ubiquitous Computing

- Imagine when computing is as cheap as paper
- Computers will be everywhere, for every need
  - transparent computing
  - transparent communication
- Mark Weiser (Xerox PARC)
- PARC Tab, Pad, Boards + infrastructure

Fels: Design of Interactive Multimodal Graphics

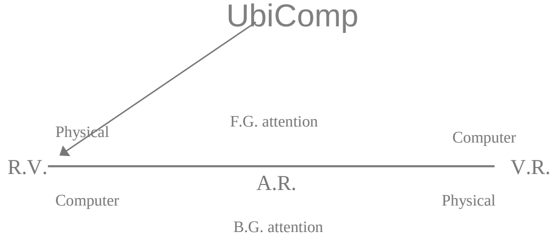
Readings: #64, #65, #66, #69, #70

UbiComp

- Computing should be in background
  - end of personal computer
  - not just portable!
- Key concepts:
  - location
    - context awareness
  - scale
- applications:
  - doors, preference forwarding, call forwarding, diaries, daily informational assistance, prosthetics

Fels: Design of Interactive Multimodal Graphics

UbiComp



Attentional mechanisms play a key role in designing multimodal interfaces.

UbiComp

- Dependencies:
  - H/W
    - displays
    - computers
    - network infrastructure
  - S/W
    - OS to handle real-time + constantly changing resources
    - Network protocols

UbiComp

- Advantages
  - computing where and when you need it
  - contextually aware devices
  - lower cognitive load
- Disadvantages
  - privacy
  - dependency
  - interaction limited to physical world metaphors

Tangible Bits

- Hiroshi Ishii's group at MIT and others
- Idea: couple virtual world to real, physical objects
  - Interactive Surfaces
  - Couple bits and atoms
  - ambient media
- Main Goals
  - grasp & manipulate foreground with physical objects
  - awareness of background using ambient media
- Dependent upon good metaphor
  - need to really do user-centred design

Tangible Computing

- Leverage *affordances* from real world
  - Bricks (Fitzmaurice, Ishii, Buxton)
  - Clearboard (Ishii & Kobayashi)
  - metaDESK, ambientRoom, and more
  - Marble Answering Machine (Bishop), Props (Hinkley), Live Wire (Jeremijenko)
- Can you think of richly afforded physical devices?
  - Doors, windows, cars, toys, dishes...

Tangible Bits: Advantages

- People already know what manipulations make sense
  - as long as metaphor is maintained, life is good
- persistence of data
- make abstract concrete
- composition is natural
- nice match of function, form and augmentation

Tangible Bits: Disadvantages

- Mismatched metaphor
  - makes task harder
- Limited to real world interactions
- Complex interactions may be difficult to express
  - looping contructs
  - boolean operations
- Mechanical failure of physical devices

Wearable Computing

- Smaller and cheaper computers can be embedded in clothing
  - available all the time
  - can have first person perspective
  - augment person's ability
- MIT/U. of Toronto group including Steve Mann, Thad Starner and others

Wearable Computing

- Wearable examples
  - video camera (glasses)
  - head's up display (glasses)
  - compute device (shoes)
  - body monitoring devices
  - communication devices
  - tracking devices
  - audio devices
  - etc.

Wearable Computing

- Applications:
  - altered realities
    - freeze frame, colour
  - augmented realities
    - extra information such as people id tags
    - prosthetics: visual, audio, memory
- Social implications?
  - New protocols possibly needed
  - security

Entertainment, Art, Music

- Music lead the push for many alt. Controllers
  - keyboards, wah-wah pedals, pitch benders
  - Theremin, Sackbut
- Artists often push boundaries of tech. to:
  - explore human emotion
  - concepts and philosophy
  - expression
- Video games drive H.I.T.
- Education - web
- Medicine - VR, tracking

Entertainment, Art, Music

- Technologies
  - video processing and integration
  - gesture sensing and recognition
    - air guitar
  - wireless applications
  - robotics
  - image processing
  - high speed graphics
  - alternate controllers of all shapes and sizes

Summary of Multimodal I/F Apps.

- VR/AR
  - immersive experience
- UbiComp
  - Tangible interfaces
- Information Spaces/WWW
  - agents
  - better GUIs
- Entertainment/Art/Music/Medicine
  - explore boundaries of expression

Multimodal Design:

- User centred and non-user centred
- Intimacy and Embodiement
  - automatic behaviour
  - sources of aesthetics
- Some examples

In general, user centred design will help to improve designs and make them usable. Non-user centred design will help create innovations that may provide new ways of doing things. For example, user-centered design around using a horse for transporation will not lead to the design of the car.

Intimacy and Embodiment

- Want interfaces that feel “good” to use
- Humans and machines intimately linked
  - degree of intimacy supported may determine success
- Types of relationships:
  - human to human
  - human to machine

See attached document. Also refer to http://www.ece.ubc.ca/~ssfels and http://www.ece.ubc.ca/~hct for additional papers and information.

Readings: #5, #14, #7, #8, #11, #14, #16, #17, #18, #20, #22

Intimacy

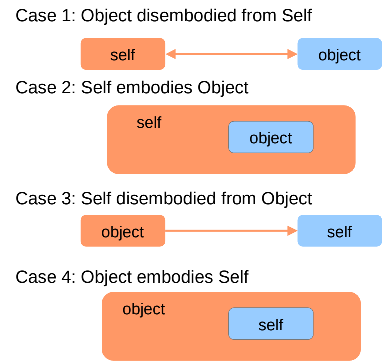
- Intimacy is a measure of *subjective* match between the behaviour of an object and the control of that object.
  - extension of “control intimacy” from electronic musical instruments analysis (Moore, 1997)
- High intimacy implies:
  - object feels like an extension of self
  - satisfaction derives from interacting with object
  - emotional expression flows
    - requires cognitive effort to prevent

Intimacy

- Contributing factors
  - consistency
  - responsiveness
  - usefullness
  - learnability
  - functionality
  - others

Fels: Design of Interactive Multimodal Graphics

Intimacy: Embodiment vs. Disembodiment



Fels: Design of Interactive Multimodal Graphics

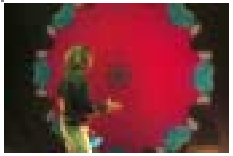
Each type of relationship has its own aesthetic that drives people. Case 1: response, Case 2: control, Case 3: reflection, and Case 4: belonging/submission. Each can be used in design.

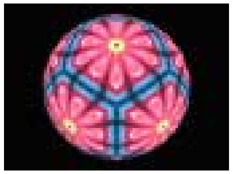
Intimacy and Embodiment Design Examples

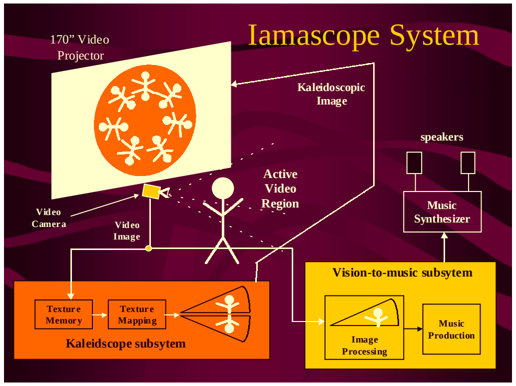
- Iamascope
- Flowfield
- Swimming Across the Pacific

Iamascope

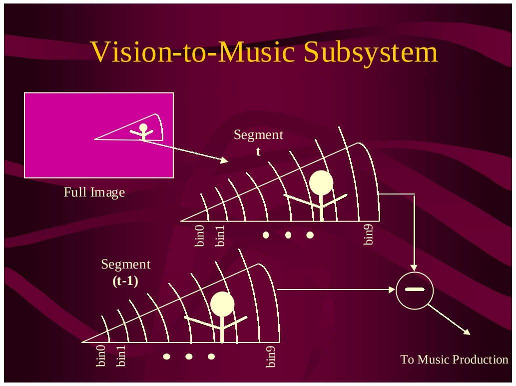
- Interactive multimedia artwork (SIGGRAPH’97)
- Participants are put inside a large kaleidoscope
- Participants movements also map to music
- Run in real-time
- Immersive, exciting, satisfying, intimate





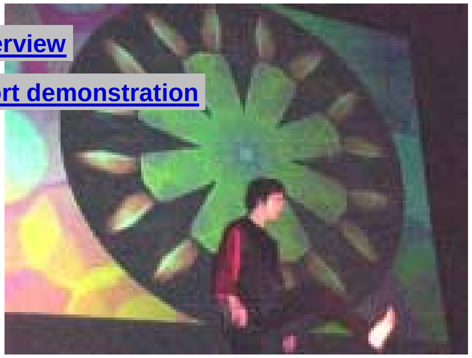


Fels: Design of Interactive Multimodal Graphics



Fels: Design of Interactive Multimodal Graphics

Iamascope Video



Overview

Short demonstration

Fels: Design of Interactive Multimodal Graphics

FlowField: Semantics of Caress

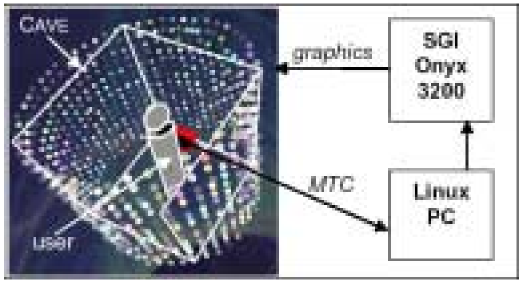
- Investigate whole hand interaction techniques for VR
  - use Tactex MTC Express work with Tim Chen and Thecla Schiphorst
- Idea:
  - allow users direct manipulation of fluid
    - use particle simulation for fluid
      - aesthetics was important
    - hand manipulation on hard surface mapped to obstructions in the flow field

Fels: Design of Interactive Multimodal Graphics

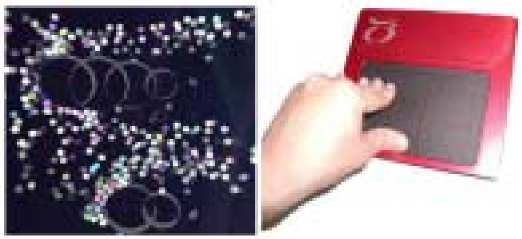
FlowField: Semantics of Caress

- Implementation
  - create a constant flow of particles
    - simulate elastic collisions
  - use CAVE for visualization of particles
    - 3D environment with stereo graphics
    - feel inside of flow
  - pressure on touch pad mapped to width of cylindrical obstruction
    - one-to-one mapping of Taxels

FlowField: Sematics of Caress



FlowField: Sematics of Caress



FlowField: Summary

- Interface is very compelling
  - easy to learn to use
  - correspondence simple to understand
  - fun
  - immersive
  - intimate

2 Hearts System

- Motivation:
  - create two person instrument
  - facilitate human-human communication
  - explore human intimacy
- Idea:
  - map two heart beats to sound and image
  - intimacy between two people function of ability to control each other's heart
- Graeme McCaig

2 Hearts System

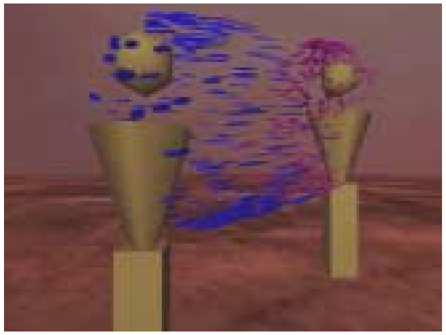
- Implementation in 2 phases
  - Phase 1:
    - navigation in musical terrain
    - one heart beat to clear change in music
    - two heart beats together push virtual ball around musical terrain
    - 2D musical score
      - patches of music connected together

2 Hearts System

- Phase 2
  - visualization of heart beat relationship
  - created in CAVE
- Implementation
  - use auras to represent heart beat information
  - two perspectives in the CAVE
    - head tracked + stereo graphics
  - music mapped as well

Example of Visualization

Example of Visualization



Fels: Design of Interactive Multimodal Graphics

2 Hearts System: Mapping

Change in Heartbeat	Graphical Effect
Beat detected	New aura particles spawned
Time-averaged rate (averaged over last 3 beats) increases/decreases	Particle color becomes more red/more blue
Instantaneous rate increases/decreases	Particle tint becomes more/less white
Users' rates become more/less similar	Aura attraction becomes stronger/weaker
User A's rate greater/less that User B's rate	Users float in direction of User B/User A
Users' average rate increases/decreases	Users levitate up/down
Users' average rate increases/decreases	Sky becomes lighter/darker

Fels: Design of Interactive Multimodal Graphics

2 Hearts System: Summary

- Work still being completed
  - Phase 1: tried at CHI NIME'01 Workshop
    - demo at the Experience Music Project
    - succeeded in getting people to interact to play with heart beat
    - positive feedback possible
  - Phase 2: Implemented in CAVE
    - split screen technique works
    - mapping is clear but probably not best
    - still doing more testing

Swimming Across the Pacific

- Motivation
  - contemporary art piece based on Swimming Across the Atlantic (Misheff, 1982)
  - Virtually swimming in an airplane across the Pacific Ocean
- Implementation:
  - build virtual swimming apparatus
  - integrate with virtual swimming terrain
- Grace Chen and Ashley Gadd

Swimming Across the Pacific

Fels: Design of Interactive Multimodal Graphics

Swimming Across the Pacific

Fels: Design of Interactive Multimodal Graphics

Swimming Across the Pacific

- Use Fastrak sensors for arms, legs, torso and head
- Use inverse kinematics + “hack” for arms to map sensors to graphics
- HMD with head tracking for perspective

Swimming Across the Pacific: Summary

- Feels a little like swimming
  – floating feeling
- General purpose VR navigation device
- Swimming allows for energy to be expended for navigation
  – aid in relative scale comparisons
  – body centric visualization

Summary

- Multimodal interfaces need to consider
  - human information processing
  - matching interface to task
    - use complementary modes where appropriate
  - intimacy and embodiment
- Plenty of research opportunities

Reading List

1. Human Information Processing, Fogel, L, Prentice-Hall Inc., 1967. (An older book, but an excellent reference)
2. Research Directions in Virtual Environments, Report of an NSF Invitational Workshop, Fuchs, H. and Bishop, G., 1992 (ftp://ftp.cs.unc.edu/pub/publications/techreports/92-027.ps.Z)
3. Human Computer Interface Design, NSF Sponsored Multi-university Course, Perry Cook, Ben Knapp, Richard Duda, Bill Verplank, Shumin Zhai, http://www-engr.sjsu.edu/~knapp/hci.html, 1998.
4. Tangible Bits: Towards Seamless Interfaces between People, Bits and Atoms, Ishii, H. and Ullmer, Brygg, Proc. of SIGCHI'97.
5. The Psychopathology of Everyday Things, in Human-Computer Interaction: Toward the Year 2000, Norman, D., pp 5- 21, 1995.
6. Ch. 14 of Human-Computer Interaction: Toward the Year 2000, Baecker, R., Grudin, J., Buxton, W. and Greenberg, S, Morgan Kaufmann, 1995.
   The World Wide Web, Berners-Lee, T., Cailliau, R., Luotonen, A., Nielsen, H., and Secret, A., pp 907-912.
   Nature and Origins of Virtual Environments, Ellis, S., pp 913—932.
   The Computer for the 21st Century, Weiser, M., pp 933-940.
7. Wearable Computing: A First Step Toward Personal Imaging, Mann, S., Computer, Vol. 30, No. 2, pp. 25-32, 1997.
8. A Conversation with Marvin Minsky About Agents, Minsky, M. and Riecken, D., Comm. of the ACM, pp. 22-29, 1994.
9. Review of Virtual Environment Interface Technology, Institute For Defense Analyses, IDA Paper P-3186, Christine Youngblut, Rob E. Johnson, Sarah H. Nash, Ruth A. Wienclaw, and Craig A. Will, 1996
10. Cubby, What You See Is Where You Act, Djajadingingrat, Ph.D. Thesis, Royal College of Art, 1998.
11. Human Interaction with Pervasive Computers, IBM System Journal, Vol 38, No. 4, pp 501-698, 1999.
    The whole journal issue is excellent.
    Classroom 2000: An experiment with the instrumentation of a living educational environment, Abowd, G., pp 508-530.
    Making Sharing Pervasive: Ubiquitous computing for Shared Note Taking, Landay J and Davis R., pp. 531-550.
    Wireless Neworked digital devices: A new paradigm for computing and communication, Zimmerman, pp. 566-574.

Lots of readings. These are from my HIT course page. They cover many aspects of human information processing, HCI design and descriptions of systems that I think represent some of the important ideas for designing interfaces (either multimodal or single mode). Please refer to http://www.ece.ubc.ca/~elec596 for annual updates.

12. A Survey of 3D Interaction Techniques, Hand, C., Computer Graphics Forum, Vol 16., No. 5, pp 269-281, 1997.
13. The Influence of Muscle Groups on Performance of Multiple Degree-of-Freedom Input, Zhai, S., Milgram, P., and Buxton, W., SIGCHI96, pp. 308-315, 1996.
14. Sensuality in Produce Design: a Structured Approach, Hofmeester, G., Kemp, J., and Blankendaal, A., SIGCHI96, pp. 428-435, 1996.
15. Half-QWERTY: A One-handed Keyboard Facilitating Skill Transfer From QWERTY, Matias, E., MacKenzie, S., and Buxton, W., INTERCHI'93, pp. 88-94, 1993.
16. The Active Badge System, Hopper, A., Harter, A., and Blackie, T., INTERCHI'93, pp. 533-534.
17. Passive Real-World Interface Props for Neurosurgical Visualization, Hinckley, K., Pausch, R., Goble, J., and Kassell, N, SIGCHI'94, pp 452- 458, 1994.
18. Bricks: Laying the Foundations for Graspable User Interfaces, Fitzmaurice, G., Ishii, H., and Buxton, W., SIGCHI'95, pp. 442-449.
19. Whole-hand Input, Sturman, D., Ph.D. Thesis, MIT Media Lab, 1992.
20. The Media Equation: How People Treat Computers, Televisions and new Media as Real People and Places, Reeves, B. and Nass, C., Cambridge University Press, 1996.
21. Multi-Level Direction of Autonomous Creatures for Real-Time Virtual Environments, Blumberg, B. and Galyean, T., Computer Graphics Proceedings (SIGGRAPH'95), pp. 47-54, 1995.
22. Instrumented Footwear for Interactive Dance, Paradiso, J., Hu, E., and Hsiao, K., XII Colloquium on Musical Informatics, 1998.
23. Optical Engineering Challenges of the Virtual Retinal Display, Kollin, J and Tidwell, Proceedings of Novel Optical Systems Design and Optimizat ion, 1995, http:/ /www.hitl.washington.edu/publications/p-95-12
24. 3D task performance in fish tank virtual worlds, Arthur, K. W., Booth, K. S., & Ware, C. (1993). . ACM Transactions on Information Systems, Special Issue on Virtual Worlds, 11(3), 239-265.
25. Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of the CAVE, Cruz-Neidra, C., Sandin, D., and DeFanti, T., Proceedings of SIGGRAPH'93, pp 135-142. Also at: http://www.evl.uic.edu/EVL/RESEARCH/PAPERS/CRUZ/sig93.paper.html (no figures).
26. The Virtual Retinal Display: A New Display Technology Using Scanned Laser Light. Pryor, Homer L., Furness, Thomas A. and Viirre, E., Proceedings of Human Factors and Ergonomics Society, 42nd Annual Meeting, pp. 1570-1574, 1998. Also at: http://www.hitl.washington.edu/publications/r-98-27//r-98-27.doc.
27. World Skin, Benayoun, M. and Barriere, J., Cyberarts'98, pp. 70-76, 1998.
28. Electric Garden, Visual Proceedings of SIGGRAPH'97, pp. 61-121, 1997.
29. Electric music: new ways to play, Paradiso, J, IEEE Spectrum, Dec. 1997. http://www.spectrum.ieee.org/select/1297/muse.html
30. The Sackbut Blues: Hugh Le Caine, Pioneer in Electronic Music, Young, G., Canadian Cataloguing in Publication Data, 1989.

31. Auditory Display: Sonification, Audification and Auditory Intefaces, ed. Kramer, G, Addison-Wesley, 1994.
32. Input/Data Acquisition system Design for Human computer Interfacing, Putnam and Knapp, http://www-ccrma.stanford.edu/CCRMA/Courses/252/sensors/sensors.html
33. Whisper: A Wristwatch Style Wearable Handset, Masaaki, F, and Yoshinobu T., proceedings of SIG CHI'99, pp. 112-119, 1999.
34. Stochastic Coding of Speech Signals at Very Low Bit Rates, Atal, B, and Schroeder, M., in Links for the Future, ed. Dewilde and May, IEEE/Elsevier Science, pp. 1610-1613, 1984.
35. Speech Recognition and Sensory Integration, Massaro, D., and Stork, D., American Scientist, Vol. 86, pp. 236-244, 1998.
36. Bishnu S. Atal, V. Cuperman, and A. Gersho, editors. Advances in Speech Coding. Kluwer Academic, Boston, 1991.
36a. New directions in low bit rate speech coding, Atal, B., Signals, Systems and Computers, 1991. 1991 Conference Record of the Twenty-Fifth Asilomar Conference on , 1991, Page(s): 933 -934 vol.2
37. An Implementation of the PSOLA/KDG Waveform Synthesis Technique, Verhelst, W.D.E., TR-733, Institute for Perception Research, Netherlands, 1990.
38. A New Model of LPC Excitation for Producting Natural-Sounding Speech at Low Bit Rates, Atal, B., and Remde, J., proc. IEEE ICASSP, pp. 614-617, 1982.
39. Code-Excited Linear Prediction CELP: High-Quality Speech at Very Low Bit Rates, Schroeder, M., and Atal, B., proc. IEEE ICASSP, pp. 937-940, 1985.
40. Linear Prediction of Speech, Markel, J.D. and Gray, A.H., Springer Verlag, New York, 1972.
41. Auditory Interfaces section, Institute of Defense Analysis (IDA), Youngblut, C., Johnson, R.E., Nash, S., Wienclaw, R., and Will, C., 1996 (http://www.hitl.washington.edu/scivw/IDA/).
42. Barriers to the Acceptance of Synthesized Speech as a Communication Aid, Rowden, C.G., Int Conf. on speech I/O: Techniques and Applications, March, pp. 220-224, 1986.
43. A Versatile Software Parallel-Formant Speech Synthesizer, Rye, J. and Holmes, J., JSRU TR-1016, 1982.
44. Articulatory Analysis and Synthesis of Speech, Parthasarathy, S., Schroeter, J., Coker, C., and Sondhi, M.M., IEEE TENCON, pp. 760-764, 1989.
45. Automatic generation of control signals for a parallel-formant speech synthesizer, Seeviour, P.M., J.N. Holmes and M.W. Judd, Proc. IEEE Int. Conf. Acoustics Speech and Signal Processing, Philadelphia, PA, 690-693. (1976)
46. Physiologically-Based Speech Synthesis Using Neural Networks, Hirayama, M., Bateson, E., and Kawato, M, IEEE trans. Fundamentals, Vol. E76-A, No. 11, pp. 1898-1910, 1993.
47. Survey of Speech Synthesis Techniques, Holmes, J., IEE Colloquium on Speech Synthesis and Devices, Apr. pp. 1-4, 1983.

48. Talking Heads, Rubin, P, and Bateson, E, http://www.haskins.yale.edu/haskins/heads.html
49. Sound and Voice Synthesis and Analysis, Cook, P., http://www.cs.princeton.edu/courses/archive/fall98/cs436/SoundVoice/sndvoic1.html
50. Ambiophonics: The Science of Domestic Concert Hall Design http://www.ambiophonics.org/index.html
51. SIGGRAPH 97 Panel on Facial Animation: Past, Present and Future, Demetri Terzopoulos, University of Toronto; Frederic Parke (moderator), Texas A&M University; Doug Sweetland, Pixar; Keith Waters, Digital Equipment Corporation; Michael M. Cohen, UC-Santa Cruz, http://mambo.ucsc.edu/psl/sig97/siggraph97-panel.html
52. Controllers for Computers and Musical Instruments, http://www-ccrma.stanford.edu/CCRMA/Overview/controllers.html
53. Video Rewrite-Driving Visual Speech with Audio, Bregler, Slaney, Covell, Siggraph'97 Video proceedings, video 27
54. Modeling of the vocal Tract in Three Dimensions, Engwall, O., proc. Eurospeech'99, Sept. pp. 113-116, 1999.
55. Synthesis Toolkit in C++, Cook, P., Siggraph, 1996. http://www.cs.princeton.edu/~prc/ - look under STK
56. Baby Ears: A Recognition System for Affective Vocalizations, Slaney, M. and McRoberts, G., Proc. of ICASSP, 1998. http://web.interval.com/papers/1997-063
57. Applying Electric Field Sensing to Human-Computer Interfaces, Zimmerman, T., Smith, J., Paradiso, J., Allport, D. and Gershenfeld, N., CHI'95, pp. 280-287.
58. WorldBeat: Designing a Baton-Based Interface for an Interactive Music Exhibit, Borchers, J., CHI97, pp. 131-138, 1997.
59. "Body Coupled FingeRing": Wireless Wearable Keyboard, Fukumoto, M. and Tonomura, Y., CHI97, pp. 147-154, 1997.
60. Virtual Chopsticks: Object Manipulation using Multiple Exact Interactions, Kitamura, Y., Higashi, T., Masaki, T., and Kishino, F., IEEE Virtual Reality, 1999, Page(s): 198 -204.
61. The Rockin' Mouse: Integral 3D Manipulation on a Plane, Balakrishnan, R., Baudel, T., Kurtenbach, G., and Fitzmaurice, G., CHI'97, pp. 311-318, 1997.
62. Beat Tracking based on Multiple-agent Architecture - A Real-time Beat Tracking System for Audio Signals, Goto, M. and Muraoka, Y., ICMAS-96, pp. 103-110, 1996.
63. Emancipated Pixels: Real-World Graphics in the Luminous Room, Underkoffler, J., Ullmer, B., and Ishii, H., SIGGRAPH'99, pp 385-392, 1999.
64. Network computers-ubiquitous computing or dumb multimedia?, Herrtwich, R.G.; Kappner, T., Autonomous Decentralized Systems, 1997. Proceedings. ISADS 97., Third International Symposium on , 1997, Page(s): 155 -159.
65. Context-aware, adaptive wearable computers as remote interfaces to 'intelligent' environments, Kortuem, G.; Segall, Z.; Bauer, M., Wearable Computers, 1998. Digest of Papers. Second International Symposium on, 1998, Page(s): 58 -65

66. Context-awareness in wearable and ubiquitous computing, Abowd, G.D.; Dey, A.K.; Orr, R.; Brotherton, J. Wearable Computers, 1997. Digest of Papers., First International Symposium on , 1997, Page(s): 179 -180
67. A tangible goal for 3D modeling, Massie, T. IEEE Computer Graphics and Applications , Volume: 18 Issue: 3 , May-June 1998, Page(s): 62 -65.
68. Attachable computer: augmentation of electric household appliances by fit-up computer, Iga, S.; Itoh, E.; Higuchi, F.; Michiaki, Y., Computer Human Interaction, 1998. Proceedings. 3rd Asia Pacific , 1998, Page(s): 51 -56.
69. An Internet protocol for flexible, scalable, and secure interaction with ubiquitous computing devices, Driessen, P.F.; Gabert, J.; Levy, M.R.; van Emden, M.H. Communications, Computers and Signal Processing, 1997. 10 Years PACRIM 1987-1997 - Networking the Pacific Rim. 1997 IEEE Pacific Rim Conference on , Volume: 1 , 1997, Page(s): 414 -416 vol.1.
70. Software design issues for ubiquitous computing, Abowd, G.D. VLSI '98. System Level Design. Proceedings. IEEE Computer Society Workshop on , 1998, Page(s): 104 -109.
71. A Comparison of Three Selection Techniques for Touchpads, MacKenzie, S. and Oniszczak, A., CHI98, pp. 336-343, 1998.
72. Development of a System for Three-Dimensional Fleshpoint Measurement of Speech Movements, Zierdt, A., Hoole, P., and Tillmann, H., http://www.articulograph.de/.
73. Manual and Gaze Input Cascaded (MAGIC) Pointing, Zhai, S., Morimoto, C. and Ihde, S., CHI99, pp. 246-253.
74. A Survey of 3D Interaction Techniques, Hand, C., Computer Graphics Forum, Vol. 16, No. 5, pp. 269-281, 1997.
75. The Influence of Muscle Groups on Performance of Multiple Degree-of-Freedom Input, Zhai, S., Milgram, P. and Buxton, W., CHI96, pp. 308-315, 1996.
76. Two-Handed Input in a Compound Task, Kabbash, P., Buxton, W., and Sellen, A., CHI94, pp. 417-423, 1994.
77. Whole-hand Input, Sturman, D., Ph.D. thesis, MIT Media Laboratory, 1992.
78. Design of Virtual 3D Instruments for Musical Interaction, Mulder, A., Fels, S. and Mase, K., GI99, pp. 76-83, 1999.