# Issues in Multimodal Human-Computer Communication

Harry Bunt  
Computational Linguistics and Artificial Intelligence Group  
Tilburg University  
5000 LE Tilburg, The Netherlands  
bunt@kub.nl

## 1 Introduction

Human communication is inherently multimodal in nature. People naturally communicate by means of spoken language in combination with gestures, mimics, and nonlinguistic sounds (laughs, coughs, sniffs,... ). They thereby use several communicative systems (‘modalities’), and several physical ‘carriers’ of the messages of these systems, such as vocal sounds and visible hand movements, (‘multimedia’), the use of which involves several perceptual channels. The various modalities are moreover not used independent of each other, but in such a way that the information communicated using one modality depends on that for which another modality is used more or less simultaneously.

Many modalities have originally developed in combination with a particular medium, but have in time become used also in combination with other media. For instance, natural language originally developed in combination with speech but later also with print and Braille, and gestures and mimics, like pointing, winks and smiles, originally developed using visible movements and facial expressions (involving the haptic-kinesthetic and visual channels), but have recently found new expression with the medium of computer screens as ‘=>’, ‘:-)’ and ‘:-('. The development of new media, in particular of digital media, has contributed much to the insight that communicative systems should be conceptually distinguished from the physical carriers they use, thus leading to the distinction between (multi)modality and (multi)media. Multimodal communication usually involves several media, but may also derive its multimodal character from the use of several communicative systems that make use of the same medium, e.g. text and graphics.

When the communicative situation precludes the use of certain media or modalities, the degree of multimodality and multimedia obviously decreases; for instance, two people who speak totally different languages can only communicate by means of gestures and mimics. It seems, though, that people naturally exploit all the available communicative resources, communicating according to what we might call the Multimax Principle: *In natural communication, the participants use all the modalities and media that are available in the communicative situation.* Natural communication is thus maximally multimodal and multimedia.
It seems quite clear why this should be so: the maximal use of modalities and media gives a maximal bandwidth for communication, allowing communication to be maximally effective, by exploiting redundancy, and efficient, by using different media/modalities simultaneously for different aspects of the information to be exchanged. Widening the communication bandwidth by arbitrary adding more and more modalities and media is of course not the way to optimize communication, however; the success of adding modalities depends on whether the cognitive processes operating on the messages conveyed in these modalities are capable of handling the information flow allowed by the bandwidth. Human cognitive processes are well-attuned to the forms of multimodality and multimedia of natural communication between people, but for new, technology-driven modalities and media, this is not always the case; the design and evaluation of new forms of multimodal human-computer communication which are well-suited to human cognitive and perceptual processing therefore forms an important area of research. The same goes for new forms of human-human communication mediated by new modalities and media. In computers, the relation between input information and internal (‘cognitive’) processing is a well-known bottleneck from the point of view of information flow (see e.g. Thimbleby, 199o), which stands in the way of developing powerful user interfaces.

2 Interaction and Human-Computer Communication

When using a computer, one interacts with it in two fundamentally different ways. On the one hand, one formulates instructions or queries, typically by typing commands and by clicking on menu items. On the other hand, one adjusts the window size, the font size in a window, and the brightness of the monitor; this is typically done not by typing commands, but by physically manipulating switches or by mouse clicks (again!) on a window edge or on items in a pop-up menu. Activities of the latter type are similar to adjusting the colours or the volume of a TV, and are not communicative in nature; instead, in these activities the user treats the computer as a physical object, comparable to a TV, a washer, or a kitchen machine. One does not ‘communicate’ with such objects, but ‘interacts’ with them as with other physical objects. By contrast, when one inputs a query or a command, one is treating the computer as an agent that tries to find answers to queries and to carry out commands, i.e. one treats the computer as a conversational partner: one communicates with the machine.

Communicating with computers is often a rather frustrating business, because computers don’t live up to the expectations created by the ‘conversation metaphor’ (Hutchins, 1989). A conversational interface by its very nature offers a mode of interaction that has certain similarities with natural human communication; there is an implicit suggestion that the main difference is one of syntax: if only one could type English instead of Unix commands, one would essentially be able to communicate in a similar way as when talking to an assistant. Computer experts know that this is not true, that there are fundamental limitations behind the syntactic obscurities and constraints of most user interfaces, but they can’t