AudiNect: An Aid for the Autonomous Navigation of Visually Impaired People, Based On Virtual Interface

Mario Salerno *salerno@uniroma2.it*
Faculty of Engineering, Department of Electronics
University of Rome,"Tor Vergata"
Rome, 00133, ITALY

Marco Re *marco.re@uniroma2.it*
Faculty of Engineering, Department of Electronics
University of Rome,"Tor Vergata"
Rome, 00133, ITALY

Alessandro Cristini *alessandro.cristini84@gmail.com*
Faculty of Engineering, Department of Electronics
University of Rome,"Tor Vergata"
Rome, 00133, ITALY

Gianluca Susi *gianluca.susi@uniroma2.it*
Faculty of Engineering, Department of Electronics
University of Rome,"Tor Vergata"
Rome, 00133, ITALY

Marco Bertola *mrc.bertola@gmail.com*
Faculty of Engineering, Master in Sound Engineering
University of Rome,"Tor Vergata"
Rome, 00133, ITALY

Emiliano Daddario *emiliano182@yahoo.it*
Faculty of Engineering, Master in Sound Engineering
University of Rome,"Tor Vergata"
Rome, 00133, ITALY

Francesca Capobianco *francesca.capobianco@yahoo.it*
Faculty of Engineering, Department of Electronics
University of Rome,"Tor Vergata"
Rome, 00133, ITALY

Abstract

In this paper, the realization of a new kind of autonomous navigation aid is presented. The prototype, called *AudiNect*, is mainly developed as an aid for visually impaired people, though a larger range of applications is also possible. The *AudiNect* prototype is based on the *Kinect* device for Xbox 360. On the basis of the *Kinect* output data, proper acoustic feedback is generated, so that useful depth information from 3D frontal scene can be easily developed and acquired. To this purpose, a number of basic problems have been analyzed, in relation to visually impaired people orientation and movement, through both actual experimentations and a careful literature research in the field. Quite satisfactory results have been reached and discussed, on the basis of proper tests on blindfolded sighted individuals.

Keywords: Visually Impaired, Blind, Autonomous Navigation, Virtual Interfaces, Kinect, AudiNect.

1. INTRODUCTION

It is well known that the traditional aids for visually impaired people are canes and guide dogs, though they can present a number of limitations. Indeed, they do not allow natural movements or extended mobility, so that external helps and additional costs can often be required. In recent years, Virtual and Augmented Reality has been applied to rehabilitation of children and adults in presence of visual impairment, so that increased autonomy and significant life quality improvement can be reached [1], [2], [3], [4].

Various localization and identification techniques have been developed and discussed in the technical literature. In particular, systems based on radio waves [5], infrared [6], ultrasound [7, 8], GPS [9, 10], RFID [9] [11] are of interest in the field. However, some actual limitations can often be pointed out. For instance, systems based on radio waves seem not very accurate in position, while GPS systems appear not quite reliable in the case of indoor applications. It has been pointed out that RFID (Radio Frequency Identification) systems can overcome some limitations of many other technologies. However, in general, all the actual needs of visually impaired people seem not always fully solved by many technical approaches [12].

The use of *Kinect* for Xbox has been recently considered in the literature [13], as a new promising approach. Besides its very low cost, the main advantage is its robustness to light, being able to work both in light and in dark environments. A significant contribution in this field is the Navigational Aids for Visually Impaired (NAVI), developed at the University of Konstanz, Germany [14]. This device can improve the navigation in indoor environments, exploiting several *Kinect* capabilities. NAVI works through a vibrotactile belt that can mimic the room layout. The *Kinect* is placed on a helmet worn by the user. In addition, a number of Augmented Reality (AR) markers is placed along the walls, in order to identify a significant path from room to room. Depth information is provided by *Kinect* and processed by a C++ software, so that some vibration motors can properly be activated (on the right, the center and the left sides of the patient waist). In this way, according to the markers location, the system can guide the user through the correct path. To this purpose, a synthetic voice is also generated as an auxiliary navigation aid. A similar project is Kinecthesia, that uses a *Kinect* placed at the waist [15, 16]. Vibrating motors are placed on a belt and activated when obstacles along the path are detected. Proper audio alarms are also generated. In particular, different vibration intensities stand for different obstacle distances. On the basis of the new technologies and solutions, the studies on optimal autonomous navigation tools seem still an open research field. In particular, new or better solutions should be very important to satisfy the typical blind issues with regard to both real time and robust interaction with the environment.

In this paper, the development of a useful tool for the autonomous navigation is presented. This tool, called *AudiNect*, is based on the synthesis of proper acoustic feedback, designed as an aid for visually impaired people. It seems able to overcome some traditional tool limitations, and it is efficacy since the acoustic feedback encode the information in a simple way to be learned by a user. Moreover, the use of the *Microsoft Kinect* provides both a low cost design (e.g., compared with GPS or RFID technologies) and the possibility to encode more information from the environment than the classical tools. Also, since a system based on this new kind of technology is robustness to any light conditions (due to the use of both IR and RGB cameras, and also thanks to a proper software, introduced in the next section), can represent a valid alternative for the autonomous navigation. Finally, the proposed system does not need to be supported by other external devices to obtain information from the environment. For example, RFID systems need AR markers located along the particular considered paths; whereas GPS systems need signals sent by GPS satellites, but urban canyon and indoor site are no suitable environment due to technological limits in GPS signals capturing [12].
The paper is organized as follows: first, we will provide the basic characteristics of the proposed model. Second, we will illustrate the operation mode of the system. Third, we will test the latter on blindfolded sighted people in order to prove its validity in real cases, that is in presence of real scenarios. Results and future developments will be discussed in the last section.

2. AUDINECT

2.1. Basic Characteristics

AudiNect is based on the use of the sense of hearing, usually very well developed in people with visual impairment. To this purpose, proper auditory feedbacks are generated to detect and identify the presence of obstacles on the path in front of the user. The auditory feedbacks consist of proper sound signals, both frequency and intensity modulated. In this way the information on the most useful path can easily be recognized. The basic visual data are acquired by the *Kinect* device. By using the infrared technology, it generates proper output stream information, whose features depend on the used Software Library. A number of 30 frames per second can be produced, representing the frontal scene depth map. A matrix of up to 640 x 480 pixels can be generated, in which the distances of identified obstacles are coded and memorized. The angular ranges, covered by the depth map, are about 60 degrees (horizontal) and 40 degrees (vertical) in front of the device. Distances between 0.8 and 4.0 meters can be detected with good accuracy. In the case of the use of OpenNI Library, information of distances out of this range are provided with less accuracy. The distance information is coded in the gray scale for the usable detectable distance in the depth map. On the contrary, no-information pixels are generated in all other cases. On the basis of the *Kinect* data, the main purpose of *AudiNect* is to synthesize the proper acoustic feedbacks by means of depth map careful analysis and interpretation. This is made by proper Digital Signal Processing software [17], based on PureData for the sound engine [18], SimpleOpenNI and Open Sound Control OSCp5 Libraries [19]. In this way, the proper acoustic information can be generated, so that they can quite be useful as an aid to the autonomous navigation. *AudiNect* can work in the range from 0.5 m to 4 m.

“Auditory display” is the term used in literature to name the use of the sound output to convey information to the user. The homonymous block in FIGURE 1 is the software component that receives numerical values computed only from visual data, and properly translates them in sound amplitudes, frequencies and timing messages sent to PureData, which is simply used as a tone generator.

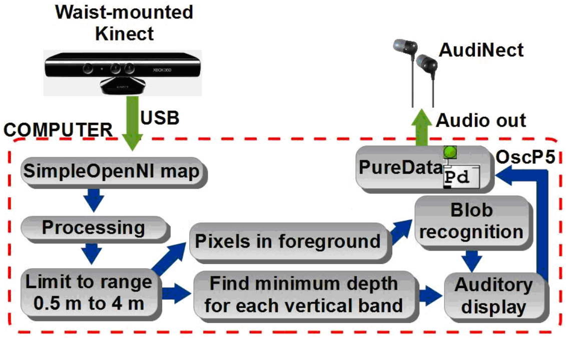

FIGURE 1: Scheme for the Proposed Navigation Prototype.

2.2. Operating Mode

The first step for the depth map processing consists of the generation of two synthesized “images”. Using Simple OpenNI library, the first image is defined on the basis of a static maximum threshold of 4 metres, the second one is derived from the first image, applying to it a dynamic maximum threshold placed at 0.5 metres beyond the pixel with the minimum depth. This pixel will be denoted as the *foreground point*. In addition, at every scan and for each of the two images, three “vertical bands” (left, middle, and right) are defined as a partition of the whole image. In correspondence to every vertical band, the first image is analyzed to identify the minimum depth of all the pixels. Furthermore, the second image is used to identify the significant obstacles, with a process called *blob recognition*. This is the only processing done working on the whole image and not in correspondence to a single band. The blobs are used to evaluate the width occupied by the obstacles. This last global information is further processed and separated in three pieces of information each relative to each band. To this purpose, all the blobs can be used, or else only some of them, as for instance the closest *N*. The blob identification and the distance are analyzed to generate the proper acoustic feedbacks, consisting of sequences of stereophonic pulses and modulated continuous tones. In correspondence to the left and right bands, using only amplitude modulation, proper impulsive sounds are generated, according to the relative obstacle presence in each band. In particular, a single impulsive sound is generated if the obstacle width is less than 33.3% of the band width, two sounds if it is in the range (33.3% - 66.7%), three if it is greater than 66.7%. Pulses of 0.1 seconds are used, adding proper stereophonic effects for the user convenience. The complete pulse train repetition occurs every 1.5 seconds.

Each block in FIGURE 2 has a well determined high-level, object-oriented meaning, which can be implemented using different low-level types. For example, the term “range” here is meant in the mathematical sense; its data structure is just a pair of integers (xMin, xMax).

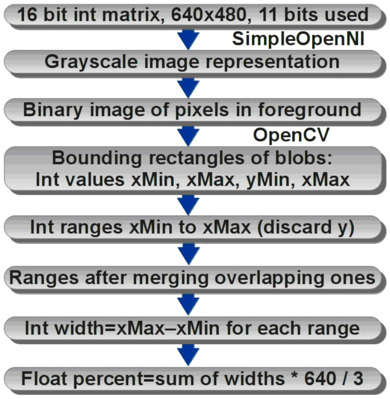

FIGURE 2: All the Intermediate Data Structures Used in Computing of the Width Percentage Occupied By Obstacles

With regard to the central band, a continuous signal is generated, using both frequency and amplitude modulation. The tone frequency is proportional to the width percentage occupied by the obstacles. In particular, when a “clear path” is identified, a high-pitched signal is produced. On the contrary, a low-pitched signal indicates that the free available space becomes lower and, if a significant obstacle is faced, the continuous signal fully stops. Thus, the total signal absence clearly denotes the presence of an obstacle in the frontal direction. In addition, the continuous signal intensity is used to indicate the minimum distance of the closest obstacle in front of the user. The impulsive sound sequence and the continuous signal frequency and amplitude modulation can easily be interpreted by the user to identify the different and usable free paths. As a consequence, it is quite easy to identify the correct path with fewer obstacles. Such a spatialization allows the user to correlate the sounds sequence with each of the three bands.

In summary, the information about the occupied band width and the minimum depth for each band are coded in the acoustic feedback, and helps the user to find the correct path.

It is important to note that, if a wall, or a large obstacle, is located less than 50 cm from the *Kinect*, the number of no-information pixels becomes quite large, within the whole depth map. If this number exceeds a fixed threshold, the acoustic feedback is temporarily stopped and a special tone is produced to indicate the presence of a significant obstacle in front of the user.

In the two graphs, shown in the following figure, the two sound layers (A: lateral; B: central), in the particular case in which the device is stopped and completely hampered, are represented. In order to avoid that central sound pulse could mask lateral sound pulses, the condition $a_2 << a_1$ is required.

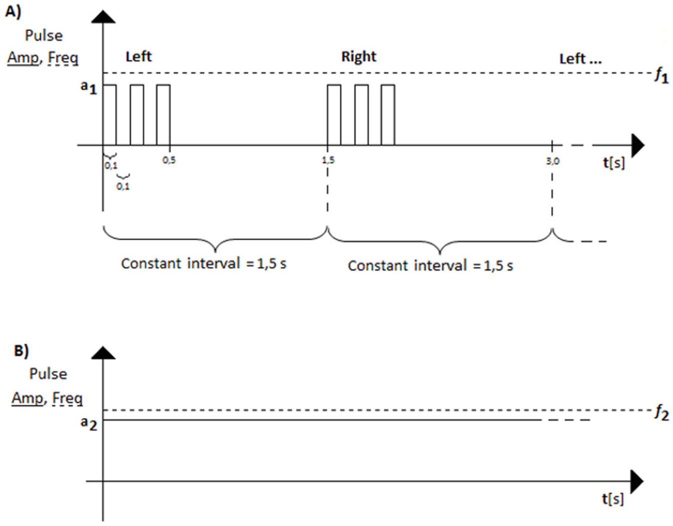

FIGURE 3: Explanation of the Acoustic Feedback Coding With Respect To the Bands.

The proposed approach has been tested on actual scenarios, as it will be shown in the next section. The test analysis appears quite satisfactory, even though some particular critical situations can be observed. The first one concerns the presence of glass surfaces or mirrors; indeed, the *Kinect* source depth map could not be quite correct for the scene representation. This problem can be overcome introducing a proper noise detector software. The second critical issue is that the *Kinect* field of view is supposed not to contain the ground plane. Indeed, in certain cases, this plane could be erroneously detected as an obstacle on the ground. This phenomenon is due to the fact that a proper algorithm for the ground plane identification is not present in the Simple OpenNI software. The use of the Microsoft SDK easily overcomes this limitation.

2.3. Actual Tests For The AudiNect Evaluation

Actual analysis of the *AudiNect* operation mode is presented in this section on the basis of a number of experimental tests, in order to discuss its actual usefulness. To this purpose, two different sets of experiments have been carried out at the University of Rome, “Tor Vergata” (*Master in Sound Engineering* Laboratories). The aim is to compare some actual walking tests on the device, so that the validity of the proposed acoustic feedback can be evaluated. In particular, the learning curves are measured by proper comparison among untrained and trained test people. The walking paths are realized by means of proper obstacle sets, placed in different ways, in a closed room. In the first set, initially untrained people are employed, and the walking tests are analyzed to evaluate the learning behaviour. In the second set, new untrained people are compared with the people already trained in the previous set, in order to make a comparison between these two groups.

a) *First set*. Five properly blindfolded people are equipped with battery-powered *AudiNect* and wireless headphones. Each person is asked to walk in a number of walking paths of similar difficulty (*walking set*). The aim was to measure the walking time behaviour (*learning curve*) in a number of similar difficulty paths, so that the validity of the acoustic guide can be evaluated. The paths have been realized in a room (5.70 m x 4.80 m), in which 13 main obstacles have been placed in random way, as shown in FIGURE 4. In addition, some other smaller obstacles are used. In order to avoid memorization, different paths have been proposed to each testing person. In addition, in order to ensure similar difficulties, the new paths were obtained by flipping the first one with respect to x and y axes. Random path sequences are used to measure the learning curve of each person.

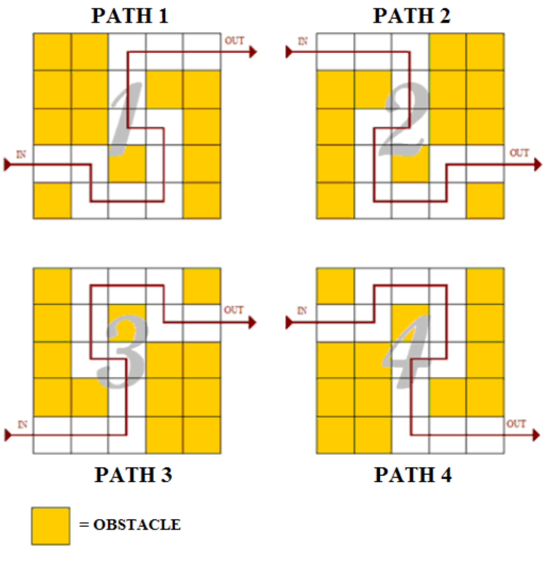

FIGURE 4: The Four Different Paths Proposed.

As the total number of trials increases, the travelling time spent to complete the path decreases, according to the behaviour of FIGURE 5. In particular, on average, a person shows the best improvements in his first trials, as shown in TABLE 1.

Trial Number	Path	Travelling Time (s)
1	3	156
2	2	117
3	1	56
4	4	42
5	2	50
6	3	41

TABLE 1: The Proposed Paths sequence To an Untrained Individual, and the Relatively Exhibited Travelling Times.

These results show the validity of the *AudiNect* approach. Indeed, the acoustic code seems easy to learn.

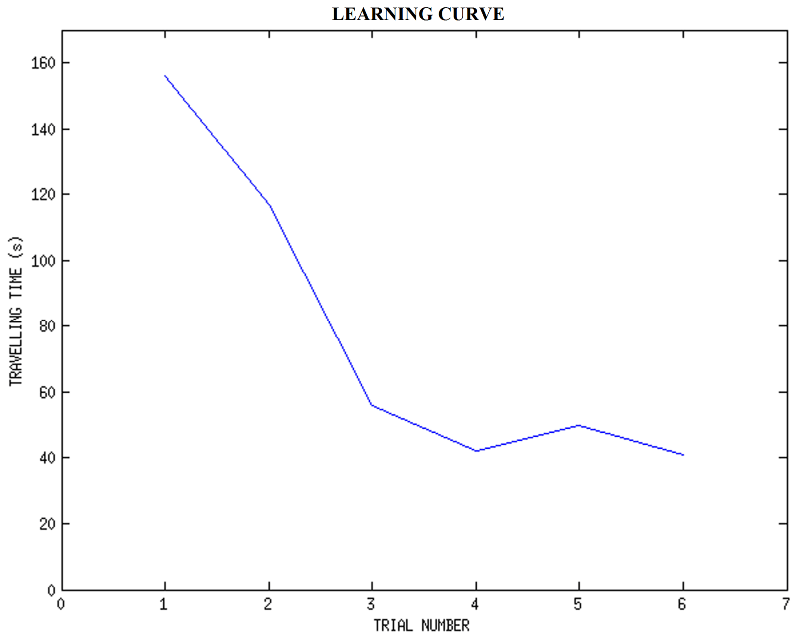

FIGURE 5: Learning Curve Determined By the Trials Performed By the Same Individual

The personal learning times and curves are shown and compared in TABLE 2 and FIGURE 6. They appear quite similar for all the involved test people.

Individual	Travelling Time (s)					
1	156	117	56	42	50	41
2	141	134	76	40	42	47
3	175	132	80	74	48	43
4	161	124	75	74	53	40
5	164	182	70	40	51	68
Trial Number	1	2	3	4	5	6

TABLE 2: Exhibited Travelling Times By the Individuals For Each Trial.

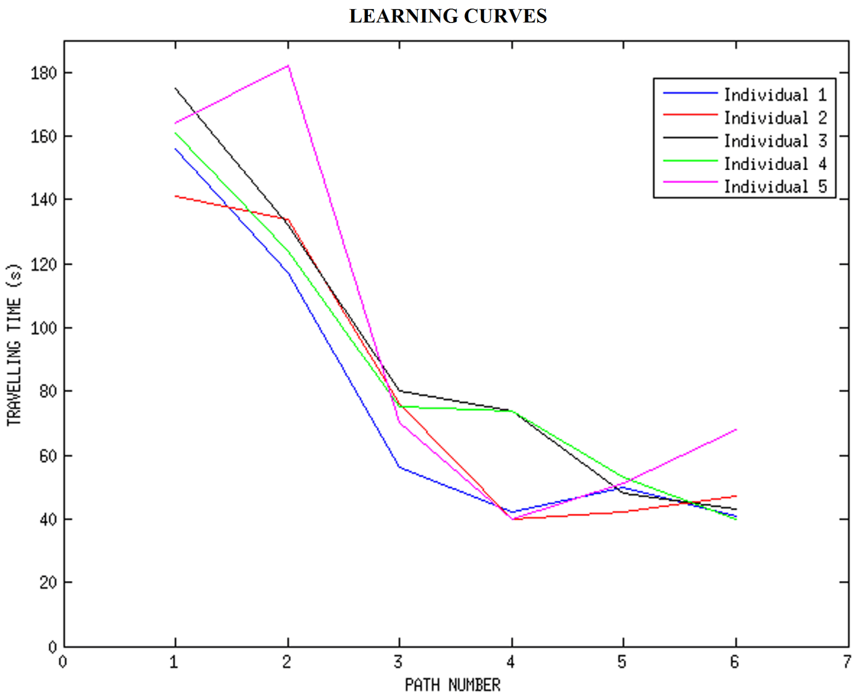

FIGURE 6: Learning Curves Determined By the Trials Performed By Five Individuals.

On the basis of the learning behaviours, the *AudiNect* approach appears quite easy to correctly be added to the human cognitive process. Thus, we can assume that the system can be used, after a little training time, as an assistance to autonomous navigation. It may be applied in the case of visually impaired people, as well as in other applications in which the direct human vision is not possible.

b) *Second set*. The same paths proposed in the case a) are still applied, proposed in a random way again. The main difference is that the second set is devoted to the comparison between untrained and trained people. In particular, people involved in the first set of experiments are now compared with new people. The following results are now obtained.

- Person “A” (already trained) travelling times (TABLE 3).

Trial Number	Path	Travelling Time (s)
1	4	71
2	3	65
3	2	56
4	1	68

TABLE 3: Travelling Times Exhibited By A Trained Person.

- Person “B” (untrained) travelling times (TABLE 4).

Trial Number	Path	Travelling Time (s)
1	1	160
2	2	121
3	3	155
4	4	100

TABLE 4: Travelling Times Exhibited By An Untrained Person.

The data comparison is shown in FIGURE 7. Note that the order of the walks is not the same for both the individuals.

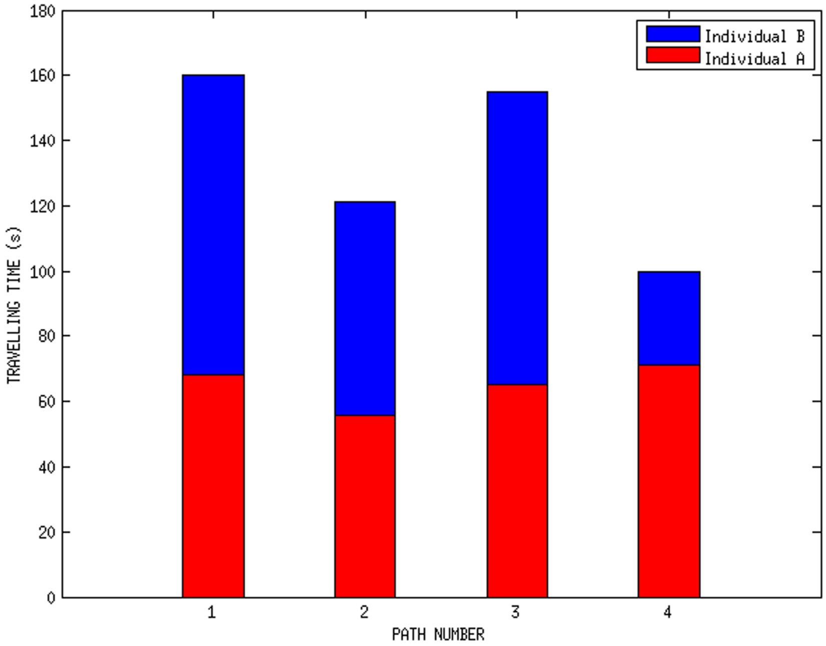

FIGURE 7: Difference Between the Travelling Times Shown By Both Trained and Untrained Individuals.

In particular, it can be noted that the travelling times of trained person "A" are quite smaller than those of the untrained person "B". On the contrary, this last person shows the most remarkable average improvements.

3. CONCLUSIONS

In this paper, a prototype for blind autonomous navigation is presented. As discussed in the introduction, the prototype is able to overcome some limitations of the systems proposed in the technical literature, both in terms of costs and in the terms of quantity of information extractable from the environment. The *AudiNect* project makes use of a recent low cost device, the *Kinect*, in order to obtain useful data related to the surrounding environment. As illustrated in the section 2.2, the data are processed by proper DSP techniques. Through PureData, it is possible to synthesize the acoustic feedback related to the obstacle detection for a given path. The system was tested on blindfolded sighted individuals. The related tests show satisfactory results on the basis of learning curves, showing a rapid adaptation of the individuals to the proposed method. This suggest that the technology is well integrated with the human cognitive processes. Indeed, the *AudiNect* let the user to easily identify the best free path.

In further works we will improve the portability of the system, by making a miniaturization of the latter using a tablet device. This can reduce the overall power consumption, increasing the autonomy of the system. In order to further improve the autonomous navigation capabilities, another future development would be to integrate a device with tactile feedback in the system. Indeed, this could be helpful to reduce the use of acoustic feedback only in critical cases of danger, thus allowing to avoid potential sound masking of the natural sound information from the environment.

4. REFERENCES

[1] K.P. Beier. "Virtual reality: A short introduction." Internet: http://www-vrl.umich.edu/intro/, Nov. 25, 2008 [Nov. 16, 2012].

[2] D. Talaba and A. Amditis. Product engineering, tools and methods based on virtual reality. Dordrecht, Springer, 2008.

[3] A.B. Craig, W. Sherman, J. Will. Developing Virtual Reality Applications: Foundations of Effective Design. Burlington, The Morgan Kaufmann Series in Computer Graphics, 2009.

[4] N.I. Ghali, O. Soluiman, N. El-Bendary, T.M. Nassef, S.A. Ahmed, Y.M. Elbarawy, A.E. Hassanien. "Virtual Reality Technology for Blind and Visual Impaired People: Reviews and Recent Advances," in Advances in Robotics and Virtual Reality, 1st ed., vol. 26, T. Gulrez and A.E. Hassanien, Ed.: Springer, 2012, pp. 363-385.

[5] L.W. Alonzi, D.C. Smith, G.J. Burlak, M. Mirowski. "Radio frequency message apparatus for aiding ambulatory travel of visually impaired persons," U.S. Patent 5 144 294, Sep. 1, 1992.

[6] D. Ross and A. Lightman. "Talking braille: a wireless ubiquitous computing network for orientation and wayfinding," in Proceedings of the seventh international ACM SIGACCESS conference on Computers and accessibility, 2005, pp. 98-105.

[7] J. Souquet, P. Defranould and J. Desbois. "Design of Low-Loss Wide-Band Ultrasonic Transducers for Noninvasive Medical Application." IEEE Tran. Sonics and Ultrasonics, vol. SU-26, no. 2, pp. 75-81, 1979.

[8] A.J. Ali and A.H. Sankar. "Artificial Guidance System for Visually Impaired People." Bonfring International Journal of Man Machine Interface, vol. 2, Special Issue 1, Feb. 2012.

[9] K. Yelamarthi, D. Haas, D. Nielsen, S. Mothersell. "RFID and GPS Integrated Navigation System for the Visually Impaired," in 53rd IEEE International Midwest Symposium on Circuits and Systems, 2010, pp. 1149-1152.

[10] N. Márkus, A. Arató, Z. Juhász, G. Bognár and L. Késmárki. "MOST-NNG: An Accessible GPS Navigation Application Integrated into the MObile Slate Talker (MOST) for the Blind," in Computers Helping People with Special Needs, Lecture Notes in Computer Science, K. Miesenberger, J. Klaus, W. Zagler and A. Karshmer, Ed.: Springer, 2010, pp. 247-254.

[11] M. Bessho, S. Kobayashi, N. Koshizuka, K. Sakamura. "Assisting mobility of the disabled using space-identifying ubiquitous infrastructure," in Proceedings of the tenth international ACM SIGACCESS conference on Computer and accessibility, 2008, pp. 283-284.

[12] U. Biader Ceipidor, E. D'Atri, C.M. Medaglia, A. Serbanati, G. Azzalin, F. Rizzo, M. Sironi, M. Contenti, A. D'Atri. "A RFID System to Help Visually Impaired People in Mobility," presented at the EU RFID Forum, Brussels, Belgium, 2007.

[13] S. Keane, J. Hall, P. Perry. Meet the Kinect: An Introduction to Programming Natural User Interfaces. New York, Apress, 2011.

[14] M. Zöllner, S. Huber, H.C. Jetter, H. Reiterer. "NAVI – A Proof-of-Concept of a Mobile Navigational Aid for Visually Impaired Based on the Microsoft Kinect," in Proceedings of thirteenth IFIP TC13 conference on Human-Computer Interaction, 2011, pp. 584-587.

[15] E. Berdinis and J. Kiske. "Kinecthesia: Using Video Game Technology to Assist the Visually Impaired." Internet: http://www.noblevisiongroup.com/ophthalmic/kinecthesia-using-video-game-technology-to-assist-the-visually-impaired, Nov. 17, 2011 [Nov. 16, 2012].

[16] E. Berdinis and J. Kiske. "Students Hack Kinect to Make the Kinecthesia Personal Radar for the Vision Impaired." Internet: http://au.ibtimes.com/articles/251028/20111116/students-hack-kinect-make-kinecthesia-personal-radar.htm, Nov. 17, 2011 [Nov. 16, 2012].

[17] C. Reas and B. Fry. Processing: a programming handbook for visual designers and artists. Cambridge, The MIT Press, 2007.

[18] Audio Programming Languages: Mathematica, Csound, Max, Pure Data, Supercollider, Comparison of Audio Synthesis Environments. Books LLC, 2010.

[19] R.T. Dean. The Oxford handbook of computer music. New York, Oxford University Press, 2009.