Toward a Mobility Aid for the Blind

Roger F Browne
Institute of Information Sciences and Technology,
Massey University, Palmerston North, New Zealand
r.f.browne@massey.ac.nz

Abstract

The development of mobility aids for the blind is a challenging task that has a number of potential solutions. Direct electrical stimulation of the visual cortex has been used to generate ‘phosphenes’ (perceived as flashes of light) that can, with training, be interpreted as low resolution images. A second approach is to use a grid of tactile stimulators to represent an array of pixels. Thirdly, images can be mapped to an interpretation suitable for aural presentation. This approach provides the opportunity to maximise the amount of processing handled by the technology, thus freeing up the user to act on the information as it is presented.

Keywords: mobility, blind aid, scene interpretation

1. Introduction

Blindness is one of the most feared of all human ailments [1]. The most common cause of blindness in the world is cataracts, which can be operated on and normal eyesight restored in a fifteen minute operation [2]. In Western countries age-related ailments such as diabetes are a common cause of blindness, and these people are unlikely to be receptive to a technologically complicated solution to their desire to lead a normal life. Thus any sophisticated mechanism designed to enhance the mobility of the blind is likely to be limited to people for whom all of the following apply:

(1) Those who cannot be helped through standard medical procedures.
(2) Those who are likely to succeed at the task of learning to use an unfamiliar (and possibly complicated) device, perhaps involving the development of a whole new set of stimulus responses.
(3) Those who can afford to make an expensive purchase or who live in a country with welfare agencies able to make such a purchase for them.

The information capacity of the optic nerve has been estimated to be approximately four million bits per second [3]. In comparison, the information capacity of the ear has been estimated to be approximately eight thousand bits per second [4]. Jacobson [3, 4] has estimated the total capacity of the human body's receptor nerve fibres to be no more than ten million bits per second. Thus the eye may represent half of the total data input bandwidth for the human body. Therefore there would appear to be no non-invasive alternative to the eyes as a means of providing visual information at its full bandwidth.

When considering the potential technological solutions to blindness there are several possible approaches.

1.1. Approach One

If one or both optic nerves are intact retinal implants can be inserted. These can take one of either of the following forms:

(1) One method is to place a retinal implant chip over the photo receptors at the back of the eye. This is referred to as a sub-retinal implant.
(2) A second method places retinal chips over the ganglion cell layer of the retina. The axons of the ganglion cells form the optic nerve and carry the signals from the photoreceptors to the lateral geniculate nucleus in the middle of the brain.

A number of approaches to retinal implants have been reported in the literature:

(1) The implants may be artificial silicon retinas.
(2) A hybrid system may consist of an artificial retina consisting of a combination of cultured nerve cells with a microelectrical mechanical system, probably based on nanotechnology.
(3) A master gene has been discovered that determines whether or not an eye will grow. This opens the potential for growing artificial eyes employing gene technology.

Clearly these are very costly and invasive procedures.

1.2 Approach Two

A surgical procedure is available that results in direct stimulation of the visual cortex using video data [1]. This visual prosthesis produces a pattern of stimulation on the visual cortex that is experienced as a spatial pattern of "phosphenes", which are perceived as flashes of light following electrical stimulation. The position of each phosphene is a function of which part of the visual cortex has been stimulated. There can be up to 512 electrodes, which must be surgically implanted. This is a very expensive procedure (US$120,000?). It has been reported [1] that some users of this visual prosthesis have been able to drive a car while relying solely on input from the prosthesis.

1.3 Approach three

An alternative to the invasive procedures described above is to generate a very simplified version of the image, converting it to a form that some other body sense (tactile or audio) can interpret. This approach will be explored in greater detail in the remainder of this paper. In particular a proposal for an approach that uses off-the-shelf hardware will be described.

2. Constraints

In the area of aids for the blind the promises of new technologies have not always been fulfilled. According to *The Institute for Innovative Blind Navigation* [2], "There is a quagmire of negativity to wade through as you attempt to get the world to pay attention to your interests." They pose the following questions:

(1) Will the proposed system make travel less stressful than using the long cane or the guide dog?
(2) Will the proposed system add unnecessary quantities of information inflow that might interfere with the usual processing of the native senses (hearing, smell, kinesthesis, tactual)?
(3) Is it portable, attractive, and a reasonable cost?

They state that, "There is no money to be made making technologies just for the few good blind travellers on the planet who wish to use advanced technologies." Last but not least, they state that blindness agencies or blind individuals must be involved.

3. Review of work on tactile stimulators

Attempts to develop sensors based on tactile stimulators as the basis of a visual prosthesis date back to approximately 1970. Early equipment tended to be based on vibrating tactile stimulators, although electro-cutaneous stimulation has shown promise.

Geldard proposed the use of the skin as a communication channel in 1957 [5], and this marked the beginning of research in a number of centres. An important step in the development of a blind aid was marked by the Second Conference on Visual Prosthesis, held at the University of Chicago in 1969 [6]. Although the main thrust of the work reported was based on the use of phosphenes a number of papers dealt with experiments in tactile sensory substitution. Attendees at the conference were very optimistic that a usable prosthesis would soon be developed. However Ferguson and Ferguson [5] state that by 1986 no significant progress had been made beyond the initial experiments, and that a fresh approach was needed. In particular, account should be taken of what Bach-y-Rita terms 'brain plasticity' [7] which he defines as "the adaptive capacities of the central nervous system – its ability to modify its own structural organization and functioning." He notes that "specific training appears to be necessary to make maximum use of a sensory substitution device." Ferguson and Ferguson [5] conclude that electro-cutaneous transfer is the preferred approach.

The physiological responses of various stimuli have been investigated by Anani et al [8], Butikofer and Lawrence [9, 10], and Szeto and Saunders [11]. An overview of early work using a bank of 400 vibrators mounted on the back-rest of a dentist's chair is given by White et al [12]. The details of an electromagnetic tactile stimulator are provided by Holmlund and Collins [13].

Note that although work on tactile stimulators has concentrated on a grid pattern for input, there is no reason why the tactile input could not be used to simply 'nudge' or 'pull' the user, much as a human helper might do while walking.

4. Review of work on auditory image representation

One product that is commercially available is called "The voice" [14]. Each image as captured by a miniature video camera is swept from left to right at a little under once per second. The pixels in each column generate a particular sound pattern, consisting of a combination of frequencies based on that specific set of pixels. The result is an auditory 'signature', effectively an inverse spectrogram, that characterises that particular image. Meijer refers to the conversion process as sonification (auralization) [14]. This technology has the potential to support 3-D stereo vision. Overall the spatial resolution will be determined by the camera resolution, the details of the inverse spectrogram algorithm, and the combination of frequency and time resolution in the cochlea. Demonstration software available from [14] uses 64 × 64 images with 16 grey levels, and 64 available waveforms. However, according to Meijer, "Mastering it all will require lots of effort and practice, perhaps comparable to learning a foreign language" [14].

5. An alternative approach to auditory image presentation

The process described in section 4 above is a simple mapping from a sequence of images to a sequence of inverse spectrograms. The resolution in the time domain is between one and two orders of magnitude lower than that of the human eye, while spatial resolution is constant across the field of view (that is, there is no equivalent of the fovea). This crude representation of an image must then be interpreted by the user, making high-level judgements before being able to apply it to mobility tasks. Furthermore, aspects of the image that are irrelevant to mobility (for instance, regions of sky) will consume part of this limited information bandwidth.

The approach outlined in this section endeavours to transfer the high-level image interpretation to the hardware, thereby:

(1) substantially reducing the volume of data that must be presented to the user, and hence alleviating problems of data overload,
(2) maximising the relevance of the data to the task of achieving mobility.

Although the mechanism for transferring the final information to the user could be via any of the senses, transfer via verbal commands is explored in this work.

The sequence of processes is as follows:

(1) Image acquisition will provide a temporal sequence, most probably in colour. Optionally stereo vision could be implemented, but as applied to mobility the advantages may not justify the added complexity and expense. The orientation of the camera will be an important parameter. This could be provided by basic instrumentation such as a clinometer, perhaps in conjunction with a magnetometer.
(2) The system will perform low-level image segmentation to highlight objects or events of relevance. This will also involve correlation within the temporal sequence. Problem knowledge will be incorporated at this stage, with expectations about traffic flow, the nature of footpaths, and so on.
(3) High-level scene interpretation applied to the processed images will produce a symbolic description of the scene. The issue of problem knowledge is clearly a major one. Neural networks may be useful at this stage, but the likelihood that a user of the system will some day encounter a scene quite unlike any in the training set is a very real one. Thus the system is likely to need ongoing learning capabilities.
(4) This symbolic description is then converted into verbal instructions appropriate to the needs of the user.

For instance, in the case of work described by Younger [15] the image stream is reduced to a stream of relevant symbolic descriptors containing sufficient information on the location and status of a pedestrian crossing to deduce whether the system user can cross the road safely. These symbolic descriptors would be internal to the system, and provided they contained sufficient data for any conceivable road-crossing activity, would be totally independent of any changes to the verbal instructions given to the user.

Clearly such a system would require some level of feedback so that

(1) the particular requirements of the user, such as a desire to cross the road or enter a particular shop, are transferred to the system,

(2) the verbal instructions are at a comfortable level of granularity, and can be readily adjusted as needs change.

For instance, if the user was walking along a footpath and was approaching a pedestrian crossing the system would need to be directed to alert the user when the crossing had been reached. The system should avoid the danger of 'information overload' for the user. A possible enhancement would be to combine the image processing with a GPS-based system so that the camera gives fine detail to augment the GPS information. For instance, the image processing system would signal the presence of a pedestrian crossing, while the GPS system might provide the name of the road.

The work to date has just 'nibbled' at the periphery of the problem, no attempt being made to design (much less fully develop) a complete system.

A response to the constraints described in Section 2 above might be as follows:

(1) The proposed system might result in the addition of a large information inflow to the user, potentially leading to sensory saturation. This is likely to be the most serious problem, at least if auditory input is used. Achieving the right granularity of verbal information, advice, and warnings is likely to be a key issue.

(2) In attempting to provide a system that will make travel less stressful than using the long cane or the guide dog, the hurdles are not just those concerned with the technology. According to *The Institute for Innovative Blind Navigation* [2], "There is great scepticism in the blindness community, and well meaning inventors have not always been welcomed with open arms."

(3) Increasing miniaturisation will give developers the opportunity to package their systems attractively so that they are unobtrusive.

(4) A system based primarily on microelectronics rather than electromechanical systems will reap the benefits of ongoing miniaturisation, although micro electromechanical systems may bridge the gap. Potential investors may be won over by a strategy based on probable future improvements in the underlying technology. *The Institute for Innovative Blind Navigation* [2] uses four criteria for deciding whether to fund projects:
    - functionality ("What is the technology designed to do?"),
    - reliability ("Does the invention work in extremes of weather?"),
    - convenience ("Is the invention attractive or ugly?"),
    - service ("Is the technology affordable?").

6. Conclusions

The application of technology to providing mobility aids for the blind is likely to reach steadily higher levels of sophistication. The extent to which these aids are accepted by the blind community is likely to be determined by issues not directly associated with the technology itself. Indeed, a highly sophisticated level of technology will not guarantee that such devices will supplant traditional low-technology mobility aids. In the meantime there is a host of very difficult technical problems to be solved, encompassing image processing, pattern recognition, and signal processing.

References

[1] Dobelle, W.H., Artificial Vision for the Blind by Connecting a Television Camera to the Visual Cortex, *ASAIO Journal* 2000; 46:3-9.

[2] Institute for Innovative Blind Navigation: http://www.wayfinding.net/, visited on 2/4/2003.

[3] Jacobson H., "The Informational Capacity of the Human Eye", *Science*, Vol 113, 16 March, 1951, pp. 292-3.

[4] Jacobson H., "The Informational Capacity of the Human Ear", *Science*, Vol 112, 4 August, 1950, pp. 143-4.

[5] Ferguson S. and Ferguson S.D., "High Resolution Vision Prosthesis Systems: Research After 15 Years", *J. of Visual Impairment and Blindness*, Vol 80 (1), Jan 1986, pp 523-527.

[6] Sterling T.D., Bering E.A., Pollack S.V., Vaughan H.G. (eds.), "Visual Prosthesis: The Interdisciplinary Dialogue", *Proc 2nd Conference on Visual Prosthesis*, held 2-4 June, 1969, University of Chicago, Academic Press, NY, 1971.

[7] Bach-y-Rita P., "Brain Plasticity as a Basis of Sensory Substitution", *J.*

Neurologic Rehabilitation, Vol 1 (2), 1987, pp.67-71.

[8] Anani A.B., Ikeda K., Korner L.M., "Human ability to discriminate various parameters in afferent electrical nerve stimulation with particular reference to prostheses sensory feedback", *Medical and Biological Enginering and Computing*, Vol 15, July 1977, pp.363-372.

[9] Butikofer R., Lawrence P.D., "Electrocutaneous Nerve Stimulation - I: Model and Experiment", *IEEE Trans. Biomed. Eng*, BME-25 (6), Nov 1978, pp 526-531.

[10] Butikofer R., Lawrence P.D., "Electrocutaneous Nerve Stimulation - II: Stimulus Waveform Selection", *IEEE Trans. Biomed. Eng*, BME-26 (2), Feb 1979, pp 69-75.

[11] Szeto A.Y.J., Saunders F.A., "Electrocutaneous Stimulation for Sensory Communication in Rehabilitation Engineering", *IEEE Trans Biomed Eng*, Vol BME-29 (4), April 1982, pp 300-308.

[12] White B.W., Saunders F.A., Scadden L., Bach-y-Rita P., Collins C.C., "Seeing with the skin", *Perception and Psychophysics*, Vol 7(1), 1970, pp 23-27.

[13] Holmlund G.W. and Collins C.C., "An Electromagnetic Tactile Stimulator", *J Biomedical Systems*, Vol 1 (3), 1970, pp. 25-40.

[14] Meijer, P., “Vision Technology for the Totally Blind”, http://www.seeingwithsound.com/voice.htm, visited on 11/8/2003.

[15] Younger, T., "Road crossing aid for the blind and visually impaired”, Final Year Project Report, Massey University, 2000