Acoustic echo modeling of people in acoustic arrays using LIDAR

Alberto Izquierdo[1]
Lara del Val[2]
Juan J. Villacorta[3]
Sergio Canseco[4]

Signal Theory and Communications and Telematics Engineering Department. Telecommunication Engineering School. University of Valladolid
Paseo Belén 15, 47011 Valladolid (Spain)

ABSTRACT

In the field of human detection using acoustic arrays, the design of beamforming and detection algorithms is of vital importance. Evidently, the acoustic echo is directly dependent on the ergonomic characteristics of the people, as well as on the clothes they are wearing. Traditional techniques use a large set of people to characterize the system and evaluate the detection and false alarm probabilities. This work proposes a different approach, where a reduced set of people is selected and a cluster of points with their ergonomic data is obtained by means of a 2D LIDAR. From this data and using a classical reflection model, the signals that would be received in an acoustic array are calculated and, using beamforming techniques, the 3D acoustic image is obtained. The work compares these synthesized acoustic images with real acoustic ones.

1. INTRODUCTION

Nowadays, there are a large number of applications in which human detection is required. One of these applications is related to the automotive sector, specifically in the task of detecting pedestrians to prevent the vehicle from colliding with them. Most systems used for pedestrian detection are based on RGB cameras. They work very effectively under adequate visibility conditions, but their performance decreases if the visibility is reduced. So, currently there are many studies trying to solve this problem using other detection systems, such as thermal cameras [1,2], LiDARs [3,4], or an array of microphones [5], or by fusing images obtained from RGB cameras with these other detection systems [6-8], as they can be complementary.

Using acoustic arrays for human detection, it is evident that the acoustic echo is directly dependent on the ergonomic characteristics of the people, as well as on the clothes they are wearing. Traditionally, these techniques use a large set of people to characterize the system and obtain reliable detection and false alarm probabilities related to their performance. But normally, it is difficult to find such a large set of people willing to participate in the corresponding tests.

This work proposes a different approach, where the design of an acoustic image simulator based on the finite elements' method is shown. With this acoustic simulator, 3D acoustic images of different people, with different physical characteristics and wearing different clothes could be simulated, and

---

[1] alberto.izquierdo@uva.es

[2] lara.val@uva.es

[3] juavil@tel.uva.es

[4] sergio.canseco@alumnos.uva.es

the could be used in the test of different designed detection methodologies or even classification algorithms based on machine learning.

This acoustic simulator uses a cluster of points with their ergonomic data, extracted by the images obtained from a set of reduced people by means of a 2D LiDAR camera. From this data, the signals that would be received in an acoustic array are calculated by using a classical reflection model. After that, the 3D acoustic image is obtained by using beamforming techniques. Finally, this work compares the synthesized acoustic images with real acoustic ones.

2. HARDWARE AND SOFTWARE RESOURCES

2.1. Hardware Resources

The system used in this work is composed of two elements: a LIDAR camera that allows obtaining a 3D image of an object by means of a point cloud and an 8x8 active acoustic array of MEMS microphones that obtains acoustic images of these objects.

One of the main hardware resources used was the Intel RealSense L515 LiDAR Camera [9], which can be observed in Figure 1. It is a depth camera designed for use in applications that require depth data captured with high resolution and accuracy. It uses a MEMS sensor that emits an infrared laser beam and captures its reflection in a mirror. Its optimal performance is obtained indoors and over a range of distances form 25 cm to 9 m. The pixels in the obtained depth image have a certain value depending on the distance to the LiDAR camera, instead of presenting color information, as a classic RGB camera obtains.

Figure 1: Intel RealSense L515 LiDAR Camera.

The acoustic imaging generation system that has been employed is based on pulse-echo techniques [10] and consists of three main blocks working together which can be observed in Figure 2. This system is based on:

- Firstly, an acoustic signal is generated by a tweeter at the frequency desired. In fact, several acoustic images of the same individual at different frequencies can be captured, with the idea of having as much information as possible for each person.
- Secondly, the acoustic signal acquisition system consists of a uniform planar Array (UPA) composed of 8x8 MEMS microphones [11], as shown in Figure 1. The microphones are spaced 2.5cm between each other and the squared shape provides the same resolution in both coordinates, azimuth and elevation.
- Finally, a National Instruments myRIO controller [12] interconnects the two previous blocks with the image capture software.

Figure 2: Acoustic acquisition system

2.2. Software Resources

To obtain the real acoustic images, a library of signals in the temporal, frequency and spatial domains has been developed, together with a program that captures the acoustic signals and obtains the 3D acoustic images by means of beamforming techniques. This program allows importing the acoustic signals synthesized by the acoustic model developed in this work. Figure 3 shows the temporal and spatial interface of the VISAM application. The programming tool used is LabVIEW 2020, from National Instruments.
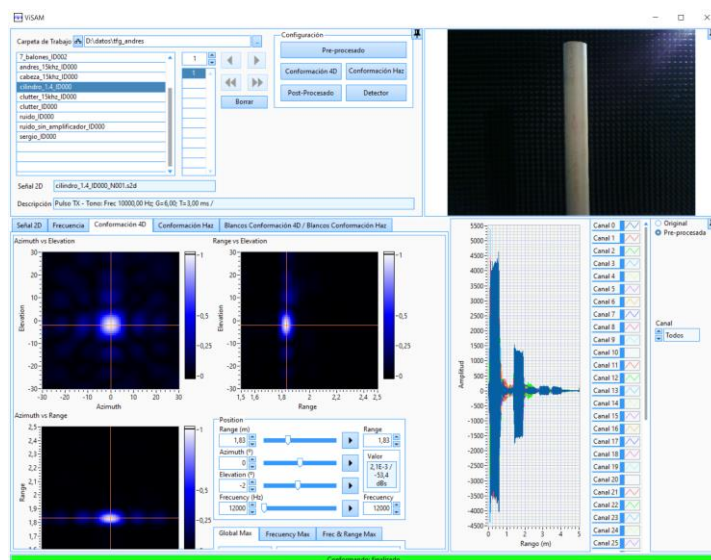

Figure 3: VISAM software user interface

3. ACOUSTIC MODEL OF REFLECTION

Assuming that by finite elements an object can be decomposed as a set of points, as shown in Figure 4, the distance between the loudspeaker and each of the microphones of the array can be estimated, assuming that each of the points in which the object under study is decomposed reflects the transmitted acoustic signal.
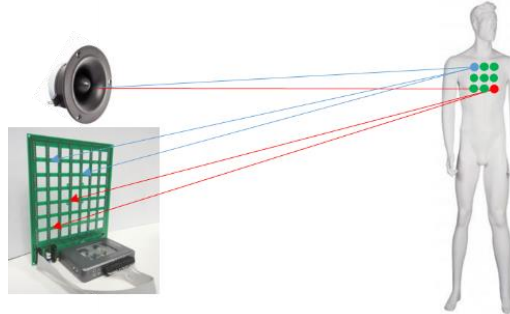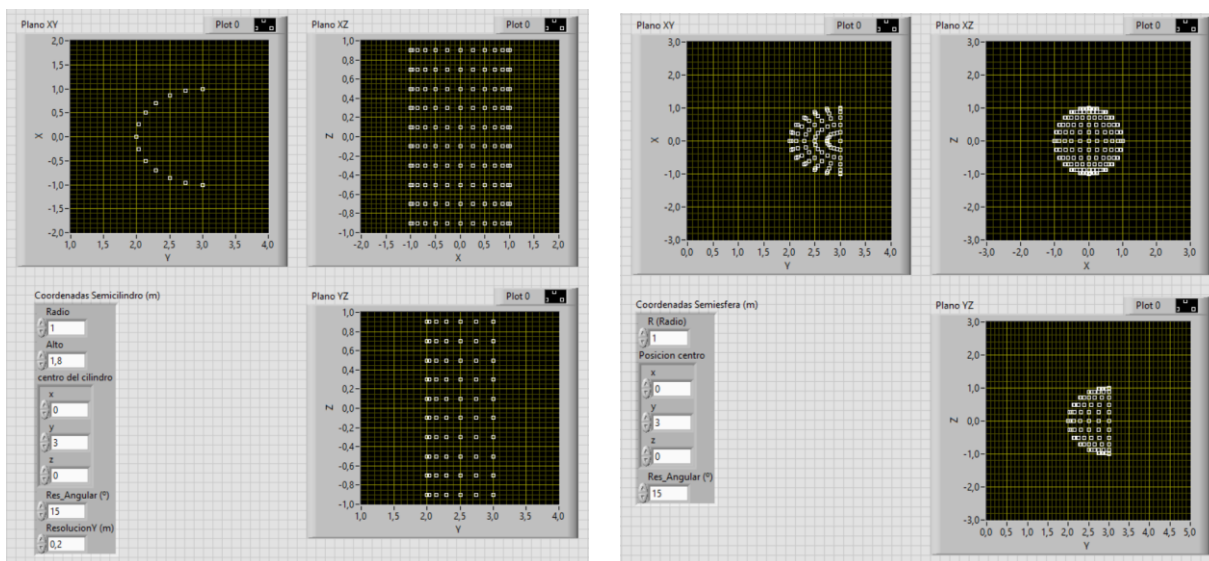
Figure 4. Scheme of the acoustic model of reflection

Two synthesis models have been developed:

a) A theoretical model of simple objects such as a hemisphere and a semi-cylinder, whose point cloud can be observed in Figure 5.

b) A real model of objects from the captured LIDAR point cloud, which is shown in Figure 6, for the case of a sphere, a cylinder and a person.

(a)   (b)

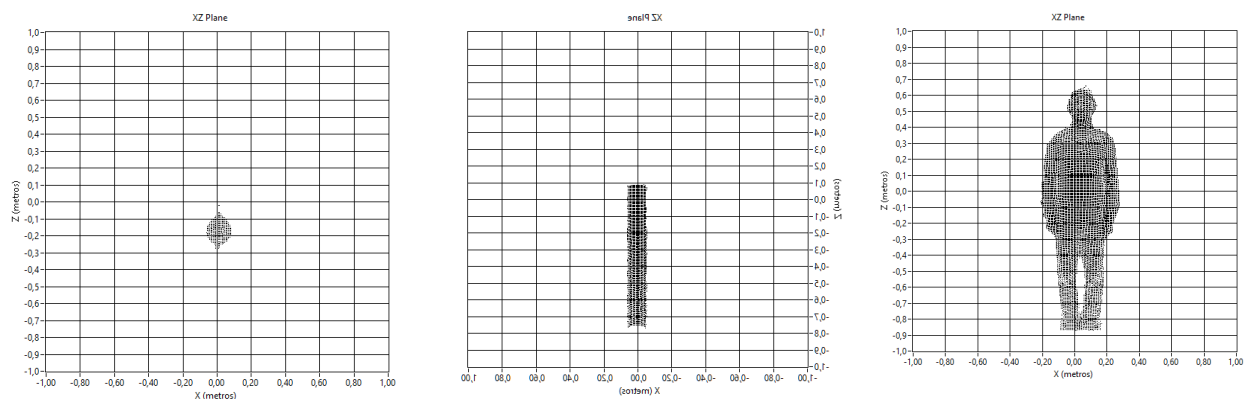Figure 5: Object generator: (a) Semi-cylinder. (b) Hemisphere.

Figure 6: LIDAR images: (a) Hemisphere (b) Semi-cylinder (c) Person.

Based on these geometrical models, the acoustic signal received by the microphone array is simulated assuming a sinusoidal pulse of fixed frequency and duration. Considering that the differential distances between the different points of the object are small, radiation and attenuation losses have not been taken into account.

4. RESULTS

These tests are based on the comparison, for different targets, of the image synthesized from a theoretical model, the acoustic image synthesized from the cloud of LIDAR points obtained from the real target, and the real acoustic image obtained from the acoustic acquisition system based on the MEMS microphone array. The images have been obtained assuming a pulse of 3ms duration and a working frequency of 10kHz. Figure 7 shows the targets used for the analysis.
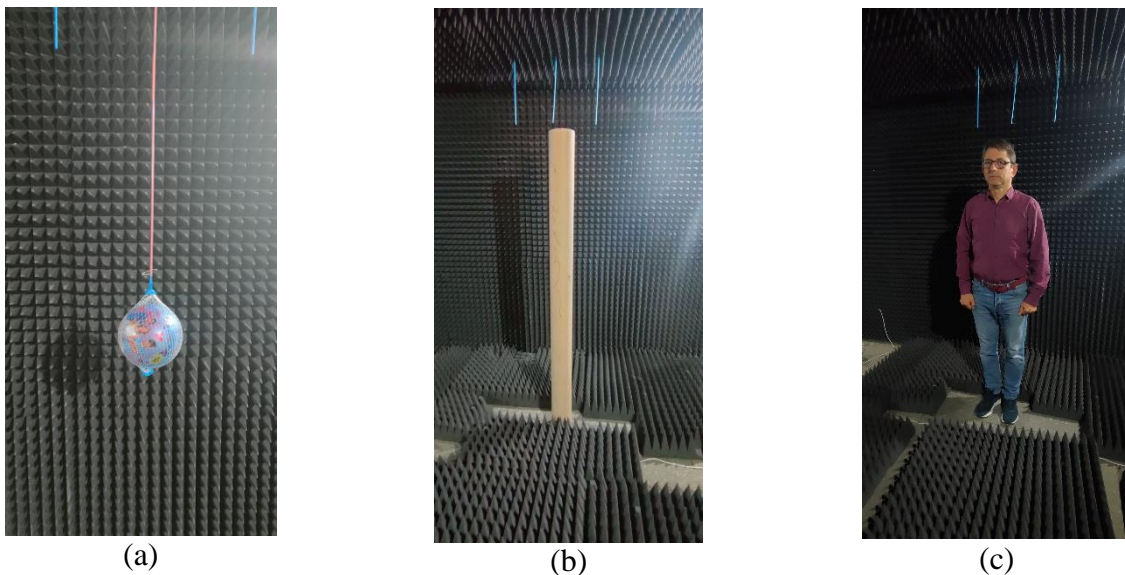

(a)   (b)   (c)

Figure 7: LIDAR images: (a) Hemisphere. (b). Semi-cylinder. (c) Person.

4.1. Simple Objects: Cylinders and Spheres

In the first scenario, the tests were based on a hemisphere, simulated with the developed software, and on the results obtained from a real sphere of 9cm radius and located 1.9m away from the acoustic array, at 0º in azimuth and 0.5º in elevation.

Figure 8 shows the corresponding acoustic images obtained in this first scenario. Figure 8a shows the acoustic image obtained from the real sphere with the acoustic acquisition system; Figure 8b shows the acoustic image obtained from the simulated hemisphere; and Figure 8c shows the acoustic image generated from the cloud of LIDAR points obtained from the real sphere.

Analyzing the images in Figure 8 it can be observed that there is a difference in height in the position of the echo produced by the sphere in Figure 8a compared to Figures 8b and 8c. This difference in height is a consequence of the fact that the LIDAR camera and the MEMS microphone array are not centered, but are separated by a vertical distance of 25cm, as can be seen in Figure 2. Except for this difference, it can be seen how similar the real acoustic image (Figure 8a) is to the image simulated from the LIDAR data (Figure 8c).
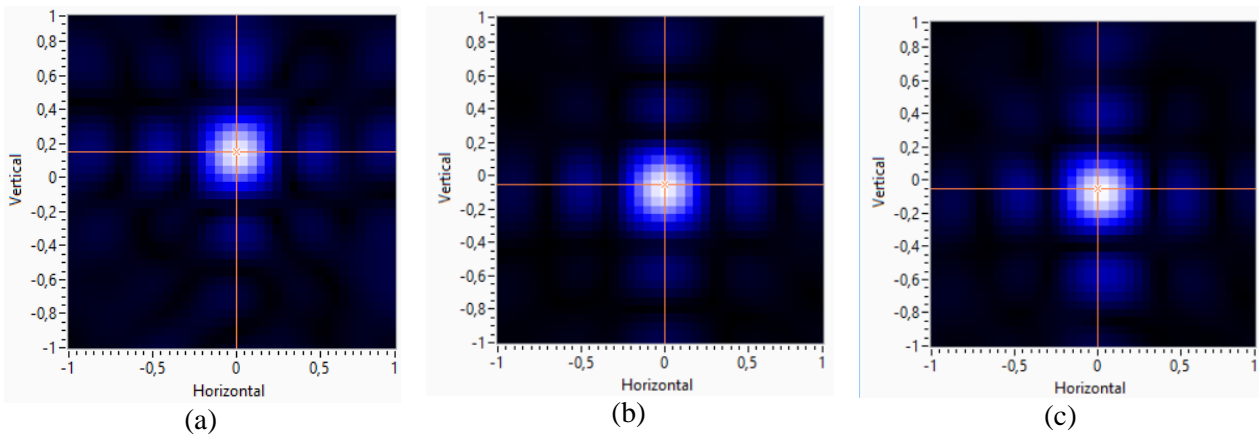
Figure 8: Acoustic Images using a: (a) Real sphere. (b) Synthetized hemisphere. (c) LIDAR points of a real sphere.

In the second scenario we will compare the results obtained from a cylinder of 1.7m high and a radius of 6cm, located 1.4m away from the microphone array and from its corresponding modellings.

Figure 9 shows the corresponding acoustic images obtained in this second scenario. Figure 9a shows the real acoustic image obtained from the cylinder; Figure 9b shows the acoustic image obtained from the simulated cylinder; and Figure 8c shows the acoustic image generated from the cloud of LIDAR points obtained from the real cylinder.

Analyzing the images in Figure 9 it can be also be observed that in this second scenario the modelling is correct. In this case the difference in height in the position of the echo produced by the cylinder is not observed due to the larger vertical dimension of the cylinder compared to the sphere.
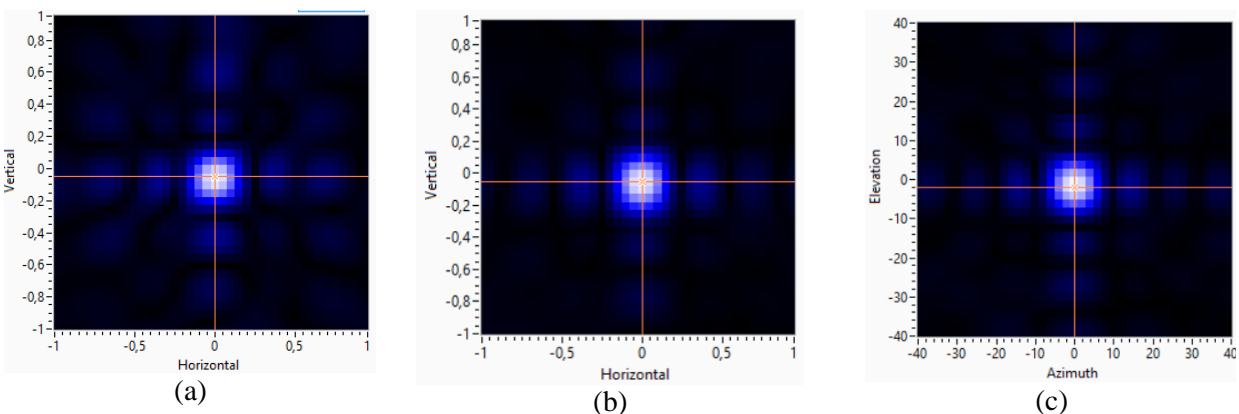

Figure 9: Acoustic Images using a: (a) Real cylinder. (b) Synthetized hemisphere. (c) LIDAR points of a real sphere.

4.2. Complex Objects: People

In the third scenario we will compare the results obtained from a 1,7m high person with his arms close to the body, located 1.5m away from the microphone array.

Figure 10 shows the corresponding acoustic images obtained in this third scenario. Figure 10a shows the real acoustic image obtained from the person; and Figure 10b shows the acoustic image generated from the cloud of LIDAR points obtained from him.
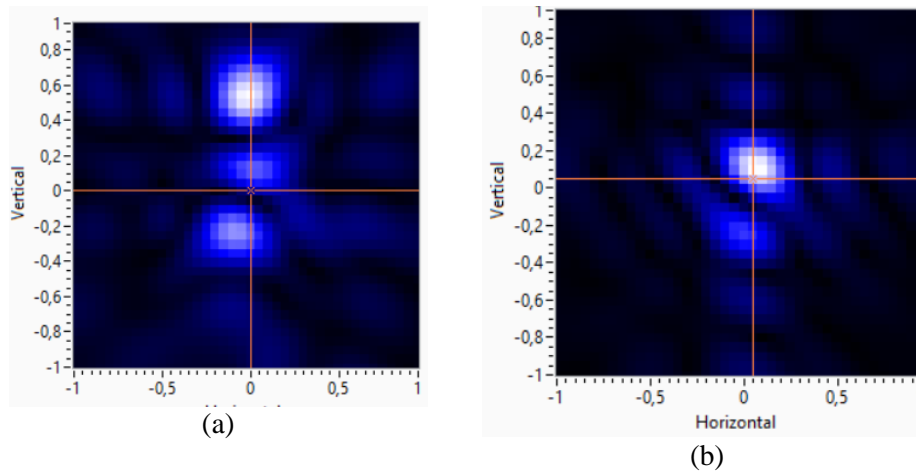
Figure 10: Acoustic Images using a: (a) Real person. (b) LIDAR points of a real person.

In Figure 10, it can be observed that there are 3 targets corresponding to the head, chest and waist of the person in both models, logically the relative amplitude of these targets has not been modelled, since the LIDAR system does not provide information about the acoustic reflectivity of the person. Again there is a slight error in height due to the different position of the LIDAR camera and the acoustic array.

5. CONCLUSIONS

In view of the results, the modelling looks promising, in that it allows for an accurate identification of the main targets into which a person can be broken down. It remains to determine the typical acoustic reflectivity of the different parts of a person in order to incorporate them into the model.

In this way, based on a LIDAR image of a person, it would be possible to emulate the signals generated on any geometry of an acoustic array and for any type of emitted signal, without needing to acquire the acoustic images with the real person. In this way, multi-frequency 3D acoustic images could be obtained by means of beamforming techniques that could be applied in the field of biometric identification, using machine learning techniques.

6. ACKNOWLEDGEMENTS

7. REFERENCES

1. Miclea, R.C.; Dughir, C.; Alexa, F.; Sandru, F.; Silea, I. Laser and LIDAR in a System for Visibility Distance Estimation in Fog Conditions, Sensors 2020, 20(21), 6322.
2. Goodin, C.; Carruth, D.; Doude, M.; Hudson, C. Predicting the Influence of Rain on LIDAR in ADAS, Electronics 2019, 8(1), 89.
3. Piniarski, K.; Pawłowski, P.; Dąbrowski, A. Tuning of Classifiers to Speed-Up Detection of Pedestrians in Infrared Images, Sensors 2020, 20(16), 4363.
4. Kwak, J.; Ko, B.C.; Nam, J.Y. Pedestrian Tracking Using Online Boosted Random Ferns Learning in Far-Infrared Imagery for Safe Driving at Night, IEEE Trans Intell Transp Syst 2017, 18(1), pp. 69-81.

5. Izquierdo, A., Del Val Puente, L. & Villacorta, J.J. Feasibility of Using a MEMS Microphone Array for Pedestrian Detection in an Autonomous Emergency Braking System. Sensors, **21(12), 4162**, (2021).

6. Shopovska, I., Jovanov, L. & Philips, W. Deep Visible and Thermal Image Fusion for Enhanced Pedestrian Visibility, *Sensors*, **19(17)**, 3727 (2019).

7. Wei, P., Cagle, L., Reza, T., Ball, J. & Gafford, J. LiDAR and Camera Detection Fusion in a Real-Time Industrial Multi-Sensor Collision Avoidance System, *Electronics*, **7(6)**, 84 (2018).

8. King, E.A., Tatoglu, A., Iglesias, D. & Matriss, A. Audio-visual based non-line-of-sight sound source localization: A feasibility study, *Applied Acoustics*, **171**, 107674 (2021).

9. Intel RealSense L515 LiDAR Camera. (2022). [Available online]: https://www.intelrealsense.com/download/7691/

10. Skolnik, M.I. *Introduction to RADAR systems*, 3rd Edition, McGraw-Hill Education, 2001.

11. Izquierdo, A., Villacorta, J.J., Del Val Puente, L. & Suárez, L. Design and Evaluation of a Scalable and Reconfigurable Multi-Platform System for Acoustic Imaging. *Sensors*, **16, 1671** (2016).

12. National Instruments. NI myRIO Hardware at a Glance. (2021). [Available online]: http://www.ni.com/product-documentation/14604/en/.