Edwards, A. D. N. (2002). Multimodal interaction and people with disabilities. (in) <http://www.wkap.nl/prod/b/1-4020-0635-7>
Multimodality in Language and Speech Systems. B. Granström, D. House and I. Karlsson, (Eds.). Dordrecht, Kluwer, pp. 73-92.

T.H.E. Editor(s) (ed.), Book title, 1—6.

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

T.H.E. Editor(s) (ed.), Book title, 1—6.

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 signal 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 one. 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 language[1]. 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 suggested as the basis of a possible form of communication (e.g. Challis, Hankinson

---

[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.).

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; 2000).

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 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, 1990and 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 it 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 build 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 straight-forward. 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 there we are considering just three channels of communication, corresponding to the available senses, there are may more *modalities* – of communication[2].

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 which 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.

might be read as meaning that Robert is a pharmaceutical investigator – or that he performs research work while under the influence of narcotics. Mapping between modalities is, therefore, not always as simple as one might hope. The text of Sentence (2) could easily be passed to a speech synthesizer, thereby mapping from text to speech, but the synthesizer would have to impose one or other of the interpretations – possibly the wrong one.

Table 1 is based on one produced by Jens Allwood showing how the different components of human dialogue are perceived by the 'listener'. Such a table might be used as a basis for remapping communication to accommodate a disabled communicator. For instance, for a deaf person, the *hearing* column is unavailable and so one of the other senses might be used instead. Vocabulary might be shifted into the Vision column, by use of text instead of speech.

Table 1. Modalities in human dialogue (after Allwood). Each row is an expressive modality while the columns indicate how the modality is perceived by the 'listener'

		Hearing	*Vision*	*Touch*	*Smell*	*Taste*
Speech	Prosody/ phonology	Ŏ	Ŏ			
	Vocabulary	Ŏ				
	Grammar	Ŏ				
Gestures	Head movements		Ŏ			
	Facial gestures		Ŏ			
	Manual gestures		Ŏ	Ŏ		
	Body movement	Ŏ	Ŏ	Ŏ		
	Posture		Ŏ			
	Touch			Ŏ		
	Smell				Ŏ	
	Taste					Ŏ
	Writing		Ŏ			

It would be constructive if such a table could be extended to all forms of communication and used as the basis of (semi-) automatic remediation of communications impairments. However, this is not practical – at least not at the current state of knowledge. For a start we would need a fine grained taxonomy of communication channels and acts. Whereas Table 1 demonstrates that psycho-linguists have a good grasp of the nature of human dialogue, there are many other forms of communication that are less well understood. An extreme example would be a painting. Who can say exactly *what* the 'message' of (say) the Mona Lisa is, never mind how that message is communicated?

Table 1 also illustrates another point that should be borne in mind. It represents interpersonal dialogue and the very size of the table demonstrates the richness of that communication. However, most communication aid devices do not span the whole table. In other words, they tend to concentrate solely on the Speech rows of the table. The assumption is that it is sufficient to generate the vocabulary and grammar of

dialogue, but that is to exclude all the other facets. Communication aids are discussed further in Section 4.3

4. EXAMPLES OF NOVEL MAPPINGS

In this section we will illustrate the potential to use technology to map between modalities in specific applications which are aimed at alleviating the affects of disabilities. These examples are of existing systems, some (commercially) available and some which are rather more experimental. To an extent they represent the state of the art, but only hint at the applications that may eventually be possible with further development of knowledge and the technology.

4.1 Screen readers

The screen reader is probably the clearest example of cross-modality mapping technology. Blind people do not have access to visual information and computers are very visual in operation. A screen reader is software which captures the information that is on a computer screen such that it can be translated into a non-visual form. That form may be auditory (mainly synthetic speech, but also non-speech sounds) or tactile – braille.

One of the important features of the screen reader is that it works with standard applications. That is to say that the blind person can use the same word processor or database package as sighted colleagues do because of the addition of the adaptation on top of the application software.

A screen reader cannot simply ‘dump’ everything on the screen into speech – because of the differences between vision and hearing listed earlier. That is to say that a computer screen generally contains a large amount of information and the user focuses on the area of interest at any time, but if all presented in speech it would be an unfocused babble to the user. The screen reader therefore must provide some form of control. That is to say that the software must provide the control that the visual system does for a sighted user. It must present only the amount of information that the user can cope with at a time – but also make it possible to access other information. For instance, on a visual screen a warning may be flashed in the periphery to attract the user’s attention and that same warning will be equally valuable to the blind user.

The miss-match between the characteristics and capacity of vision and hearing lead to difficulties for the screen reader designer. The simplest approach would be to filter out information. However, it is not for the designer to decide what the blind user shall have access to. If information has been provided in the visual interface, it is there for a reason, and the blind user must be allowed access to it. The designer does have to decide, though which is the most salient information, that must be presented immediately to the user, and what is less important information that will not be presented in parallel, masking the primary message. This secondary information may be made more difficult to access (i.e. in response to an explicit request from the user), but still present. The problem with this approach is that it complicates the interface between the user and the screen reader. This compounds

the usability problem, because the screen reader is itself represents an addition to the complexity of the application being used. There is a difficult trade-off then between giving users as much power as possible but to do so in a way that they can utilize it, avoiding making it too complex to use. Human-computer interaction principles of simplicity, consistency and so on become even more critical. (Such principles are exemplified by guidelines such as Smith and Mosier, 1984, though they tend to be visually oriented).

In the early days of interactive computers, once control mechanisms had been devised, the problem of the screen reader designer was relatively simple – because the main mode of display was textual. In other words, the mapping from text to speech was easily achieved, through speech synthesis technology (Edwards, 1991). However, the advent of the graphical user interface (GUI) raised rather more complex problems. Now visual objects and properties other than text were significant and it was rather more difficult to convert these into non-visual forms.

The history of approaches to this problem is presented in Edwards, 1995 and eventually resulted in commercially available screen readers for common GUIs. Some of the design questions and trade-offs are documented in Mynatt and Weber, 1994.

4.2 Non-visual diagrams

While GUIs contain graphical elements, a particular problem for blind people is access to graphics in general, to diagrams and pictures. A number of approaches have been investigated experimentally, but none has yet been adopted as practical.

One potential mapping is into a tactile form. As discussed below, this can be achieved quite readily for static pictures, it is not practical for dynamic, changing graphics. The use of (computer) screens to display graphics is increasing and so it would be attractive to have some tactile counterpart: a tactile screen that can quickly switch between different pictures, or even display animations. Many developers have thought of this idea – but none has been able to find a practical solution.

Most of the devices are based on moving pins. A pin raised proud of the others can be felt. A row of such pins becomes a tactile line and so on. The practical problems are caused by the fact that the device is electro-mechanical and hence prone to physical and mechanical problems. To achieve good resolution the pins must be small and hence must be manufactured to high precision, making manufacture very expensive. Reliability is a problem. A mechanical fault may cause a pin to stick which may render the whole device practically useless. Refresh rates will be slow too because of mechanical delays.

A commercially available pin matrix device is available from Metec, in Germany, and has the following features. It consists of 7200 pins (120 x 60) in a 37.2 x 18.6cm rectangle. That gives a resolution of 3.2 pins per cm or 8 ‘dots’ per inch. The device can be used to display braille, giving 15 lines of 6-dot braille cells. Pins are individually addressable, which means that software can be optimized to reduce the area refreshed. This is important because it takes of the order of 21.6 seconds to refresh the whole display. The price of one of these devices is of the order of EUR 60,000.

Static tactile representations are rather more easy to produce, and modern computing and printing technology has a role to play. The simplest way to produce tactile graphics is using what is known as *swell paper*. This is paper with a plastic coating. The plastic is heat sensitive such that it expands on heating. So, diagrams can be produced by photocopying a black-and-white picture onto the paper and then putting the paper through a heating machine. The black areas absorb more heat and swell to make a raised area.

The development of this technology has made it almost as easy (and cheap) to produce tactile diagrams as visual ones and some guidance exists as to how best to design such diagrams (e.g. Hinton, 1996). However, it is apparent that many designers have not appreciated the fundamental differences between sight and touch and the implications for good design of tactile diagrams. Often tactile diagrams are simply visual diagrams printed on swell paper. They may appear comprehensible and attractive to visual inspection, but may be of little use to the blind person. Challis (Challis, 2000; Challis and Edwards, 2000) is developing guidelines for tactile diagram design which will avoid these errors.

Examples of the kind of error that can be made in this kind of visual to tactile mapping is the use of blank space. A plain white region of a visual map can meaningfully depict an empty area. However, if that same map is printed on swell paper and the reader places a finger on that blank area then no information is conveyed. There is no indication of what that area depicts nor of what direction to move the finger in order to find useful information. That kind of feedback could have been provided if there had been some form of texture in that area.

Challis's work seems to suggest that good, effective tactile diagrams may be difficult to interpret visually. For instance, he has been developing a tactile representation of music for blind musicians, and Figure 1 shows his representation of the score shown in conventional notation in Figure 2.

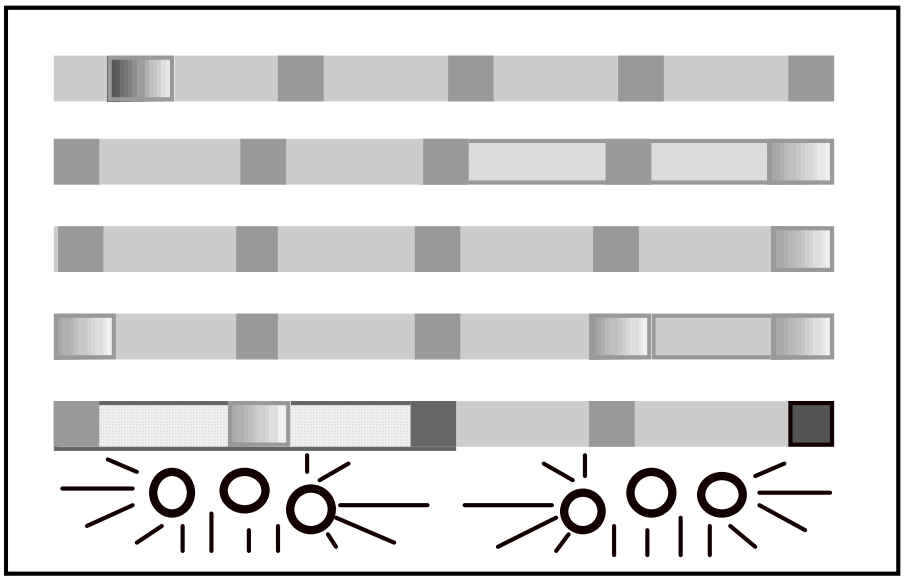

Figure 1. A Weasel overlay, representing the piece of music shown in Figure 2.

Figure 2. Common music notation representing the same piece of music as in Figure 1.

Sounds have also been used as a means of representing non-visual diagrams. Bennett (Bennett and Edwards, 1998; Bennett, 1999) experimented with the representation of simple 'box and line' diagrams (Figure 3) using speech and the non-speech sounds known as earcons (Blattner, Sumikawa *et al.*, 1989; Brewster, Wright *et al.*, 1995). Rigas (Rigas, 1996; Rigas and Alty, 1997) developed Audiograph, which represented the shape and position of elements of diagrams using musical encodings. The confusingly-named Audiograf (Kennel, 1996) is a prototype reader for diagrams that have hierarchies and connections. It is based on a model of audio-tactile exploration which assumes that there is a cycle of user interaction with the interface, through which the user enhances knowledge with regards to areas of uncertainty or interest by interrogating the system, In this way information is incorporated into the user's knowledge base.

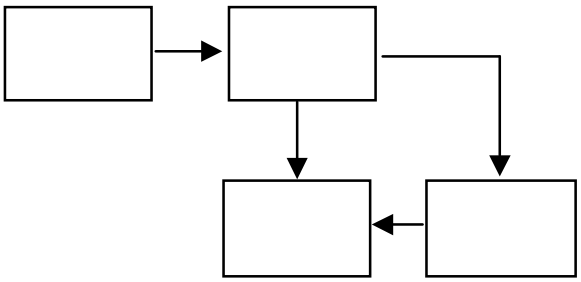

Figure 3. A simple box-and-line diagram, that might be converted into an auditory form following the approach of Bennett.

Examination of pictures is one problem, but beyond that it would be most valuable if blind people could create their own graphics. Kurze, 1996 has tackled this through a multi-modal approach, using both auditory and tactile feedback. Pictures are created using a heated stylus on swell paper, so that a tactile picture is created but at the same time audio feedback is generated.

A different approach to cross-modality mapping has been taken by Blenkhorn and colleagues to make software engineering graphical notations accessible to blind programmers. Dataflow diagrams are a graphical representation of software. There is an alternative, text-based representation in which components and the links between them are transformed into a matrix, known as an N-squared chart. This representation contains the same information, but it cannot be said to be perceptually equivalent. That is to say that there is greater cognitive effort on the user to extract the same information from the N-squared diagram. Generally sighted programmers would prefer the graphical representation over the N-squared diagram. However, the latter is easier to transform into a usable non-visual form, based on text to speech translation, that is the basis of the *Kevin* tool (Blenkhorn and Evans, 1994).

4.3 Communication aids

The use and generation of language is an important human characteristic (even if experiments with chimpanzees, dolphins and other animals may suggest it is not uniquely human). According to models such as that of Patterson and Shewell, 1987) the mechanisms of production of written and spoken language are intimately related. For instance, they suggest that a writer ‘hears’ an utterance internally before writing it on paper. Therefore, it is a moot point as to whether communication via an external device (usually referred to as *alternative and augmentative communication*, or AAC) involves any mapping between modalities.

However, there are evidently differences in the means of expression and comprehension of language. For instance, there are some people who lose the ability to use written form of language as a result of brain injuries (usually caused by trauma, as in road accidents, or due to a stroke). In some cases they retain an ability to recognize and manipulate other representations of language, such as pictures. This

is the basis of the *Lingraphica* communication aid (Steele and Weinrich, 1986; Sacks and Steele, 1993). There is some reason to suggest that such *aphasic* people benefit from languages based on multiple representations, the more modes included the better.

One of the most popular AAC systems is based on the *Minspeak* 'language'[3] as available on the Prentke Romich *Liberator* (Baker, 1982). Minspeak uses pictorial symbols which are selected in such a way as to generate words, which are spoken with a synthetic voice – and can be displayed visually as text. Minspeak can operate at different levels. At its simplest, each picture represents a single word or phrase and utterances are created by stringing together a set of these components. There are 128 pictures to chose from. Such a limited range of utterances might be sufficient for a new user of the system who has a limited vocabulary, but for every-day speech a much wider vocabulary is required. This can be achieved by allowing multiple selections of pictures. That is to say, if two selections are required to select a word, then the vocabulary rises to 128^2 = 1,6384 words. In practice the device may be programmed such that commonly used words and phrases are available through a single selection, while multiple selections are required for less frequently used ones.

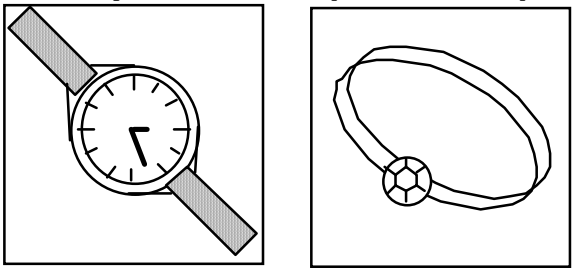

Figure 4. Icon-like pictures that might be used in Minspeak. The watch might be associated with a number of time-related concepts, while the ring might be associated with circular concepts (such as the number 0) as well as items which are expensive – associated with diamond rings.

In order to use Minspeak, the user must, of course, learn the mapping from sequences of pictures to language utterances. The Liberator can be programmed entirely by the user (or more likely their teacher or speech therapist). This means that it can be matched to the abilities of the individual. For some users it may be appropriate to use pictures which are accurate representations of real-world objects to generate the name of that object. For instance, a photograph of the user's mother could be used to generate the word 'mother'. This is the simplest form of mapping. However, to extend the range of utterances available, the pictures become more

[3] Some people (e.g. Strong, G. W. (1995). An evaluation of the PRC Touch Talker with Minspeak: Some lessons for speech prosthesis design. (in) *Extra-Ordinary Human-Computer Interaction: Interfaces for Users with Disabilities*. A. D. N. Edwards (Ed.) New York, Cambridge University Press**:** pp. 47–57.) argue that Minspeak is not truly a language in that it does not have a defined syntax and semantics.

abstract and more generalized. The way that this is achieved in Minspeak is through the concept of 'semantic compaction'. That is to say the attachment of as many associations as possible to each picture. As is illustrated in the example below, this often relies on the use of puns, which evidently assist in memorization. Selection is also structured, so that choosing one symbol will set a theme, while the second selection will then specify the particular object or concept within that theme.

For instance Figure 4 shows two pictures that might appear on a Liberator. The watch might be used for concepts associated with time. Selecting it twice could represent 'today', which is easier to remember if one thinks of 'two-day'. The ring can be associated with circular objects (including the number 0). The word 'tomorrow' could be watch + ring. That could be explained – and remembered – as watch setting the time theme and then the ring is reminiscent of the three letter Os in the word tOmOrrOw. the A more detailed example of generating an utterance using Minspeak is available in Chapter 4 of Edwards, 1991. Skilled Minspeak users can generate speech as quite fast rates and it would be interesting to investigate the way they map internally from symbols to speech and to see whether the process maps into the Patterson and Shewell (*op. cit.*) model of language production.

Since Minspeak replies on selection of symbols a variety of different techniques can be used to make those selections. The obvious one is to press keys. Other users may find it easier to select by pointing at the pictures using an infra-red source which can be attached to whichever part of the body they can best control (e.g. the head). Or, finally, selection may be made using a scanning system, so that a picture is selected by pressing a switch when it is highlighted.

One limitation of current communication aids becomes apparent when one considers them within a multimodal context. That is that designs concentrate solely on the verbal content of inter-personal communication. Linguists such as Allwood recognize that there is more to it than that (as exemplified in Table 1), but AAC designers tend to think of the task as being a simple translation from one modality of communication into speech. That is to exclude the other important modalities (bodily gestures, facial gestures, non-verbal utterances *etc.*) as if they are mere decorations to accompany the speech, when in fact they carry a lot of information. Researchers at Dundee University (Newell, Arnott *et al.*, 1995) are an exception in that they have looked at means of including non-verbal utterances in augmented communications.

The commonest mode of output of AAC devices is synthetic speech, which most closely matches natural dialogue. Another group of people who find spoken communication difficult or impossible are those who are deaf. Many such people rely on sign language. In the past there has been some controversy over the status of sign languages, but it is now generally accepted that they are true languages. Though they are equal in status to spoken languages, they do not resemble them. That is to say, for instance, that British Sign Language (BSL) is not a word-to-gesture translation of English. Indeed, although Britons and Americans speak the same language, American Sign Language (ASL) is entirely different from BSL; deaf British signers cannot communicate directly with Americans.

Again there is little point in getting into a debate as to the modes of communication involved, but it is apparent that technology may have a role in bridging the communication gaps that exist between signers and others (i.e. the vast

majority of the hearing population). Research is under way in a number of institutions into the possibility of automatic sign language to speech translation. However, Edwards (Edwards, 1998) suggests that there is a long way to go before this will be achieved. Translation in the other direction, from text (or speech, via voice recognition) into sign (rendered by an animated, cartoon-like avatar) should also become possible.

Multimodal technology is already playing a role, though, in other ways. Hitherto a problem with sign language is that it has not been easy to capture it in disembodied formats. That is to say that whereas (most of) speech can be translated into written text, there has been no equivalent for sign. Paper sign language dictionaries have relied on static photographs, but, because signs are often dynamic, these have had to be augmented by arrows and similar notations. Now, with the ready availability of video, properly expressive dictionaries are possible and these are easily made available over the web. (See, Lapiak, for instance).

Speech recognition technology has greatly improved in recent years. Now recognition of connected speech is available at a low price. The recognition levels are often still frustratingly low for many users, but they are good enough as an alternative to the keyboard for many people who cannot use the keyboard for one reason or another. This would include people with physical impairments which preclude them from using a keyboard. Ironically a growing cause of such impairments is the increasing use of information technology, leading to repetitive strain injuries (RSI[4]) caused by typing.

An intriguing possibility for the exploitation of translation between modalities, is the combination of speech-to-text and text-to-speech which might assist communication for people with speech impairments. Some (speaker-dependent) speech recognition techniques rely more on consistency of the speaker than conformity to recognized language. Therefore, such a system can be trained that a particular sound has a specified meaning, regardless of whether that sound resembles the pronunciation of the word – as long as that sound is consistently reproduced by the speaker. Thus, a person who speaks in an idiosyncratic manner that (untrained) listeners find hard to understand, might speak to a machine that would translate into synthetic speech which is easy to comprehend. Experiments have been carried out which suggest that this approach would be viable (Edwards and Blore, 1995; Schulz and Wilhelm, 1992), though it has yet to be implemented in practice.

Speech technology may have a particular role for the large group of people who have problems with written language, due to dyslexia. For those who have difficulty writing, speech input may be a useful alternative. This is being used increasingly, though Elkind and Shrager, 1995 demonstrated that this may not be as efficient an alternative as is assumed. The problem is that speech recognizers are less than 100% accurate. In case of doubt the system will present the user with a menu of possible interpretations (near homophones) of the word spoken. Yet, what task is a dyslexic person likely to find more difficult than recognizing the correct spelling of a word?

---

[4] At present a variety of different terms are being used to describe this phenomenon. Another term is ‘work-related upper limb disorders’ or WRULD.

Speech output has a role, though. Many dyslexics find that they can spot errors when their text is read out to them (even in a synthetic voice) that they would not spot if they read the text themselves.

4.4 Further examples

There is quite a large range of assistive devices available for people with disabilities, which rely on multiple modalities. This list will only increase in size as the technology develops. Some more of those currently available are described in this section.

Optacon
Braille is one example of translation of (visual) text into a tactile form. Modern technology for optical character recognition, translation and printing (embossing) can facilitate the mapping, but not in real time. The Optacon is a device which achieves a visual-to-tactile translation automatically as the user reads. A miniature television camera is scanned across a printed page and the shape of the letter under the camera is reproduced on a pad of vibrating pins. The user holds a finger on that pad and can feel the shape of the letter. In this way a (trained) blind user can read printed text without assistance.

Although it has been a successful device and very popular with its users, the Optacon is no longer manufactured.

Speech viewers
Deaf people who try learn to communicate orally have difficulty in speaking because they cannot hear their own voices well. Their training can be supported by the use of devices which will present them with a visual representation of their speech sounds. Typically the speech therapist of teacher will speak into the device, producing a sample visual representation and then the deaf person will attempt to produce a similar pattern. Alternatively this approach may take the form of a game. The deaf person (usually a child) will achieve a good score by producing vocal sounds of the appropriate kind. IBM's SpeechViewer is an example of such a device.

Text-to-sign
People who are both deaf and blind effectively only have their tactile senses for receiving communication. They may communicate with a signer by feeling the person's hands. For people who do not know sign language experimental technology (known as *Ralph*, Jaffe, 1994) has been developed whereby a robot hand will perform the signs of fingerspelling. Thus, the non-signer can type input which is displayed on the robot hand which is felt by the deaf blind person.

Sound graphs
Graphs are a very powerful visual representation of quantitative information. Many of the same properties can be captured in a representation in sounds. Usually the x-axis is represented by time and the height of a curve on the y-axis by the pitch of a

note. One of the first publications of this idea is Mansur, Blattner *et al.*, 1985, one implementation is described in Edwards and Stevens, 1993 and a version that is currently available commercially is the Audio Graphing Calculator (AGC), available from *ViewPlus* (http://www.ViewPlusSoft.com). An evaluation of the usefulness of sound graphs can be found in Bonebright, Nees *et al.*, 2001.

5. A METHODOLOGY FOR MULTIMODAL DESIGN

The above discussion should have established the proposition that multimodal technology has a valuable role to play in the remediation of some of the effects of disabilities. A remaining question, though, is how best to make use of this opportunity. That is to say that there is a existing fount of knowledge as to how to design conventional, limited-modality interfaces and systems, but there is less known about how to design a good multimodal system. This is the question that has been addressed by Mitsopoulos in his development of a methodology for multimodal design.

It is beyond the scope of this chapter to outline the methodology in full, but it will be outlined here. Full details can be found instead in Mitsopoulos, 2000, while different aspects are described and examples worked through in Mitsopoulos and Edwards, 1998; Mitsopoulos and Edwards, 1999a; Mitsopoulos and Edwards, 1999b. The motivation behind the methodology is to support the designer of a multimodal interface. It does not present simple guideline but steps that the designer can go through. The methodology has been applied to the transformation of visual computer interfaces into auditory counterparts, but should be equally applicable to other modalities and to designing interfaces from scratch as well as adaptation.

The methodology works at three levels:

- conceptual
- structural
- implementation

At the conceptual level, the designer must consider the *tasks* to be undertaken, in terms of the information required by the user. That information can be classified according to the dimensions suggested by Zhang, 1996: *nominal*, *ordinal* and *ratio*.

Having thus ascertained the information required to support the tasks, the structure that will support the tasks can be identified. The structural components of an auditory interface are *streams* as described in Bregman, 1990. These are the auditory counterparts of the *objects* that make up any visual scene.

Now, having designed the structure, it is necessary to choose components which will implement it. In the case of an auditory interface, that means choosing sounds that will be perceived as belonging to the streams identified. Cognitive models, such as Interacting Cognitive Subsystems (ICS, Barnard and May, 1994) can give some guidance at this level – though the auditory and tactile contribution to cognition is less well developed in this model as yet.

6. CONCLUSIONS

The world is a sensually rich environment. That humans have thrived within it is to a significant extent due to the ability of the senses to capture information and of cognition to process that information. The richness of the environment enhances communication as it allows for overlap and redundancy.

The senses are not omnipotent, though; we cannot hear the whole range of auditory frequencies nor see the whole electromagnetic spectrum, and we do not necessarily understand all that we can perceive. Technology has enabled us to extend our inherent capabilities, though. For instance, we use the communication properties of radio waves by transforming them to and from signals we can perceive.

Some people's window onto the world is further restricted as a result of physical, sensory or cognitive impairments – to the extent that they are said to be disabled, but once again technology has a role to play. It can again broaden that window so as to reduce the deleterious affect of the impairment. To do this most effectively we need to imitate nature and to ensure that as many modes of communication are recruited as possible, in such a way that they reinforce and support each other. It is only recently, with development of digital information and communication technology (ICT) that this has become a realistic proposition. This introduces new and exciting possibilities. It is likely that technology will redefine what we mean by 'disability'. Most people who wear spectacles or eye glasses would not class themselves as visually disabled, though without the glasses they would find common tasks difficult. So it may be that in the future people who rely on multimodal ICT will be able to operate just as efficiently as their unaided peers and so not be classed or categorized as any different.

Department of Computer Science, University of York, York, UK, YO10 5DD

REFERENCES

.*ASL Dictionary Online*. http://www.bewellnet.com/dario/asl_dictionary_online_practical.htm.

Baker, B. (1982). Minspeak. *Byte* **7**(9): pp. 186–202

Barnard, P. and May, J. (1994). Interactions with Advanced Graphical Interfaces and the Deployment of Latent Human Knowledge. (in) *Interactive Systems: Design, Specification and Verification.* F. Paterno (Ed.) Heidelberg, Springer-Verlag**:** pp.

Bennett, D. (1999). *Presenting diagrams in sounds for blind people*, DPhil thesis, University of York, Department of Computer Science,

Bennett, D. J. and Edwards, A. D. N. (1998). Exploration of non-seen diagrams. (in) *Proceedings of ICAD '98 (International Conference on Auditory Display)*, S. A. Brewster and A. D. N. Edwards (Eds.), Glasgow, British Computer Society.

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.

Blattner, M. M., Sumikawa, D. A. and Greenberg, R. M. (1989). Earcons and icons: Their structure and common design principles. *Human-computer Interaction* **4**(1): pp. 11–44

Blenkhorn, P. and Evans, D. G. (1994). A method to access computer aided software engineering (CASE) tools for blind software engineers. (in) *Computers for Handicapped Persons: Proceedings of the 4th International Conference, ICCHP '94*, W. L. Zagler (Ed.) pp. 321-328, Springer-Verlag.

Bonebright, T. L., Nees, M. A., Connerley, T. T. and R, M. C. G. (2001). Testing the effectiveness of sonified graphs for education:A programmatic research project. (in) *ICAD 2001*, J. Hiipakka, N. Zacharov and T. Takala (Eds.), Espoo, Finland, Helsinki University of Technology.

Bregman, A. S. (1990). *Auditory Scene Analysis*. Cambridge, Massachusetts:, MIT Press.

Brewster, S. A. (1994). *Providing a structured method for integrating non-speech audio into human-computer interfaces*, DPhil Thesis, University of York, Department of Computer Science,

Brewster, S. A., Wright, P. C. and Edwards, A. D. N. (1995). Experimentally derived guidelines for the creation of earcons. (in) *Adjunct Proceedings of HCI'95: People and Computers*, G. Allen, J. Wilkinson and P. Wright (Eds.) pp. 155–159, Huddersfield, British Computer Society.

Buxton, W. (1989). Introduction to this special issue on nonspeech audio. *Human-Computer Interaction* **4**(1): pp. 1-10

Challis, B., Hankinson, J., Evreinova, T. and Evreinov, G. (1998). Alternative textured display. (in) *Computers and Assistive Technology, ICCHP '98: Proceedings of the XV IFIP World Computer Congress*, A. D. N. Edwards, A. Arato and W. L. Zagler (Eds.) pp. 37–48, Vienna & Budapest, Austrian Computer Society.

Challis, B. P. (2000). *Design principles for tactile communication within the human-computer interface*, DPhil thesis, University of York, Department of Computer Science,

Challis, B. P. and Edwards, A. D. N. (2000). Design principles for tactile interaction. (in) *First International Workshop on Haptical Human-Computer Interaction*, S. Brewster (Ed.) pp. 98-101, Glasgow, British Computer Society.

DigiScents (2000). *Digiscents: A revolution of the senses*. http://www.digiscents.com/.

Edwards, A. D. N. (1991). *Speech Synthesis: Technology for Disabled People*. London:, Paul Chapman.

Edwards, A. D. N. (1995). The rise of the graphical user interface. *Information Technology and Disabilities* **2**(4), (http://www.isc.rit.edu/~easi/itd/itdv02n4/article3.html).

Edwards, A. D. N. (1998). Progress in sign language recognition. (in) *Gesture and Sign Language in Human-Computer Interaction*. I. Wachsmuth and M. Frölich (Eds.), Berlin, Springer**:** pp. 13–21.

Edwards, A. D. N. and Blore, A. (1995). Speech input for persons with speech impairments. *Journal of Microcomputer Applications* **18**: pp. 327–333

Edwards, A. D. N. and Stevens, R. D. (1993). Mathematical representations: Graphs, curves and formulas. (in) *Non-Visual Human-Computer Interactions: Prospects for the visually handicapped*. D. Burger and J.-C. Sperandio (Eds.), Paris, John Libbey Eurotext**:** pp. 181-194.

Elkind, J. and Shrager, J. (1995). Modeling and analysis of dyslexic writing using speech and other modalities. (in) *Extra-ordinary Human-Computer Interaction: Interfaces for Users with Disabilities*. A. D. N. Edwards (Ed.) New York, Cambridge University Press**:** pp. 145–168.

Hinton, R. (1996). *Tactile Graphics in Education*. Edinburgh:, Moray House Publications.

Jaffe, D. L. (1994). *Ralph: A fourth generation fingerspelling hand*. http://guide.stanford.edu/Publications/dev2.html.

Kennel, A. R. (1996). Audiograf: A diagram reader for the blind. (in) *Proceedings of Assets '96*, pp. 51–56, Vancouver, ACM.

Kurze, M. (1996). TDraw: A computer-based tactile drawing tool for blind people. (in) *Proceedings of Assets '96*, pp. 131–138, Vancouver, ACM.

Lapiak, J. A. Handspeak: A sign language dictionary online., http://dww.deafworldweb.org/asl/.

Mansur, D. L., Blattner, M. and Joy, K. (1985). Sound-Graphs: A numerical data analysis method for the blind. *Journal of Medical Systems* **9**: pp. 163-174

Mayes, T. (1992). The 'M' word: Multimedia interfaces and their role in interactive learning systems. (in) *Multimedia Interface Design in Education*. A. D. N. Edwards and S. Holland (Eds.), Berlin, Springer-Verlag. **76:** pp. 1-22.

Mitsopoulos, E. (2000). *A principled approach to the design of auditory interaction on the non-visual user interface*, DPhil thesis, University of York, Department of Computer Science,

Mitsopoulos, E. and Edwards, A. D. N. (1999a). A methodology for the specification of non-visual widgets. (in) *Adjunct Conference Proceedings of HCI International '99*, H.-J. Bullinger and P. H. Vossen (Eds.) pp. 59-60.

Mitsopoulos, E. N. and Edwards, A. D. N. (1998). A principled methodology for the specification and design of non-visual widgets. (in) *Proceedings of ICAD '98 (International Conference on Auditory Display)*, S. A. Brewster and A. D. N. Edwards (Eds.), Glasgow, British Computer Society.

Mitsopoulos, E. N. and Edwards, A. D. N. (1999b). A principled design methodology for auditory interaction. (in) *Proceedings of Interact 99*, M. A. Sasse and C. Johnson (Eds.) pp. 263-271, Edinburgh, IOS Press.

Mynatt, E. D. and Weber, G. (1994). Nonvisual presentation of graphical user interfaces: Contrasting two approaches. (in) *Celebrating Interdependence: Proceedings of Chi '94*, C. Plaisant (Ed.) pp. 166-172, Boston, New York: ACM Press.

Newell, A. F., Arnott, J. L., *et al.* (1995). Intelligent systems for speech and language impaired people: A portfolio of research. (in) *Extra-Ordinary Human-Computer Interaction: Interfaces for Users with Disabilities*. A. D. N. Edwards (Ed.) New York, Cambridge University Press**:** pp. 8–102.

Oakley, I., McGee, M. R., Brewster, S. A. and Gray, P. D. (2000). Putting the feel in 'look and feel'. (in) *The Future is Here: Proceedings of Chi 2000*, T. Turner, G. Szwillus, M. Czerwiniski and F. Paternò (Eds.) pp. 415-422, The Hague, NL, ACM Press Addison-Wesley,.

Patterson, K. and Shewell, C. (1987). Speak and spell: Dissociations and word-class effects. (in) *The Cognitive Neuropsychology of Language*. G. S. Max-Coltheart and R. Job (Eds.), London, Lawrence Erlbaum Associates**:** pp. 273–294.

Pitt, I. J. (1996). *The Principled Design of Speech-Based Interfaces*, DPhil Thesis, University of York,

Rigas, D. (1996). *Guidelines for auditory interface design: An empirical investigation*, unpublished PhD Thesis, Loughborough University, Department of Computer Science,

Rigas, D. I. and Alty, J. L. (1997). The use of music in a graphical interface for the visually impaired. (in) *Proceedings of Interact '97, the International Conference on Human-Computer Interaction*, S. Howard, J. Hammond and G. Lindegaard (Eds.) pp. 228-235, Sydney, Chapman and Hall.

RNIB (1996). *This is Moon*, RNIB. http://www.rnib.org.uk/braille/moonc.htm.

Sacks, A. H. and Steele, R. (1993). A journey from concept to commercialization - Lingraphica. *OnCenter Technology Transfer News*(5), (http://guide.stanford.edu/Publications/issue5.html#ling).

Schulz, B. and Wilhelm, B. (1992). Access to computerized line drawings with speech. (in) *Computers for Handicapped Persons: Proceedings of the 3rd International Conference, ICCHP '92*, W. L. Zagler (Ed.) pp. 461-465, Vienna, Osterreichische Computer Gesellschaft.

Smith, S. L. and Mosier, J. N. (1984). *Design guidelines for user-system interface software*, Report ESD-TR-84-190, USAF Electronics Division, (http://www.info.fundp.ac.be/httpdocs/guidelines/ ??).

Steele, R. D. and Weinrich, M. (1986). Training of severely impaired aphasics on a computerized visual communication system. (in) *Proceedings of Resna 8th Annual Conference*, pp. 320–322.

Stevens, R. (1996). *Principles for the design of auditory interfaces to present complex information to blind computer users*, DPhil Thesis, University of York, UK,

Strong, G. W. (1995). An evaluation of the PRC Touch Talker with Minspeak: Some lessons for speech prosthesis design. (in) *Extra-Ordinary Human-Computer Interaction: Interfaces for Users with Disabilities*. A. D. N. Edwards (Ed.) New York, Cambridge University Press**:** pp. 47–57.

UN (1981). The United Nations Declaration on the Rights of Disabled Persons. *Unesco Courier* **1**: pp. 6–7

Youngblut, C., Johnson, R. E., *et al.* (1996). *Review of Virtual Environment Interface Technology*, Report IDA Paper P-3186, Institute for Defense Analyses - IDA, (http://www.hitl.washington.edu/scivw/IDA/).

Zhang, J. (1996). A representational analysis of relational information displays. *International Journal of Human-Computer Studies* **45**: pp. 59–74