Desktop conversations — the future of multimedia conferencing

M Russ

Both the recent explosion of interest in the Internet and the ongoing development of video-on-demand services have focused attention on the retrieval and delivery aspects of telecommunications networks. Yet the telephony network is fundamentally based on a much more basic human need — two-way communication. In fact, people normally communicate using conversation — interactive speech plus additional information in the form of non-verbal cues like body language.

The underlying technologies and standards to enable the easy exchange of information in a conversational way are now converging — which brings together the fields of conferencing, multimedia, and telephony into a unified whole. This could well be the catalyst to turning data communications into the dominant conversational medium of the future.

This paper looks at the ways that the desktop may soon rival the telephone as the natural route for carrying out a media-rich conversation with a remotely located person, and focuses on one example of a desktop conferencing application — Passepartout.

1. Introduction

The history of human beings documents an ongoing desire to communicate at distance. The close-up interaction of body language works at very short distances, while the spoken word can work to large audiences over longer distances. The spoken word is an attempt to translate thoughts into a code which effectively conveys meaning to a listener, and this ‘translation and coding’ element forms a common thread through all communications theory. Where the spoken word is too quiet, then shouts can be used over longer distances at the cost of a significant loss of both intelligibility and speed of transmission.

Where audio signals fail, then visual methods can be used. Shouting is frequently accompanied by gesticulation — normally arm-waving and exaggerated body movements. The arm waving works much better if the length of the arms and the position of the hands can be emphasized, which results in the use of flags, and ultimately, in telegraph systems using large mechanical arms mounted with lights. The electrical analogue of these telegraph systems shares the same name, and uses simple electrical signals to convey the information — either via multiple wires and an indication device, or by using a time-based coding scheme like Morse code. Extrapolations of this technology to telegrams and telex merely return to the written word as the coding method, and the physical transport is replaced with an electrical equivalent.

The telephone returns to the spoken word, but uses electrical signals to transmit an audio signal across distances. Facsimile reuses the technology of the telephone to provide the transmission of pictures, and although this does not seem significantly different to a telegram or telex message to users of languages which use the small number of characters like the Roman or Cyrillic letter forms, for users of pictorial languages with large numbers of characters like Chinese or Japanese the facsimile is a very different communications device indeed. The videophone combines a telephone with, typically, a video picture of the two participants — although it is sometimes also possible to replace the video with the ability to share a common view of a picture or document.

All of the spoken or visual systems mentioned so far allow communication between groupings of people — either ‘one-to-many’, or ‘many-to-many’. In contrast, the electrical systems have been ‘one-to-one’, where two participants can communicate either by coded words or by audio. Extending this ‘communication between two points’ to multiple points uses the concept of a gathering together of people to talk — a conference. Teleconferencing is the process of linking together several participants in a multi-point conversation using telecommunications, and it can use audio and video as well as additional data (text or pictures). Just as the telephone allows people to talk even though they
are geographically separated, so teleconferencing allows a conference to be held when the participants are in different locations.

To arrive at a desktop teleconferencing system, the way that the telephone and the desktop work needs to be considered, and then the standards infrastructure that is required to make a broadly applicable product, which exploits the merging of telephone and computer, needs to be looked at.

2. **Telephony — two-way interaction**

The telephone appears such a natural device for providing conversational interaction that it is now difficult to envisage a world without it. But the original intention of the telephone was to provide access to audio signals at a distance — which is where the name comes from. The word ‘telephone’ is made up of two Greek parts — tele meaning ‘remote’ and phone meaning ‘voice’. Comparisons between the radio and the telephone as a way of listening to music or speech appear unlikely in a modern context, but the use of the telephone to carry out conversations which are disjointed by distance is a serendipitous result of the technology, not its original intended usage.

Although it tends to be used mainly for person-to-person communication, it is possible to use the telephone to connect several people together — this is called audio-conferencing. Compared to a face-to-face situation where several people are gathered around a table, audio-conferencing provides no visual information about who is speaking, who is shuffling, or who keeps glancing at their watch. Telephone calls to more than one other person, with or without video, tend to be used mostly for business purposes. Adding video pictures to the telephone can provide some of the information that is lost when using just audio connections. Despite this, the videophone has only slowly moved from speculative fiction to real-world usage, whereas the telephone is almost ubiquitous.

3. **The desktop — one-way retrieval**

The desktop is the word used for the graphical user interface (GUI) that is presented on the screen of many personal computers. It is called ‘the desktop’ because it uses some of the elements of a real desktop to provide a metaphor for controlling the computer. The wastebasket or trash-can is perhaps the most obvious example — to throw something away you put it on top of a small picture or ‘icon’ that represents the wastebasket, and the item is then deleted by the computer. Documents are represented as icons which have the appearance of documents — although they are still stored as data files inside the computer. Storage locations inside the computer are represented by icons which appear either as pictures of the storage device itself (a floppy or hard disk, perhaps) or as a folder to indicate several documents filed in one place. The contents of an icon or folder can be displayed in an on-screen ‘window’ area — the equivalent of opening up the folder and looking at a catalogue or directory of the contents. To move documents from one place in the computer to another, they are dragged from one icon, folder or window to another.

Because many of the desks found in IT-aware companies are equipped with a telephone and a networked computer, the desktop is a familiar interface to many business users. Unlike the telephone, the desktop has only been in common usage for a few years, and so its use is still developing, although by adopting the metaphor of a real desktop it carries with it a number of associative characteristics. The difference between resources which are local to the computer, and resources which are provided by a local area network (LAN), is hidden from the user — they appear as icons or windows, and can be manipulated in exactly the same way. This is being extended to wider network resources like the Internet, where future operating systems and GUIs will provide effectively ‘transparent’ access to the WWW.

The main focus of the desktop is one-way — often just the retrieval or publishing of items of information. Just as sending a letter necessarily involves a time-delay while it is in transit, processed and a response sent back, so desktop systems tend to have corresponding delay mechanisms. E-mail merely shortens the time in which a reply may be received, but provides no guarantee that it will be any faster than conventional mail. The Internet provides access to huge quantities of information, but the inherent interaction is normally with a remote database rather than a remote person, and so is characterised by the download of large quantities of text and pictorial information, while the upload of control signals normally occupies significantly less bandwidth.

Business users who have a desk equipped with a telephone and networked computer are also likely to be users of conferencing. At the moment, conferencing is synonymous with audio or video conferencing, and, while the desktop may be involved in the set-up of the conference, it is rarely used interactively other than as a way of displaying the video picture. When collaborative tools are provided, the use of interactive working with another person is unfamiliar in the context of a desktop, and so is under-utilised. The desktop metaphor needs to be extended so that it also becomes associated with interactive bi-directional usage instead of uni-directional retrieval or publishing activities. This will involve a change in the public perception of the telephone and the computer, but it also requires an amalgamation of their different standards and methods of working. Computer/telephony integration (CTI) is one aspect of this merging process [1].