Abstract

Cognitive Science and Psychology can provide us with information about the ways in which users perceive, think, and act in complex perceptual environments. This can be used to design perceptual interfaces based upon human perceptual mechanisms as well as to design data displays and input devices that are optimized for human characteristics.

Both of these applications serve the function of “closing the loop” between the operator and the interface by building a richer set of linkages between events in one domain (operator or system) and operations in the other. This paper explores several recent theories of perceptual cognition that may help to inform the design of perceptual interfaces.

1. Spatial indexing and cognition

By concentrating on the nature of the linkages between perceptual events and cognitive structures, we derive models of information processing that rely upon the perceptual world to provide much of its’ own representation-- so-called deictic cognition models. The minimal mechanism for a deictic perceptual/cognitive system is a set of pointers or attentional tokens that serves to link mental operations (ranging from simple visual routines such as collinearity to more complex conceptual structures) and individual perceptual events. These pointers (called FINSTs, for FINgers of INSTantiation) were first used in computer vision applications by Pylyshyn, Elcock, Marmor, and Sander (1978). They have subsequently been used to explain human perceptual and cognitive processing in a wide variety of domains (Pylyshyn, 1989, Pylyshyn, Burkell, Fisher, Sears, Schmidt, and Trick, 1993). Please see http://ruccs.rutgers.edu/faculty/pylyshyn.html for more information.

In tests with human subjects, display items that are FINSTed have the potential to be accessed and acted upon by cognitive processes in parallel, with higher priority than
unFINSTed items. We use this perspective to predict the emergent knowledge structures that are produced by the interaction of dynamic display events and cognitive processes, and to design displays that are optimized for the characteristics of users’ mental processes (see Fisher, Conway, and, Groeneboer, 1997 and Fisher, Agelidis, Dill, Tan, Collaud, and Jones, 1997). They have also served to inform the design of computer vision systems that process information in a way that is similar to humans (Eagleson, 1991), tracking local motion to aid in gesture recognition and robotic vision. Additional information about this line of research is available on http://www.csd.uwo.ca/~eagleson/

2. Cognitive architecture of sensory integration

One way in which the ways in which FINSTs may be used by humans and computers is to improve performance in rich perceptual scenes containing multiple events and multiple channels of sensory information about those events. In a sparse perceptual world, the number of possible combinations of sensory primitives is limited, and the process of matching visual and auditory stimuli (for example) that are products of a single event is straightforward. In order to parse a complex perceptual world, however, multiple streams of sensory information must be indexed and partitioned into a number of discrete transmodal event constructs.

Evidence from functional Neurophysiology and perceptual Psychophysics suggests that human information processing occurs in a highly modularized fashion, with multiple processes operating on the same primary sensory stimuli. If event partitioning and sensory integration are to take place within a modular cognitive architecture, each input module must perform its own indexing, matching and fusion operations independently on the same input data. These operations can draw upon any of a set of obvious matching cues: co-occurrence in time, similar location, and categorical “goodness of fit”.

The computational isolation of perceptual modules creates the possibility of conflicts between the representations they form of the same external events. Our studies on the integration of visual and auditory information in bimodal speech perception (Fisher, 1991, Fisher and Pylyshyn, 1994) support the robustness of the integration of visual and auditory speech despite differences in location of the two components and subjects’ attention to one or the other modality. In other measures, the apparent location of an auditory source was found to be “captured” by a visual distractor in the familiar “ventriloquism effect”; however, this effect was stronger for motor measures than for cognitive ones. The result of this was paradoxical, with subjects reporting two simultaneous events, one visual and one auditory, but being able to point to only one of them.

Conflicts between representations within perceptual modules such as this are easy to produce in the laboratory, but do not obviously affect our daily life. This is due to the fact that events in the physical world are themselves consistent, and the perceptual system can exploits these consistencies to coordinate the actions of the various perceptual modules. As our studies document, this can take the form of a recalibration of auditory perceptual space to bring it into agreement with visual stimuli. Please see www.cs.sfu.ca/~brfisher/personal/brfisher.work.html for more information on this topic.

Extending our work into the development of interfaces that pick up and respond to cues in the world (such as speech perception, computer vision, and tracking) can be done by applying some of the same methods that our perceptual systems use to minimize the processing
demands of dealing with complex sensory environments. This can include architectures that are highly modularized (perhaps agent-based) but which rely on the consistencies of the perceptual world to coordinate their behaviour.

3. Challenges for the future

Our interest in creating perceptual interfaces is largely due to their perceived ability to help computers to become more responsive to users’ needs, “closing the loop” between user and system for improved performance. However, this increase in responsiveness may carry with it some undesirable consequences. As computers become more responsive to the state of the user through their ability to perceive, they run the risk of supporting a change in the user’s model of the interaction from one of the control of a dynamic artifact to that of an interaction with an intelligent agent. This change may carry with it a great increase in expectation on the part of the user for human-like patterns of interaction.

Another focus of our research applies psycholinguistics models (Grice 1975; Clark and Wilkes-Gibbs 1986) to address the pragmatics of social interaction and communication over limited bandwidth channels. Face-to-face communication is to a great degree mediated by conversants’ knowledge of their common ground of shared referents, which is largely created and maintained by the use of non-verbal cues such as expression and gesture.

Grice’s theory of conversational implicature states that conversants disambiguate their language (a process known as “grounding”; Clark, and Brennan 1991) by assuming that both parties are making a good-faith effort to achieve the goal of communication.

While the application of this work has been limited to improving computer-mediated communication between users in telelearning and decision support systems, it may provide a structure for the exploration of users’ expectations for interaction with systems that appear to exhibit intelligent responses to users’ behaviour.

4. Conclusion

The quest for natural and compelling interaction with computers will require us to leap Psychological hurdles as well as technical ones. Continued exploration in this field will benefit greatly from input from Psychological theories, and may well have a major impact on how those theories develop in the future.

This research was supported by the Institute for Robotics and Intelligent Systems, a Canadian Network of Centres of Excellence.

References:


