

A. D. N. EDWARDS

MULTIMODAL INTERACTION AND PEOPLE WITH DISABILITIES

1. MODALITIES, COMPUTERS AND PEOPLE WITH DISABILITIES

What is the connection between computers, multiple modalities and people with disabilities? A traditionally scientific chapter might start out with definitions of these terms. However, there is problem in doing that in this case, which is that only one of them – ‘computer’ – is at all easy to define.

Daily activities of people take the form of interactions between them and their environment (where ‘environment’ is meant in a broad sense, encompassing other people as well as the physical surroundings). These interactions occur through an ‘interface’ which uses the physical, cognitive and sensory functions of the person. If any of those functions is impaired to the extent that the person finds forms of interaction difficult or impossible, then that person is said to be disabled (UN, 1981). The degree to which that disability handicaps the person depends on the extent to which the impaired function can be supported or substituted.

Before the discussion becomes too abstract, let us consider some examples. A person with a hearing impairment has difficulty interacting with the auditory component of the environment. A hearing aid may help them to continue to operate in an auditory mode, which amounts to supporting the impaired channel. Yet, if the impairment is so severe that the hearing aid cannot assist, then they may still take part in conversation by substituting non-auditory channels. That is to say that the visual channel can be used to pick up the visible cues of speech (lip and tongue movements, facial expressions *etc.*) and hence substitute for the auditory channel.

This example is a good one because it illustrates how compensation may take place at a human level (the physical mechanisms of speech production) or by the application of technology. Often people with disabilities can be accommodated within human interactions because of the richness of the interaction, but where they cannot, technology has an increasing role to play – and that improvement is largely due to the broadening of the technology to exploit more modes of interaction.

The excitement about the technology in this area is that it is opening new opportunities. The technology can make some tasks easier for people with disabilities and in many cases can make things possible that were previously impossible. This chapter will describe a number of examples of this. It is the extension of the technology into new modalities of interaction that is making new possibilities viable.

Speech-reading (the current, more accurate term for ‘lip-reading’) is an example of a mapping from one communication channel to another, in this case from the auditory to the visual. The availability of multimodal technology facilitates this kind

of mapping with the aid of technology. That is to say that speech-reading makes use of inherent redundancy in speech communication, but where such redundancy is not present, such mappings can be created technologically. This is why multimodal technology is such an important development for people with disabilities.

2. COMMUNICATION AND THE SENSES

Communication takes place via the five senses:

vision,
touch,
smell,
taste,
hearing.

Each of the sensory channels has particular characteristics, strengths and weaknesses. In many ways, vision is primary. A large proportion of the brain is devoted to visual processing and studies have shown (Mayes, 1992) that when conflicting signals are presented on the visual channel and another one, it is the visual one that will tend to be believed. Vision is very powerful, so that large amounts of information can be presented visually at any time and the real power of vision comes from the fact that it is possible to focus attention very precisely. There may be many objects and events in any visual scene and the viewer may have the impression of taking in all that information at once. In truth, the field of attention is very narrow – but that attention can be switched very quickly. Thus an event in the periphery of vision will attract attention and the eyes will be shifted to focus on it. In this section, the primacy of vision in current human-computer interfaces is discussed as well as the possibility of shifting some communication to the other senses, where appropriate.

Touch is an interesting case. In some ways it is an under-regarded form of communication. The only formal tactile languages are those used by blind people. Braille is the best-known one, but there is also the less-known Moon language¹. While sighted people may think that their use of tactile communication is negligible touch-typing has become a major component of communication in this computer-oriented age. The majority of computer users are untrained and cannot truly touch type, but nevertheless they do rely on tactile feedback as part of their typing activity. There are also many other situations in which people rely largely on tactile feedback in interacting with switches, buttons and such-like (e.g. secondary controls in a car, such as heating and radio switches, which are activated without diverting visual attention from the road, the primary task).

The tactile senses are generally not only associated with pressure and feedback from physical contact, but also with sensations of temperature. This has been

¹ *Moon is used almost exclusively in the UK. It is based on tactile shapes that are more akin to printed letters and therefore more easily learned by people who have lost their sight later in life, after having had experience of visual reading. (RNIB (1996). This is Moon, RNIB. <http://www.rnib.org.uk/braille/moonc.htm>).*

suggested as the basis of a possible form of communication (e.g. Challis, Hankinson *et al.*, 1998) but (at least for the present) this is not practical, not the least because of health and safety considerations.

We usually associate touch with the cutaneous feedback from the skin (mainly of the fingers) in contact with objects. There is, however, another, related form of bodily feedback, usually referred to as kinaesthetic. That is the information that we have about our limbs and other body parts in terms of the awareness of our muscles. The combination of tactile and kinaesthetic can be referred to as *haptic* (Oakley, McGee *et al.*, 2000).

Smell is another important form of communication. The exact level to which people use it is disputable. There is clear evidence of its having a large influence on interactions between animals, but many people would prefer to suggest that human behaviour has risen above the influence of pheromones. Yet, even if smell does not play a part in inter-personal communication, it can carry some very important messages. People are generally very sensitive to smells as warnings: the presence of a fire, that food has gone off and such-like. Also, it is suggested that ambient smells can have an important effect on people's moods, which could turn out to be influential in commercial situations, such as consumer e-commerce web sites. Hitherto technology has not existed to control olfactory messages; it has not been possible to generate smells of particular types – though work is proceeding in that area (Youngblut, Johnson *et al.*, 1996).

Taste is very closely related to smell. In fact, taste is quite a crude sense and most of the sensations that we attribute to taste are in fact the results of the smells accompanying them. We can only distinguish four primary tastes: sweet, bitter, salty and sour and the richer sensations that we derive when we drink a glass of wine, for instance, are in fact generated by the aroma of the wine in the glass just below our nose. Like smell, it is not really possible to generate tastes on demand and there is the further complication that to be tasted a substance must come in contact with the inside of the mouth – which raises a wide range of health and safety considerations!

Technological constraints imply that in technology-mediated communication it is practical to use the senses of vision, hearing and touch. Physical, sensory and cognitive impairments may mean that one or more of these senses is unavailable or inefficient. It is the role of technology to supplement or replace the lacking function. Taking one form of information to make it accessible via a different channel implies a *mapping*. That is a main theme of this chapter – the technological facility to map information between different modalities in order to accommodate the needs of users with disabilities.

It may be said that the designers of modern computer interfaces exploit the power of vision, in making maximum use of visual displays. Another view would be that such designers are lazy; if more information is required, they will slap another 'widget' onto the display, so leaving it to the user to cope with this extra complexity. With more thought, there might be better ways of presenting the new information, ways that will not increase the visual complexity and the user's task load. It is to be expected (and hoped) that in the future designers will be aware of the possibility of using different channels when appropriate. For instance, Brewster (Brewster, 1994) demonstrated that by analysing human-computer interfaces, in terms of *events*,

status and *modes* it was possible to identify information that was hidden from the user, which then could be presented in an auditory form.

While sight is generally assumed to work well in processing simultaneous sources of information, hearing is usually assumed to not be good at such parallel processing. This is not necessarily true, though. Massively parallel information can be presented in sounds - if they are designed in the right way. Once again it is more a question of attention switching. Buxton's example (Buxton, 1989) is of driving a car, when the driver might be engaged in a conversation, but at the same time may have the radio on, be monitoring auditory signals from the car (turn indicator clicking, note of the engine *etc.*) and be aware of external signals, such as an ambulance siren. In the event of a significant change to the auditory scene (such as a traffic report on the radio or the onset of a 'clunking' noise in the engine), the driver may have to withdraw from the conversation to switch attention to the alternative event. Another popular example is known as the 'cocktail party effect'. In a busy room with conversations all around, it is possible to have a dialogue with another person without interference. Yet the auditory system still monitors the ambient sounds, so that, for instance, if the person's name is spoken by someone elsewhere in the room, their attention will be drawn to that and away from their current conversation.

This processing of different sources of sound is known as auditory streaming (Bregman, 1990) and is mentioned again in Section 5. One difference from visual attention is that sound is not directional, so that it is not possible to focus exclusively on one sound to the same extent. Hearing has a degree of directional discrimination, but this is not very precise in humans, who have their ears on the side of their heads and which cannot be turned independent of the head (unlike some animals). Thus it is a natural reaction to turn the head in the direction of a sound in order to locate it or to listen to it.

Another important difference with hearing is that sound is inherently transient; it exists in time. It is not possible to review or re-examine a sound. The only mechanisms for doing this are dependent on memory. For instance, eye tracking experiments show that the process of reading visual text is not a simple left-to-right serial scan, but involves moving back and forth, revising and reinforcing words read. By contrast, spoken words are lost as soon as they are spoken. All that remains is an internal representation, the form of which depends on the amount of information presented and on time. (See Chapter 2 of Pitt, 1996, for more details).

Working visually one has a broad field of view, but the ability to focus on a narrow portion of that input. By contrast, tactile communication is inevitably narrowly focused. That is assuming that tactile communication takes place through the fingertips, which are the most convenient means. The 'field of view' of the fingertips is very narrow, and it is not possible to build up a larger picture by moving the fingers around, in the way that visual pictures are built by rapid movements of the eyes. Use of the tactile senses can be improved by training, so that people can learn to some extent to build more complete pictures by tactile exploration.

The tactile sense has very low resolution. The number of different surface textures that can be recognized by most individuals is small. The number can be

increased by using different materials (e.g. rubber, leather, paper and aluminium) but it is usually not practical to produce tactile materials (effectively collages) using such materials.

Impairments which affect one sense or communication channel can be alleviated by substitution of a different channel. That implies mapping information from one form to another. The above discussion has illustrated that the channels have different inherent characteristics. Hence such mappings are not always straightforward. Before we go on to examine such mappings, though, it is necessary to clarify a further point, which is that channels are not simple, uni-dimensional entities.

3. MODALITIES

It is important to realize that although here we are considering just three channels of communication, corresponding to the available senses, there are many more *modalities* – of communication².

As an example, there are a variety of forms of visual communication, and printed forms may themselves be subdivided into textual and pictorial. Mappings need not be only between channels, but may also be from one modality to another. For instance, textual, written communication is not usable by someone who is illiterate, but pictures may be. (See the examples of picture-based communication in Section 4.3).

Even within one modality, important variations exist. For instance, there is more than one style of writing; the full, emotional message of a poem is different from the dry, factual information within a technical manual.

In principle, the same information can be communicated in different modalities. In practice such mappings are not pure. That is to say that in translation to another modality, the meaning is usually altered, albeit subtly. For instance, speech and writing is based on words, but speech includes elements of intonation and prosody that are mostly lost in text. Thus, the simple utterance

(1) It's raining.

might be a statement, but if spoken with a certain intonation (a rising pitch, in British English), could be transformed into a question. A writer using those words as a question would signal the intention with a question mark, but there are other, more subtle variations that cannot be captured grammatically. For instance (an example borrowed from Stevens, 1996),

(2) Robert does research on drugs.

² 'Multimodality' is the topic of this book – and yet few authors agree exactly as to the meaning of the word *modality*. (See, for instance, the discussion in Blattner, M. and Dannenberg, R. B. (1992). *Introduction: The trend toward multimedia interfaces*. (in) *Multimedia Interface Design*. M. Blattner and R. B. Dannenberg (Eds.), New York, ACM Press, Addison-Wesley; pp. xvii-xxv.). In despair of finding consensus on a definition, this chapter does not attempt any new definitions, but attempts to at least be self-consistent.



<http://www.springer.com/978-1-4020-0635-7>

Multimodality in Language and Speech Systems

Granström, B.; House, D.; Karlsson, I. (Eds.)

2002, X, 243 p., Hardcover

ISBN: 978-1-4020-0635-7