Available online at www.sciencedirect.com
ScienceDirect
Applied Soft Computing 7 (2007) 257–264
Applied Soft Computing
www.elsevier.com/locate/asoc

Fuzzy image processing scheme for autonomous navigation of human blind

G. Sainarayanan *, R. Nagarajan, Sazali Yaacob

School of Engineering and Information Technology, Universiti Malaysia Sabah, Kota Kinabalu 88999, Malaysia

Received 9 May 2003; received in revised form 15 December 2004; accepted 20 June 2005

Abstract

The main objective of this work is to develop an electronic travel aid to assist the blinds for obstacle identification in their navigation. This navigation assistance for visually impaired (NAVI) system presented in this paper consists of a single board processing system (SBPS), a vision sensor mounted headgear and a pair of stereo earphones. The image environment in front of the blind is captured by the vision sensor. The image is processed by a new real time image processing scheme using fuzzy clustering algorithms. The processed image is mapped onto a specially structured stereo acoustic patterns and transferred to the stereo earphones in the system. Blind individuals were trained with NAVI system and tested for obstacle identification. Suggestions from the blind volunteers regarding pleasantness and discrimination of sound pattern were also incorporated in the prototype. The proposed processing methodology is found to be effective for object identification and for producing stereo sound patterns in the NAVI system.

Keywords: Electronic travel aid; Blind navigation; Fuzzy clustering; Pattern recognition; Sonification; Vision substitution; Image processing

1. Introduction

The loss of eyesight is one of the most serious misfortunes that can befall a person. The visual information forms the basis for most navigational tasks and so with impaired vision an individual is at a disadvantage, because appropriate information about the environment is not available. The number of visually handicapped persons worldwide would double from the present 45 million by 2020 [1]. There are a quarter of a million registered blind people in the UK. However, the UK has nearly one million people entitled to register as visually impaired and about 1.7 million are with vision difficulties [9]. This represents over three percent of the UK population. The vision aid for blinds had been under extensive research with restricted achievement since 1970's.

Electronic travel aids (ETA) are electronic devices developed to assist the blind for autonomous navigation. Early ETAs use ultrasonic sensors for the obstacle detection and path finding. Recent research efforts are being directed to produce new navigation systems in which digital video cameras are used as

* Corresponding author.
E-mail address: jgksai@ums.edu.my (G. Sainarayanan).

1568-4946/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.asoc.2005.06.005

vision sensors [2–4]. Peter Meijer [5] presented The voice in 1992 in which sine wave generator is used for sound producing. The image pixels captured by the camera are scanned from left to right and column by column. The top portion of the image is transformed into high frequency tones and the bottom portion into low frequency tones. The intensity of the pixel is transcoded into loudness.

All the earlier works in the direction of capturing the image of environment and mapping the image to sound, do not undertake any image processing efforts to provide the information of the objects in the scene [3,5]. Instead, the captured image is directly sonified to sound signals. In general, background fills more area in the image frame than the objects, and hence the sound produced from the unprocessed image will contain more information on the background. It is also observed that the background is usually of light colors and the sound produced on it will be of high amplitude compared to the objects in the scene. This may be one of the reasons for blinds finding difficulties in understanding the sound produced from camera based earlier ETAs.

In this paper, a pattern clustering method is proposed for object identification and applied towards the development of Navigation Assistance for Visually Impaired (NAVI) system. Human auditory system has enhanced frequency and intensity discrimination. It is talented even to infer sound patterns like music or speech in exceptionally noisy environment. Several studies have also indicated that the blind individuals are better than sighted individuals at auditory discrimination. With this anticipation, a procedure by which visual information is given to blind in terms of sound patterns is presented.

2. Developed NAVI system model

The model constructed for this vision substitution system has a vision sensor mounted headgear, a pair of stereo earphones and a single board processing system (SBPS) in a specially designed vest. The user has to wear the vest. The SBPS is placed in a pouch provided at the backside of the vest. SPBS selected for this system is PCM-9550F with Embedded Intel® low power Pentium® MMX 266 MHz processor, 128 MB SDRAM, 2.5'' light weight hard disk, two Universal serial bus and a RTL 8139 sound device chipset, all assembled in Micro box PC-300 chassis. The weight of SBPS is 0.7 kg. Constants 5 and 12 V supply for SBPS are provided from a set of rechargeable batteries placed in the front packets of the vest. Vision sensor selected for this application is a digital video camera, KODAK DVC325. A blind individual carrying the headgear and processing equipment in the vest is shown in Fig. 1. The work is progressing to miniaturize the size of the equipment, so as to be more convenient for the blind individual to carry.

3. Fuzzy based image processing

Digital video camera mounted in the headgear captures the vision information of scene in front of the blind user and the image is processed in the SBPS in real time. The processed image is mapped to sound patterns. Image processing should be properly designed to have effective sonification. Since the processing is done in real time, the time factor has to be critically considered. The image processing method should require less computation. In industrial vision system applications, there can be a priori knowledge on the features of objects to be detected, such as contour or size; thus with the known features, the object of interest is identified by eliminating the background [7]. In the proposed vision substitutive system, the features of objects to be identified are undefined, uncertain and time varying [10]. The classical methods for object identification and segmentation cannot be used in this application. The main effort in the NAVI system is to identify the

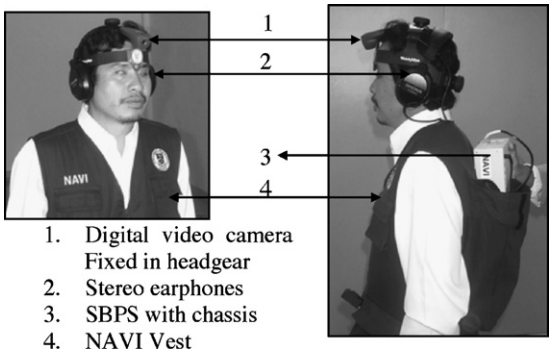

Fig. 1. Blind volunteer with NAVI system.

objects in the scene in front of the blind. Unless the task is automated, it will be very difficult for the blind user to understand the environment and navigate without collision. The important requirement of the blind is to identify the size of the objects and to discriminate the objects from background. During sonification, that is to be discussed in a later section, the amplitude of sound generated from the image directly depends on the pixel intensity. In 8-bit gray-scaled image, the pixel value of white color is maximum at 255 and black is minimum at zero. If the pixel value is related to amplitude of sound, the image pixels of light color produces sound of higher amplitude than darker pixels. Acoustic pattern set of pixels with bright colors in the dark background is easy to identify than the dark pixels over bright background. It can be felt that the background of the most real world pictures are of bright colors than the object. If the image is transferred to sound without any enhancement, it will be a complex task to understand the sound, which is the major problem faced in early works. There can be a possibility that the background may also have some important features and these features will be eliminated if a total elimination of background is undertaken. Hence, an effort is made in this paper to suppress the background instead of elimination and also to enhance the object of interest in order to impart more consideration to the object. Human vision system creates the concentration of vision only on the region of interest, while other regions are considered as background and are given less consideration and focus. Generally, the focused object will be in the center of vision and this property human vision is incorporated in the proposed method. This is the property of iris in the human vision system. There are more chances of object to be in the central (iris) area of human vision; if not, the human aligns the object to the central area by moving the iris or by turning the head. In the case of blind, moving the head is appropriate since the central area of vision is fixed to the camera. This is one of the main aspects to be considered during object identification.

3.1. Feature extraction and clustering

The main aim of this work is to suppress the background and to enhance the object; for this, the gray levels of the object and background have to be identified. Image used for processing is of 32×32 pixel size and of four gray levels, namely black (BL), white (WH), dark gray (DG) and light gray (LG). Feature extraction is the most critical part in image processing for blind. The extracted features should represent the object with limited data. The type of features extracted from image for object identification or classification depends on the application and also mainly on the computational time. In this work four features are extracted from each gray level. Each image is considered to have four feature vectors namely X_{BL}, X_{DG}, X_{LG} and X_{WH}, each with four features such as $X_{BL} = [x_1, x_2, x_3, x_4]$; $X_{DG} = [x_1, x_2, x_3, x_4]$; $X_{LG} = [x_1, x_2, x_3, x_4]$; $X_{WH} = [x_1, x_2, x_3, x_4]$

where,

x_1 represents the number of respective gray pixel in the image, this is histogram value of the particular pixel.
x_2 represents the number of respective gray pixel in the central area of the image. Iris area is the central area of human eye, which maintains a concentration of vision. This concentration is distributed towards the boundary in a non-linearly decreasing function. The central area of the image obtained by the camera is considered here as iris area and thus models the human eye. Generally the object of interest will be in the center of human vision.
x_3 represents the pixel distribution gradient. Its value depends on the location of the particular gray level pixels in the image area. x_3 is calculated by the sum of the gradient values assigned to the pixel location. The gradient value increases towards the center with Gaussian function. So that, pixel of a particular gray level in the center has comparatively higher value than the pixels of same gray level in outer area.
x_4 represents the gray value of the pixel. Generally most of the background in the real world are of light colors than the objects.

3.2. Architecture of FLVQ

Learning vector quantization (LVQ) is an effective neural network for classification. A fuzzy based LVQ (FLVQ) is considered in this work for identifying objects from background. The architecture of FLVQ is similar to Kohonen self organizing map [8]. The objective of the algorithm of FLVQ network is to identify the output node that is nearest to the input

vector and update the weight accordingly. In the NAVI classification problem, there are only two output nodes-one belonging to object class and the other to background class. In conventional LVQ training algorithm, the weight is updated by:

If $T = C_J$

$$w_J(\text{new}) = w_J(\text{old}) + \alpha_t[X - w_J(\text{old})]; \quad (1)$$

If $T \neq C_J$

$$w_J(\text{new}) = w_J(\text{old}) - \alpha_t[X - w_J(\text{old})]; \quad (2)$$

The learning rate α_t updation, in every iteration is by

$$\alpha_{t+1} = \frac{\alpha_t}{k}; \quad t \geq 0$$

If the winning cluster (C_J) is same as the defined target T, the weight (W) is updated with positive learning rate α; while winning cluster is not the defined target T, the weight is updated with negative learning rate. The procedure here is crisp in nature. A feasible connection between batch Fuzzy c Means (FCM) clustering algorithm and sequential LVQ can be made [6]. The crisp determination of learning rate can be replaced by fuzzified learning rate, by having membership both to winning and non winning clusters. Learning rate α_t is replaced by fuzzy membership value $\alpha^{ik,t}$ computed using FCM clustering algorithm [6]. Even though this approach is innovative, proper choice of fuzzification factor 'm', which determines the level of fuzzification, is essential.

Clustering algorithm for FLVQ is as follows:

Let $X = \{x_1, x_2, x_3, \ldots, x_n\}$ be the input data to be clustered,T be the maximum iteration; E_t the termination measure = $\|w_t - w_{t-1}\|$; ε the termination criterion; m_0 the initial $m < 7$; m_f the final $m > 1.1$; t the iteration count; c is the number of clusters = 2

Step 0: Initialize the weight matrix W

Step 1: Do Step 2 to Step 4 until ($t > T$ or $E_{t-1} \leq \varepsilon$)

Step 2: Calculate m_t

$$m_t = m_o + t\frac{m_f - m_o}{T} \quad (3)$$

Step 3: Calculate the learning rate $\alpha_{ik,t}$, for $k = 1$–n

$$\alpha_{ik,t} = \left(\sum_{j=1}^{c} \left(\frac{||x_k - w_{i,t}||}{||x_k - w_{j,t}||} \right)^{2/m_t - 1} \right)^{-m_t} \quad (4)$$

Step 4: Update the weight, for $i = 1$–c

$$w_{i,t} = w_{i,t-1} + \frac{\sum_{j=1}^{n} \alpha_{ik,t}(x_k - w_{i,t-1})}{\sum_{s=1}^{n} \alpha_{is,t}} \quad (5)$$

In this experimentation, the simulated images representing basic geometrical shapes are created using MS-Paint and real life images are colleted both from indoor and outdoor environment for training and testing the network. A set of 250 data are extracted from simulated as well as real life images. The extracted data have to be clustered into two classes namely object class and background class (that is $c = 2$, in the above algorithm). The number of input nodes are 4 (x_1, x_2, x_3, x_4). The efficiency of the network and speed of convergence depend on m_t and hence on the m_o and m_f. They determine the learning rate at a particular iteration 't'. In the initial stage, the learning rate is fuzzified to maximum level and the fuzzification is reduced as 't' increases. In the conventional method the m_t is reduced linearly as in Eq. (3). In this work, an exponential variation of m_t is found to be suitable.

Table 1
Results of training and testing with LVQ with crisp learning rate

S.no.	Learning rate; α	Updation parameter; k	Number of data for training	Number of data for testing	Number iteration for convergence	Percentage of classification for trained data	Percentage of classification for untrained data	Percentage of classification for trained and untrained data
1	1.0	2	150	250	35	83.33	80.0	82.0
2	1.0	1.5	150	250	61	82.0	78.0	80.4
3	2.0	2	150	250	41	88.0	83.0	86.0
4	2.0	1.5	150	250	48	86.0	80.0	83.6
5	3.0	2	150	250	38	85.33	80.0	83.2
6	3.0	1.5	150	250	59	84.67	74.0	81.2

Table 2
Results of training and testing with FLVQ

S.no.	m_o	m_f	Number of data for training	Number of data for testing	Number iteration for convergence	Percentage of classification for trained data	Percentage of classification for untrained data	Percentage of classification
1	6	4	150	250	48	86.67	82.0	84.8
2	6	3	150	250	42	89.33	87.0	88.4
3	6	2	150	250	37	92.0	91.0	91.6
4	5	4	150	250	57	86.0	84.0	85.2
5	5	3	150	250	46	87.33	84.0	86.0
6	5	2	150	250	41	88.67	87.0	88.0

Eq. (3) is modified as

$$m_t = m_f + k_z(m_o - m_f) \quad (6)$$

where $k_z = 2/1 + e^{qt}$; t is the iteration number and q is the slope parameter. The target values for clustering are fixed a priori. FLVQ network is trained with 150 data. The trained network is then tested for all 250 data. The data are clustered to the class with minimum euclidean distance with its final weight values. In Table 1, the convergence of LVQ for different values of α and k is presented and it is compared with the percentage of classification, when tested with 250 image data. The maximum classification accuracy achieved was only 86.0%

Results of the training and testing for FLVQ and FLVQ with proposed fuzzification variation method in Eq. (6) are shown in Tables 2 and 3, respectively.

By repeated experimentation as in Table 2, optimal value for m_o and m_f were found to be 6 and 2 respectively. Maximum performance of FLVQ with linear variation was found to be 91.5% as shown in Table 2. Maintaining $m_o = 6$ and $m_f = 2$, an exponential variation is applied. With $q = 2$, the classification performance was increased to 97.6%.

The trained network is implemented for online object and background detection. On detection, the object pixels are enhanced, while the background pixels are suppressed with following algorithm. Results of object enhancement and background suppression in two experiments are shown in Fig. 2. In the first experiment, the object (bag) is dark and in the second, the object (door) is light gray. In both figures, the objects are identified as white with background as black. Several experimentations reveal that the object in the output image is classified as white and background as black. The output image is thus object enhanced and background suppressed.

4. Image to sound conversion

The processed output image matrix is to be transferred into sound patterns. The frequency of the sound produced was designed within the human audible range of 20 Hz to 20 kHz. Since the human auditory system is more sensitive to lower frequency than high frequency, the frequency band is selected to be in the low frequency range [4]. The vertical position

Table 3
Results of training and testing with FLVQ with proposed fuzzification

S.no.	z_o	z_f	q	Number of data for training	Number of data for testing	Number iteration for convergence	Percentage of classification for trained data	Percentage of classification for untrained data	Percentage of classification
1	6	2	1	150	250	23	95.33	95.0	95.2
2	6	2	2	150	250	19	97.33	98.0	97.6
3	6	2	3	150	250	25	94.66	93.0	94.0
4	6	2	4	150	250	29	92.67	90.0	91.6
5	6	2	5	150	250	25	93.33	91.0	92.4
6	5	2	6	150	250	31	92.67	91.0	92.0

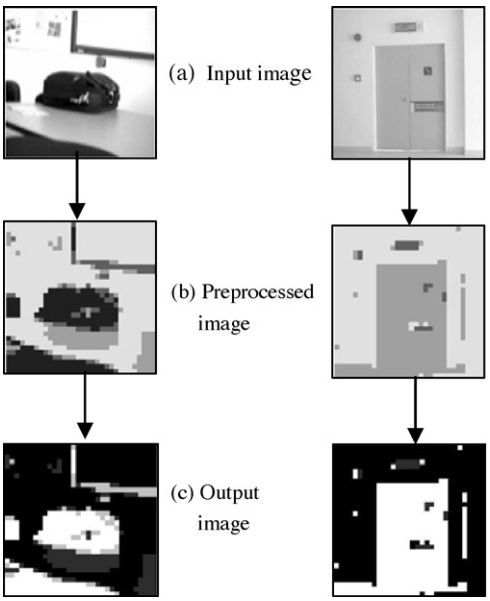

Fig. 2. Results of object enhancement and background suppression.

of the pattern is inversely related to pitch and the pixel intensity is converted into loudness of the sound. The frequency variations in the vertical position are designed to be audibly differentiable. The sound pattern produced is given by

$$S(j) = \sum_{i=1}^{m} I(i, j) \sin 2\pi f(i) t; \quad j = 1, 2, \ldots, n$$

where $S(j)$ is the sound pattern for column j of the image; $I(i, j)$ the (i, j)th element of object enhanced and background suppressed image; $t = 0$ to D, D depends on the total duration of the acoustic information of the image; $f(i)$ is the frequency of the ith row of image matrix; m the number of rows; n is the number of columns in the image matrix.

The sine wave with the designed frequency is multiplied with gray scale of each pixel of each column and summed up to produce the sound pattern. The sound pattern from each column is appended to produce the sound for whole image. The scanning of picture is performed in such a way that stereo sound is produced. In this stereo type scanning, the sound patterns created from the left half side of the image is given to left earphone and sound patterns of right half side to right earphone simultaneously. The scanning is performed from leftmost column towards the center and from right most column towards to center.

Then, the sound pattern to the left earphone, $S_L = S(1)$ to $S(n/2)$ is appended from the left side the sound pattern to the right earphone $S_R = S(n)$ to $S(n/2)$ is appended from the right side, where n is total number of columns.

The importance of the image processing stages undertaken in NAVI can be illustrated by comparing the sound in 3D form for an image with and without image processing. It is important to note that, by the human auditory nature, it is easy to identify and differentiate a high amplitude sound in the middle of low amplitude noise, compared to low amplitude sound in between high amplitude noise [11]. The background of image acquired by the camera is assumed to take more image area and is of light color compared to that of object. The NAVI image processing makes background to be darker than object. If the proposed image processing methodology is not undertaken, the background is transformed to

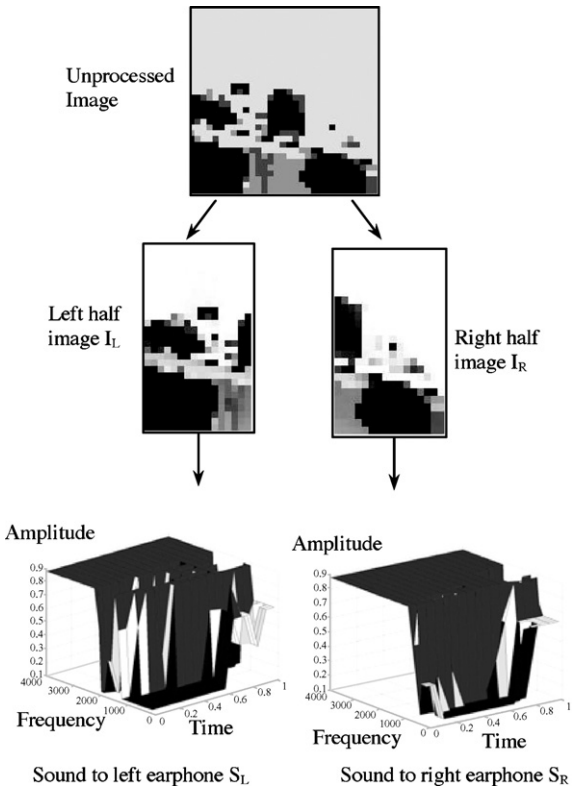

Fig. 3. 3D plot of sound from unprocessed image.

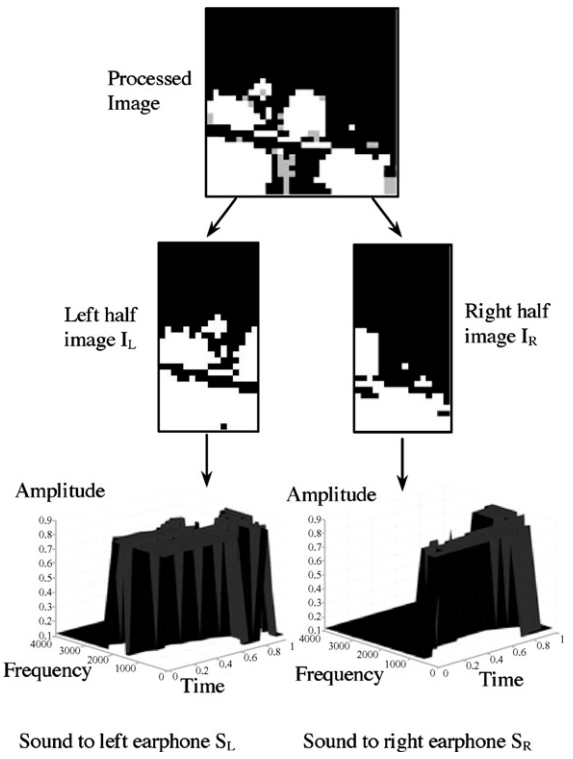

Fig. 4. 3D plot of sound from processed image.

high amplitude sound compared to that of object and therefore the features of the background will predominate over the object. The distribution of the frequency and the amplitude in the sound produced from the unprocessed image and processed images are shown in Figs. 3 and 4. The image considered is split into left half and right half namely I_L and I_R, respectively. Distribution of sound S_L to the left ear phone and S_R to the right earphone are shown in three dimensional plot (3D), in which x axis represents the time after the starting of sound, y-axis represents the frequency and z axis represents the amplitude. In Fig. 3, the sound from the background predominates the sound produced from the objects. This may cause confusion for the blind user to discriminate the object from the background. In Fig. 4, the sound from the objects predominates over the sound from the background, and hence easier for discrimination. From the above examples and discussions, the importance and necessity of the proposed object identification module can be acknowledged.

5. Conclusion

The developed prototype NAVI hardware and software were tested on blind persons. The blind persons were trained with some basic geometric shapes through computer simulation and were asked to identify obstacles of indoor environment. The volunteer is tested to locate the obstacle in the indoor environment. It was encouraging to note that, the blind is also able to identify the objects moving with a nominal speed. Work is continued to train the blind person in identifying the outdoor scene through the sound pattern produced by this prototype. The image processing designed for NAVI is found to be suitable for this application. However objects with textured surface are not enhanced evenly. In this research, information regarding depth of the object is not considered. However by comparing the sound patterns from relative distances between the blind person and the object, information regarding the nearness of objects can be manipulated by the blind person after getting an experience with the developed scheme. That is, an object is 'perceived' bigger through the variation in sound pattern as the blind moves near to the object.

Acknowledgment

Authors wish to thank Ministry of Science, Technology and Environment, Malaysia for funding the research through Universiti Malaysia Sabah: IRPA code: 03-02-10-0004.

References

[1] World Health Organization (WHO), 1997. Blindness and Visual Disability, Part I of VII: General Information, Fact Sheet 142.
[2] F. Wong, R. Nagarajan, S. Yaacob, A. Chekima, N. Eddine, Electronic travel aids for visually impaired—a guided tour, in: Conference in Engineering in Sarawak, Proceedings, 19–20, May, Malaysia, 2000, pp. 377–382.
[3] C. Capelle, C. Trullemans, A real-time experimental prototype for enhancement of vision rehabilitation using auditory substitution, IEEE Trans. Biomed. Eng. 45 (10) (1998) 1279–1293.
[4] Fish, R.M., "Auditory display for the blind", US Patent No. 3,800,082 (1974).

[5] P.B.L. Meijer, An experimental system for auditory image representations, IEEE Trans. Biomed. Eng. 39 (2) (1991) 112–121.
[6] J.C. Bezedek, J. Keller, R. Krishnapuram, N.R. Pal, Fuzzy Models and Algorithms for Pattern Recognition and Image Processing, Kluwer Academic Publishers, Boston, 1999.
[7] N.R. Pal, S.K. Pal, A review on image segmentation techniques, Pattern Recognit. 26 (1993) 1277–1293.
[8] L. Faussent, Fundamentals of Neural Networks, Prentice Hall, New Jersey, 1994.
[9] National Federation of the Blind (NFB). 2002. http://www.nfb.org/default.htm.
[10] G. Sainarayanan, R. Nagarajan, S. Yaacob, Incorporating certain human vision properties in vision substitution by stereo acoustic transform, in: Proceedings of IEEE Sixth International Symposium on Signal Processing, ISSPA, 13–16 August, Malaysia, 2001.
[11] Perrott, Discrimination of the spatial distribution of concurrently active sound sources: some experiments with stereophonic arrays, J. Acoustic Soc. Am. 76 (6) (1984) 1704–1712.