DOI: http:/dx.doi.org/10.18180/tecciencia.2015.19.7

TECCIENCIA

Acoustic Tracking System for Autonomous Robots Based on TDE and Signal Intensity

Sistema de Rastreo Acústico para Robots Autónomos basado en TDE e Intensidad

Edgar D. Lasso L.[1], Alejandra Patarroyo Sánchez[2], Fredy H. Martínez S.[3]

[1] *Universidad Distrital Francisco José de Caldas, Bogotá, Colombia, edlassol@correo.udistrital.edu.co*
[2] *Universidad Distrital Francisco José de Caldas, Bogotá, Colombia, apatarroyos@correo.udistrital.edu.co*
[3] *Universidad Distrital Francisco José de Caldas, Bogotá, Colombia, fhmartinezs@udistrital.edu.co*

Received: May 25 2015 *Accepted: August 6 2015* *Available Online: Sept 7 2015*

Abstract

This article details the development and evaluation of an autonomous acoustic localization system for robots based on Time Delay Estimation (TDE) and signal intensity, principally aimed at robotic service applications. Time Delay Estimation is carried out through an arrangement of two microphones. The time delay criteria are supported with the signal intensity of a third microphone (coplanar arrangement), which permits discerning precisely the location of the source. This third microphone also feeds a voice identification system, which lets the system respond only to specific voice commands. The prediction algorithm operates by comparing the sensed TDE against the theoretical values of the acoustic propagation model, results that are then weighted according to the signal's mean intensity. A broad set of laboratory experiments is reported on a real prototype that support the system's performance, showing average errors of Azimuth of 18.1 degrees and elevation of 7.6 degrees. Particularly, the analysis conducted for the estimation permits defining the necessary and sufficient conditions to establish in real time a single position in the space of origin, with sufficient precision for autonomous navigation applications

Keywords: Voice identification, acoustic localization, estimated time delay.

Resumen

Este artículo detalla el desarrollo y evaluación de un sistema de localización acústico autónomo para robots basado en TDE e intensidad de la señal, principalmente orientado hacia aplicaciones de robótica de servicios. La estimación del tiempo de retardo se realiza mediante un arreglo de dos micrófonos. El criterio del tiempo de retardo se apoya con la intensidad de la señal de un tercer micrófono (arreglo coplanar) que permite discernir de forma precisa la localización de la fuente. Este tercer micrófono alimenta también un sistema de identificación vocal, que permite que el sistema responda sólo a comandos vocales específicos. El algoritmo de predicción opera comparando el TDE sensado frente a los valores teóricos del modelo de propagación acústica, resultados que luego son ponderados de acuerdo a la intensidad promedio de la señal. Se reporta un amplio conjunto de experimentos en laboratorio sobre un prototipo real que soportan el desempeño del sistema, mostrando errores promedio en azimut de 18.1 grados y de elevación de 7.6 grados. En particular, el análisis desarrollado a partir de la estimación permite definir las condiciones necesarias y suficientes para establecer en tiempo real una posición única en el espacio de origen, con suficiente precisión para aplicaciones de navegación autónoma.

Palabras clave: Identificación vocal, localización acústica, tiempo estimado de retardo

1. Introduction

The design of artificial systems, like robots, involves incorporating systems that permit interaction with the environment, that is, state feedback. One of these possible systems is related to radiated signal sensors and, specifically for robots interacting with human beings, acoustic signal sensors. In human beings, the three-dimensional sensation of the auditory system is related to the difference in amplitude and time in which the signal is received in each ear (Time Difference of Arrival - TDOA) [1]. This means

***Corresponding Author.**
E-mail: edlassol@correo.udistrital.edu.co

How to cite: Lasso L., Edgar D.; Patarroyo S., Alejandra; Martinez S., Fredy H..; ***Acoustic Tracking System for Autonomous Robots Based on TDE and Signal Intensity***, TECCIENCIA, Vol. 10 No. 19., 43-48, 2015, **DOI:** http:/dx.doi.org/10.18180/tecciencia.2015.19.7

that to determine the direction of the origin of sound, it is essential to have information with respect to three elements:

1. Time delay and Hass effect
2. Wavelength
3. Masking

Time delay or Time Difference of Arrival (TDOA) permits inferring the position of a determined sound source according to the difference of intensity, given the different paths established up to the sensor. This strategy has been widely used in autonomous navigation applications [2] [3] [4] [5] [6] [7]. Also, the Haas effect explains how the brain reacts to various sounds, how it chooses a sound, and what happens to the rest. The wavelength is related to the frequency of mechanical oscillation, and the range is determined by the ear's bandwidth (approximately from 20 Hz to 20 kHz in men). Masking is an effect produced when listening to two or more different sounds and the strongest covers the rest making them act as an echo.

Currently, much research works with variants of the TDOA strategy linked to certain design geometries to develop artificial systems of acoustic localization and tracking. A good example is [8] and [9], which carry out a general formulation contemplating an arbitrary number of microphones, and propose to optimization algorithms to solve the non-linear definition of the source estimation. Reference [10] also proposes a system for source localization using a set of three microphones (always over the same plane) geometrically triangulated. With this type of architecture, a robot can estimate the direction and elevation of a source in real time with a minimum error.

Other authors worked on the problem of localization of the sound source, focusing specifically on correction through masking and the Hass effect. An example of this is the work by [11], which uses multiple speakers arbitrarily for the artificial system to recognize a sound source among many others, with all interacting amongst each other. In real environments, many sound sources interact at the same time; for this, Cho, Choi, and Ko [12] implement a set of microphones, a remissive microphone, and some algorithms, with which they manage to find the sound source in a real environment where abundant noisy sound sources exist.

Reference [13] uses sensory microphones, which facilitate localization in environments with noise and less estimation time, allowing the artificial system to respond better upon locating the source. A current application of these sound source localization artificial systems is found in video-conference rooms. This system is important for the location through speaking. An example of this is analyzed in [14], a work that documents a video-conference application that manages to improve the video and solid emission quality. This permits using a limited number of cameras in the video-conference room. By applying this video and audio duo, it is possible to make a map of great acoustic pressure.

The research problem is defined by considering the existence of a sound source located on a navigation environment point, p_S $\in\mathbb{R}^3$ (specific sound source). The signal emitted, *x(t),* is modelled in the environment as a function of intensity over $\mathbb{R}^3$. With m being the signal mapping in the environment in the form m: $\mathbb{R}^3 \longrightarrow [0,1]$ in which m(p) provides the signal intensity in point p$\in\mathbb{R}^3$ of the environment.

The intensity functions, thus, modelled may generally be as complex as those measured in practice for audio or radio signals in different propagation media [15] [16] The signal emitted is detected by both microphones (microphone *a* and microphone *b*) located in points P_a and P_b, respectively, and separated from each other by the distance (d). It is assumed that sound propagation is in straight line without obstacles and constant rate of propagation, so that the time of signal arrival to a point *p*, tm, is described by equation (1)

$$t_m = \frac{\|p_S - p\|}{x(\dot{t})} \tag{1}$$

Where $x(\dot{t})$ is the propagation rate of the signal emitted and the time delay between microphones *a* and *b* is equivalent to:

$$t_{a,b} = t_b - t_a = \frac{\|p_S - p_b\| - \|p_S - p_a\|}{x(\dot{t})} \tag{2}$$

The time delay measured between both microphones is indicated by $\widehat{t_{a,b}}$. In $\mathbb{R}^3$, the equation (2) for different values of p_S corresponds to a hyperboloid, whose rotation axis (symmetry axis) is defined by the line joining points *a* and *b* (**Figure 1** and **Figure 2**). Through construction, each point belonging to the plane perpendicular to the line joining points a and b and which crosses its central point at *d/2* of distance from each point has a value of $t_{a,b} = 0$. This microphone, part of the voice identification system, permits selecting between the front and rear signals.

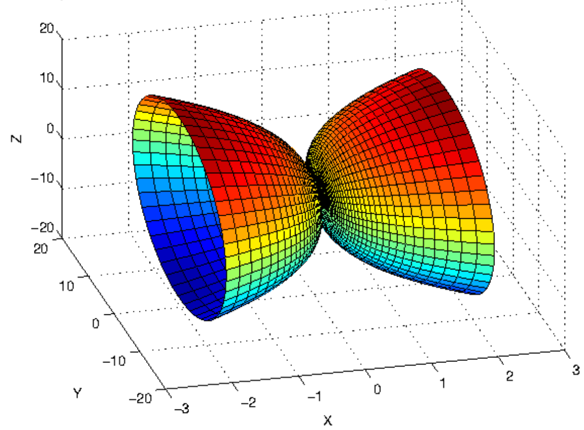

Figure 1 Time delay behavior over $\mathbb{R}^3$ for two microphones separated by a distance, d, and located over the X axis (dimensions in cm).

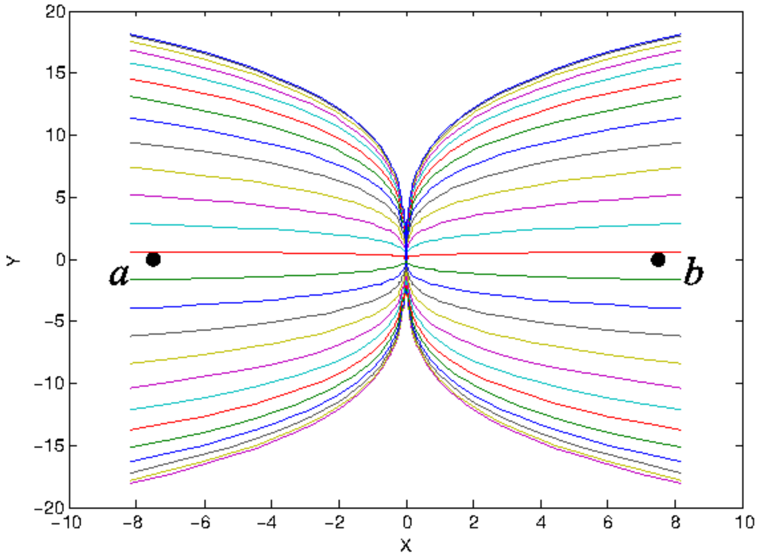

Figure 2 Time delay behavior over the X-Y plane of signals of different intensity for two microphones located over the X axis separated by a distance d = 15 cm (dimensions in cm).

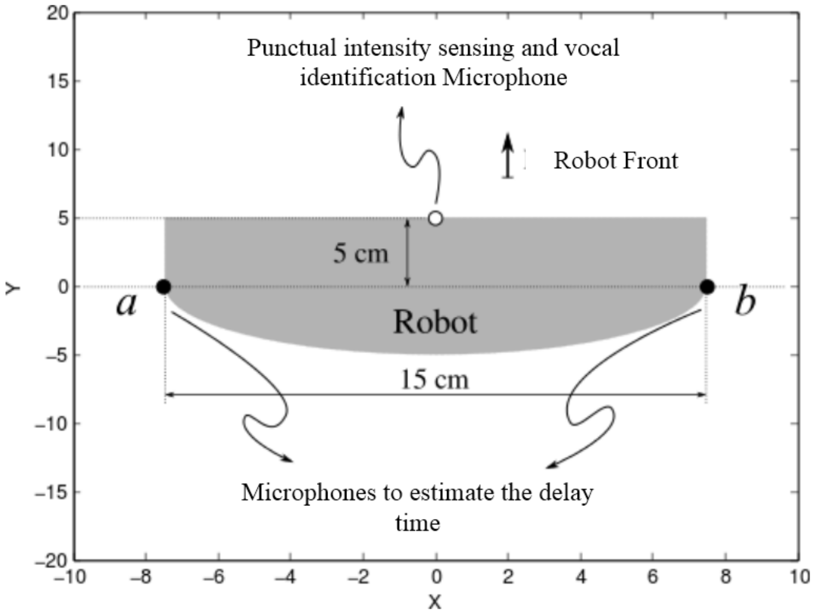
Punctual intensity sensing and vocal identification Microphone
Robot Front
5 cm
Robot
15 cm
a
b
Microphones to estimate the delay time
20
15
10
5
0
−5
−10
−15
−20
−10
−8
−6
−4
−2
0
2
4
6
8
10
Y
X

Figure 3 View of the robotic prototype plant, detailing the location of the microphones to estimate time delay and the voice identification system microphone (dimensions in cm).

This work focused on autonomous robotics applications in dynamic indoor environments; particularly, in robot-human interaction applications that seek to improve robotic response against human commands. A sound source localization system is proposed for the time delay estimation (TDE), using a two-microphone arrangement, as most proposals do for these types of problems.

The scheme is complemented with the estimation of distance from the average intensity of the signals detected and discrimination of the emitting source through the user's voice identification. The localization information is used to coordinate the robot's movements in a way that it simulates paying attention to the human.

2. Methodology

For voice identification, the Easy VR module by Tigal KG was used [17]. This module was programmed with a series of commands in Spanish, like: *Hola* (hi), *Adiós* (bye), *Activar* (activate), *Mírame* (look at me), *Derecha* (right), *Izquierda* (left), *Aléjate* (leave), *Aquí* (here).

These commands are recognized by the module so that it does not respond to other acoustic stimuli.

Upon recognizing the command, it analyzes the acoustic information captured regarding intensity, calculating the TDE ($\widehat{t_{a,b}}$) and defining a possible location of the source. Finally, this information is transmitted to the control unit to activate corresponding movements in the prototype. To evaluate the performance of the functions, a particular identification test was designed and applied

2.1 Acoustic sensor

The acoustic sensor system is an autonomous system, of decentralized operation (it receives on and off signals and provides source localization information and code of the voice order), which is in charge of identifying the user's voice, identifying the voice command, identifying intensity values on the microphones, calculating the TDE of the sound source, and estimating its location.

It incorporates two omnidirectional microphones specific to calculate the TDE (ADMP401 from Analog Devices), a voice recognition unit with its own microphone (Easy VR from Tigal KG), and a low-cost control unit supported by a microcontroller (ATmega 2560 from Atmel) (**Figure 4**).

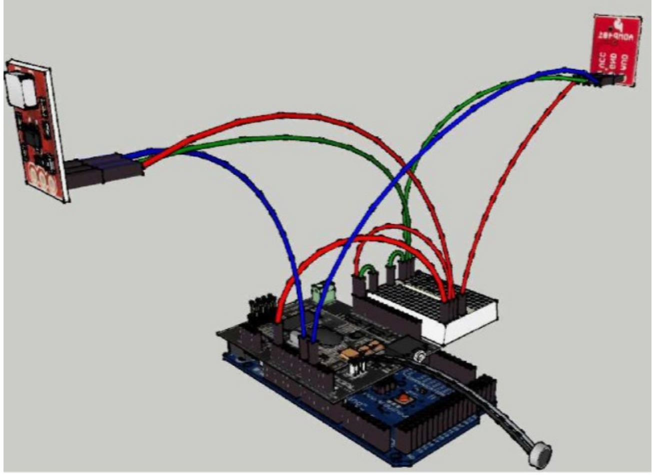

Figure 4 Acoustic sensor scheme. The autonomous operation sensor incorporates two omnidirectional microphones, a voice recognition unit with its own microphone, and a control unit

2.2. Performance test definition

A performance test was proposed to conduct the experiments, which included origin identification, along with grammar and phonetic identification. For origin identification, an algorithm was designed from the signal width detected by the ADMP401 sensors.

After the audio signal is emitted by the source, it is detected by the ADMP401 microphones and a coded signal from 0 to 1023 bits is sent, depending on the audio signal's bandwidth. The control unit receives this information, analyzes and compares it to determine the TDE value.

With the TDE value, the control unit estimates the source localization. To calibrate the system, eight different sounds were used: applause, finger snapping, the words Hola (hi), Aquí (here), A (to), Hey, Derecha (right), and Izquierda (left). For grammar and phonetic identification, the Easy VR module was used, which operated as the system's initial acoustic filter.

The words *Aquí* (here), *Giro* (turn), and *Derecha* (right) were stored in the module, each coming from a different person, with evidently distinct frequency range and voice amplitude. That is, the word "*Aquí* (here)" came from an adult male, the word "*Giro* (turn)" belonged to an adult female and, finally, the word "*Derecha* (right)" was dictated by the voice of a 12-year old child. With this initial filter, the localization process was only conducted in case the voice identification module detected one of the commands stored. If the comparison resulted correct, the module sent a confirmation signal to the control unit, which will begin with the signal comparison and TDE determination process.

Performance evaluation consisted of repetitive tests of the system assessing its consistency in voice assessment and determination of the distance to the source. In each case, and for each command established and stored in the device's internal memory, 30 repetitions were applied with three different voices declared randomly. Only 10 of the repetitions coincided with the real signal. The device does not carry out any action in case of false signals, thus, showing that the signal emitted is different from the signal stored.

A formal evaluation protocol was developed using real data under laboratory settings to validate the prototype's performance. The system's configuration was the same in all cases: a rectangular room 3 m wide by 3.6 m long; in the center of the room, the prototype is localized comprised of two omnidirectional microphones in contraposition separated 15 cm from each other; a voice recognition unit with omnidirectional microphone; and a control unit supported by a microcontroller (ATmega 2560 from Atmel).

Everything was mounted on a robotic head articulated by two servomotors (pan/tilt movement capacity). The sound source (test user voice) was localized in different points of the room at an effective distance measured from the punctual source to the central point of the two microphones. The Azimuth angle, θ, was measured on the plane of the microphones, while the elevation angle, ϕ, is measured with respect to its plane

3. Results

The final summary of the results measured in the laboratory is shown in Table 1. For the final performance evaluation, 30 exercises were carried out with each of the three words, but with only 10% of the orders registered by the voice identification system.

The rate of correct responses during the voice identification process was 100% while angle errors were about 18.1 degrees for Azimuth and 7.6 degrees for elevation.

Table 1 Results of the performance evaluation. The first row shows the average angular error in degrees, the second row corresponds to the standard deviation of the values.

	Summary of results
Average angular error	**18.1**
Standard deviation (σ)	**6.84**

Table 2 and Table 3 show greater detail of the error behavior against the distance of sound source localization and azimuth and elevation angles.

These tables summarize the results of only one type of voice for the evaluation word "Aquí (here)". Note that the average errors for azimuth and elevation are highly dependent on the distance of the sound source to the microphones. They grow considerably (up to 28 degrees in azimuth) when the distance from the sound source is in the order of 10 times the separation between microphones (d).

A similar effect on error was observed when the sound source tried to align with the microphones (when the azimuth angle increases)

Table 2 Errors in azimuth angle for sound source estimation

Distance from the source (m)	-90	-80	-60	-40	-20	0	20	40	60	80	90
0,5	20,6	20,8	18	13,4	8,8	7,8	8,7	13,1	17,5	20,2	21
1	24	23,1	20	15	9,9	8,9	9,7	14,5	19,4	22,4	23,3
1,5	28	27	23,3	17,4	11,5	10,3	11,4	17	22,6	26,1	27,2

Table 3 Errors in elevation angle for sound source estimation

Distance from the source (m)	-90	-80	-60	-40	-20	0	20	40	60	80	90
0,5	7,1	6,8	5,9	5,8	5,6	5	5,5	5,6	5,7	6,6	6,9
1	7,9	7,6	6,6	5,9	5,3	4,7	5,1	5,7	6,4	7,4	7,6
1,5	9,2	9,2	7,7	7,5	7,3	6,5	7,2	7,3	7,4	8,6	8,9

4. Conclusions

This article presented the implementation and evaluation in the laboratory of a localization system of sound sources based on TDE and identification of voice commands. The algorithm was supported by the measurement of acoustic signal intensities at different points and estimated times to the comparison of values.

The comparison permitted estimating the TDE. A third microphone provided additional intensity information for precise sound source localization.

The acoustic propagation model assumed a punctual source of the sound and distance from the sound source to the microphones comparable in magnitude to the separation between them (d). Thinking of human-robot interaction applications, the system included a user voice identification module and commands.

A large number of experiments were conducted to verify the system's performance under different operating conditions, the majority including user-voice identification and commands. The low mean error and standard deviation values obtained permit testing the system's adequate operation.

Acknowledgments

This research was funded by Research Project code 1-72-398-13 from Universidad Distrital Francisco José de Caldas (Bogotá D.C., Colombia), with technical support in prototype development and evaluation from John B. Parada R. and Claudia M. Galvis L., research group members

References

[1] M. Yang, J. D. R., Z. Xiong y J. Chen, «Numerical study of source localization using the TDOA method,» de *Radio Science Meeting (USNC-URSI NRSM), 2014 United States National Committee of URSI National*, Boulder, 2014.

[2] B. Huang, L. Xie y Z. Yang, «TDOA-Based Source Localization With Distance-Dependent Noises,» *IEEE Transactions on Wireless Communications,* vol. 14, nº 1, pp. 468-480, 2014.

[3] W. Panlong, G. Qiang, Z. Xinyu y B. Yuming, «Maneuvering target tracking using passive TDOA measurements,» de *Control Conference (CCC)*, Nanjing, 2014.

[4] S. Zhang, H. Jiang y K. Yang, «Detection and localization for an unknown emitter using TDOA measurements and sparsity of received signals in a synchronized wireless sensor network,» de *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Vancouver, 2013.

[5] S. Hara, D. Anzai, T. Yabu y L. Kyesan, «A Perturbation Analysis on the Performance of TOA and TDOA Localization in Mixed LOS/NLOS Environments,» *IEEE Transactions on Communications,* vol. 61, nº 2, pp. 679-689, 2013.

[6] T. Wang, X. Chen, N. Ge y Y. Pei, «Error analysis and experimental study on indoor UWB TDoA localization with reference tag,» de *9th Asia-Pacific Conference on Communications, APCC.*, Denpasar, 2013.

[7] Y. Bin y K. Martin, «A graph-based approach to assist TDOA based localization,» de *Proceedings of the 8th International Workshop on Multidimensional Systems (nDS)*, Erlangen, 2013.

[8] X. Alameda-Pineda, R. Horaud y B. Mourrain, «The geometry of sound-source localization using non-coplanar microphone arrays,» de *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, 2013.

[9] X. Alameda- Pineda y R. Horaud, «A Geometric Approach to Sound Source Localization from Time-Delay Estimates,» *IEEE/ACM Transactions on Audio, Speech, and Language Processing,* vol. 22, nº 6, pp. 1082-1095, 2014.

[10] H.-Y. Gu y S.-. S. Yang, «A sound-source localization system using three-microphone array and crosspower spectrum phase,» de *International Conference on Machine Learning and Cybernetics (ICMLC)*, Xian, 2012.

[11] N. T. Greene y G. D. Paige, «Influence of sound source width on human sound localization,» de *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, San Diego, 2012.

[12] H. Cho, J. Choi y H. Ko, «Robust sound source localization using a Wiener filter,» de *IEEE 18th Conference on Emerging Technologies & Factory Automation (ETFA)*, Cagliari, 2013.

[13] J. Orchard y Y. Hioka, «Localisation of a sound source in different positions using Kinect sensors,» de *IEEE Sensors Applications Symposium (SAS)*, Queenstown, 2014.

[14] J. Tuma, P. Janecka, M. Vala y L. Richter «Sound Source Localization,» de *13th International Carpathian Control Conference (ICCC)*, High Tatras, 2012.

[15] F. H. Martinez Sarmiento, D. M. Acero Soto y M. Castiblanco Ortiz, «Autonomous navigation strategy for robot swarms using local communication,» *Tecnura,* vol. 18, nº 39, pp. 12-21, 2014.

[16] K. Taylor y S. M. LaValle, «I-Bug: An intensity-based bug algorithm,» de *IEEE International Conference on Robotics and Automation, ICRA '09.*, Kobe, 2009.

[17] VeeaR Easy VR, *User Manual Release 3.3 TIGAL KG,* 2012.