Neurocomputing 74 (2011) 1182–1190

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Toward local and global perception modules for vision substitution

Guido Bologna *, Benoît Deville, Juan Diego Gomez, Thierry Pun

Computer Science Department, University of Geneva, Route de Drize 7, 1227 Carouge, Switzerland

ARTICLE INFO

Available online 19 October 2010

Keywords:
Sensorial substitution
Colour sonification
Mobility experiments
Stereo-vision
Sound spatialisation

ABSTRACT

Although retinal neural implants have considerably progressed they raise a number of questions concerning user acceptance, risk rejection, and cost. For the time being we support a low cost approach based on the transmission of limited vision information by means of the auditory channel. The *See ColOr* mobility aid for visually impaired individuals transforms a small portion of a coloured video image into sound sources represented by spatialised musical instruments. Basically, the conversion of colours into sounds is achieved by quantisation of the HSL colour system. Our purpose is to provide blind people with a capability of perception of the environment in real time. In this work the novelty is the simultaneous sonification of colour and depth, the last parameter being coded by sound rhythm. The main drawback of our approach is that the sonification of a limited portion of a captured image involves limited perception. As a consequence, we propose to extend the local perception module by introducing a new global perception module aiming at providing the user with a clear picture of the entire scene characteristics. Finally, we present several experiments to illustrate the limited perception module, such as: (1) detecting an open door in order to go out from the office; (2) walking in a hallway and looking for a blue cabinet; (3) walking in a hallway and looking for a red tee shirt; (4) avoiding two red obstacles; (5) moving outside and avoiding a parked car. Videos of experiments are available on http://www.youtube.com/guidobologna.

1. Introduction

In recent years significant progress in miniaturisation of electronic devices has been accomplished. The implantation of electrodes in the human body is becoming more and more a reality. Regarding neural implants, a camera embedded in glasses collects visual information and sends it to a computer that converts the images to electrical signals, which are then transmitted to the implant and interpreted by the brain. In the Argus II project the implants allowed several blind individuals to recognise objects and obstacles [39]. Moreover, one person was able to read large prints.

One of the first hopes of retinal implants is that they will be used to aid blind people to navigate and orient themselves [40]. However, body-embedded devices raise a number of questions, such as long-term user acceptance, risk of rejection, and cost. We support a non-invasive and low cost approach based on the transmission of limited vision data, merely by the auditory channel. It is worth noting that with neural implants it is not still possible to elicit the sensation of colour and depth. Here, we specifically focus on these crucial parameters, by sonification.

Several authors proposed special devices for visual substitution by the auditory pathway in the context of real time navigation. The "K Sonar-Cane" combines a cane and a torch with ultrasounds [22]. Note that with this special cane, it is possible to perceive the environment by listening to a sound coding depth.

"TheVoice" is another experimental vision substitution system that uses auditory feedback. An image is represented by 64 columns of 64 pixels [28]. Every image is processed from left to right and each column is listened to for about 15 ms. In particular, every pixel gray level in a column is represented by a sinusoidal wave sound with a distinct frequency. High frequencies are at the top of the column and low frequencies are at the bottom.

Capelle et al. [13] proposed the implementation of a crude model of the primary visual system. The implemented device provides two resolution levels corresponding to an artificial central retina and an artificial peripheral retina, as in the real visual system. The auditory representation of an image is similar to that used in "TheVoice" with distinct sinusoidal waves for each pixel in a column and each column being presented sequentially to the listener.

Gonzalez-Mora et al. [18] developed a prototype using the spatialisation of sound in the three dimensional space. The sound is perceived as coming from somewhere in front of the user by means of head related transfer functions (HRTFs). The first device they achieved was capable of producing a virtual acoustic space of $17 \times 9 \times 8$ gray level pixels covering a distance of up to 4.5 m.

* Corresponding author.
E-mail addresses: guido.bologna@unige.ch (G. Bologna), benoit.deville@unige.ch (B. Deville), juan.gomez@unige.ch (J. Diego Gomez), thierry.pun@unige.ch (T. Pun).

0925-2312/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2010.07.022

This paper presents on-going work of the *See ColOr* (*See*ing *Col*our with an *O*rchestra) project. For the mobility of visually impaired individuals, we consider that colour perception is very important, as it can facilitate the recognition of potential landmarks. The See ColOr interface encodes a limited number of coloured pixels by spatialised musical instrument sounds, in order to represent and emphasize the colour and location of visual entities in their environment [7–11]. The basic idea is to represent a pixel as a directional sound source with depth estimated by stereo-vision. Finally, each emitted sound is assigned to a musical instrument, depending on the colour of the pixel. In previous works we introduced the sonification of colours by means of instrument sounds, as well as experiments related to image comprehension, recognition of coloured objects and mobility [10].

This work focuses on the simultaneous sonification of colour and depth. Generally, distance to objects is a crucial parameter for mobility, while colour allows a user to determine potential landmarks. In our experiments, depth is captured by a stereoscopic camera used by a well trained blindfolded individual. We perform five experiments, for which depth is a crucial parameter: (1) detecting an open door in order to go in and out; (2) walking in a hallway and looking for a blue cabinet; (3) walking in a hallway and looking for a red tee shirt; (4) avoiding two red obstacles and passing through; (5) moving outside and avoiding a parked car. The resultant videos are available on http://www.youtube.com/guidobologna.

Another contribution in this article is a new See ColOr module architecture. Specifically, as the current See ColOr mobility aid is based on the sonification of a limited portion of a captured image, we propose to extend the current local perception module by including a new module aiming at providing the user with the perception of the current image, as a whole. In the following sections, Section 2 depicts several systems helping blind people in navigation and exploration tasks, Section 3 presents examples of sensorial coding techniques of colour, Section 4 describes colour sonification in See ColOr, Section 5 summarises our past experiments, Section 6 explains the simultaneous sonification of colour and depth, Section 7 introduces a global perception module and an alerting system, and Section 8 illustrates several new experiments, followed by a discussion (Section 9) and a conclusion (Section 10).

2. Systems aiming at helping visually impaired individuals

A large number of tools have been created so far to help bind people in different tasks, such as the perception of texts [6], pictures or chart data [29,36,37]. Here, we limit our description to those devices related to navigation and exploration, as it is the main topic of this work. White canes and dogs are established aids for mobility. They are both used by visually impaired persons for near space perception tasks, such as orientation and obstacle avoidance. Nevertheless, a dog has a significant cost of more than 14 k$ and can assist mobility ten years, on average.

According to Hersh, mobility aids can be classified by the nature of the assistance provided (mobility or orientation), the space which will be explored and the complexity of the technology [19]. The main classes are:

- traditional low-tech aids for near space exploration;
- electronic travel aids of medium-tech for near space exploration;
- high-tech electronic orientation aids for large space exploration;
- mobility aids for near/far space navigation.

Two essential constituents of this classification are "exploration" and "navigation". Exploring a particular place means discovering the main components or looking for something in it. A navigation task for a visually impaired person involves making a decision on which course to follow between a starting point and a destination point. Note that obstacles should be avoided for both tasks.

An example of the first class is the white cane, while in the second we find among others, several variants of a cane having laser or ultrasound sensors that provide the user with distance measurements translated into tactile or auditory data. Examples include: *LaserCane* [5], *GuideCane* [12], *UltraCane* [20], and *Télétact* [17]. Other examples without encoded depth information, but with a camera capturing video images are the prototypes built by Bach-y-Rita et al. [3], which transmit the gray levels of an image to a tactile surface positioned on the skin. Similarly, BrainPort is a recent prototype, for which the tactile information is transmitted to the tongue [2]. Note however that only a camera is used; thus it is not possible to estimate the distance to obstacles.

High-tech electronic orientation aids for large space exploration assist visually impaired individuals in tasks such as self-localisation and space perception by verbal description. The Global Positioning System (GPS) is a fundamental ingredient of the prototype aids belonging to this class. However, the GPS provides an inaccurate localisation (about 10–15 m) and the signal is absent indoor and underground. The *Mobic* travel Aid, [30,31], the *Sextant System* [23], and Loomis' *Personal Guidance Systems* [24], are examples of this class of assistance. A more recent system is the *BrailleNote GPS* [32]. In the future it will be possible to obtain better localisation, as the forthcoming Galileo Global Positioning System will provide two levels of precision: 1 m for commercial applications and 5 m for public use.

For the last class of mobility aids, the *GUIDO Robotic SmartWalker* is a representative example that serves as support and navigation aid [34]. This prototype builds maps of the close environment with depth captured by a laser. The assistant robot can calculate the path from one point to another and can also avoid obstacles. Another example is represented by *Talking Signs* [25], which is an information system consisting of infrared transmitters conveying speech messages to small receivers carried by blind travelers. A user can get to the destination by walking in the direction from which the message is received. *Drishti* is an integrated indoor/outdoor blind navigation system [33]. Indoors, it uses a precise position measurement system, a wireless connection, a wearable computer and a vocal communication interface to guide blind users. Outdoors, it uses a differential GPS to keep the user as close as possible to the central line of sidewalks. Moreover, Drishti uses a Geographical Information Systems (GIS), in order to provide the user with a speech description of the close environment.

3. Sensorial colour coding

Colours of particular points of interest in images could be described to blind people by voice. However, the typical time duration for names of colours given by voice is approximately a second. Note that it would be difficult to remember the name of hundreds of different hues. Moreover, the sequential enumeration of the colour of several pixels would take too much time for a real-time system for mobility. This approach would become even worse with the addition of depth.

Recently, the research domain of colour sonification has started to grow [16,35,14]. A number of authors defined sound/colour mappings with respect to the HSL colour system. HSL (Hue, Saturation, Luminosity) is a symmetric double cone symmetrical to lightness and darkness. HSL mimics the painter way of thinking with the use of a painter tablet for adjusting the purity of colours. The *H* variable represents hue from red to purple (red, orange, yellow, green, cyan, blue, purple), the second one is saturation,

which represents the purity of the related colour and the third variable represents luminosity. The H, S, and L variables are defined between 0 and 1.

Doel defined colour/sound associations based on the HSL colour system [16]. In this sonification model, sound depends on the colour of the image at a particular location, as well as the speed of the pointer motion. Sound generation is achieved by subtractive synthesis. Specifically, the sound for grayscale colours is produced by filtering a white noise source with a low pass filter with a cutoff frequency that depends on the brightness. Colour is added by a second filter, which is parameterised by hue and saturation.

Rossi et al. [35] presented the "Col.diesis" project. Here the basic idea is to associate colours to a melody played by an instrument. For a given colour, darker colours are produced by lower pitch frequencies. Based on the statistics of more than 700 people, they produced a table, which summarises how individuals associate colours to musical instruments. It turned out that the mapping is: yellow for vibraphone or flute; green for flute; orange for banjo or marimba; purple for cello or organ; blue for piano, trumpet or clarinet; red for guitar or electric guitar.

Capalbo and Glenney [14] introduced the "KromoPhone". Their prototype can be used either in RGB mode or HSL mode. Using HSL, hue is sonified by sinusoidal sound pitch, saturation is associated to sound panning and luminosity is related to sound volume. The authors stated that only those individuals with perfect pitch perform well. In RGB mode the mapping of colours to sounds are defined by pan, pitch, and volume. For instance, the grayscale from black to white is panned to the centre, with black being associated to the lowest pitch sound. Blue and yellow are mapped to the left, with blue being associated with a pitch lower than yellow. Similarly, green and red are related to sounds heard at the right. Finally, the intensity of each colour is mapped to the volume of the sound it produces.

In one of their experiments, Capalbo and Glenney [14] illustrated that the use of colour information in a recognition task outperformed the performance of "TheVoice" (cf. Section 1). Specifically, the purpose was to pick certain fruits and vegetables known to correlate with certain colours. One of the results was that none of the three subjects trained with "TheVoice" could identify any of the fruit, either by the shape contours or luminance, while with the "KromoPhone" individuals obtained excellent results.

Meers and Ward [26,27] proposed the ENVS system which codes colours by means of electro-tactile pulses. Note also that they consider colour perception very important, as it can facilitate the recognition of significant objects that can serve as landmarks when navigating the environment. As stated by the authors, delivery of all colours by frequencies proved too complex for accurate interpretation. Consequently, in the ENVS prototype only an eight-entry lookup table was implemented for mapping eight colours.

It turns out that if we would like to use one of the sensorial colour coding described above, we would come across several limitations. Specifically, we would like to use a system that reacts in real-time; thus, the sonification used in the Col.diesis project would be unsuitable, since the sonification of a single colour last many seconds [35]. The KromoPhone colour sonification is very reactive; however, because of the pan colour coding, only a single pixel could be sonified. In other words, the spatial coding of more than a pixel would not be possible. As we would like to represent simultaneously a reasonable number of sonified pixels with also their corresponding spatial positions, we also dispose of this colour sonification scheme. A similar argument yields the same conclusion for Doel's system, which is also much more oriented toward the exploration of static pictures. The colours/electro-tactile mappings of the ENVS system present the advantage to represent ten pixels, simultaneously. However, in terms of quantity of information the tactile sensorial channel represents 0.1 kbps, while audition is about 10 kbps [38]. Thus, we prefer to represent colours by sounds transmitted to the auditory pathway. Finally, we would like also to represent colour and depth, simultaneously. With a tactile interface it is unclear to us on how to achieve an efficient coding in real-time.

4. Colour sonification in See ColOr

Relative to the HSL colour system, we represent the Hue variable by instrument timbre, because it is well accepted in the musical community that the colour of music lives in the timbre of performing instruments. Moreover, learning to associate instrument timbres to colours is easier than learning to associate for instance, pitch frequencies to colours. The saturation variable S representing the degree of purity of hue is rendered by sound pitch, while luminosity is represented by double bass when it is rather dark and a singing voice when it is relatively bright. From our practice, learning this sonification framework requires a number of training sessions spanning several weeks. Note however, that learning Braille is much more demanding that learning See ColOr sonification.

With respect to the hue variable, the corresponding musical instruments are based on an empirical choice:

1. oboe for red ($0 \le H < 1/12$);
2. viola for orange ($1/12 \le H < 1/6$);
3. pizzicato violin for yellow ($1/6 \le H < 1/3$);
4. flute for green ($1/3 \le H < 1/2$);
5. trumpet for cyan ($1/2 \le H < 2/3$);
6. piano for blue ($2/3 \le H < 5/6$);
7. saxophone for purple ($5/6 \le H \le 1$).

Note that for a sonified pixel, when the hue variable is exactly between two predefined hues, such as for instance between yellow and green, the resulting sound instrument mix is an equal proportion of the two corresponding instruments. More generally, hue values are rendered by two sound timbres whose gain depends on the proximity of the two closest hues.

The audio representation h_h of a hue pixel value h is

$$h_h = g h_a + (1-g) h_b \tag{1}$$

with g representing the gain defined by

$$g = \frac{h_b - H}{h_b - h_a} \tag{2}$$

with $h_a \le H \le h_b$ and h_a, h_b representing two successive hue values among red, orange, yellow, green, cyan, blue, and purple (the successor of purple is red). In this way, the transition between two successive hues is smooth.

The pitch of a selected instrument depends on the saturation value. We use four different saturation values by means of four different notes:

1. C for ($0 \le S < 0.25$);
2. G for ($0.25 \le S < 0.5$);
3. B flat for ($0.5 \le S < 0.75$);
4. E for ($0.75 \le S \le 1$).

When the luminance L is rather dark (i.e. less than 0.5) we mix the sound resulting from the H and S variables with a double bass using four possible notes (C, G, B flat, and E), depending on luminance level. A singing voice with also four different pitches (the same used for the double bass) is used with bright luminance (i.e. luminance above 0.5). Moreover, if luminance is close to zero,

the perceived colour is black and we discard in the final audio mix the musical instruments corresponding to the H and S variables. Similarly, if luminance is close to one, thus the perceived colour is white we only retain in the final mix a singing voice. Note that with luminance close to 0.5 the final mix has just the hue and saturation components.

The sonified part of a captured image is a row of 25 pixels relative to the central part of the video image. We take into account a single row, as the encoding of several rows would need the use of 3D spatialisation, instead of simple 2D spatialisation. It is well known that rendering elevation is much more complicated than lateralisation [4]. On the other hand, in case of 3D spatialisation it is very likely that too many sound sources would be difficult to be analysed by a common user. Two-dimensional spatialisation is achieved by the convolution of monoaural instrument sounds with filters encompassing typical lateral cues, such as interaural time delay and interaural intensity difference [4]. In this work we use filters available in the public CIPIC database [1].

5. Previous experiments without depth sonification

In the first step of the See ColOr project, we performed several experiments with six blindfolded persons who were trained to associate colours with musical instrument sounds [8]. As shown in Fig. 1, the participants were asked to identify major components of static pictures presented on a special paper lying on a T3 tactile tablet (http://www.rncb.ac.uk/page.php?id=872) representing pictures with embossed edges. Specifically, this tablet makes it possible to determine the coordinates of a contact point. When one touched the paper lying on the tablet, a small region below the finger was sonified and provided to the user. Colour was helpful for the interpretation of image scenes, as it lessened ambiguity. As an example, if a large region "sounded" cyan at the top of the picture it was likely to be the sky. Finally, all participants to the experiments were successful when asked to find a bright red door in a picture representing a churchyard with trees, grass and a house.

The work described in [9] introduced an experiment during which ten blindfolded individuals participants tried to match pairs of uniform coloured socks by pointing a head mounted camera and by listening to the generated sounds. Fig. 2 illustrates an experiment participant observing a blue socket. The results of this experiment

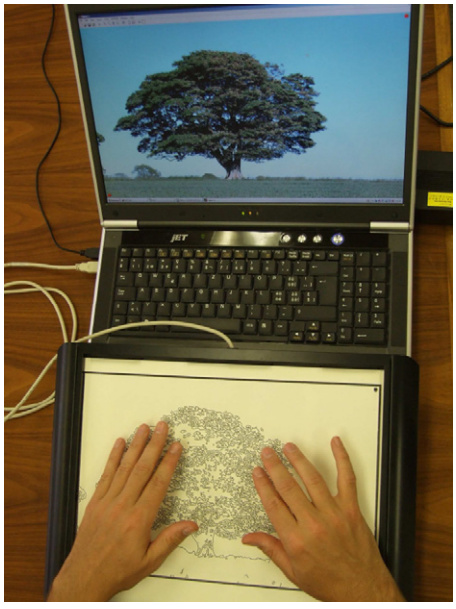


Fig. 1. Example of an embossed picture on the T3 tactile tablet.

Fig. 2. A blindfolded subject observing a blue socket. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 3. A blindfolded individual following a red sinuous path. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

demonstrated that matching similar colours through the use of a perceptual (auditory) language, such as that represented by instrument sounds can be successfully accomplished.

In [11] the purpose was to validate the hypothesis that navigation in an outdoor environment can be performed by "listening" to a coloured path. As shown by Fig. 3, we introduced an experiment during which ten blindfolded participants and a blind person were asked to point the camera toward a red sinuous path painted on the ground and to follow it for more than 80 m. Results demonstrated that following a sinuous coloured path through the use of the auditory perceptual language was successful for both blind and blindfold participants. A video entitled "The See ColOr project" illustrates several experiments on http://www.youtube.com/guidobologna.

6. Depth sonification

We use two stereoscopic cameras with colour. The first is denoted as the STH-MDCS2 (SRI International: http://www.videredesign.com/) and the second is the "Bumblebee" (Point Grey: http://www.ptgrey.com/). An algorithm for depth calculation based on epipolar geometry is embedded within both the

stereoscopic cameras. The resolution of images is 320×240 pixels with a maximum frame rate of 30 images per second.

Our See ColOr prototype presents two sonification modes that render colour and depth. The first replicates a crude model of the human visual system. Pixels near the centre of the sonified row have high resolution, while pixels close to the left and right borders have low resolution. This is achieved by considering a sonification mask indicating the number of pixel values to skip. As shown below, starting from the middle point (in bold), the following vector of 25 points represents the number of skipped pixels:

[15 12 9 7 5 3 3 2 2 1 1 1 **1** 1 1 1 2 2 3 3 5 7 9 12 15]

In the first mode, depth is represented by sound duration. The mapping for depth D is given by:

- 90 ms for undetermined depth;
- 160 ms for $(0 \leq D < 1)$;
- 207 ms for $(1 \leq D < 2)$;
- 254 ms for $(2 \leq D < 3)$;
- 300 ms for $D > 3$

Note that it is possible to have points of undetermined depth, especially in homogeneous areas like walls, for which the depth algorithm is unable to determine landmark points related to the calculation of the disparity between the left and right images.

The second mode sonifies only a pixel of a particular area of 25 adjacent points in the middle of the image. Specifically, we first determine among these 25 points the greatest number of contiguous points labelled with the same hue. Then, we calculate the centroid of this area and the average depth. Points with undetermined depth are not considered in the average depth calculation. The final sonification presents only a spatialised sound source representing the average colour and the average depth.

In the second mode, depth between 1 and 4 m is sonified by sound duration (the same sonification scheme explained above), while after 4 m the volume V starts to decrease by following a negative exponential function given by

$$f(V) = V \exp(-kD) \tag{3}$$

with k a positive small constant.

7. Toward a multi-module prototype

An important drawback in see ColOr is the perception restricted to a small portion of a video image. In order to go one step further, we propose to implement a multi-module architecture including a module that will allow the user to perceive the current image scene, as a whole. In other words, the user will be able to explore the main components of an image and to determine their relative positions (if depth is well defined).

As shown by Fig. 4, in the future we will have a prototype with specialised modules:

- the local perception module;
- the global perception module;
- the alerting system.

The global perception module and the alerting system are under development.

7.1. The global perception module

Since the local module provides the user with the auditory representation of a row containing 25 points of a captured image, a main issue is the tunnelling vision phenomenon, which reduces the user perception only to a small portion of the captured image. The global module will rectify this deficiency by providing a perception of the nearby space that provides the end user with global perception. This will be achieved by means of a small touchpad, such that the new prototype will allow an almost complete perception of the environment. The key idea behind this module is related to the work carried out with use of the T3 tactile tablet (cf. Section 5). Basically, the purpose of several experiment participants was to perceive the main components of images by exploring with their fingers the image surface. The results were very encouraging with simple images [8], while the majority of pictures representing perspective views of landscapes resulted very difficult to understand; the reason most likely being that pixel depth was not sonified at all. With the simultaneous sonification of colour and depth it is plausible to expect that users could develop consistent 3D maps of their environment.

When the user will want to explore the nearby space, he/she will rapidly scan the touchpad with one or more fingers; the principle here is that finger movements replace eye movements. In this way the user will not have to move his/her head. The finger contact point will activate a sound coding the corresponding data representation. Using a multi-touch device, the maximal number of sonified points will be equal to 10, corresponding to the number of fingers. The spatial position of each finger on the touchpad will be spatialised on the azimuth plane but not on elevation. We assume that the user will be aware of the vertical position of the touched points. The future prototype will be based on cheap components like webcams and portable touchpads or smartphones. Non-invasive and low-cost elements will be likely to be adopted by the visually impaired community.

Another functionality of the global module will be the sonification of a depth map. As shown by Fig. 5, this corresponds to the top view of the current scene with depth represented vertically. This map will not represent colour, therefore sonification will be

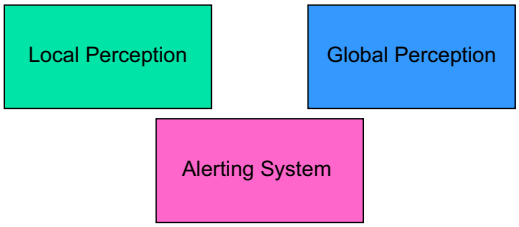


Fig. 4. Decomposition of the See ColOr mobility aid into three distinct modules.

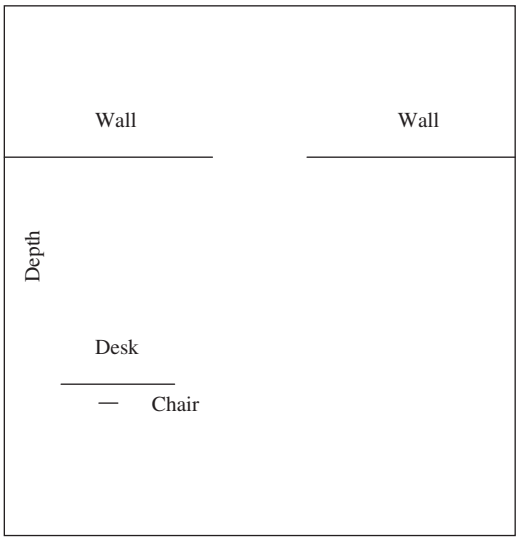


Fig. 5. An example of a top view map. Depth is represented on the vertical axis.

achieved with a particular sound different from classical instrument timbres. From time to time, the user will inspect this diagram represented on a touchpad, in order to determine the location of obstacles. Typically, fingers contact points will play the role of the eye and it will trigger a particular sound, depending on the presence/absence of an obstacle at this point coordinate.

The global module will also include another sub-module for the sonification of a compass. Indeed, it is well known that blind individuals tend to loose their orientation in large indoor environments, such as shopping centres, buildings, etc. This sub-module would be useful in situations for which natural colours' landmarks would be symmetric, and thus ambiguous. Since a compass is useful for navigation, we will provide a visually impaired person with such a tool. For instance, when the user will want to determine the North direction, a signal beacon will start to emit a 2D spatialised sound. Because of frequent front/back perception mistakes, characteristic sounds will disambiguate front/back positions.

7.2. The alerting system

The purpose of the alerting system is to warn the user when an obstacle is potentially in his/her trajectory. Roughly, when the number of points presenting a distance below 1 m is increasing over a given number of frames the user should stop walking. Note also that the alerting system will run simultaneously with respect to the local perception module or the global perception module. The sonification of an obstacle incidence will be performed by a voice warning.

We are developing this module based on a specific model of bottom-up saliency-based visual attention, namely the conspicuity maps [21]. A conspicuity map contains information about regions of an image that differ from their neighbourhood. Each conspicuity map is built according to a specific feature, which can consist of colours, orientations, edges, etc. We have combined the depth gradient feature with distance, illumination intensity, and two colour opponencies [15]. The purpose of the depth gradient conspicuity map is to detect objects that come towards the blind user, and that need to be avoided. We showed that the use of the depth gradient is an important feature in a mobility aid system [15]. It obviously helps in the cases where objects might disturb the movements of a blind user.

As depth is unreliable in a number of situations, the information provided by this module will be to some extent inaccurate. Experiments will be carried out in the future, in order to assess what will be the failure rate in a number of different situations with varying obstacle size and luminosity. Note also that with the advent of small cheap ultrasonic sensors or infrared sensors it will be worth to combine the two approaches. For instance, an ultrasonic sensor measuring a few centimetres and costing about 30$ could be worn on a belt and could be used outside to measure distance to obstacles.

Natural sounds are also important for blind people to determine landmarks (e.g. cars, birds, fountains, etc.). Therefore, sometimes the local/global perception modules should be switched off. At this point the alerting systems could still be active, in order to warn the user in case of threat situations.

8. Experiments with the local perception module

Although the local perception module limits the sonification to a small portion of a video image, five tests involving mobility or search were done. The experiments were performed by a very well trained blindfolded individual, who is very familiar with this colour sonification model, but not with depth sonification. Although in the long term we will aim at complementing the white cane of blind people by a miniaturised version of our prototype, this person relied only on the See ColOr interface. The reason is that we wanted to be sure that our prototype represented the only sensing tool. All the videos described here are available on http://www.youtube.com/guidobologna.

In the first video entitled "Going out from the office" and in the second video entitled "Going into the office" we aim at demonstrating that it is possible to perceive an open door and to pass through it. Fig. 6 illustrates a picture of this experiment, which is performed with the second sonification mode. The brown door is sonified by a viola and the rhythm is fast when the user is close to it. Note also that the user decided to move when slow sound rhythms or low volume sounds were discerned, indicating distant obstacles.

The third video entitled "Find a red tee shirt with sounds of musical instruments" illustrates the same individual in a successful search task (see Fig. 7). It is worth noting that here depth is often undetermined when the camera is pointed toward the floor or the white walls. Note also that the user trusted the depth information related to the trumpet sound representing the blue–cyan cabinets. The red tee short is sonified by oboe and when the user was close to it the rhythm frequency increased.

As shown by Fig. 8, in the video entitled "Avoiding two red obstacles with sounds of musical instruments" the blindfolded individual purpose is to detect two persons wearing a red tee shirt and then to pass through the open space. In this experiment the

Fig. 6. A blindfolded individual looking for an open door.

Fig. 7. A blindfolded individual looking for a red tee shirt. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 8. A blindfolded individual looking for two red obstacles represented by two individuals. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 9. A blindfolded individual looking for a blue cabinet. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

user selected the second sonification mode. At the beginning a low volume oboe sound represented one of the two tee shirts at a distance of about 6 m. The blindfolded individual then decided to go toward this colour. Note that as soon as the user was close to the red tee shirts the sound rhythm was fast and also the volume was high. Since no obstacles lied between the two red persons the volume was equal to zero, when the camera was pointed toward this direction. The blindfolded individual realised that he went too close to one of the "obstacles" and corrected himself by going backwards. Finally, our user perceived that he was well positioned to go between the two red "obstacles" and decided to go through.

In the video entitled "The blue cabinet" the user switched to the first sonification mode (with all 25 points sonified by colour and depth). Here the goal was to find a blue cabinet sonified by a piano playing a medium pitched tone. Fig. 9 illustrates the cabinet scene. This mode is more complex than the previous one, since more than a colour can be present in the current sonified frame. Here the distance to the floor is defined, since the floor is textured. Note also that the brown doors are sonified by viola sounds. From time to time, our experiment participant wished to ask to the computer the depth of the middle point of the sonified row. With the use of a mouse button the computer answered

Fig. 10. A blindfolded individual walking outside and avoiding parked cars.

with a voice saying numbers in French. "One" means distance between 0 and 1 m; "two" means distance between 1 and 2 m, etc. At the end of the video the user reached and indicated the cabinet.

In the last video entitled "Walking outside" the user walked outside. Fig. 10 illustrates the experiment participant in front of a car. He used again the second mode with only a sonified sound. The sound of the ground is rendered by a singing voice or a double bass, depending on its gray level. Suddenly, the user found in his trajectory a parked car and he avoided it.

9. Discussion

After the experiments the blindfolded person was asked to give his impressions of the two different sonification modes. The first impression is that the second mode (with the decreasing volume) felt "relaxing" as compared to the first mode. The second mode is valuable in large areas (for instance, outside). Moreover, in some situations, it will be very useful to switch from the second mode to the first, as the first mode gives more precision and to some extent, peripheral view. A sonified compass could be also very useful, as it is very easy to lose orientation. Finally, while the first mode provides to some extent limited global information, a "global module" would be helpful in order to get a clear picture of the close environment geometry.

Experiments regarding "open door detection" were also performed by Meers and Ward [26]. Specifically, depth was coded by the amount of electrical stimulation felt by fingers, which is directly proportional to the distance of objects in the direction pointed by each finger. After training, blindfolded individuals were able to avoid obstacles and also to identify features like an open doorway. Therefore, we obtained similar results by means of the auditory pathway. The authors stated that 10 range readings delivered to the fingers may not seem like much environmental information; however the integration of this information over time contributes to produce a consistent 3D picture.

Recently, the *Eye-Project* was funded in Australia (42 M$). Patients symptomatic of degenerative vision loss such as retinitis pigmentosa and age-related macular degeneration should benefit from this approach. In 2011 the purpose will be to experiment retinal implants with 100 electrodes, while in 2013, 1000 electrodes will be inserted into the eyes of several individuals. Authors of this project hope that the first prototype will allow blind individuals to detect large obstacles and move more independently. The second advanced prototype in 2013 aims at recognizing letters and at distinguishing human faces and expressions.

At least, the See ColOr mobility aid with its local perception module allows a trained individual to detect large obstacles and to move independently in the absence of small obstacles. Thus, this is very encouraging since the See ColOr approach involves low-cost components. With the addition of the global module, as well as the alerting system we expect trained individuals to move independently in unknown environments. Moreover, we know for sure that the See ColOr interface will not allow trained individuals to read or to discern human expressions. However, specialised modules could be developed for these challenging tasks. One question that arises regarding retinal electrodes is whether an individual with only one or two implanted electrodes could be able to determine depth and colour, the first parameter being absolutely crucial for mobility. Finally, even if neural implant will present in a few years a resolution of 1000 points in the fovea region, this will represent a local perception vision aid. Since the mechanism of peripheral vision is also very important for mobility, we really wonder how this question will be solved by retinal implants. More financial efforts should be engaged in non-invasive mobility interfaces for vision substitution.

10. Conclusion

We presented the colour and depth sonification model of the See ColOr mobility aid. A See ColOr prototype was tested by a well trained individual. He successfully (1) detected an open door in order to go in and out; (2) walked in a corridor with the purpose to find a blue cabinet; (3) moved in a hallway with the purpose to locate a red tee shirt; (4) detected two red obstacles and passed through them; (5) walked outside and avoided a parked car. In the future, we would like to measure in a more systematic way whether the use of our prototype allows users to locate objects and to avoid obstacles of different sizes. Thus, we will perform experiments with more participants (including blind individuals), in order to obtain more robust statistics. Finally, our multi-module prototype for visual substitution puts forward several novel aspects. Particularly, by proposing an architecture presenting a local perception module and a global perception module, See ColOr will imitate the visual system by providing the user with essential cues of vision.

References

[1] V.R. Algazi, R.O. Duda, D.P. Thompson, C. Avendano, The CIPIC HRTF Database, in: Proceedings of the WASPAA'01, New Paltz, NY, 2001.

[2] A. Arnoldussen, C. Nemke, R. Hogle, K. Skinner, BrainPort plasticity: balance and vision applications, in: Proceedings of the Ninth International Conference on Low Vision, 2008.

[3] P. Bach-y-Rita, K. Kaczmarek, M. Tyler, J. Garcia-Lara, Form perception with a 49-point electrotactile stimulus array on the tongue, J. Rehab. Res. Dev. *35* (1998) 427–431.

[4] R. Begault, 3-D Sound for Virtual Reality and Multimedia, Boston A.P. Professional, 1994, ISBN: 0120847353.

[5] M.J. Benjamin, N.A. Ali, A.F. Schepis, A laser cane for blinds, in: Proceedings of the San Diego Biomedical Symposium, vol. 12, 1973, pp. 53–57.

[6] J.C. Bliss, Reading machines for the blind, in: G. Gordon (Ed.), Active Touch: The Mechanism of Recognition of Objects by Manipulation, A Multidisciplinary Approach, Pergamon Press, Oxford, UK, 1978.

[7] G. Bologna, M. Vinckenbosch, Eye tracking in coloured image scenes represented by ambisonic fields of musical instrument sounds, in: Proceedings of the International Work-Conference on the Interplay between Natural and Artificial Computation (IWINAC), (1), Las Palmas, Spain, June 2005, pp. 327–333.

[8] G. Bologna, B. Deville, T. Pun, M. Vinckenbosch, Transforming 3D coloured pixels into musical instrument notes for vision substitution applications, in: A. Caplier, T. Pun, D. Tzovaras (Guest Eds.), Eurasip Journal of Image and Video Processing, 2007, Article ID 76204, 14 pp. (Open access article).

[9] G. Bologna, B. Deville, M. Vinckenbosch, T. Pun, A perceptual interface for vision substitution in a color matching experiment, in: Proceedings of the International Joint Conference on Neural Networks, Part of IEEE World Congress on Computational Intelligence, Hong Kong, June 1–6, 2008.

[10] G. Bologna, B. Deville, T. Pun, On the use of the auditory pathway to represent image scenes in real-time, Neurocomputing 72 (2009) 839–849 Elec tronic publication: 2008, Article DOI: 〈http://dx.doi.org/10.1016/j.neucom.2008.06.020〉.

[11] G. Bologna, B. Deville, T. Pun, Blind navigation along a sinuous path by means of the See ColOr interface, in: J. Mira, J.M. Ferrández, J.R. Álvarez, F. Paz, F.J. Toledo (Eds.), Proceedings of the Third International Work-Conference on the Interplay between Natural and Artificial Computation: Part II: Bioinspired Applications in Artificial and Natural Computation (Santiago de Compostela, Spain, June 22–26, 2009), Lecture Notes In Computer Science, vol. 5602, Springer-Verlag, Berlin, Heidelberg.

[12] J. Borenstein, I. Ulrich, The guide cane, a computerized travel aid for the active guidance of blind pedestrians, in: Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Albuquerque, April 21–27 1997, pp. 1283–1288.

[13] C. Capelle, C. Trullemans, P. Arno, C. Veraart, A real time experimental prototype for enhancement of vision rehabilitation using auditory substitution, IEEE T. Bio-Med. Eng. 45 (1998) 1279–1293.

[14] Z. Capalbo, B. Glenney, Hearing color: radical plurastic realism and SSDs, in: Proceedings of the Fifth Asia-Pacific Computing and Philosophy Conference (AP-CAP 2009), Tokyo, Japan, October 1–2 2009.

[15] B. Deville, G. Bologna, M. Vinckenbosch, T. Pun, See Color: seeing colours with an orchestra, in: D. Lalanne, J. Kohlas (Eds.), Lecture Notes in Computer Science, vol. 5440, Springer-Verlag, Berlin, Heidelberg2009, pp. 251–279.

[16] K. Doel, Soundview: sensing color images by kinesthetic audio, in: Proceedings of the International Conference on Auditory Display, Boston, MA, USA, July 6–9, 2003.

[17] R. Farcy, Une aide électronique miniature pour les déplacements des déficients visuels en intérieur: Le "Tom Pouce Light, in: Proceedings of the Handicap 2008, Paris (in French).

[18] J.L. Gonzalez-Mora, A. Rodriguez-Hernandez, L.F. Rodriguez-Ramos, L. Dfaz-Saco, N. Sosa, Development of a new space perception system for blind people, based on the creation of a virtual acoustic space, in: Proceedings of the International Work Conference on Artificial and Natural Networks (IWANN), Alicante, Spain, June 1999, pp. 321–330.

[19] M. Hersh, M.A. Johnson, Assistive technology for visually impaired and blind people, Spinger, ISBN 9781846288661, 2008.

[20] B.S. Hoyle, S. Dodds, The UltraCane mobility aid at work—from training programmes to in-depth use case studies, in: Proceedings of the Conference on Visual and Hearing Impairments (CVHI), Granada, 19–21 July 2006.

[21] L. Itti, C. Koch, E. Niebur, A model of saliency-based visual attention for rapid scene analysis, IEEE Trans. Pattern Anal. Mach. Intell. 20 (1998) 1254–1259.

[22] L. Kay, A sonar aid to enhance spatial perception of the blind: engineering design and evaluation, Radio Electron. Engr. 44 (1974) 605–627.

[23] C.M. La Pierre, Navigation System for the Visually Impaired, Carleton University, Canada, 1993.

[24] J.M. Loomis, J.R. Marston, R.G. Golledge, R.L. Klatzky, Personal guidance system for visually impaired people: comparison of spatial displays for route guidance, J. Vis. Impairment Blindness 99 (4) (2005) 219–232.

[25] J.R. Marston, Towards an accessible city: empirical measurement and modeling of access to urban opportunities for those with vision impairments, using remote infrared audible signage, Ph.D. Thesis, Department of Geography, University of California Santa Barbara, 2002.

[26] S. Meers, K. Ward, A vision system for providing 3D perception of the environment via transcutaneous electro-neural stimulation, in: Proceedings of the Information Visualisation, Eighth International Conference, July 14–16, 2004, IV, IEEE Computer Society, Washington, DC, pp. 546–552.

[27] S. Meers, K. Ward, A vision system for providing the blind with 3d colour perception of the environment, in: Proceedings of the Asia-Pacific Workshop on Visual Information Processing, Hong Kong, December 2005, pp. 102–108.

[28] P.B.L. Meijer, An experimental system for auditory image representations, IEEE Trans. Bio. Eng. 39 (2) (1992) 112–121.

[29] J. Pasquero, V. Hayward, STRESS: a practical tactile display systems with one millimeter spatial resolution and 7000 Hz refresh rate, in: Proceedings of the Eurohaptics, Dublin, Ireland, 2003.

[30] H. Petrie User requirements for a GPS-based travel aid for blind people, in: Proceedings of the Conference on Orientation and Navigation Systems for Blind Persons, 1–2 February 1995, RNIB, UK.

[31] H. Petrie, V. Johnson, T. Strothotte, A. Raab, S. Fritz, R. Michel, MoBIC: designing a travel aid for blind and elderly people, J. Navigat. 49 (1996) 45–52.

[32] P.E. Ponchilla, E.C. Rak, A.L. Freeland, S.J. LaGrow, Accessible GPS: reorientation and target location among users with visual impairments, J. Visual Impairment Blindness 101 (7) (2007) 389–401.

[33] L., Ran, S., Helal, S., Moore Drishti: an integrated indoor/outdoor blind navigation system and service, in: Proceedings of the Second IEEE International Conference on Pervasive Computing and Communications (PerCom'04), 2004, p. 23.

[34] D. Rodriguez-Losada, F. Matia, R. Galan, Building geometric feature based maps for indoor service robots, J. Robot. Autonomous Systems 54 (7) (2006) 546–558.

[35] J. Rossi, F.J. Perales, J. Varona, M. Roca, Col.diesis: transforming colour into melody and implementing the result in a colour sensor device, in: Proceedings of the Second International Conference in Visualisation, Barcelona, Spain, July 15–17, 2009.

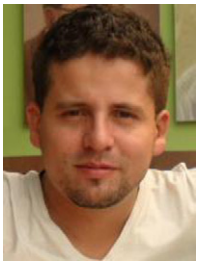
[36] R. Velázquez, E. Pissaloux, M. Hafez, J. Szewczyk, Tactile rendering with shape memory alloy pin-matrix, IEEE Trans. Instrumentat. Meas. 57 (5) (2008) 1051–1057.

[37] C.R. Wagner, S.J. Lederman, R.D. Howe, A tactile shape display using RC servomotors, in: Proceedings of the 10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, Orlando, USA, 2002.

[38] T.P. Way, K.E. Barner, Automatic visual to tactile translation, part I: human factors, access methods and image manipulation, IEEE Trans. Rehabil. Eng. 5 (1997) 81–94.
[39] D.D. Zhou, R.J. Greenberg, Microelectronic visual prostheses, in: D. Zhou, E. Greenbausm (Eds.), Implantable Neural Prostheses 1: Devices and Applications, Springer, ISBN 978-0-387-77260-8, 2009.
[40] E. Zrenner, Will retinal implants restore vision ? Science 295 (2002) 1022–1025.

Guido Bologna is a senior researcher at the University of Geneva, Switzerland. He received his Ph.D. in artificial intelligence in 1998 at the University of Geneva. Subsequently, he worked as a postdoc at the Queensland University of Technology, Brisbane; at the National University of Singapore and at the Swiss Institute of Bioinformatics. He has authored or co-authored more than 50 full papers in refereed journals, books and conferences. His current research interests, concern: multimodal interfaces for blind users, machine learning, and bioinformatics.

Benoît Deville is a Ph.D. student in computer science at the University of Geneva, Switzerland. He received two M.Sc. degrees from the University of Strasbourg, France, in 2003, and from the University of Reims, France, in 2005. His research interests concern: multimodal interfaces for blind users, saliency detection, image simplification, and video processing.



Juan Diego Gomez is a Ph.D. student in Computer Science at the University of Geneva, Switzerland. He has received two M.Sc. degrees, in Computer Vision/Artificial Intelligence from the Autonomus University of Barcelona, and in Computer Science from University 'Rey Juan Carlos' of Madrid, Spain. His research interests concern: multimodal interfaces for blind users, machine learning, and medical imaging.

Professor Thierry Pun is head of the Computer Vision and Multimedia Laboratory, Computer Science Department, University of Geneva. He received his Ph.D. in image processing for the development of a visual prosthesis for the blind in 1982, at the Swiss Federal Institute of Technology, Lausanne, Switzerland. After working at the National Institutes of Health, Bethesda, USA, he joined the University of Geneva, Switzerland in 1986, where he is a full professor at the Computer Science Department. He has authored or co-authored about 300 full papers as well as eight patents. His current research interests, related to multimodal interaction and multimedia information systems, concern: multimodal interfaces for blind users, physiological signals analysis for emotion assessment and Brain–Computer Interaction, data hiding, and information retrieval systems.