AUDITORY DISPLAY OF COARSE OPTICAL IMAGERY: CONCEPT FOR A REHABILITATION AID FOR BLIND SPATIAL ORIENTATION

Aries Arditi

Visibility Metrics LLC,
49 Valley View Road, Chappaqua, NY, USA
arditi@visibilitymetrics.com

ABSTRACT

We introduce a concept for a rehabilitation aid for blind persons that will present, on a sonic display, coarse optical information obtained from a spectacle-mounted camera. The aid will serve blind persons who have no light sense or who can at most detect ambient light. The approach is to map luminous intensity to loudness of continuous tones of distinct timbre representing a small number of directions relative to that of the user's head.

1. INTRODUCTION

A long-pursued goal has been to develop a sensor-based device that can provide sufficient information for blind people to perform basic tasks that able-sighted people perform visually, including orienting and traveling in space independently. There have been many attempts (see [1]), but none has enjoyed wide adoption. We believe that in most cases, this is because they seek to provide the user either with a) a means for simple obstacle detection, a function that is largely redundant in that it is already well served by the long cane or the dog guide; or b) full object recognition capability—a far too ambitious goal in that the task of recognizing objects that are translated to other senses with significantly lower bandwidth, is simply too complex, laborious and cognitively demanding for blind users to engage in, especially during the already attentionally demanding task of navigating without vision.

2. CONCEPT

We describe here a new kind of device that we believe will be helpful to blind persons' spatial orientation. The device is intended to have more modest perceptual goals but be more practical, than those addressed by earlier attempts at vision substitution. Rather than translating input to a surrogate sense for the purpose of object identification, we do so to give blind users simple sonic cues to help identify their heading while walking and stationary, and sonic signatures for remembering gross features of the environment, including the rough location of lights and large high contrast objects. The aid will convey very low-resolution information to supplement already available auditory and tactile environmental cues. It will do so responsively, simply and naturally, and without imposing a heavy cognitive load on the user. Such information, we believe, can improve a blind person's sense of egocentric orientation (direction of the head in relation to environmental features), and therefore mobility.

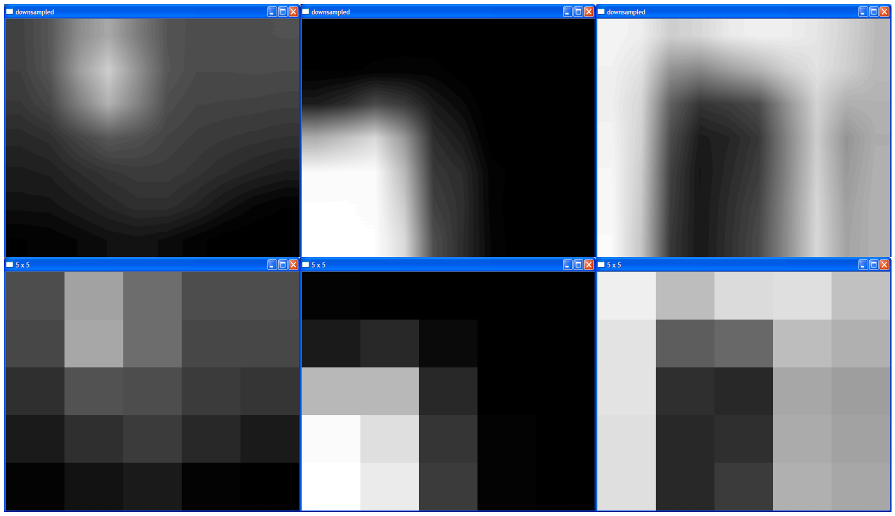
Figure 1: Low cost pinhole camera glasses. The camera is mounted on the nose bridge.

The device we envision will be inexpensive, requiring minimal hardware: a pair of low cost camera glasses such as those shown in Figure 1, a simple image and sound processor (possibly that in a smartphone), and a pair of bone-conduction headphones. This hardware will present to the user, a sonic display of information about light and dark in the environment, as captured by the camera.

We seek to provide only the most rudimentary capabilities to the functionally blind, i.e. those who either have no visual response to exogenous light whatsoever (known as those with "no light perception"; NLP), and those who can at most detect changes in ambient light level ("bare light perception"; BLP). In doing so, the system will give users with NLP the equivalent of BLP *plus* the ability to perceive approximate locations of lights and large high contrast objects around them. It will also provide improved BLP to users who already have BLP, in that it will allow perception of the lower luminances that even inexpensive camera sensors can detect, *and* it will improve response latencies to light increments and decrements, which appears to be diminished in many of those with BLP [2, 3].

We believe the system can be used for perceiving bright lights such as lamps and light fixtures and other environmental lighting. Since the camera can be steered with head motion, users should also be able to localize those lights, as well as large high contrast spatial edges. A user who can locate lights and edges in a room can then more easily learn, through ordinary

This work was supported by NIH grant R43EY023483 and FHWA cooperative agreement DTFH61-12-H-00002 with its content solely the responsibility of the author.



(a) (b) (c)

Figure 2: images taken with camera glasses, (top row) low-pass filtered to show crude image features and (bottom row) downsampled to show average gray values within 5 x 5 grid. Gray values like those in the bottom images will be mapped to tones through sonification (see text). (a) ceiling light fixture in an office; (b) corner of a computer monitor; (c) open doorway

exploration with touch and other aids such as a long cane, to locate important features of the room such as the door, in relation to those lights and edges.

3. EARLIER RELATED WORK

A well-studied system using similar hardware to that described here, called the vOICe ([4, 5]) displays down-sampled camera-based input as hopefully recognizable "soundscapes" of pure tones where the brightness of image pixels is coded as tone amplitude and the vertical locations coded by frequency. 64-pixel vertical "lines" of sound are swept from left to right in raster fashion, with each representing a horizontal location. Earlier related systems used musical tones [6] instead of pure tones, or frequency coding for both dimensions [7]. Studies with the vOICe have demonstrated that subjects without visual input can localize and identify some objects [8–10] under favorable conditions of contrast and lighting.

There are several limitations of the vOICe and similar sound-based vision substitution systems, however. First, learning to use them is very time-consuming even for simple localizations and discriminations [5,11,12]. Second, success may depend to a degree on musical ability [13]. Third, usability may require a very quiet environment due to the necessity of hearing subtleties in its complex soundscapes. Finally, such systems cannot display motion and dynamic imagery effectively because the imagery is displayed in auditory "snapshots" that sweep over the scene, each taking one second or more to render. This last drawback limits the responsiveness of such systems to alert users of obstacles and events in a timely fashion since the effective frame rate of the imagery is less than 1 Hz. The vOICe, however, remains promising for allowing object identification in static images, and it has successfully demonstrated the sonic display as an effective and inexpensive way to present spatial information to aurally intact blind persons.

Compared to auditory displays, presenting spatial information on a two-dimensional biological surface such as the skin, the tongue or the retina would seem to require less cognitive effort on the part of the user, since such displays are inherently spatial. Indeed various vision substitution attempts have portrayed down-sampled and processed imagery on the skin of the back [14], the finger [15], the tongue [16], and the retina itself [17], by way of arrays of vibrotactile, electrotactile or electrical stimulating elements. The only devices of this type that have as yet approached commercial viability have costs that range from $10,000, as in the case of BrainPort, a device that displays head-mounted camera images on the tongue [18], to more than $100,000 for a retinal prosthesis [17], which displays imagery in a retinal electrode array display.

4. APPROACH

Our approach, which has substantially different goals than that of the vOICe and the other existing tactile and retinal vision substitution devices, is partially illustrated in Figure 2. The images in the top row of the figure are low-pass filtered and intended to convey to the reader the gross environmental features that blind users would have access to through the device. The bottom row shows one simple implementation for how such information can be processed for subsequent display through sound – by downsampling to a 5 x 5 grid. With this extremely coarse spatial resolution (in this illustrative case only 25 locations), the sonic display can be temporally far more compact and likely more easily comprehended than one displaying a large number of spatial locations (e.g. 4,096 in the vOICe system). In the best case, for example, information from all spatial locations might be simultaneously displayed and perceived, perhaps by identifying each location with a sound of unique pitch (coding elevation), distinct timbre (coding azimuth), while luminous intensity will be coded by loudness/amplitude. Sound spatialization may also be employed to further reinforce perception of azimuthal location. Since blind people already depend on environmental sounds for cues about spatial layout, a device of this nature should minimize masking of those sounds by using bone conduction headphones. This concept is obviously in its very early stages, but we are now embarking on a program of prototype development and assessing usefulness to blind users in enhancing their spatial orientation.

5. REFERENCES

[1] U. R. Roentgen, G. J. Gelderblom, M. Soede, and L. P. de Witte, "Inventory of electronic mobility aids for persons with visual impairments: A literature review," *J. Vis. Impair. Blind.*, vol. 102, no. 11, pp. 702–724, 2008.

[2] M. Bach, M. Wilke, B. Wilhelm, E. Zrenner, and R. Wilke, "Basic Quantitative Assessment of Visual Performance in Patients with Very Low Vision," *Invest. Ophthalmol. Vis. Sci.*, vol. 51, no. 2, pp. 1255–1260, Feb. 2010.

[3] E. C. Hall and A. Arditi, "Temporal processing in very low vision," *Vis. Impair. Res.*, vol. 2, no. 1, pp. 43–48, 2000.

[4] P. Meijer, "Seeing with sound -- the vOICe." [Online]. Available: http://www.seeingwithsound.com.

[5] J. Ward and P. Meijer, "Visual experiences in the blind induced by an auditory sensory substitution device," *Conscious. Cogn.*, vol. 19, no. 1, pp. 492–500, Feb. 2010.

[6] J. Cronly-Dillon, K. Persaud, and R. Gregory, "The perception of visual images encoded in musical form: a study in cross-modality information transfer," *Proc. R. Soc. Lond. B Biol. Sci.*, vol. 266, no. 1436, pp. 2427–2433, 1999.

[7] C. Capelle, C. Trullemans, P. Arno, and C. Veraart, "A real-time experimental prototype for enhancement of vision rehabilitation using auditory substitution," *Biomed. Eng. IEEE Trans. On*, vol. 45, no. 10, pp. 1279–1293, 1998.

[8] M. Auvray, S. Hanneton, and J. K. O'Regan, "Learning to perceive with a visuo – auditory substitution system: Localisation and object recognition with 'The vOICe,'" *Perception*, vol. 36, no. 3, pp. 416 – 430, 2007.

[9] J.-K. Kim and R. J. Zatorre, "Generalized learning of visual-to-auditory substitution in sighted individuals," *Brain Res.*, vol. 1242, no. 0, pp. 263–275, Nov. 2008.

[10] M. J. Proulx, P. Stoerig, E. Ludowig, and I. Knoll, "Seeing 'Where' through the Ears: Effects of Learning-by-Doing and Long-Term Sensory Deprivation on Localization Based on Image-to-Sound Substitution," *PLoS ONE*, vol. 3, no. 3, p. e1840, Mar. 2008.

[11] R. W. Massof, "Auditory assistive devices for the blind," in *In ICAD*, 2003, pp. 271–275.

[12] E. Striem-Amit, M. Guendelman, and A. Amedi, "'Visual'acuity of the congenitally blind using visual-to-auditory sensory substitution," *PloS One*, vol. 7, no. 3, p. e33136, 2012.

[13] D. Brown, T. Macpherson, and J. Ward, "Seeing with sound? Exploring different characteristics of a visual-to-auditory sensory substitution device," *Perception*, vol. 40, no. 9, pp. 1120 – 1135, 2011.

[14] P. Bach-y-Rita, C. C. Collins, F. A. Saunders, B. White, and L. Scadden, "Vision substitution by tactile image projection," *Nature*, vol. 221, pp. 963–4, 1969.

[15] J. G. Linvill and J. C. Bliss, "A direct translation reading aid for the blind," *Proc. IEEE*, vol. 54, no. 1, pp. 40–51, 1966.

[16] P. Bach-y-Rita, K. A. Kaczmarek, M. E. Tyler, and J. Garcia-Lara, "Form perception with a 49-point electrotactile stimulus array on the tongue: a technical note," *J Rehabil Res Dev*, vol. 35, pp. 427–30, 1998.

[17] M. Humayun, J. Dorn, L. da Cruz, G. Dagnelie, J. Sahel, P. Stanga, A. Cideciyan, J. Duncan, D. Eliott, E. Filley, A. Ho, A. Santos, A. Safran, A. Arditi, L. Del Priore, R. Greenberg, and Argus Study Group, "Interim Results from the International Trial of Second Sight's Visual Prosthesis.," *Ophthalmology*, vol. 119, no. 4, pp. 779–788, 2012.

[18] Y. Danilov, M. Tyler, and others, "Brainport: an alternative input to the brain," *J. Integr. Neurosci.*, vol. 4, no. 4, pp. 537–550, 2005.