User Interface Issues in Multimedia

John Fulcher
University of Wollongong, Australia

A BRIEF HISTORY OF COMPUTER USER INTERFACES

Much has changed in computer interfacing since the early days of computing—or has it? Admittedly, gone are the days of punched cards and/or paper tape readers as input devices; likewise, monitors (displays) have superseded printers as the primary output device. Nevertheless, the QWERTY keyboard shows little sign of falling into disuse—this is essentially the *same* input device as those used on the earliest (electromechanical) TeleTYpewriters, in which the "worst" key layout was deliberately chosen to slow down user input (i.e., fast typists). The three major advances since the 1950s have been (1) the rise of low cost (commodity off-the-shelf) CRT monitors in the 1960s (and in more recent times, LCD ones), (2) the replacement of (text-based) command line interfaces with graphical user interfaces in the 1980s, and (3) the rise of the Internet/World Wide Web during the 1990s. In recent times, while speech recognition (and synthesis) has made *some* inroads (i.e., McTeal, 2002; O'Shaughnessy, 2003), the QWERTY keyboard and mouse remain the dominant input modalities.

MULTIMEDIA USER INTERFACES

Over the years the term "audio-visual" (AV) has segued into the more modern one of "multimedia" (MM), reflecting not only the incorporation of various I/O modalities, but also the implication of interactivity between user and system. Nevertheless, the older term still reflects the primary focus of MM systems today, these being sight and sound—the two *primary* modalities. Barfield (2004) views sound as being the "forgotten child" of MM, characterizing it as active, non-localized, transient, and dynamic, in contrast to graphics, which he characterizes as passive, localized, permanent, and static. Now speech is inherently temporal in nature, whereas vision is spatial—hence synchronization is a fundamental consideration in multimodal interfaces, in order that users do not suffer cognitive overload.

Multimodal Interfaces

Ordinarily when we speak of multimodal interfaces, we mean the concurrent arrival of user input via more than one modality (sense). It is possible in some situations, however, that *sequential* operation is more appropriate—in other words, switching modalities where appropriate for improved clarity of user input to a system.

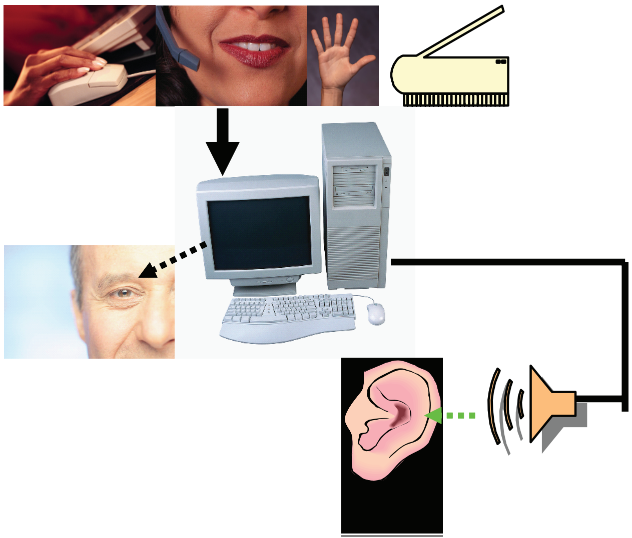
Now AV speech is inherently bi-modal in nature, which means that visual cues (such as eye/lip movement, facial expression, and so on) play an important role in automatic speech recognition—ASR (Potamianos, Neti, Gravier, Garg, & Senior, 2003). The key issues in this context are (a) *which* features to extract from lip movements, and (b) how to *fuse* (and synchronise) audio and visual cues.

Another example of a bi-modal user interface is the speech-gesture one developed by Sharma et al. (2003) for crisis management in emergency situations. Taylor, Neel, and Bouwhuis (2000) likewise discuss combining voice and gesture, whereas Smith and Kanade (2004) focus specifically on extracting information from video. Many other multimodal interfaces abound in the literature (e.g., Booher, 2003; Bunt & Bevin, 2001; Jacko & Sears, 2003; McTeal, 2002; Yuen, Tang, & Wang, 2002).

Figure 1 shows a system that incorporates three input modalities—(1) sight (hand gesture + scanned text and/or images), (2) sound (voice input via microphone), and (3) touch (via mouse + keyboard), together with the two most commonly used (i.e., AV) output modalities (sight and sound).

We have just seen how data (information) fusion needs to be considered in bi- (multi)modal user interfaces. This is especially important in lip synching in the case of combining the two dominant modalities (that is, sight and sound). This is a common post-produc-

Figure 1. Common interface modalities



tion activity in film production, as well as a factor that needs to be taken into account in designing multi-user role-playing games for the Internet.

Multimedia Interfaces

Modern day MM applications incorporate (formatted) text, images, drawings (graphics), animation, video, and sound (including speech)—so called "rich content" (Dix, Finlay, Abowd, & Beale, 2004; Li & Drew, 2004). Accordingly, a text-only, dot point MS PowerPoint presentation does *not* qualify as an MM system (it is a *mono*-media one!).

Current generation MM systems can be classified into CD-ROMs/DVDs (such as encyclopaedia) vs. Web sites (Barfield, 2004). There is also currently a trend toward product convergence—more specifically, present day personal digital assistants, mobile (cellular) phones, portable games machines, MP3 players, digital cameras, GPS, and the like could well become integrated into the *single* mobile, portable, Internet-enabled, multimedia, wireless, (in other words, ubiquitous and pervasive) devices of the future (Li & Drew, 2004). Furthermore, given the inherent limitations of such devices—small displays, limited memory and bandwidth—we necessarily need to consider compression, of images (JPEG), video (MPEG), and sound (MP3) alike.

Accordingly, Oquist, Goldstein, and Chincholle (2004) analyze desktops, laptops, palmtops, and handheld devices along the dimensions of (1) levels of portability (stationary, seated, standing, or moving), (2) attentiveness (primary, secondary, minimal), and (3) manageability (stable, unstable, unbalanced—in other words, 1- vs. 2-handed).

Seffah and Javahery (2004), in their discussion of multiple user interfaces, characterize the interaction styles of MUI—graphical vs. Web vs. handheld. The fundamental characteristics of abstraction, cross-platform consistency, uniformity, user awareness of potential trade off, and conformity to default UI standards are highlighted. They further advocate the use of the User Interface Markup Language (Abrams, Phanouriou, Batongbacal, Williams, & Shuster, 1999) for the development of device-independent UIs.

Lastly, we need to keep in mind the extensive use being made of MM for purposes of data/scientific visualization, in which once again, technical considerations go hand-in-hand with HCI, psychology, graphic design, and so forth (Brodie et al., 1992; Earnshaw & Wiseman, 1992; Spence, 2000; Ware, 1999; Wolff & Yaeger, 1993).

Fundamentally, scientific visualization is a two-step process: (a) firstly mapping (analogue) real-world data into a numerical (digital) representation thereof, and (b) thence into graphical and/or animated form, for display to the user, in order for them to make better informed decisions. Such systems also tend to incorporate user interactivity, in order to uncover more meaning/information previously buried in the raw data (Bonneau, Ertl, & Nielson, 2006; Chen, 2003; Fortner & Pervukhin, 1995; Jensen, 2005; Keller & Keller, 1993; Nielson, Hagen, & Muller, 1997). Examples can be readily found in the fields of (2D/3D) fluid flow, surface heat transfer, geo-satellite images (GIS), weather patterns, molecular/chemical interactions, DNA structure, and MRI scans, to name but a few (http://math.nist.gov/mcsd/savg/vis/; http://www.opendx.org/). In many cases, the resulting colors are not real, but rather intended to aid the user in their conceptual understanding.

MM USER INTERFACE DESIGN

Now the goal of interaction design can be viewed as users performing the correct action(s) at the appropriate time(s). This will necessarily involve design tradeoffs (compromises). For instance, (1) use of controls that are analogs of real-world objects can simplify the map-

Related Content