On How to Achieve Visual Sensory Substitution

Dragos Moraru, and Costin A. Boiangiu

Abstract— This paper tries to find the answer to a question that throughout the time many writers and philosophers have debated on: is it possible to genuinely replace a human sense with the help of other senses and some clever technology? Starting with the 80's, technology capable of aiding damaged human senses started to emerge, the first commonly used device being the cochlear implant. Visual prosthesis (retinal implants and brain implants) as well as sensory substitution devices give artificial sight to blind persons (although the number of patients which benefited from such implants is relatively low compared to the number of cochlear implant users). This paper investigates existing technologies able to aid persons with a visual sense deficit, either by implantation of a visual prosthesis or by substituting the defective sensory modality with another functional sensory modality.

Keywords—Image to sound, medicine and technology, multi-sensory, substitution of senses.

I. INTRODUCTION

FROM a structural point of view, an artificial system to substitute the visual sense of a blind person with the auditory sense consists of three components: a sensor or sensor array that collects visual information from the environment (e.g. a video camera), an electronic system that processes this information and converts it to audible signal, and an electroacoustic transducer (speaker, headphones) that converts this electronic signal into sound to be heard by the person.

Alternatively, visual patterns could be converted into tactile stimuli patterns, or auditory and tactile stimuli simultaneously. As of 2015, many implantable visual prosthesis and visual sensory substitution devices have been designed.

II. CURRENT STATUS

Concern for the creation of artificial human senses able to replace damaged natural senses became possible with the development of technology, the development of electronic systems able to detect luminous and acoustic signals and especially the development of digital computing systems, capable of performing advanced signal processing [1]. While devices capable of receiving and recording light and sound were already developed and the processing power of computing systems had increased, the problem of achieving a connection between the electronic part and the human brain remained an issue.

Neuroscience and medical studies have shown the existence of many correlations between information patterns obtained from the senses and the generated responses of different areas of the cerebral cortex. Based on this information, detailed maps showing the locations on the cerebral cortex that are specialized in processing information from the senses and the role of each area in processing a particular type of specific information was created.

For a person to be able to regain sight artificially, the sense itself can be partially recovered through an artificial visual prosthesis or by the use of another unaffected sensory modality through which information corresponding to the affected sensory modality will be transmitted, thus achieving a substitution of the senses.

A. *Making visual prostheses*

Depending on which part of the patient's visual system had been affected and the technology available, a connection can be made between a visual prosthesis and the optic nerve (in patients with affected retinas, but with a healthy optic nerve) or directly to the brain (when the patient does not have a functional optic nerve or in cases where this type of prosthesis is more likely to succeed). In most cases, primary sensory information is obtained by the use of a video camera.

The first attempt of an implantable visual prosthesis occurred in 1982 when João Lobo Antunes, a Portuguese surgeon implanted an experimental electronic device into the eye of a blind patient. Two years later, a team of researchers led by Dr. William H. Dobelle developed a cerebral visual prosthesis [2]. The device consisted of a grid of 68 electrodes implanted in the brain of the patient, directly stimulating the visual cortex. Visual information from the environment is gathered by an assembly containing an external video camera and an ultrasonic distance measuring system. The information received is processed by a computer system and sent to the network of electrodes. Following stimulation of the visual cortex, the patient perceives an array of black and white dots. Although the image resolution is quite low, it is sufficient to restore some sense of sight to the patient and allows the perception of an object outline or the reading of large print. The visual acuity obtained is comparable to the visual acuity of the peripheral area section of the retina of a person with normal sight or of a person suffering from a high degree of uncorrected myopia.

Most subsequent research to achieve brain visual prostheses relied on the principle established by Dr. Dobelle, with various improvements in information transfer from the outside world

D. Moraru is with the Computer Science Department from "Politehnica" University of Bucharest, Romania (e-mail: dragos,moraru@cti.pub.ro)

C. A. Boiangiu is with the Computer Science Department from "Politehnica" University of Bucharest, Romania (e-mail: costin.boiangiu@cs.pub.ro)

(camera) to the implanted prosthesis, such as wireless signal and electricity transmission systems [8][23]. Newer research also aims to achieve a durable implant and increased image resolution. Researchers from the "Laboratory of Neural Prosthesis", Illinois Institute of Technology developed a device similar to that of Dr. Dobelle's, but with intracortical electrodes. This type of implant, at least in theory, has the advantage of a better resolution.

Many research projects have focused on the development of implantable visual prosthesis on the retina [3]. The idea that underpins their achievement is that in many cases, the partial or total loss of the sense of sight is due to the destruction of the first retinal layer (retinal photoreceptors) [4]. In many patients, this is the only cell layer affected, while the rest of the retina remains healthy and/or partly functional. Thus, a retinal implant is intended to stimulate these healthy layers of cells (ganglion cell layer). [5] A team of researchers from the MIT University have developed a retinal prosthesis implant consisting of a network of electrodes that stimulate the ganglion cell layer. Subsequently, the team has improved their design, by usage of wireless transmission of the information received from the external camera.

Fig. 1 Retinal implant and external power supply and camera. Courtesy of Second Sight

A similar commercial device was developed by a US company, Second Sight. The implant partially restores the patient's vision, allowing him to visualize the surrounding space, simple object recognition and reading of large printed letters. [6] Other research teams have developed a retina implantable prosthesis capable of detecting light. It partially reproduces the normal functioning of a healthy retina (an artificial retina). This prosthesis is implanted on the retina and it contains a network of sensors that convert light stimuli into electrical impulses (similar to a normal retina). Such a prosthesis is, for example, the one developed by a German team at the University of Tübingen. The device is implanted on the retina, but unlike other devices, it contains a network of photodiodes that convert light into electrical impulses which further stimulate the photoreceptors. Natural photoreceptors however are more sensitive than the implanted photodiodes. Since the photodiodes are not sensitive enough to be stimulated by normal levels of visible light exposure, an external power supply is used to amplify current. The device enjoyed some success and patients using it could recognize simple objects and read large letters. According to some patients' claims the device enabled useful functional vision. The sensory perception obtained allowed simple navigation around the room, object outline detection and detection of light source direction and position. Other researchers have developed implantable photovoltaic retinal prosthesis [7], but in this case the implantation procedure proved difficult, because of the difficulty in power transmission to the implant from the power supply system.

A. Substitution of the visual sense

Implantable visual prostheses described in the previous chapter have the advantage that they give the patient real visual perceptions. Thus a patient who was previously completely blind or with a very low vision would benefit from functional vision restoration and conscious visual perception and can still use his visual cortex even if the retina is non-functional. Unlike an implant, substituting the visual sensory modality with another functional sensory modality does not imply surgery and in theory, a completely blind person could recover some function of the visual system without taking the risk of implantation. Also, substitution of senses could theoretically be used when the patient has a non-functional visual cortex (an implant would not work in this case) or if the patient is suffering from other diseases that would make surgery very risky.

Substitution of the senses aims only to recover the functionality of the damaged sense with no direct stimulation of the retina or brain, so it cannot produce real visual experience (sensations). [24]

1) The scientific basis for sensory substitution

It is now known that in response to loss of function or training, the brain can change both anatomically (structurally) and physiologically (functionally), neural pathways and synapses undergoing anatomical changes. This phenomenon is known as neuroplasticity [9]. The brain can rewire and adapt to the changing environmental conditions to which it is subjected and when trained in the use of other senses in processing visual information, the brain can adapt structurally and functionally to process the visual information received. [10] To test this theory, many researchers have tried substituting visual sense with other senses (usually auditory or tactile sensory modality).

In terms of information, an image can be represented by a 2D array (matrix) in which each element is the value of a distinct pixel. Each frame can be converted directly into tactile stimuli by a direct mapping between each pixel in the image and an array of touch sensitive pixels, placed over a fixed area of the skin. This is possible because the body has a large number of tactile sensors in the skin, and for a picture with low resolution, a 1:1 mapping between a pixel in the image and a point on the skin can be achieved. For images with higher resolution this is not possible because the "resolution" of the sense of touch is not as great as that of the visual sense. Also the brain's ability to process signals from different sensory modalities varies. In addition, the human sense of touch resolution is not uniform across the skin. There are areas of high sensitivity, such as fingers or tongue with a large number of receptors, and there are areas where the sense of touch is much diminished. If a higher tactile resolution is desired, a skin area with high sensitivity must be selected (or a skin area of a larger size).

Unlike tactile, in the case of the auditory sense, the image frame mapping will not be a 1:1 pixel mapping. Physically, the audio signal represents vibration of air molecules, a one-dimensional signal (in time domain), as opposed to the visual/tactile signal, where each frame is a matrix of x * y pixels where x and y determine the resolution of the image. But the human ear does not receive sounds in this unprocessed form. Inside the ear, the sounds are frequency separated so that auditory information is received and processed by the brain largely depending on their frequency. Frequency only carries part of the information, but frequency alone can be considered as a two-dimensional signal, considering the range of frequencies and time axes.

2) vOICe Technology

vOICe technology is a system that performs substitution of the visual sense with the auditory (sound-image conversion). A video camera is attached to the head of the user. The camera takes images of the environment. These are processed and turned into sounds that are heard by the user through headphones. [12]

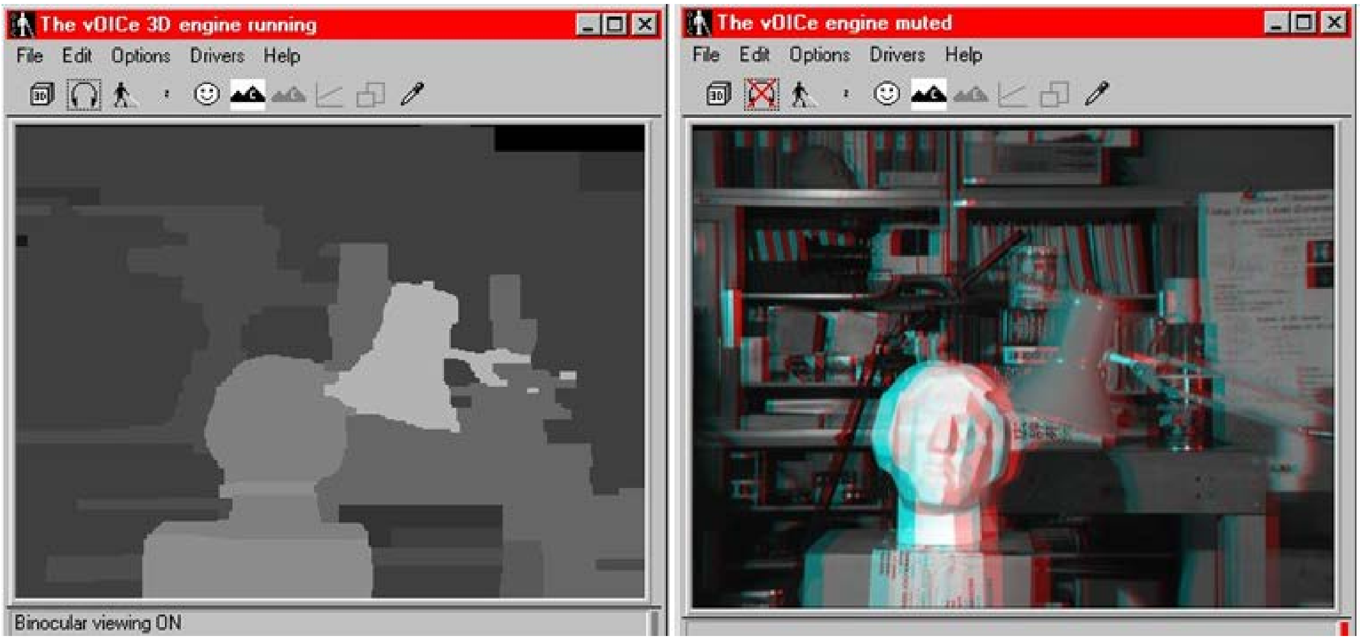

Fig. 2 Stereoscopic vOICe engine

Conversion is based on the notions discussed above. The sound signal is considered two dimensional. The x-axis is the sound frequency, and the y-axis is time. In general, a frame is considered to have a period of 1 second. Each image frame is scanned from left to right, and at every moment a sound signal with corresponding frequencies is generated. This is a direct light-sound mapping. The brighter the pixel, the louder it will be heard. Later versions tested for a reversed (dark loudness) version. [13] At the end of a frame conversion, the whole process is run again for the next frame.

Changes in the visual processing system occurred, the brain adapting to the stimuli obtained by means of the auditory sense. [12] [14] The experiments were performed on subjects with normal visual function, but the subjects were blindfolded, so as to be able to use only auditory information.

Subjects would walk around a room filled with common objects, with the purpose of identifying simple shapes and navigating an area with obstacles. Previous studies indicated that, the success rates obtained when subjects were made aware of the exact operation of the sensory substitution system in the beginning of the experiment did not vary significantly compared to the success rates obtained when subjects did not know any details on how the system worked. The success rate in locating objects was much better than that of exactly identifying objects. Even after repeated training for object recognition, the success rate did not increase significantly, although the object recognition time was significantly decreased.

Other studies followed people for several months, as they practiced with the visual substitution system. Brain imaging showed an activation of brain areas normally used in pattern recognition and spatial processing, which confirms the validity of the neuro-plasticity phenomena. An interesting consequence was that after more practice with the device, subjects whether blind or having normal visual perception, reported that they could perceive sound signals as being image-like, objects having a visual spatial distribution, and the entire process of identification and recognition similar to visual perception. [13]

Despite the positive results achieved, the system cannot provide the same performance in terms of visual acuity as other methods can, such as implantable prostheses or other methods of visual substitution. Reading and identifying letters is very difficult, if not impossible with such raw input.

Although the image resolution could in theory be increased by the simultaneous transmission of several frequencies, or by scanning the image from the left and right at a slower rate, this does not necessarily benefit the patient in any practical way. While a low resolution proves to be insufficient to provide functional vision in real world cases, high resolution would create difficulties in information integration in the brain. In order to achieve the best possible results with this system, a compromise between resolution and scanning speed is necessary. Compared to other types of substitution systems, vOICe and in general all visual-auditory substitution systems have the disadvantage of a slow frame rate. This is normal because the visual-auditory substitution system must sweep the visual signal on one axis, while visual-tactile systems can process a full frame (2D array of points) at once.

Improved system performance can be achieved by image preprocessing. The system could detect and extract relevant features and could differentiate them from background noise, while at the same time increasing the difference in contrast, so that objects become clearly identifiable and can be more easily isolated from the rest of the image by the brain.

1) Prosthesis for Substitution of Vision by Audition (PSVA)

The equipment is similar to the one used in vOICe: a video camera captures images, they are processed and the sound is heard in the user's headset. A pixel-frequency association is performed on the vertical axis, similar to the vOICe system, but in this case, an inverse correlation between the visual and auditory model is achieved. A simplified visual model of the retina is mapped over an inverted linear model of the cochlea. The authors argue that a system based on this type of mapping is feasible and that there are practical advantages in using such devices for visual-auditory substitution. A prototype that allows optimization of the operations specific to this type of substitution process has been developed. [20]

2) The EyeMusic system

It works much like the vOICe system, with the difference that the vertical axis height is represented by the musical notes of a musical instrument. [15] As in the vOICe system, a bright pixel will generate a sound with higher amplitude. [16] The novelty it brings is that it also allows color to be identified. [17] For each color represented, the corresponding musical notes are generated by a different instrument, thus, the user can, for example, identify a red apple in a bowl of green apples.

Despite improvements, EyeMusic suffers from the same weaknesses and limitations as the vOICe. Identifying objects (compared to identifying their location) is quite difficult, even for a trained user.

Despite receiving more complex sensory information, the brain cannot process the input properly and identifying outlines and shapes of objects is not practical (although as in the case of vOICe, patients can learn to recognize and process information in a manner which is similar to visual processing). In addition, as with other types of visual and auditory substitution, the identification and processing of information is very slow compared to other methods, such as visual prosthetic implants.

3) The Vibe

The brain has the ability to perceive stimuli received from the senses as separate objects located in different regions of space. [18] When, for example, a human being looks at an object, reflected light from the object reaches the eye (the retina), where it stimulates the photosensitive cone and rod cells. Although the actual stimulus occurs inside the retina, the image is not perceived as such, but as an object in space, situated at a specific distance from the observer. Similarly, when we hear a noise, the conscious perception is not of vibration inside the ear, but of a remote sound produced at some distance from the hearer that comes from an identifiable direction. This natural ability of the brain of distal attribution of stimuli is not limited to information from the senses and it can be extended to other areas of perception.

In visual-auditory substitution, it is generally intended that an auditory stimulus artificially generated and heard in the headphones be perceived as a distal effect produced in space. Such a technique is used in the software "The Vibe". It is believed that through the generated sound, spatial perception of the environment can be obtained. Practical advantages of such a system would include orientation in space, movement and avoiding obstacles in the environment rather than identifying an object or performing other complex tasks. The algorithm divides the image received from a camera into localized areas of pixels called receptive fields. Each receptive field is considered a virtual sound source in the visual space and will sound accordingly. For each of these fields, the algorithm calculates an average brightness to be used when generating the sinusoidal sound. The frequency of sound is calculated based on the pixel coordinates of the center of gravity of each receptive field. [19]

After numerous studies, observations on how the remote system can produce distal perception of stimuli were obtained. At first, subjects reported hearing something with no distal significance, which could not be associated with remote stimuli. However, after training with the device, subjects reported and demonstrated that they had acquired the ability of perceiving sounds differently than other random noise and could associate them with distant information (distal perception). Sounds were no longer perceived as local sounds, but as stimuli located in a three dimensional space, caused by objects located in space, each object emitting corresponding sounds. Such results are not unique to The Vibe system. Similar results were obtained with visual-tactile substitution systems.

4) *Visual-tactile substitution systems*

These systems use the tactile receptors in the skin to substitute the visual sensory modality. Because a visual (image) matrix of pixels can be mapped directly to a sensitive area of the skin, these systems allow for a higher resolution and a more precise mapping of each pixel. While visual-auditory substitution din not allow for the exact location of the pixel to be perceived and interpreted by the user, visual-tactile substitution systems allow the user to more precisely localize the stimulated spatial area. For this reason, such systems have the potential to enable handwriting recognition, object outline recognition, shape identification and precise localization, although skin cannot provide the necessary resolution for driving a car or moving in a complex, real-world environment. [21] In such cases, the perception can be improved using methods similar to those used in visual-auditory substitution systems, such as image simplification (reducing resolution up to the point that performance would start to fall) to retain only relevant details that matter and can be effectively identified, or visual feature extraction. It is recalled that the skin sensory organ has variable sensitivity. This would imply that a determining factor for the success of a visual-tactile substitution system is to choose an area with sufficient sensitivity and an appropriate region size for stimulation. The tactile resolution in a given area depends not only on the number of receptors in that area, but it is also limited by the quality of the connections with the brain and it depends on the brain's ability to process and identify independent stimuli received from different areas. In addition to the actual resolution of the area of skin used, the resolution of the system is limited by the number of stimulation elements. This is a concern since electrode arrays cannot be made to resolutions that would match a video camera resolution (such high pixel resolution would not be useful anyway because the skin does not have such high resolution).

Visual-tactile substitution systems can be divided into two categories: general systems for use in various situations and specialized systems that are intended for a specific purpose (e.g. for reading printed text).

Fig. 3 Seeing with tongue. Courtesy of Wicab

The first visual-tactile substitution system known in the scientific literature was developed by Bach-y-Rita and included a video camera for capturing images and a vibrational tactile stimulation system (the image was transduced into tactile vibrations) located on the back of the user. Using this system, it was possible to identify simple objects as well as their spatial orientation by means of contrast difference. Research was continued using different skin areas. Ultimately, Bach-y-Rita concluded that the most effective method is to stimulate the tongue, because the human tongue has high spatial resolution and the device could be greatly reduced in dimensions compared to the older versions that stimulated other parts of the body. Although the initial results were promising, further development of the project lead to the conclusion that such a system, although useful as a text reader, has severe practical limitations when used in real-world situations. In such situations, large amounts of information processing is necessary, which is much more than the brain can reliably process and interpret using tactile modality (despite neuroplasticity phenomena taking place), compared to a much greater capacity of visual modality processing. The mere fact that a certain sense has a higher spatial or informational resolution does not mean that it allows the perception of nonspecific information better than another sense with lower resolution, but specialized in processing information of that certain type. For example, although the amount of visual sensory information the brain can process simultaneously is much greater than the auditory bandwidth, for a deaf person, it is almost impossible even with intense training to be able to recognize words in a graphical (visual) representation of sound (spectrogram), because the brain is not adapted to recognize such information using visual sensory modality.

As with visual-auditory substitution systems, it is not clear how much of the phenomenon of neuroplasticity can be exploited in order to use nonspecific brain areas for visual processing.

In general, visual-tactile substitution systems, under specific use-cases can achieve better performance. A relevant example is Optacon [22], invented by Professor John Linvill of Stanford University, for optical character recognition on paper or on a monitor, by a blind person. Light signals are converted into a binary touch signal (each pixel is either white or black). The vibrations are perceived by the subject by placing his fingers on the touch surface. Unlike general purpose visual-tactile substitution devices, which stimulate a fixed surface on the skin, in the case of Optacon the surface and its position varies (the user moves his fingers on the device). System resolution is 24 rows and 6 columns (the required resolution was determined experimentally). The user moves his fingers from left to right as if reading a written text. It is important for proper function of the device that the vertical resolution, which gives the number of rows to be higher. Although the device also works with a single column, there was a significant improvement in performance with an increased number of columns. The system was steadily improved from the original version, but was eventually abandoned in favor of modern text to speech software which is much faster and easier to use.

III. CONCLUSIONS

Although image-to-sound substitution systems developed so far can offer a simple visual perception to a blind person [25][26], providing the brain with information that can be perceived and interpreted in a manner similar to natural vision (due to the phenomenon of neuroplasticity), their applicability has not been demonstrated in real and practical cases. On the contrary, studies indicate a weak brainpower in identifying details, shapes, outlines, based on auditory and tactile stimuli, without real visual cues.

By its nature, the brain has a greater capacity for processing visual information than other types of sensory modality. Therefore, it is much harder to develop a system capable of providing a detailed and conscious perception of the environment using other sensory modalities, regardless of how information is acquired and processed and the number of neurons directly stimulated. From this point of view, implantable visual prosthetics, through direct connection to the visual cortex or retina is the only available solution that generates authentic visual sensations, so that in theory, they will tend to have the edge over any visual substitution methods in future research. [11]

Currently, the exact limits of neuroplasticity in allowing the usage of other senses for visual processing are not clear. Studies conducted so far had not been conclusive and the possibility of using a substitution system to produce real visual perceptions remains an open question.

REFERENCES

[1] Bach-y-Rita, Paul, and S. W. Kercel. *"Sensory substitution and the human–machine interface."* Trends in cognitive sciences 7.12 (2003): 541-546.

[2] , W. H Dobelle, M. G. Mladejovsky, and J. P. Girvin. *"Artificial vision for the blind: Electrical stimulation of visual cortex offers hope for a functional prosthesis."* Science 183.4123 (1974): 440-444.

[3] K. Kyuma, et al. *"Artificial retinas---fast, versatile image processors."* Nature 372 (1994): 197-198.

[4] A. Y. Chow, et al. *"Subretinal Artificial Silicon Retina Microchip for the Treatment of Retinitis Pigmentosa: 3 1/2 Year Update."* Investigative Ophtalmology and Visual Science 45.5 (2004): 3398.

[5] J. D. Dorn, et al. *"The detection of motion by blind subjects with the epiretinal 60-electrode (Argus II) retinal prosthesis."* JAMA ophthalmology 131.2 (2013): 183-189.

[6] M. P. Barry, and G. Dagnelie. *"Use of the Argus II retinal prosthesis to improve visual guidance of fine hand movements."* Investigative ophthalmology & visual science 53.9 (2012): 5095-5101.

[7] L. Wang, et al. *"Photovoltaic retinal prosthesis: implant fabrication and performance."* Journal of neural engineering 9.4 (2012): 046014.

[8] P. Walter, et al. "Cortical activation via an implanted wireless retinal prosthesis." Investigative ophthalmology & visual science 46.5 (2005): 1780-1785.

[9] L. B. Merabet, et al. *"What blindness can tell us about seeing again: merging neuroplasticity and neuroprostheses."* Nature Reviews Neuroscience 6.1 (2005): 71-77.

[10] J. P. Rauschecker, *"Compensatory plasticity and sensory substitution in the cerebral cortex."* Trends in neurosciences 18.1 (1995): 36-43.

[11] C. Lenay, et al. *"Sensory substitution: limits and perspectives."* Touching for knowing (2003): 275-292.

[12] 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-LONDON- 36.3 (2007): 416.

[13] http://www.seeingwithsound.com, Accessed on 15 June 2015.

[14] A. Amedi, et al. *"Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex."* Nature neuroscience 10.6 (2007): 687-689.

[15] S. Abboud, et al. "EyeMusic: Introducing a “visual” colorful experience for the blind using auditory sensory substitution." Restorative neurology and neuroscience 32.2 (2014): 247-257.

[16] http://brain.huji.ac.il/site/em.html, Accessed on 15 June 2015.

[17] G. Hamilton-Fletcher, and J. Ward. *"Representing colour through hearing and touch in sensory substitution devices."* Multisensory research 26.6 (2013): 503-532.

[18] J. H. Siegle, and W. H. Warren. *"Distal attribution and distance perception in sensory substitution."* Perception 39.2 (2010): 208.

[19] S. Hanneton, M. Auvray, and B. Durette. *"The Vibe: a versatile vision-to-audition sensory substitution device."* Applied Bionics and Biomechanics 7.4 (2010): 269-276.

[20] C. Capelle, et al. *"A real-time experimental prototype for enhancement of vision rehabilitation using auditory substitution."* Biomedical Engineering, IEEE Transactions on 45.10 (1998): 1279-1293.

[21] Bach-y-Rita, Paul, and B. Hughes. *"Tactile vision substitution: some instrumentation and perceptual considerations."* Electronic spatial sensing for the blind. Springer Netherlands, 1985. 171-186.

[22] J. G. Linvill, *"Research and Development of Tactile Facsimile Reading Aid for the Blind (The Optacon). Final Report."* (1973).

[23] S. Raghavan, G. A. Kumar, *"Microstrip Patch Antenna for a Retinal Prosthesis"*, 7th WSEAS Int. Conf. on Instrumentation, Measurement, Circuits and Systems (IMCAS '08), Hangzhou, China, April 6-8, 2008

[24] A. F. Rodriguez-Hernandez, C. Merino, O. Casanova, *"Sensory Substitution for Visually Disabled People: Computer Solutions"*, WSEAS Transactions on Biology and Biomedicine Volume 7, 2010

[25] V. Tiponut, Z. Haraszy, D. Ianchis, I. Lie, *"Acoustic Virtual Reality Performing Man-machine Interfacing of the Blind"*, 12th WSEAS International Conference on Systems, Heraklion, Greece, July 22-24, 2008

[26] D. Moraru, C. A. Boiangiu, *“About Visual Sensory Substitution”*, Proceedings of the 3rd International Conference on Acoustics, Speech and Audio Processing (ASAP '15), Salerno, Italy, June 27-29, 2015, WSEAS Press, pp. 115-124.