

Sensing Super-position: Human Sensing Beyond the Visual Spectrum

David A. Maluf
650-604-0611
David.A.Maluf@nasa.gov

John F. Schipper
650.604.5286
John.F.Schipper@nasa.gov

NASA Ames Research Center
Moffett Field, CA 94035

NASA Ames Research Center
Moffett Field, CA 94035

Abstract— The coming decade of fast, cheap and miniaturized electronics and sensory devices opens new pathways for the development of sophisticated equipment to overcome limitations of the human senses. This paper addresses the technical feasibility of augmenting human vision through Sensing Super-position by mixing natural human sensing. The current implementation of the device translates visual and other passive or active sensory instruments into sounds, which become relevant when the visual resolution is insufficient for very difficult and particular sensing tasks.

A successful Sensing Super-position meets many human and pilot vehicle system requirements. The system can be further developed into cheap, portable, and low power taking into account the limited capabilities of the human user as well as the typical characteristics of his dynamic environment. The system operates in real time, giving the desired information for the particular augmented sensing tasks.

The Sensing Super-position device increases the image resolution perception and is obtained via an auditory representation as well as the visual representation. Auditory mapping is performed to distribute an image in time. The three-dimensional spatial brightness and multi-spectral maps of a sensed image are processed using real-time image processing techniques (e.g. histogram normalization) and transformed into a two-dimensional map of an audio signal as a function of frequency and time.

This paper details the approach of developing Sensing Super-position systems as a way to augment the human vision system by exploiting the capabilities of the human hearing system as an additional neural input. The human hearing system is capable of learning to process and interpret extremely complicated and rapidly changing auditory patterns. The known capabilities of the human hearing system to learn and understand complicated auditory patterns provided the basic motivation for developing an image-to-sound mapping system.

The human brain is superior to most existing computer systems in rapidly extracting relevant information from blurred, noisy, and redundant images. From a theoretical

viewpoint, this means that the available bandwidth is not exploited in an optimal way. While image-processing techniques can manipulate, condense and focus the information (e.g., Fourier Transforms), keeping the mapping as direct and simple as possible might also reduce the risk of accidentally filtering out important clues. After all, especially a perfect non-redundant sound representation is prone to loss of relevant information in the non-perfect human hearing system. Also, a complicated non-redundant image-to-sound mapping may well be far more difficult to learn and comprehend than a straightforward mapping, while the mapping system would increase in complexity and cost. This work will demonstrate some basic information processing for optimal information capture for head-mounted systems.

TABLE OF CONTENTS

1. INTRODUCTION 1
2. MULTI-SPECTRAL HUMAN SENSING 2
3. TECHNICAL APPROACH 3
4. SUPER-SENSING 4
5. SUPER-POSITION 5
6. RELATED WORK 7
REFERENCES 7

1. INTRODUCTION

Humans rely heavily on vision to sense the environment in order to achieve a wide variety of goals. The expected payoff is to augment human sensory visual system, which is deficient flight in many respects.

An auditory technology, often integrated within helmet-mounted display systems technology has been limited to the use of radio communication. Modern research in stereoscopy had demonstrated the use of stereo sounds for human perception of the environment. In addition to sensor technology advancement, there has been a tremendous effort towards the development of newer display modalities. Such devices can be readily integrated in the pilot head-mounted systems to provide both pilotage and targeting imagery mappings.

Over the past 30 years, there have been innumerable articles and scientific papers, which address the design and performance of helmet- and head-mounted systems. A large

portion of this paper is the result of a research capability to augment the current state-of-the-art in head mounted display systems (implicitly), rather than propose a new system. With the fielding of various human factors in military systems, research within this area of mapping visual to audio stereoscopy has been fairly new but effective [6]. While this paper is intended to provide a fairly comprehensive overview of a particular project, the technology and its interface for a human perception, has numerous potential beyond the proposed ideas.

2. MULTI-SPECTRAL HUMAN SENSING

Since the time of Aristotle (384-322 BC), human kind has been interested in perceiving what is beyond its "vision", but it was not until Roentgen (1845-1923) discovered X-Rays, which enabled him to see inside living tissue, that "vision" beyond the naked eye entered a new era [7]. In the following years imaging and sensing techniques have developed so rapidly that astronomy, medicine and geology are just few of the areas where sensing beyond vision has been found useful. Finally, whether in electromagnetic, optical or acoustic sensing, the main intention of this research is to augment our understanding of the surrounding world [7].

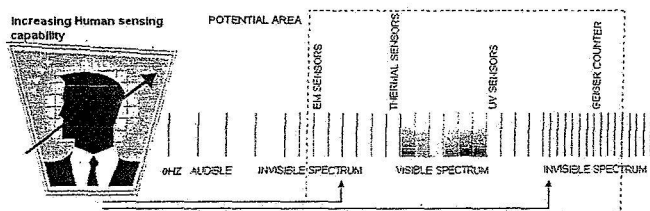


Figure 1 – Human augmented sensing spectrum.

Figure 1 is an illustration of human augmented sensing capability beyond the visible spectrum. "Human perceives and interprets the world through their sensing. Altering the human sensing of the world will alter his perception of it." [7].

The field of computer vision has already resulted in ample research ramifications in image interpretation, where visual perception aspects have gained a large influence. Numerous mathematical models for representation of the neurophysiological world, whether partial or complete, have been established into concepts and paradigms [8]. The immense complexity and evolution of the computational aspects of vision has now given way to an increased perception, comprehension and understanding of the surrounding world. Despite advances in computational vision, this paper discusses again the nature of the human perception. The focus will be on the evolution of ideas rather than on models, and, likewise, the emphasis is placed more often on the classical foundations of the field human perception instead of the current approaches in computer

vision (e.g. camera registration, and pattern recognition). Thus, the intention is to consider the human perception of the physical world achieved by correlating auditory sensing and visual sensing.

Generally² speaking, this paper describes the concept of humans acquiring information distantly about specific targets using alternative sensing approaches (multi-spectral). In this paper contrasts and distinguishes the ability to penetrate specific targets with the intent to reconstruct in detail the morphology of the target into knowledge (computer-based processing) from the "end-effect" perception (human perception) of physical world (desired result). Finally, what mainly constitutes sensing is the end effect of particular energy spectrum and the source is arbitrary in all natural dimensions – temporal, spatial, energy spectrum etc. -

Objectives

This paper describes a human-centered computing multimodal interfaces needs and deficiencies and it addresses the *super-position* of human sensing (visual and audio). It is tied with the requirements of human-vehicle system and human-environment interactions.

A number of tools have been created to help address these deficiencies such as heads-up displays. However, the tools tend to specialize on only one of the limitations. Moreover, these tools do not capitalize on, to any significant extent, other human senses to augment the human visual system and to address the deficiencies described. Humans and primates rely heavily on vision to sense and react to the environment in order to achieve various goals namely, provide a capacity to sense beyond the human visible light range of the electromagnetic spectrum..

The approach can also be applied to provide aural signal components representing shape signatures, sizes and estimated separation distances for objects that cannot be seen, or that are seen very imperfectly, because of signal interference, signal distortion and/or signal attenuation by the ambient environment. This may occur in a hazardous environment where fluids present provide an opaque, darkened or translucent view of objects in the environment, including moving or motionless persons and objects that present a hazard.

This interference may also occur in an airborne environment in which rain, snow, hail, sleet, fog, condensation and/or other environmental attributes prevent reasonably accurate visual perception of middle distance and far distance objects. A visual image component that is likely to experience interference can be converted and presented as a sequence of audio signal attributes that can be more easily or more accurately perceived or interpreted by an operator

² Or profoundly.

of an aircraft (airborne or on the ground). The audio signal attributes may be extended to include estimated closing velocity between the operator/aircraft and the not-yet-seen object.

3. TECHNICAL APPROACH

Sensing Super-position as a way to augment the human vision system exploits the capabilities of the human hearing system. It is known that the human hearing system is capable of learning to process and interpret extremely complicated and rapidly changing auditory patterns, such as speech or music in a noisy environment. The available effective bandwidth, on the order of 15 kHz, corresponds to a channel capacity of several thousands of bits per second. The known capabilities of the human hearing system to learn and understand complicated auditory patterns provided the basic motivation for developing an image-to-sound mapping system.

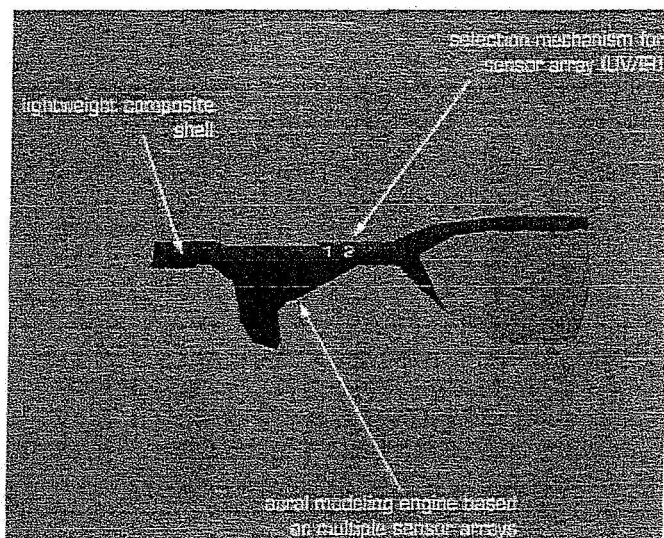


Figure 2 –Superposition as a head mounted device.

Figure 2 illustrated the derived device from an original sketch. This illustration demonstrates the concept of the the superposition device showing multiple sensory super-impositions. External miniaturized multi-spectral sensors signals are translated into sounds targeting the same human visual field.

The human brain is far superior to most existing computer systems in rapidly extracting relevant information from blurred, noisy, and redundant images. From a theoretical viewpoint, this means that the available bandwidth is not exploited in an optimal way. While image-processing techniques can manipulate, condense and focus the information (e.g., Fourier Transforms), keeping the mapping as direct and simple as possible might also reduce the risk of accidentally filtering out important clues. After all, especially a perfect non-redundant sound representation is prone to loss of relevant information in the non-perfect

human hearing system. Also, a complicated non-redundant image-to-sound mapping may well be far more difficult to learn and comprehend than a straightforward mapping, while the mapping system would increase in complexity and cost. This paper will demonstrate some basic information processing for optimal information capture for head-mounted display systems.

Technically, it is difficult to obtain unambiguous high-resolution input using scanning sonar, while any commercially available low-cost camera will do. It should be added, however, that this statement only holds if the camera input is supplemented by depth clues through changes in perspective, to resolve ambiguities in distance. Further, relative depth information can be derived from evolving relative positions of the viewer and his environment, combined with knowledge about the typical real size of recognized objects. Augmenting the mapped-to-audio visual motion with mapped-to-audio binocular vision lends itself to providing image sensing beyond the concepts of binocular extension, which in overall mimics the way current visual system operates. Research in computer and machine vision is still active in motion and stereoscopy operating concurrently or standalone as real time systems.

In order to increase the image resolution obtainable via an auditory representation, mapping is performed to distribute an image in time. A somewhat related method has been applied by the British in the design of tactile displays. In this paper, the three-dimensional spatial brightness and color maps of a visual image is processed using real-time image processing techniques (e.g. histogram normalization) and transformed into a two-dimensional map of an audio signal as a function of frequency and time. The same basic functionality of mapping was considered in [2].

The proposed translation breaks down to three logical operations in the general image to sound mapping of grayscale sensor-based images as in thermal sensors. Each operation dealing with one fundamental aspect or dimension of the world projection to sensors: The first aspect is concerned with the first dimension (horizontal to the visual field) and establishes the image scan vector. The second operation is concerned with second image dimension (vertical to the visual field) and establishes the mapping the sound pitch. The third operation is concerned with the pixel intensity.

Stereoscopy

Extracting distance information using a single camera is difficult because the distance information in a single camera view is essentially ambiguous and requires much a priori knowledge and motion (multiple views) about the physical world to derive distances from recognized objects. By comparing the slight differences in images obtained from two different simultaneous viewpoints, the distances to nearby objects can be understood. However, the level of visual resolution in stereoscopic vision is exponentially

increased with dynamic scenes (motion) -- where strong correlation between the slight differences captured through stereo and motion are sensed.

The stereoscopic processing used in the proposed technology is in contrast to so-called anaglyphic often used in computer vision. Instead, two images are created from two distinct viewpoints from a single spectral signature. For thermal sensing, the two differently views are not altered. The proposed stereoscopy process is pushed to the brain level in sounds waves, but similar enough to the biological stereoscopic vision. Stereoscopy is an overlay process achieved by the human brain. In simulated environment however, sighted viewers can then again see the actual scene in 3D by looking at the anaglyph through red-green glasses: the red filter in front of the left eye blocks the green or cyan component and transmits only the left-eye red image while the green filter in front of the right eye transmits only the right-eye image. The human brain subsequently combines the slightly different views into one perceived three-dimensional view. Distance information is then apparent from something called "disparity", the small visual displacements of the red and green/cyan color components for nearby objects. Whereas the left-eye and right-eye view may coincide perfectly for far-away objects, the mismatch for a nearby object tells how close this object is. This is the principle upon which the proposed 3D stereoscopic sensing super-position is based on.

4. SUPER-SENSING

Since the human brain can combine slightly different views into one perceived three-dimensional view, distance information and resolution size (level of details) are then apparent from the stereoscopic disparity, i.e., the small visual displacements in two images. Whereas the left-eye and right-eye views may coincide indistinguishably for far-away objects, the mismatch for a nearby object tells how close this object is. However, the human eye is limited to acuity that is limited by the density of the cones and rods in the human eye. Since the brain can analyze the difference in displacements, the device can use a denser pixel distribution in a CCD layout and map this additional stereo vision information to spatialized sound.

Challenges with the technology development

There is no suitable hardware commercially available for the visor prototype, but an electronics engineer should be able to construct stereo sensor systems ("3D thermal sensors") from standard commercially available components. In the first prototype, we intend to use two regular thermal CCD cameras and research adequate disparity angles for the camera resolution and associate each camera with one eye and ear. We plan to represent individual intensity components directly as a function of reverberation or sound intensity. For example, it would

render a bright red object with a strong reverberation and no intensity with any reverberation effects. An alternate approach to reverberations is the pixel intensity would be represented by the amplitude of the signal. In such case, it would render a bright red object with a strong loud signal and no intensity with any sounds.

The coming decade of fast, cheap and miniaturized electronics and sensory devices opens new pathways for the development of sophisticated equipment to overcome limitations. To miniaturize this system, the work lies ahead in the process of dismantling and rebuilding a common low resolution CCD-based thermal cameras (no cooling needed) where the actual CCD array measures less than 1/4 inch in size.

While the research aspect of the proposed technology will be based on a computer with anaglyphic simulations, the actual miniaturization will take a harder pathway where electronic circuitry will used to map the CCD sensor array signals to sounds. In all cases, use of a head mounted sensors is strongly recommended for most consistent head sensory feedback, and achieve sensors alignment with the visual axis. In the scope of this paper, the outcome of the simulation environment will set the course of the actual build. It is planned that two prototypes will be constructed.

Geometric Variation and impact on Human Perception

In this paper, one research aspect is planned to attempt a variation on geometric masked images and measure human sensitivity to feature changes in the field of view. This exercise is very tempting since it answers a fundamental science question and compensate for a natural human sensing deficiency. There is an inherited deficiency in human perception that lies in the fact that human visual field is horizontal in respect of the position of the human eyes. Horizontal feature (lines) sensing and localization in visual field is difficult when the feature lies in the middle and not in the foveated region.

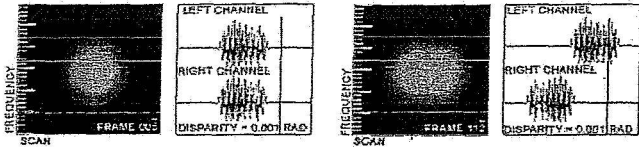


Figure 3 – A 3D rendering and audio tracks.

Figure 3 demonstrate of a spherical field motion in 3D and the corresponding sound tracks. The first image from the left reflects a 3D spherical field at a distant in space. The image is computer generated using the dual anal glyphic Cyan/Red filter for immersive 3D perception and simulates two stereoscopic cameras. The second image from the left reflect the sounds tracks, namely the left and right channels. The overlap of the blue and red spheres implies minimal shifts in the left and right channels. The third and fourth images reflect the 3D sphere at closer distance and hence a

wide disparity angle between the red and blue spheres with the corresponding sound tracks and shifts. (Note the scan is horizontal, continuous, and symmetrical to central axis. The pitch is vertical).

5. SUPER-POSITION

An example of supplementing, or replacing, visual signals by aural signals is presented here.

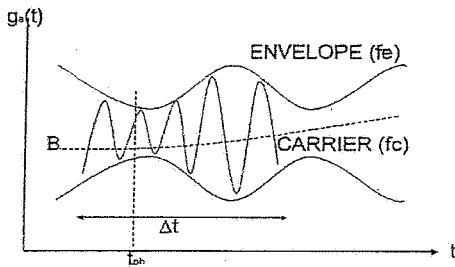


Figure 4 – characteristics of an audible signal.

Figure 4 graphically illustrates seven signal parameters that can be used to collectively characterize an undulating, audibly perceptible signal having a single information-bearing (envelope) frequency and a single carrier frequency.

This signal can be characterized by: an envelope frequency f_0 ; a carrier frequency f_c ; an envelope frequency phase ϕ_e at a selected time, $t = t_{ph}$; a carrier frequency phase ϕ_c at the selected time, $t = t_{ph}$; a baseline function $b(t)$; a signal amplitude $a(t) \approx a_0 \sin\{f_c t + \phi_e\}$ relative to the baseline curve B ; and a time interval (duration) Δt for the signal. The human ear may be able to distinguish the phase difference, $\Delta\phi = \phi_e - \phi_c$, but cannot distinguish the absolute phases, ϕ_e and/or ϕ_c . Thus, the maximum number of parameters for the signal shown in Figure 1 that may be distinguished by the human ear is six, if the (absolute) selected time, $t = t_{ph}$, is not included. These six signal parameters may be used to audibly represent a corresponding visual component of an image, such as vertical location of the visual component (relative to a fixed two-dimensional or three-dimensional coordinate system), horizontal location of the visual component, component brightness; overall component brightness, and component predominant hue (color) or wavelength. Optionally, these audible signal parameters can be presented simultaneously or sequentially, for any corresponding visual image component that is so represented. In a sequential presentation, one or more additional audible signal parameters may be included, if the information corresponding to the additional parameter(s) is necessary for adequate representation of the image

component.

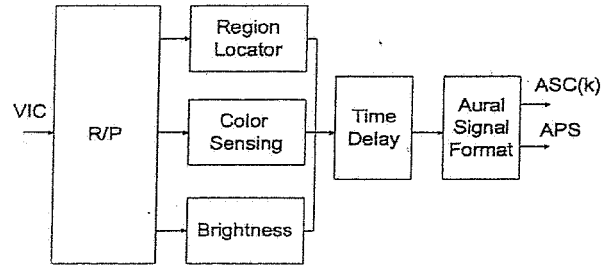


Figure 5 – Transforming an image to audio signals.

Figure 5 schematically illustrates a mapping device used to transform selected visual image components associated with a region to an audible signal with audibly perceptible parameters. One or more visual image regions, represented as an assembly of visual signal components, is received and analyzed by a first signal receiver/processor ("R/P"). The first R/P analyzes a received visual signal component and provides one or more (preferably all) of the following visual signal characterization parameters: vertical location of the (center of the) region; horizontal location of the (center of the) region, region predominant hue (color) or wavelength, region average brightness and region peak brightness, using a region locator mechanism, a region predominant (or average) color sensing mechanism and a region brightness sensing mechanism. Output signals from the locator mechanism, from the color mechanism and from the brightness mechanism are received by a second R/P, which provides a collection of audible signal parameters, illustrated in Figure 4, that are determined, in whole or in part, by the output signals from the first R/P.

As an example: the predominant or average color output signal from the color sensing mechanism can be used to determine the carrier frequency f_c ; the brightness output signal from the brightness mechanism can be used to determine the envelope relative amplitude, a_0 or $a(t)$; the vertical and horizontal location output signals from the locator mechanism can be used to determine time duration Δt (if the visual image component locations are indexed by a one-dimensional index), or to determine time duration Δt and envelope frequency f_e (if the visual image region locations are indexed using a two-dimensional index). Generally, the four visual signal parameters can be assigned to four of the six audibly perceptible signal parameters (Figure 4) in ${}^6_4 = (6 \cdot 5 \cdot 4 \cdot 3) / (4 \cdot 3 \cdot 2 \cdot 1) = 15$ distinguishable ways. More generally, N visual signal parameters can be assigned to M ($\geq N$) audibly perceptible signal parameters in M_N distinguishable ways.

The analysis performed by each of the mechanisms is not instantaneous, and the associated time delays may not be the same for each analyzer. For this reason, an overall time delay

$$\Delta t(o1) \geq \min\{\Delta t(\text{location}), \Delta t(\text{color}), \Delta t(\text{brightness})\} \quad (1)$$

is preferably imposed, using a time delay mechanism, before an aural signal (or aural signal sequence) incorporating the one, two, three, four, five or six aural signal parameters is aurally displayed. If the aural signal parameters are displayed sequentially, rather than simultaneously or collectively, this time delay might be reduced or eliminated. The overall time delay is implemented by a time delay mechanism, which incorporates an appropriate time delay value for each of the aural signal parameters received from the R/P. A signal formation mechanism (optional) forms and issues either: (1) an audibly perceptible, ordered sequence of the set of 1-6 aural signal components $ASC(k)$, $k=1, \dots, 6$, (or a subset), or (2) a collective audibly perceptible signal APS incorporating the set (or a subset) of the aural signal components. The output signal from the signal formation mechanism is perceived by the human.

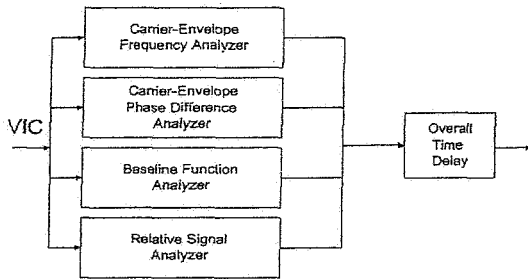


Figure 6 – Transforming an image to audio signals.

The R/P, illustrated in Figure 6, includes one or more of the following; a carrier/envelope frequency (f_c) analyzer; an envelope-carrier frequency phase difference ($\Delta\phi$) analyzer and baseline function ($b(t)$) analyzer that estimate the phase difference at a selected time and determines the baseline function; and a relative signal amplitude (a_0 or $a(t)$) analyzer, relative to the baseline function at a corresponding time.

The analysis performed by each of these analyzers is not instantaneous, and the associated time delays may not be the same for each analyzer. For this reason, an overall time delay

$$\Delta t(o2) \geq \min\{\Delta t(f_c), \Delta t(\Delta\phi), \Delta t(b), \Delta t(a)\} \quad (2)$$

is preferably imposed before an aural signal incorporating the one, two, three, four or five converted visual signal parameters is aurally displayed. If the converted visual signal parameters are aurally displayed sequentially, rather than simultaneously, this time delay might be reduced or eliminated. The overall time delay is implemented by a time delay mechanism, which incorporates an appropriate time delay for each of the aural parameters received from the R/P. A signal formation module forms a single composite

aural signal representing an aural image component, and issues this component as an output signal.

The signal shown in Figure 5 may be represented in a form

$$\begin{aligned} F_{VIC}(t) &= b(t) + a_0 \sin\{f_c(t-t_\phi) + \Delta\phi\} \sin\{f_e(t-t_\phi)\} \\ &= b(t) + a_0 \{\cos\{(f_c-f_e)(t-t_\phi) - \Delta\phi\} - a_0 \{\cos\{(f_c+f_e)(t-t_\phi) \\ &\quad + \Delta\phi\} \end{aligned} \quad (3)$$

The carrier/envelope frequency analyzer forms a sequence of correlation signals, computed over a time interval of length T ,

$$C1 = (1/T) \int F_{VIC}(t) \sin\{f_{es}t\} dt, \quad (4A)$$

$$C2 = (1/T) \int F_{VIC}(t) \cos\{f_{es}t\} dt, \quad (4B)$$

at each of a spaced apart sequence of “translated” carrier frequencies, f_{cs1} in a selected carrier frequency range, $f_{e1} \leq f_{cs} \leq f_{e2}$, where f_{cs} is not yet known, and provides an estimate of two spaced apart frequencies $f_{cs1} = f_c + f_e$ and $f_{cs2} = f_c - f_e$, associated with the VIC, where the correlation combination, $C1^2 + C2^2$, has the highest magnitudes. The envelope and carrier frequencies are then determined from

$$f_c = (f_{cs1} + f_{cs2})/2, \quad (5A)$$

$$f_e = (f_{cs1} - f_{cs2})/2. \quad (5B)$$

The envelope-carrier phase difference $\Delta\phi$ and relative amplitude a_0 are determined by computing the correlations

$$(1/T) \int F_{VIC}(t) \sin\{f_c(t-t_\phi)\} dt = a_0 \cos\Delta\phi, \quad (6A)$$

$$(1/T) \int F_{VIC}(t) \cos\{f_e(t-t_\phi)\} dt = a_0 \sin\Delta\phi, \quad (6B)$$

from which the quantities a_0 (≥ 0) and $\Delta\phi$ are easily determined. The baseline function $b(t)$ is then determined from

$$\begin{aligned} b(t) &= F_{VIC}(t) - a_0 \{\cos\{(f_c-f_e)(t-t_\phi) - \Delta\phi\} \\ &\quad - a_0 \{\cos\{(f_c+f_e)(t-t_\phi) + \Delta\phi\}. \end{aligned} \quad (7)$$

The frequency difference (f_c-f_e) and frequency sum (f_c+f_e) values are distinguished from each other in a normally functioning human auditory system if the sum-frequency difference $2f_e$ is at least equal to a threshold value, such as 500 Hz.

The approach can also be applied to “enrich” image detail or manifest more clearly some image details that are not evident where the region is viewed solely with reference to visible light wavelengths. For example, some details of a region may be hidden or muddled when viewed in visible

wavelength light but may become clear when the region is illuminated with, or viewed by an instrument that is sensitive to, near-infrared light (wavelength $\lambda = 0.7 - 2 \mu\text{m}$) or mid-infrared light ($\lambda = 1 - 20 \mu\text{m}$) or ultraviolet light ($\lambda \leq 0.4 \mu\text{m}$). These hidden details can be converted to aural signal parameter values that are more easily audibly perceived as part of a received aural signal. Operated in this manner, the approach can separately compensate for a relatively narrow (or relatively broad) visible wavelength sensitivity of the viewer into a relatively narrow (or relatively broad) auditory frequency sensitivity of the same viewer. Operated in this manner, the visible wavelength sensitivity of a first (visual image) viewer of the image component can be adjusted and compensated for electronically by adjusting the aural wavelength range of one or more of the aural signal parameters, before the transformed aural signal is received by the same viewer or by a different viewer.

6. RELATED WORK

John Zelek et al. of the School of Engineering at the University of Guelph, Canada, have developed a stereo-vision system for the blind using a tactile display for showing the nearest obstacles [18][19][20]. This work above advances research in image processing and segmentation techniques to level of target recognition.

Another player in the domain mapping images to sounds is Yoshihiro Kawai as described in [21][22]. Yoshihiro's work focuses on object 3D spatial localization prior mapping to stereo sounds.

However, what is the most relevant to the work proposed in this research will rely on neurophysiological mapping work encompassed in the literature such as the Sound Graph method as described in [10]. For example, Douglass et. al demonstrates that mathematical concepts such as symmetry, monotonicity, and gradients could be determined using sound. Additional work in sound mapping work is found under [6]. A clear distinction the proposed work has, is the mapping the invisible spectrum to the human perception bringing forth the human multi-sensing nature to level not achieved before.

Competing technology & Current Status

This technology competes with the approach of mapping recognized objects to English using natural language (speech). Sensing super-position warrants a one to one mapping of natural stereoscopy and time varying scenes to audio signals.

To maintain focus on the potential and interests, the developmental prototypes, includes thermal sensing. For now, Thermal sensing super-imposition capability demonstrates a new way of perception in human sensing.

REFERENCES

- [1] G. Von Bekesy, "Experiments in Hearing," New-York: McGraw Hill Company, 1960.
- [2] S. A. Dallas and A. L. Erickson, "Sound Pattern Generator," Patent Int. Pub. No WO 82/00395, Int. App. No. PCT/US81/00847.
- [3] W. H. Dobbelle, M. G. Mladejovsky and J. P. Girvan, "Artificial Vision for the Blind: Electrical Stimulation if Visual Cortex Offers Hope for a Functional Prosthesis," Science, Vol. 183 pp. 440-444, February 1, 1974.
- [4] R. M. Fish, "An audio Display for the Blind," IEEE Transactions on Biomedical Engineering, Vol. 23, pp. 144-154, 1976.
- [5] R. O'Hea "Optophone Design: Optical-to-Auditory Vision Substitution for the Blind," Ph.D. Thesis, the Open University, 1994.
- [6] P. B. L. Meijr, "An experimental System for Auditory Image Representations," IEEE Transactions on Biomedical Engineering, Vol. 39 No. 2, pp. 112-121, 1992. Reprinted in the 1993 IMIA Yearbook of Medical Informatics, pp. 291-300.
- [7] D. A. Maluf, "World Modeling in RADAR, A Regularization by Synthesis", Ph.D. Dissertation, McGill University, 1995.
- [8] R. Hummel and S. W. Zucker, "On the foundation of relaxation labeling processes", IEEE Transaction on Pattern Analysis and Machine Intelligence, 3 May 1983.
- [9] J. Y. Lettvin, H. R. Maturana, W. S. McCulloch and W. H. Pitts, "What the frog's eye tells the frog's brain? Proc. Ubst Radio Engrs, 47:1940-1951, 1959.
- [10] Douglass L. Mansur, Meera M. Blattner and Kenneth I. Joy, "Sound Graphs, A Numerical Data Analysis Method for the Blind," Journal of Medical Systems, Vol. 9, pp. 163-174, 1985.
- [11] J. K. Andrew and D. Moore, Trends in Neurosciences, 14:1, 31-37. 1991.
- [12] F. L. Wightman and D.J. Kistler, "Headphone simulation of free-field listening, Stimulus synthesis and Psychophysical validation", J. Acoustic. Soc. Am. 85 (2), Feb. 1989.
- [13] T. T. Takahashi, C. H. Keller, "Representation of Multiple Sounds Sources in the Owl's Auditory Map." Journal of Neurosciences, 14(8) 4780-4793, 1994.

- [14] L. Kay, Air sonar with acoustical display of spatial information. In Busnel, R-G and Fish, J.F., (Eds), *Animal Sonar Systems*, 769-816 New York Plenum Press.
- [15] Z. Zhang and O. Faugeras. *3D Dynamic Scene Analysis*. Springer-Verlag, 1992.
- [16] S. B. Pollard, J. Porrill, J.E.W. Mayhew, and J.P. Frisby. Disparity gradient, lipschitz continuity, and computing binocular correspondences. In J.E.W. Mayhew and J.P. Frisby, editors, *3D Model Recognition from Stereoscopic Cues*, pages 25-32. MIT, 1991.
- [17] N. Molton, S. Se, J.M. Brady, D. Lee, and P. Probert. A stereo vision-based aid for the visually impaired. *Image and Vision Computing*, 1997.
- [18] C. Dunk and J. Zelek, "A Stereo-vision System for the Visually Impaired," (Technical Report 2000-41x-1, School of Engineering, University of Guelph).
- [19] S. Areibi and J. Zelek, "A Smart Reconfigurable Visual System for the Blind," *Conf. Smart Systems and Devices (SSD)*, Hammamet, Tunisia, March 27-30, 2001.
- [20] J. Zelek, D. Bullock, S. Bromley and Haisheng Wu, "What the Robot Sees & Understands Facilitates Dialogue," *Human-Robot Interaction, 2002 AAAI Fall Symposium (American Association for Artificial Intelligence)*, November 15-17, 2002, North Falmouth, Massachusetts, USA.
- [21] Y. Kawai, M. Kobayashi, H. Minagawa, M. Miyakawa and F. Tomita, "A Support System for Visually Impaired Persons Using Three-Dimensional Virtual Sound," *Int. Conf. Computers Helping People with Special Needs (ICCHP 2000)*, pp. 327-334, Karlsruhe, Germany, July 17-21, 2000.
- [22] Y. Kawai, F. Tomita, "A Visual Support System for Visually Impaired Persons Using Acoustic Interface," *IAPR Workshop on Machine Vision Applications (MVA 2000)*, pp.379-382, Tokyo, Japan, Nov. 28-30, 2000.