



Parameter extraction of 3D acoustic images using a nonlinear optimization technique

Lara del Val¹
Alberto Izquierdo²
Juan J. Villacorta³
Andrés Martín⁴

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

ABSTRACT

In biometrics, classification techniques are based on parameter extraction from a large data set, such as fingerprints. Specifically, when using 3D acoustic images, it is necessary to extract a set of meaningful parameters. This work assumes that each significant target will have an acoustic image characterized by the 2D radiation pattern of the array and the envelope of the transmission pulse used. Under this assumption the final acoustic image can be synthesized by a linear combination of the significant targets in the scene. The work uses a non-linear optimization algorithm that obtains the parameters (power, range, elevation and azimuth) for each of the significant targets from a 3D acoustic image. Specifically, the algorithm is applied to the parameterization of 3D acoustic images of people, as a prior step to the use of classification algorithms based on machine learning.

1. INTRODUCTION

Using our fingerprint to unlock our mobile phone is a gesture that we can repeat hundreds of times through the same day. Fingerprint is, undoubtedly one of the most common uses of biometrics [1]. There are many elements that can be used for biometric systems, such as the voice [2], the hand geometry [3], or the retina [4], but along with fingerprint, one of the most widely implemented methods is facial recognition [5].

However, because of the global pandemic caused by COVID-19, the use of masks has become an everyday occurrence. This is a hindrance to face recognition as important biometric information is hidden and it decreases the accuracy of the system [6,7]. Besides, studies have shown that the performance of multimodal biometric systems is far superior to single-modal ones [8], so the search for new recognition methods may be a promising idea.

Each of these biometric systems is essentially a pattern recognition system. It consists of extracting, from the raw data, a set of mathematical parameters or features that unequivocally identify that individual [9]. Under the same idea, just as the finger yolk lines uniquely identify a person, the acoustic echoes reflected by each person should be sufficiently representative for their identity as confirmed at previous research [10].

¹ lara.val@uva.es

² alberto.izquierdo@uva.es

³ juavil@tel.uva.es

⁴ andres.martin@alumnos.uva.es



The original input information is usually preprocessed with the aim of transforming the original variable space to a new one significantly reducing the variability within each person class. Thus, the pattern recognition problem is expected to be easier to solve [11].

This is exactly what it is intended to do. The aim of this work is to synthesize the information contained in 3D acoustic images into a set of parameters or targets meaningful to each person as a prior step for a machine learning algorithm. In fact, if training the system with all the information contained in the 3D acoustic images, the computational burden will be unacceptable and consequently, the identification system will be useless.

2. 3D ACOUSTIC IMAGES OF PEOPLE

The acoustic imaging generation system that has been employed is based on pulse-echo techniques [12] and consists of three main blocks working together:

- 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 Knowles SPH0641LU4H-1 MEMS microphones [13], 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 [14] interconnects the two previous blocks with the image capture software.



Figure 1: Acoustic acquisition system

3D acoustic images of people were taken inside an anechoic chamber, in order to reduce reflections from interfering objects that disturb the parametric analysis of the images. Analyzing the obtained acoustic images, 7 or 9 main significant areas can be roughly identified in them. These "areas" are tended to come from the echoes reflected from different parts of a person body: head, shoulders, chest, waist, and knees. In addition, depending on the aperture of the arms, two additional targets can be identified slightly further away from the shoulders. As it is shown in Figure 2, these described "areas/parts of the body" can be observed as independent relative maximum values of the image captured with a higher resolution array. So, the acoustic response of a human body could be represented as the linear combination of the acoustic response of each of these 7 parts/targets defined.

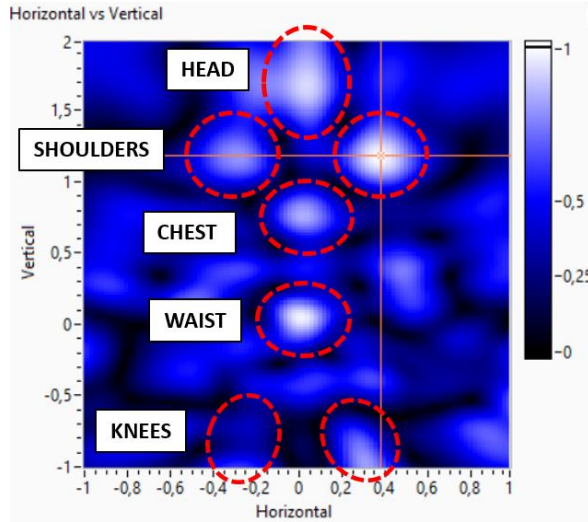


Figure 2: 3D acoustic image acquired from a person (front view).

Intensity and position of each of these areas/targets depend on the person, and this is certainly good news because in this way, it is reasonable to think that intensity and position of each target response suppose sufficiently differentiating parameters.

3. PARAMETER EXTRACTION OF THE 3D ACOUSTIC IMAGES

3.1. Mathematical Parametric Model

Assuming a plane wave with a direction of arrival θ and a linear array with N sensors separated a distance d , the output signal of a beamformer can be characterized by its spatial response which is represented by its beampattern $|F_i(\theta)|$:

$$|F_i(\theta)| = \left| \frac{\sin\left[\frac{N \cdot \omega \cdot d}{2 \cdot c} (\sin \theta_i - \sin \theta)\right]}{\sin\left[\frac{\omega \cdot d}{2 \cdot c} (\sin \theta_i - \sin \theta)\right]} \right| \quad (1)$$

where θ_i represents the steering angle, ω is the working frequency and c the propagation speed of sound [15].

Equation 1 represents the 1D-beampattern of a linear array for a singular target. For the 2D case, the response for each target is the product of the two 1D-beampatterns in both coordinates (Azimuth θ and Elevation φ). The adjustment of the 2D model to a 3D one is somewhat more complicated. For the 3D case, the array response for each target is the product of the 2D-beampattern with the range response. The range response of a target has been represented as a triangular function with its maximum value placed at the range position of the target. This triangular shape is due to the combined effect of the emitted acoustic signal being a rectangular pulse, and the adapted filter of the spatial beamforming.

Finally, it can be assumed that the array response to a human body can be represented as a linear combination of the 3D-beampatterns received separately from each of the 7 or 9 defined targets/parts of the body. So, the parametric model of the acoustic response of a person can be finally built on the basis of the intensity and the position of these 7 or 9 defined targets. From now on, this parametric model will be called $s[\theta, \varphi, r, \rho]$, where ρ is the estimated parameter vector. For the proposed model, the azimuth and elevation parameters are sweep from -30 to 30° , while the range is swept from 150 to 250 centimeters.



3.2. Parameter Estimation Algorithm

Once the mathematical parametric model is defined, the next step of this work can be defined as a target estimation problem. The problem is posed under the classical estimation approach, i.e. assuming that the value of the estimated parameter vector ρ corresponds to an unknown deterministic constant and without knowing a priori information. With this idea in mind, and after studying various estimators, the least squares estimator (LSE) has been considered as a good solution [16].

The idea is to consider that the signal model $s[\theta, \varphi, r, \rho]$ is appropriate. In this way, the rest of the contributions (reflections, etc.) are taken into account only as an imperfection of the model and without assuming anything about them.

Thus, a cost function $J(\rho)$ is defined which results from the sum of the squared differences between the simulated image $x[\theta, \varphi, r]$ and the parametric model $s[\theta, \varphi, r, \rho]$.

$$J(\rho) = \sum_{r=150}^{250} \sum_{\varphi=-30}^{30} \sum_{\theta=-30}^{30} (x[\theta, \varphi, r] - s[\theta, \varphi, r, \rho])^2 \quad (2)$$

This cost function represents the sum of all errors made. Therefore, $J(\rho)$ can be interpreted as a measure of the quality of the calculated parameters since, the more similar they are to those which originated the observation, the more similar $x[\theta, \varphi, r]$ and $s[\theta, \varphi, r, \rho]$ and the lower the value of $J(\rho)$ will be. Therefore, as the model improve, better the estimator will be.

Under this idea, the least squares estimator $\hat{\rho}_{LSE}$ is defined.

$$\hat{\rho}_{LSE} = \arg \min_{\rho} J(\rho) \quad (3)$$

The solution would be quite simple and known for a linear case, however, the model $s[\theta, \varphi, r, \rho]$ is far from being linearly dependent on the parameters to be estimated, so finding the solution is not so simple. In this way, the original estimation problem has been transformed into a nonlinear optimization problem, and it was decided to use the LabVIEW "Constrained Nonlinear Optimization VI" routine [17] to solve it. The routine is based on an iterative optimization Sequential Quadratic Programming (SQP) method, which iterates by using a gradient descent strategy [18].

But the problem is not completely solved in this way, as the optimization algorithm needs to start from appropriate initial parameters in order not to fall into a local minimum of the cost function or give rise to a convergence error. Reasonably, the targets will be located at local maxima of the acoustic image so finding the local maxima for estimating the initial parameters seems to be a wise choice.

The first step in the implemented strategy is to search local maxima. Moreover, for the cases in which the optimization algorithm does not converge to a sufficiently good MSE, the parameters are reset by using a heuristic search method [19]. It is based on a randomized strategy that randomly selects neighboring solutions for a feasible solution and an established neighborhood.

4. RESULTS

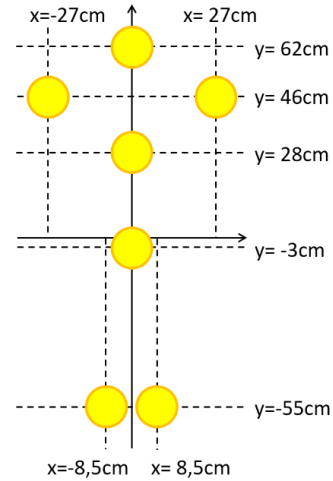
Consistent with the model presented in Section 2, a particular scenario has been simulated, representing the acoustic response of a real person with arms close to the body. Each of the 7 targets simulated represent one of the 7 areas/parts of the body observed in the acoustic imagers. The chosen arrangement is shown in Table 1, and its schematic azimuth-elevation view can be observed in Figure 3.

To analyze the performance of the estimation algorithm, 1000 simulations have been performed on a set of synthetic images placed on the position of the targets decided according to Table 1. The images have been contaminated with different seeds of uniform random noise given by $U(0,0.01)$ and consistent with the noise statistics (mean and standard deviation) observed in real images.

Table 1: Arrangement of the de 7-areas model of a person.

Target	Azimuth	Elevation	Range
Head	0°	18°	194 cm
Left shoulder	-7°	14°	195 cm
Right shoulder	7°	14°	195 cm
Chest	0°	8°	192 cm
Waist	0°	-1°	190 cm
Left knee	-3°	-14°	193 cm
Right knee	3°	-14°	193 cm

Figure 3: Schematic front view of the 7-areas model of a person.



The average MSE obtained for these 1000 simulations is $4,36E-5$, which its value is of the order of noise statistics and the Cramer-Rao bound. The results obtained for each estimation of the target position are shown Figure 4. These results show that the estimation behavior is not the same for the different simulated body parts. It can be observed that the higher errors are associated to the estimation of head and shoulders, and the best estimation is obtained for the waist. It can also be observed that the algorithm shows a better behavior estimating azimuth position of the targets. Figure 5 represents the position estimations of these 7 simulated targets.

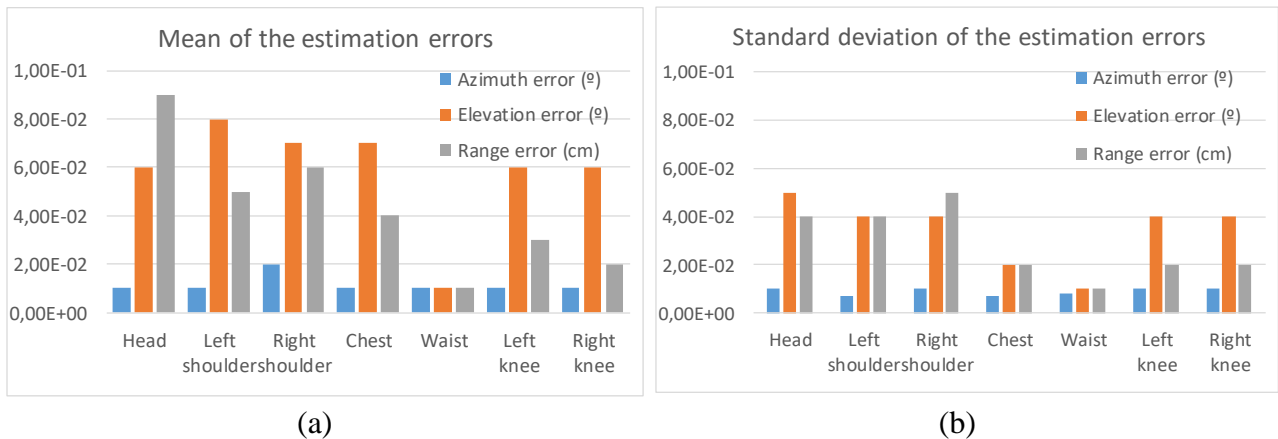


Figure 4: Statistical parameters of the error obtained in each target estimation. (a) Mean. (b) Standard Deviation

Finally, the consistency of the estimation algorithm in a real environment has been evaluated with a real scenario where 7 balls have been hung from the acoustic chamber ceiling, simulating a real person with arms close to the body. Figure 6a shows the 3D acoustic image obtained with the acoustic acquisition system from the 7-balls model, and Figure 6b shows the synthesized 3D acoustic image constructed from the estimated position of these 7 balls. The images are quite similar in terms of MSE. Specifically, the value reached was $3,50E-4$ which seems to be acceptable.

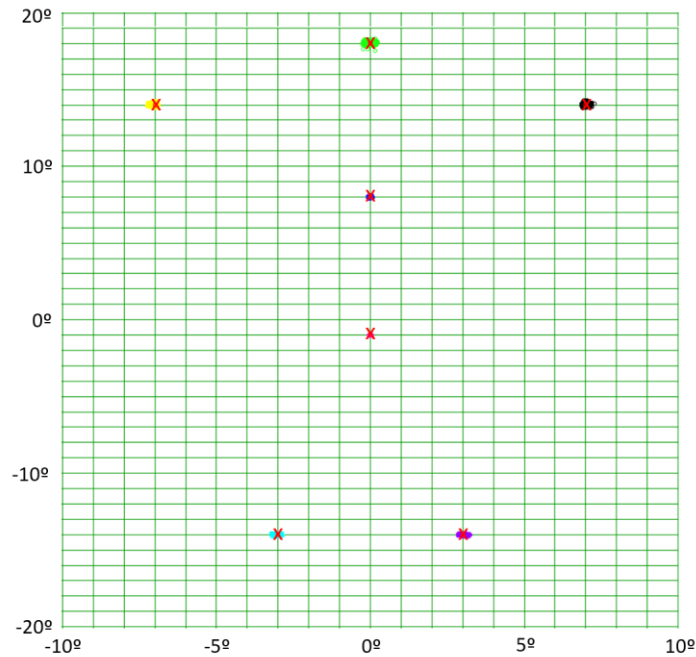


Figure 5: Simulated (red cross) vs Estimated (colored circles) target positions (azimuth-elevation)

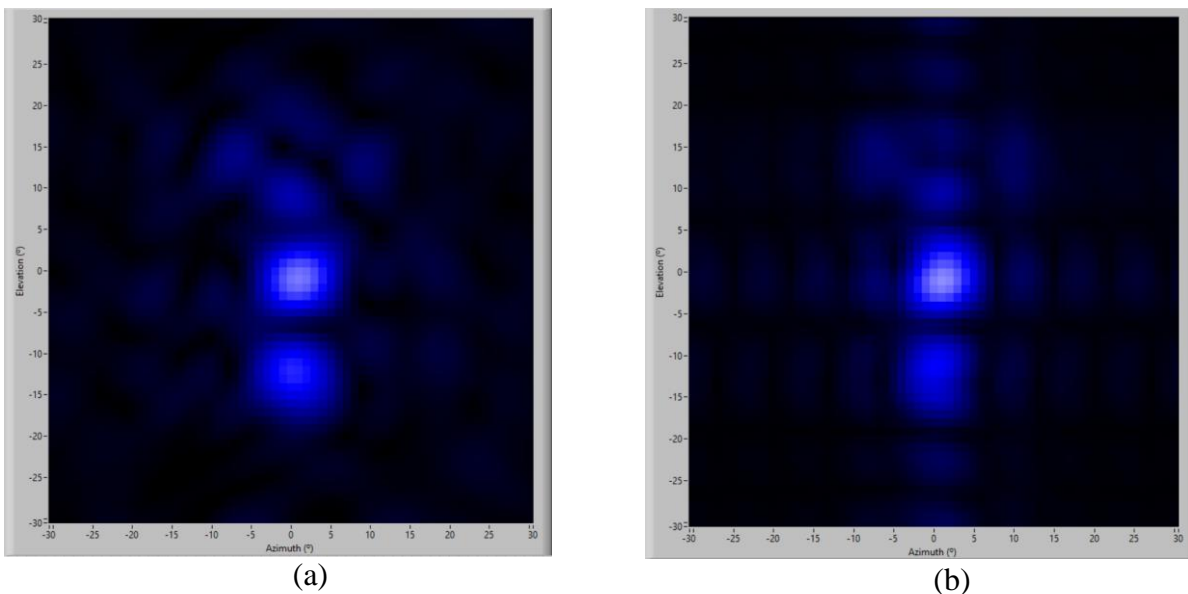


Figure 6: 3D acoustic images (azimuth-elevation). (a) Acquired with the acoustic array from the 7 balls. (b) Synthesized from the estimated position of the 7 balls.

However, when the system estimates the position of the targets/balls, the results have not been as accurate as they were supposed to be, as can be observed in Figure 7. The ball corresponding to the head is not properly detected and the azimuth resolution is not sufficient to discriminate one leg from the other considering them as a single target.

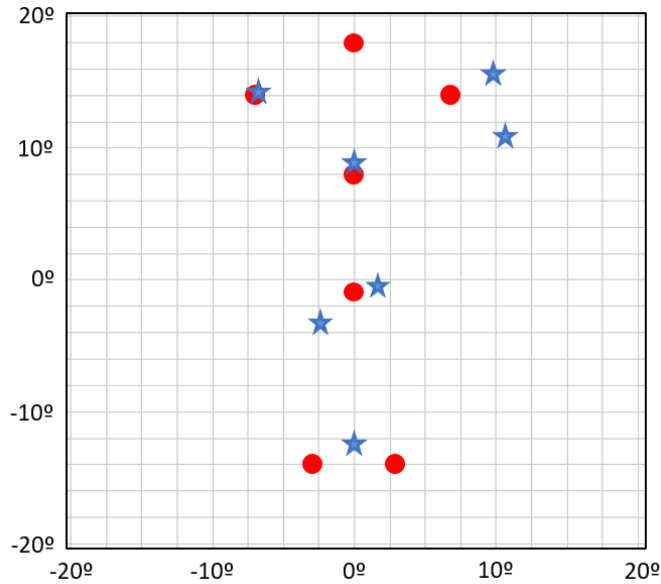


Figure 7: Original (red circle) vs Estimated (blue star) target positions (Azimuth and Elevation)

Looking at the results obtained it can be observed that the designed estimation algorithm works well for the synthetic model. However, for the real images, the result was not as expected. The system converges at a solution where the MSE is lower compared with the solution where the position and amplitude of all targets match with the ones expected, ($3,50E-4$ vs $4,36E-5$ for the case shown). This is a problem since, as mentioned in Section 3, the basis of the LSE estimation method is that the parametric model is considered to be appropriate, so the minimum value of the MSE between the observations and the parametric model would be the one where the targets estimator is most accurate. All this suggests that the parametric model is not as accurate as required.

5. CONCLUSIONS

The work described in this paper suggests that although the proposed parametric model is a good approximation to the problem and it is a good starting point, it is not as accurate as expected. This can be explained due to different reasons. Initially, the proposed model is an ideal model which considers the spatial response of the sensors and the spatial response of the tweeter to be omnidirectional, but in fact this is not correct. For example, the speaker could be radiating more energy at some angles than others and in a different way at certain frequencies. Another reasons could be that the speaker is not centered with the array so the delay in transmitting and receiving the pulses should also be a factor to be considered, or that the assumption that the overall acoustic image can be obtained by summing the response of a set of targets is not correct.

However, as a final conclusion, if the parameters obtained are sufficiently representative of the captured images, which may be reasonable, since the mean square error is not very high, the parameters are probably good enough to be used by a machine learning algorithm. Therefore, future lines of this research could attempt to optimize the parametric model according to the phenomena discussed above, and test it with acoustic images of real people.



6. ACKNOWLEDGEMENTS

This research was funded by Ministerio de Ciencia, Innovación y Universidades, grant number RTI2018-095143-B-C22.

7. REFERENCES

1. Onyesolu, M. & Ezeani, I. ATM Security Using Fingerprint Biometric Identifier: An Investigative Study. *International Journal of Advanced Computer Science and Applications*, **3(4)**, 68-72 (2012).
2. Fegade, S., Chaturvedi, A. & Agarwal, M. Voice Recognition Technology: A Review. *International Journal of Advanced Research in Science, Communication and Technology*, **8(1)**, 31-34 (2021).
3. Faundez-Zanuy, M. Biometric verification of humans by means of hand geometry, *Proceedings of IEEE 39th Annual 2005 International Carnahan Conference on Security Technology*, pp. 61-67, Las Palmas, Spain, 11-14 October 2005.
4. Fahreddin, S. & Selin, U. Biometric Retina Identification Based on Neural Network. *Procedia Computer Science*. **102**, 26-33 (2016).
5. Mohammad S.M. Facial Recognition Technology. *SSRN Electronic Journal*, **2020**, 1-18 (2020).
6. Guo, Y. Impact on Biometric Identification Systems of COVID-19. *Scientific Programming*. **2021**, Article ID 3225687, 1-7 (2021).
7. Song, Z., Nguyen, K., Nguyen, T., Cho, C. & Gao, J. Spartan Face Mask Detection and Facial Recognition System. *Healthcare*, **10 (1)**, 87, 1-24 (2022).
8. Budak, C., Çalışkan, A. & Acar, E. Comparison of Multiple Biometric Identification with a Single Biometric Identification System, *Proceedings of the 1st International Engineering and Technology Symposium*, Batman, Turkey, 3-5 May 2018.
9. Jain, A.K., Ross, A. & Prabhakar, S. An Introduction to Biometric Recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, **14(1)**, 4-20 (2004).
10. Izquierdo, A., Del Val, L., Jiménez, M.I. & Villacorta, J.J. Performance Evaluation of a Biometric System Based on Acoustic Images, *Sensors*, **11(10)**, 9499-9519, (2011).
11. Bishop, C.M. *Pattern Recognition and Machine Learning*, 1st Edition, Springer, 2006.
12. Skolnik, M.I. *Introduction to RADAR systems*, 3rd Edition, McGraw-Hill Education, 2001.
13. 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).
14. National Instruments. NI myRIO Hardware at a Glance. (2021). [Available online]: <http://www.ni.com/product-documentation/14604/en/>.
15. Van Trees, H. L. *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*, Wiley & Sons, 2002.
16. Kay, S. *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice Hall, 1995.
17. National Instruments. Constrained Nonlinear Optimization VI. (2022) [Available online]: https://zone.ni.com/reference/en-XX/help/371361R-01/gmath/constrained_nonlinear_optimization/
18. Hooke, R.H.W. *Optimization Theory. Chapter 4. Sequential Quadratic Programming*. (2006) [Available online]: https://www.math.uh.edu/~rohop/fall_06/
19. Li, J. & Rhinehart, R.R. Heuristic random optimization. *Computers & Chemical Engineering*. **22**. 427-444 (1998).