HCI and the 3C Convergence

Position Paper for the CHI 1999 Workshop on the Development of an HCI Research Agenda

Helen M. Meng
Department of Systems Engineering and Engineering Management
The Chinese University of Hong Kong
Shatin, N. T.,
Hong Kong, China
hmmeng@se.cuhk.edu.hk

I. Introduction

This paper attempts to explore the implications of the "3C convergence" (convergence of content, computers and communications) on the research agenda of HCI. We continue to witness increasingly powerful computers unite in a networked environment, where wireline and wireless communications enjoy broadening bandwidths, which permit the transmission of multimedia content. The interaction between humans and the 3C's is becoming a daily staple.

For the purpose of illustration, we can draw a few examples from the Hong Kong environment [1]. Hong Kong has implemented fully digital fiber network since 1993, and the available bandwidth allows deployment of interactive services such as on-demand movies, music and radio, bundled with high-speed Internet access (1.5Mbps), via a telephone line and a television set-top box.[1] Chinese (in addition to alphanumeric) pagers are pervasive, and approximately one out of every two people carry a mobile phone. Pagers and mobile phone screens display real-time information such as news headlines, weather, stock trades, etc. throughout the day. As we can see, HCI issues are abundant in the deployment of such information and communication services. In the following, we will present some areas which may be of interest to the workshop.

II. HCI and Content

Content is continuously created, processed, stored, retrieved and displayed. As text, graphics, speech, video, audio and data integrate horizontally in various multimedia applications, we may explore the development of a unified framework or a *unified user interface* to support the integration of multimedia.

Another interesting aspect related to content is *internationalization*. The bilingual populace in Hong Kong speaks/reads both English and Chinese. In speaking Chinese, Hong Kong has two predominant dialects – Putonghua and Cantonese. From the perspective of textual display, Chinese may be encoded in different forms, such as GB or Big5.[2] Electronic information in Hong Kong often exist in both English and Chinese, which may mean two different versions by language, or three different versions by the encoding scheme. In many cases, the text is "*mixed-*

---

[1] Hong Kong Telecom's iTV™ service, http://www.htk.com.

[2] Big5 encodes the traditional Chinese character set used in Taiwan ROC, and GB encodes the simplified Chinese character set used in mainland China. There is overlap between the two character sets. Hong Kong is situated at the crossroads as far as character sets is concerned, and hence utilizes both encodings.

lingual", i.e. the running text may be primarily in English but includes Chinese terminologies,[3] or primarily in Chinese but include English terminologies,[4] or the two languages may be mixed syntactically as well as lexically.[5] "Mixed-linguality" creates problems in decoding, since Big5 or GB requires 16 bits per character instead of the standard 8 bits. It is very cumbersome that a given document needs to be duplicated into versions in English, Chinese Big5 and Chinese GB, with a significant amount of redundancies. It will be desirable if we can discover elegant methods to support *multilinguality* or "mixed-linguality" for the interface. This may involve technologies such as translation, transcoding, meta-data (Chinese XML), etc.

Aside from textual displays, we also need to support three languages in spoken language interfaces. At the Chinese University of Hong Kong, our group[6] has developed three monolingual speech recognizers, together with one trilingual speech recognizer, all of which supports information enquiry for real-time financial information [2]. Since every user has individual preferences in choosing which language to use and when to use it, language flexibility is best supported by the trilingual speech recognizer constituting a human-computer speech interface. Similar issues also arise for speech synthesis, as the language chosen for speech input generally matches that for speech output.

III. HCI and Communications

Network bandwidth has direct impact on the latencies in system response times and display rates, which are critical elements in HCI. The issue becomes particularly important in situations such as streaming video or audio, where the transmission rate needs to match up or exceed the display rate.

Bandwidths may vary across a wide range e.g. from the telephone modem to the Ethernet, ATM, etc. Hence some networks backbones can support high data rates, while others will experience long delays. As an example, we can consider a sizeable image map on an information server, which may be retrieved by *local* clients residing on the same highspeed backbone as the server, but also *remote* clients accessing through a telephone modem. It should be desirable to enable all clients access the entire image within a reasonable amount of time. If we apply the "information-seeking mantra" presented in [3] – "Overview first, zoom and filter, then details on demand," then we should consider providing a compact version of the same image, trading off size with resolution, to tailor for remote clients accessing over narrow bandwidths. This should be followed by further details of the image provided on-demand. One research question here is: How can we provide "multiresolution" information displays, which may be scaled up or down to accommodate for bandwidth availability?

Regarding the transmission of the speech signals for spoken language interfaces, the landscape is also changing rapidly. Previous research has focused on microphone and wireline telephone speech. However, the emergence of various wireless telephone standards (e.g. D-AMPs, CDMA, GSM, PCS, dual GSM and PCS, all of which are used in Hong Kong) as well as voice-over-IP (VoIP) applications calls for further speech technology developments. For example, handling interrupted

[3] This is found in the local (Hong Kong) English newspapers.
[4] This is found in the local (Hong Kong) Chinese newspapers.
[5] This is found locally in everyday conversations, radio shows, etc.
[6] The research interest of the author is in the area of spoken language applications, which involve speech recognition, natural language processing and spoken dialog design.

transmissions due to cellular dropouts and blind spots, or the effect of packet losses and delays in VoIP may be relevant research problems.

IV. HCI and Computers

Computers are appearing in many shapes and forms – there are large server machines, desktop machines, as well as portable laptops and handheld PCs. Much variation exists in terms of computing power, storage space, screen size, keyboards, cameras, microphones and other peripheral interaction devices. The availability of computing power and storage space can help alleviate the demands on network bandwidth. Screen sizes vary from that of the desktop monitors (or larger), to the few-inch screens on handheld PCs, to the screens of pagers or mobile phones which can only display a few lines of text. One research issue is how we can design the interface of a given application to migrate gracefully from one computer to another. For example, when visual information display is constrained by a small screen size, the information transmission needs to be diverted to other input and output modalities for effective human-computer communication.

We can consider again the case of a spoken language interface (SLI). It may be "displayless" (where the user communicates with the system solely over the telephone channel) or "displayful" (where the user is equipped with the telephone augmented with a computer screen display). A previous system named WHEELS [4] serves as an example, where the user can converse with the system in order to browse through a several thousand electronic automobile classifieds. In the "displayless" version, WHEELS relies heavily on speech synthesis to provide the user with automobile descriptions; whereas in the "displayful" version, the entire automobile description (original text of the classified) is simply shown on the computer screen. We may wish to pursue this research direction, and explore how different interaction modalities may be combined in a flexible manner, to enhance the *portability* of the *effective* interface across computing platforms with varying resources.

An HCI Research Agenda

In the above, we have present the general areas considered to have critical implications on HCI, namely content, communications, computing and their convergence. In the following, we present several related issues specific to the area of spoken language interfaces, which is the author's area of expertise:

On Content:

- What mechanisms should we use to effectively express content in terms of spoken language, to be presented as synthesized speech to the user? How should human-computer spoken dialogs be designed in order to enhance such spoken presentation of information?
- How should we design spoken language interfaces for multi-lingual (or "mixed-lingual") applications?[7]

On Communication and Computing:

- How can we adapt speech technologies to support the voice inputs over different transmission protocols?

[7] Speech is especially desirable for the Chinese language, whose ideographic nature renders typing a cumbersome input method.

- Depending on the availability of network bandwidth, how can we seamlessly migrate between a “displayless” and a “displayful” interface? A visual display generally involves images, and hence has higher demands on bandwidth availability. Image rendering will also require more computation and storage as well. On the other hand, a displayless interface has lower demands on bandwidth availability, as it can be implemented entirely on a single phone line. However, the information delivery rate along the audio channel alone becomes much slower.
- For a given application, how can we implement the “displayful” spoken language interface which is portable from a desktop computer-telephony setup to a handheld device with a tiny screen and wireless voice services?

We have hereby presented some key issues concerning spoken language interfaces in relation to the 3C's, and look forward to an opportunity to further discuss them at the Workshop.

References

[1] Zue, V., H. Meng, D. Tennenhouse, J. Seid and T. Wilcox, “*Made By Hong Kong*,” Industry sector report on Information Technology, ed. S. Berger and R. Lester, Oxford University Press, 1997.

[2] Meng, H. et al, "Scenario-based Design of a Speech Recognition System," Technical Report ERG4910D, Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, February 1999.

[3] Shneiderman, B., “Designing the User Interface: Strategies for Effective Human-Computer Interaction,” 3rd edition, Addison Wesley, 1998.

[4] Meng, H. et al, “WHEELS: A Conversational System in the Automobile Classifieds Domain,” Proceedings of the International Conference on Spoken Language Processing, 1996.