

A computer vision-based sensory substitution device:  
See ColOr



Dr Juan Diego Gomez  
*Université de Genève*



**Abstract:** See ColOr is a Sensory Substitution Device (SSD) for the blind that converts color and spatial location of image points into 3D instrument sounds. Additionally, unlike related works, See ColOr integrates computer vision methods to produce reliable knowledge about the physical world. Importantly, See ColOr is made of affordable technology relative to retinal implants. Hence, it is accessible for developing countries and low-income communities, where blindness indexes continue to grow. Experiments conducted in South America using See ColOr confirm that our SSD is learnable, functional, and provides easy interaction. In moderate time, blind individuals can attain as much visual information as to partially restore: spatial awareness, ability to find someone, location of daily objects, and skill to walk safely avoiding obstacles. Our encouraging proposal opens a door towards autonomous mobility of the blind.

#### **Introduction:**

The World Health Organization estimates the world blind population at 39 million persons, which roughly corresponds to 0.56% of the total world population. More precisely this represents an incidence ranging from 0.5% to 1.4% in the developing countries and of 0.3% in the whole of the industrialized countries [1]. As for low-income countries, in nearly 90% of the cases a blind individual can no longer work and his/her life expectancy drops down to 1/3 that of a matched peer, in age and health. Back to the global picture, blindness is likely to double in the next 15 years because of ageing. In the world, people aged over 60 account for 58% of blind. In Switzerland for instance, 10000 people are affected by blindness and 80% of them are more than 60 years old. Fortunately, on a worldwide scale approximately 50% of blindness could be prevented. Nonetheless, without effective and major intervention, the projected increase in global blindness to 76 million by 2020 will be regrettably reached. [1].

At large, blind individuals are adept at traveling with help of traditional mobility aids (i.e. white canes and guide dogs). The limitations of these tools, however, become apparent in numerous daily life situations, creating a strong urge to seek aid from others. For example, exploration of new environments is particularly demanding, and handling unexpected needs or noticing serendipitous discoveries remain challenging. Moreover, there are also retinal implants intended to restore some functional vision lost after damage of the photoreceptors. Unfortunately, clinical trials reveal that these neuroprosthesis still suffer from very limited resolution (i.e.  $6 \times 10$  electrodes). Therefore, basic visual tasks remain challenging or impossible such as, objects recognition, navigation in unknown environments, or people identification. Let alone clinical risks of invasive treatments and let elevated prices. This scenario has given rise to a proliferation of SSDs, which are made up of an

optical sensor (camera) connected to a processing device (computer) that systematically converts visual features into tactile or auditory responses.




More specifically, sensory substitution refers to the mapping of stimuli of one sensory modality into another. This is usually done with the aim of bypassing a defective sense, so that associated stimuli may flow through a functioning sense [3]. In general, sensory substitution argues that when individuals go blind or deaf, they do not actually lose the ability of seeing or hearing, rather they become incapable to convey external stimuli to the brain. When the working of the brain is not affected, in most of the cases a person who lost the ability to retrieve data from their eyes could still create subjective images by using data conveyed from other sensory modalities (e.g. auditory pathway) [3]. This idea is rooted in multisensory perception theories holding that sensory inputs in our brain are never processed separately [2]. As a consequence, for example, what we see, somehow and somewhat, is always influenced by what we hear, and vice versa. In fact, this is the idea central to neurological behavior that neuroscientists have termed cross-modal transfer [4]. They state that due to sensorial interconnection in our brain, visual-like experiences might be elicited through senses others than vision [4].

### **Our hypothesis:**

Besides converting color and spatial location into sounds, See ColOrs is, at the best of our knowledge, the first aid system that integrates computer vision into Sensory Substitution. We do so because vision is a phenomenon that entails both, sensation and perception [5]. Sensation is the low-level -biochemical and neurological- feeling of external visual information as it is registered (sensed) by the eyes. The visual sensation alone does not imply the coherent conception (or understanding) of external visual objects [5], [6]. Therefore, following a sensation, perception appears as the mental process that decodes the sensory input (sensation) to create awareness or understanding of the real-world [5], [6]. In short, we perceive the world through sensations, though we derive sense from it (vision comes into being) only when perception takes place [5]. In this work, we argue that current SSDs have been intended to provide a substitute to sensation, while the perceptual experience has been left mostly unattended. The underlying problem is that the human visual system is known to be capable of  $4.3 \times 10^6$  bits per second (bps) bandwidth [7]. Yet, senses intended as substitutes can hardly reach  $10^4$  bps at most (i.e. hearing) [7]. In this light, even though a cross-modal transfer may apply, it is hard for mapping systems to overcome the large sensory mismatch between vision and other sensory pathways: *if hearing does not even provide room enough to convey visual sensations; actual visual perceptions are therefore very unlikely.*

### **See Color:**

In terms of hardware, the See ColOr prototype makes use of a 3D sensor (Microsoft Kinect), a light laptop, a tactile tablet (iPad or iPhone), and bone-phones to conduct sound through vibrations (not covering the ears). See ColOr's aim is to enlarge legibility of the nearby environment as well as to facilitate navigating towards desired locations, general exploration, and serendipitous discoveries. Three main modules were developed, in order to replicate a number of mechanisms present in the human visual systems. These modules are:

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

In the **global module** the camera image is made accessible on a tactile-tablet interface that makes it possible for the user to compare different points and explore the scene in a broader context. A user may rapidly scan an image sliding one or more fingers on this tablet. The finger movements are intended to mimic those of the eyes. A finger touch triggers a sound that codes the color and position of the corresponding pixel. To this end, we use an empirical relation between colors and instruments, and virtual sound sources that give the illusion of sounds coming from three-dimensional locations. Alternatively, instead of using a tactile screen to sense the fingertip movements, we let the user enter his hand into the camera picture. Accordingly, we can track his fingertip using a pink marker (see Figure 1), and assess the portion of the image being pointed for sonification. By and large, the global module is intended to promote a more proactive interaction to selectively explore, to discover points of interest, make comparisons, and, in general, enjoy a greater sense of independence.



Figure 1. Blindfolded individual exploring a scene using the global module with (left) and without (right) a tactile tablet. In both cases, the sound being activated is that of green (plant leaves). Also, the sound is heard as though coming from the actual plant location in space (virtual source). Note that without a tablet, the object (e.g. plant) will also emit sound if touched (tactile sensation augmented by hearing).

Since the global module is limited to describe local portions using low-level features such as color and position, it might fail to reveal cognitive aspects which often determine regions of interest within a picture. The purpose of **alerting system** is to warn the user whenever a hazardous situation arises from obstacles lying ahead. Once the system launches a warning, the user is expected to suspend the navigation not to bump into an obstacle. This allows the blind persons finding a safe, clear path to advance through. Roughly, when potential obstacle in the video presenting a distance below one meter continues to approach over a given number of frames, the user must be alerted. It is worth noticing that the alerting system is an autonomous algorithm that demands no user intervention and runs in parallel to the rest of the modules. Thus, users will focus on the exploration without loss of safety.



Figure 2. While a blind individual explores an environment (probably with the global module), the alerting system senses if any object approaches dangerously (e.g. a crash can, or a person). Note that the alerting system says nothing about the nature of the object; it just triggers an alarm indiscriminately.

See CoIOr also uses computer vision techniques to process higher visual features of the images in order to produce acoustic virtual objects. Actually, we recognize and then sonify objects that do not intrinsically produce sound, with the purpose of revealing their nature and location to the user. The **recognition module** is a detecting-and-tracking hybrid method for learning the appearance of natural objects (and faces) in unconstrained video streams. First of all, there is a training phase to learn the distinct appearance of an object of interest (scale, rotation, perspective, etc.). This is an off-line process carried out by regular sighted individual. Subsequently, visually impaired people can be informed about the presence of learned objects (in real time) during exploration. Overall, this module allows the blind noticing serendipitously discoveries; seeking a specific target; and avoiding obstacles as well.



Figure 3. The recognition module informs about the presence of previously learned objects by voice, e.g. Trash! Printer!. Note that in the rightmost frame, one of the objects became dangerously close hence, detectable by the alerting system.

Additionally, See CoIOr has a hybrid component between the global and the recognition module, which acts as a powerful engine for text recognition based on Artificial Neural Networks. Roughly, we will have a user scan a camera image on the tactile tablet (e.g. iPad, iPhone, or just pointing with the pink marker). If it happens that the user touches (or points) a letter, See CoIOr will immediately spell out the name of that letter making use of our text recognition method. This will allow the user to build words as he scans the content of the image, in quite similar way to the Braille system. We do not recognize automatically all the text in the image because this will overwhelm the user with information likely undesirable. Also, if a blind person (seeking an exit) is told the word “Exit”, for the system detected that word into the current scene. This person will gain no spatial awareness of the location of the actual exit. In sharp contrast, it is far more useful if the seeker is provided also with a rough location of the exit such as right, left, front, bottom-left etc.



Figure 4. Blindfolded individuals detecting bus numbers with the text recognition engine. At the rightmost, we can see closely the way it works: a first neural network classify whether the finger is touching on text. If so, a second network classifies the particular letter being touched.

**Testing:**

Another key aspect of See CoLoR is that its current prototype is made of relatively affordable technologies. This fact will benefit low-income countries, where blindness indexes continue to grow due to poor medical accessibility in rural areas. Likewise, it makes See CoLoR more practical and better situated in terms of social impact. Accordingly, we traveled all the way to a developing country (Colombia South America). We did so, with the belief that it is incumbent on all researchers to make the effort to reach out to the community, to engage with the population they are serving. Thus, in pursuit of more work inclusion for individuals with disabilities, we conducted over 180 experiments with 25 blind individuals, during several days. These subjects (many congenitally blind) were legally blind individuals, meaning they had visual acuity of less than 20/400. Their ages ranged between 25 and 50, and just few of them had educational level above high school. Although See CoLoR has gone through many tests, we reserve this space to highlight these ones due to their implicit social impact.



Figure 5. Some blind participants who agreed to have pictures taken for this project.

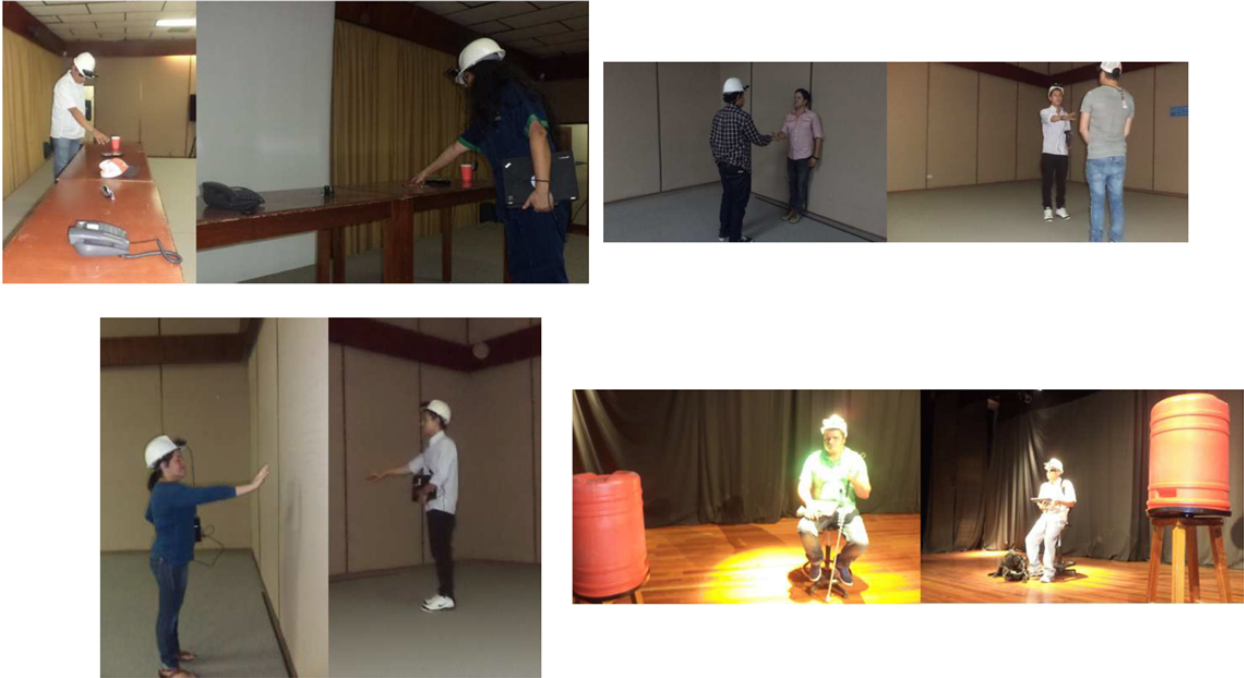


Figure 6. Experiments conducted with blind individuals. Top-left: they find and grab a specific object out of many. Top-right: they find, approach and shake the hand of a person within a 50 square-meter room. Bottom-left: They walk straight toward a wall and stop without colliding, yet being able to reach it by hand. Bottom-right: Using the red sound and the three-dimensional sonification, they can spin in a chair until they detect a surrounding target. All these experiments were conducted in relatively efficient times without white-canes, nor guide dogs. Likewise, no collision was registered due to the alerting system's intervention.

By and large, See CoLoR enjoyed the acceptance and the appreciation of the community that put it to the test. As a matter of fact, the enthusiasm and blissfulness of this South American blind people, will linger in our memories through a series of videos: video<sup>1</sup>, video<sup>2</sup>, video<sup>3</sup> and video<sup>4</sup>. Further, we think it is worth quoting some of their expressions: *“suddenly all this amount of sonified colors becomes a symphony one has to master”*; *“one feels being a kind of musician and therefore one needs to further develop the auditory capabilities”*; *“This is simply something I have been always dreaming of”*; *“transmitting sound without covering my ears really rocks and helps”*; *“this system will definitely improve the quality of my life”*; *“I was born blind, so if you ask me, knowing the colors through sounds is an experience beyond my imagination”*; *“near an otherworldly experience”*; *“I did not know technology had reached this far”*; *“learning is not hard at all to me”*.

In the end, what all this means to us is that in the middle of so many uncertainties, our research is but a little fire lighting up the vast obscurity around us. An obscurity that makes us feel, from time to time, like trapped within the darkness of blindness.

<sup>1</sup> <https://www.youtube.com/watch?v=2rNTWTpu1-8>

<sup>2</sup> <https://www.youtube.com/watch?v=4309ojboYhk&noredirect=1>

<sup>3</sup> <https://www.youtube.com/watch?v=FJyXftwwzks>

<sup>4</sup> [https://www.youtube.com/watch?v=QZ\\_t\\_BNWS7M](https://www.youtube.com/watch?v=QZ_t_BNWS7M)

## **Discussion:**

### ***General***

We cannot stress enough the need to helping the blind and the visually impaired to gain a more independent life in a daily basis. Certainly, a great deal of potential help lies nowadays in the overlap of empirical research on sensory substitution and technology strides. This latter comprises computer science and many of its branches such as robotic, computer vision or artificial intelligence. As researchers, this calls on us to keep trying our best in breaking new ground and pushing the boundaries of the knowledge in these areas. This thesis began by offering insight into what vision/blindness implies in humans. It follows that actual vision embraces both, sensation and perception. The former more related to the sensing or acquisition of visual cues found in the outer world. Whereas the latter has more to do with the coherent interpretation of such information, allowing us to derive sense out of the world. In terms of English psychologist Nicholas Humphrey, sensation is evidently related to “what is happening to me”; while, perception is evidently related to “what is happening out there”; something way more complex. For instance, the redness as we see it arises from one’s sensation, yet contemplating and understanding a rose being red is quite another thing. It is perception indeed. In the one hand, therefore, eyes and optic nerve are typically regarded as receptors that enable mere sensations. In the other hand, the brain gathers the makings of a meaningful visual experience as such (i.e. qualia), hence its association with actual perception.

The implications that follow the previous statements are such that they have given rise to the theory of sensory substitution and multisensory perception. The chief idea is that since the working of the brain is not affected in most of the cases of blindness (only the eyes), people who lose the ability to retrieve data from their eyes could still create subjective images by using data conveyed from other sensory modalities. In other words, elicitation of visual experiences in eye blindness might still be possible, provided that visual sensations somehow can reach the visual cortex of the brain. To do so, firstly, sensations from the visual space are to be mapped into another sensory modality space (e.g. sounds or tactile sensations). Thus, defective eyes might be bypassed using a substituting sensory pathway. That this mapped or encoded information will reach the visual cortex and not elsewhere in the brain, is a fact rooted in the idea of natural brain plasticity. As a consequence, cortical re-mapping or reorganization happens when the brain is subject to either neural lesions or training. This latter training of course, turns out to be central to sensory substitution. In this thesis, we studied quite a number of cases that endorse such an idea. Furthermore, we considered clinical accounts that show activity in the visual cortexes of congenital blind individuals who underwent rehabilitation, using sensory substitution devices. In light of this, the present document explores among the most relevant sensory substitution devices, from Paul Bach-y-Rita's first attempts up to cutting-edge developments in this field. Then, the conclusion was drawn that the most used modality to substitute vision is the auditory pathway. This is mostly the case owing the fact that the capability of the auditory sense to transmit information is the second greater in humans, only overtaken by that of the vision itself. Nonetheless, even though the ear is known to be capable of transmitting 10 Kbps, such capacity still lies far away from that of the vision, which may reach up to 1000 Kbps.

Looking back throughout the evolution of SSDs, it became clear that they all have tried to improve their sound outputs or sonic codes, as though ignoring the fact that no sound can lead to a full vision-like experience. In other words, they have focused on designing sensations, yet they do little for aiding the perceptual experience as such. It might be true that if an optimal audio-based



sensation were to reach the visual area of the brain, the perception will occur naturally. Yet, building an optimal sensation out of sounds is not possible at all due to the large sensory information rate mismatch between vision and hearing. Notwithstanding, state-of-the-art SSDs keep on making more complex sounds such as soundscapes, in the hope to improve the perceptual experience. In theory it should improve (though never enough), yet in practice those sounds are bound to be confusing and even uncomfortable. In these instances, we argue that more needs to be done in order to enhance the perceptual experience of a SSD's user. By no means, however, the visual-to-sound encoding must be abandoned in SSDs. Quite the opposite; we promote additional, complementary and never exclusionary techniques to cope with the actual issue of the ears being unable to convey sufficient information as to create visual-like experiences. Our thesis is, hence, that the coding into sound of basic visual cues accompanied by computational methods that model higher perceptual levels of the visual system will lead us to a SSD: functional, ease to use, and suitable for mobility and exploration tasks. Such higher perceptual levels of vision cannot be better modeled by others than computer vision techniques.

Some would argue that the use of computer vision to recognize and then, communicate objects to the user triggers just visual imagery rather than actual vision via sensory substitution. For instance, letting a user know about the presence of a tree turns out to be general symbolic mapping mediated by the concept 'tree', whereas a soundscape may convey specific information of the scene: tree's type, perspective, location and so forth. However, we may add others who claim that "The only difference is that whereas imagining finds its information in memory, seeing finds it in the environment. Thus, one could say that vision is a form of imagining, augmented by the real world." As a consequence, 'normal' vision is itself constrained by top-down knowledge. This being known, it would be unpractical to deny to this knowledge a role in visual sensory substitution. Top-down knowledge provides the kind of information that sighted individuals achieve from their visual systems, typically without conscious effort. Furthermore, this computer-vision-based strategy will prevent the blind from spending 70 hours of training (and more) in recognizing an object that in any case, will never look as real as expected. In this order of ideas, we put forward thereafter in this document the concept of See CoLoR.

Finally, as reflecting on the philosophical side of our work, we also grew very much interested in knowing how vision (a nonphysical phenomenon) comes into being, out of physical activity in a physical brain: why we draw a total blank on the nature of this transformation? More vividly, "*how is that the water of the brain becomes the wine of vision?*" to quote again English psychologist Nicholas Humphrey as he compares such a transition with a miracle. Likewise, let us raise then a question that arguably fits better this work: are researchers on visual substitution bound to succeed in eliciting visual consciousness artificially, or not? Given the intricacy inherent to this question there would be little point pursuing an answer, were it not for all breakneck strides we saw through this thesis: congenital blind people gaining brain activity in the visual cortex after electric tongue stimulation, or auditory inputs. Also, others have come to adopt sighted-like behaviors by means of mere skin stimulation, or audio-trajectory-playback. In the end, what all this means to us is that in the middle of so many uncertainties, our research is but a little fire lighting up the vast obscurity around us. An obscurity that makes us feel, from time to time, like trapped within the darkness of blindness.

## **Bibliography**

- [1] World Health Organization. (2010, March) WHO. [Online]. <http://www.who.int/mediacentre/factsheets/fs282/en/>
- [2] Ghose, D & Wallace, M 2012, 'Impact of response duration on multisensory integration', *Neurophysiol*, vol. 108, no. 9, pp. 2534-2544.
- [3] Bach-y-Rita, P 2003, 'Seeing with the brain', *International Journal of Human Computer Interaction*, vol. 15, no. 2, p. 285–295.
- [4] Amedi, A, Amir Amedi's Lab, The Hebrew University of Jerusalem. Available: <http://brain.huji.ac.il/site/em.html>. [Accessed 12 March 2013].
- [5] Humphrey, N 1992, *A History of the Mind: Evolution and the Birth of Consciousness*, New York: Simon and Schuster.
- [6] O'Regan, JK 2011, *Why Red Doesn't Sound Like a Bell: Understanding the feel of consciousness*, Boston: Oxford University Press.
- [12] Vincent, W 1998, *Perceptual Constancy: Why Things Look as They Do*, Cambridge: Cambridge University Press.
- [7] Way, T & Barner, K 1997, 'Automatic visual to tactile translation, part I: human factors', *IEEE Transactions on Rehabilitation Engineering*, vol. 1, no. 5, pp. 81-94.