Time-of-Flight Selection for Improved Acoustic Sensor Localization Using Multiple Loudspeakers

Wallace A. Martins, Leonardo O. Nunes, Diego B. Haddad, Luiz W. P. Biscainho, Markus V. S. Lima, Maurício do V. M. da Costa, Bowon Lee.

***Abstract*— This paper proposes a time-of-flight (TOF) selection algorithm for acoustic sensor localization. The proposed technique performs the least-squares (LS) estimation of the position of microphones (sensors) combined with a TOF selection scheme, which allows one to overcome the localization performance inaccuracy induced by reverberation. Indeed, the proposed TOF selection takes reverberation effects into account in order to enable automatic detection and, if necessary, re-estimation of erroneous TOFs. A key feature of the proposal is that all computations can be performed in the sensor nodes. Simulation results show that the proposed algorithm attains high accuracy, is virtually insensitive to reverberation time, and performs satisfactorily even under very low signal-to-noise ratios (SNRs). For instance, it achieves an average position-estimation error of around 0.4 cm for a room with reverberation time of 400 ms and SNR of -10 dB. This mean error is almost two orders of magnitude smaller than that obtained by a standard LS method without TOF selection.**

***Keywords*— Acoustic sensor localization, time-of-flight, least-squares, reverberation, noise, estimation**

I. Introduction

The task of localizing acoustic sources (sound source localization—SSL) within 3-D Euclidean spaces is at the same time very challenging and quite useful in a number of applications [1], [2]. As the name suggests, SSL algorithms focus on finding the active agents of the acoustic environment, namely the sound sources, the role of which can be played by a person, an undesired interferer, a loudspeaker (acoustic actuator), etc.

When compared to the SSL problem, a relatively less studied problem is the so-called acoustic sensor localization (ASL), which can be regarded as the dual of the SSL problem [3], [4]. Indeed, ASL algorithms focus on finding the passive agents of the acoustic environment, namely the acoustic sensors, the role of which can be played by a person carrying some device with built-in microphone, or simply a stand-alone microphone. Precise detection of acoustic sensor position can be useful in several applications such as indoor navigation in public places and precise calibration of microphone arrays.

Among ASL algorithms, those based on adaptations of the generalized cross-correlation (GCC) method are by far the most employed [3]. However, they exhibit high sensitivity to errors in the estimation of some algorithm parameters such as wave-sound propagation speed, loudspeaker positions, and, even more importantly, the time-of-flight (TOF) associated with each pair loudspeaker-microphone. In fact, the TOF estimation procedure is strongly affected by reverberation effects, ubiquitous in indoor applications. As a consequence, TOF estimation errors turn out to be the main source of performance degradation in ASL algorithms.

The first goal of this paper is to present a mathematical framework which allows one to estimate the positions of acoustic sensors by adapting an ASL algorithm from the literature [3], namely by taking the loudspeaker positions as known a priori. The paper also aims to propose a new algorithm that allows one to select which TOF estimates best match the available data (wave-sound propagation speed and loudspeaker positions) and the resulting position estimation. The proposed algorithm uses signals acquired by a single microphone, thus allowing its execution at the sensor node itself.

Notation: The symbols $\mathbb{R}$ and $\mathbb{N}$ denote the field of real numbers and the set of natural numbers, respectively. The set of non-negative real numbers is represented by $\mathbb{R}_+$. Vectors and matrices are denoted by lowercase and uppercase boldface letters, respectively. The Euclidean norm is denoted by $\|\cdot\|$, $\mathbf{I}_N$ represents the $N \times N$ identity matrix, and the transpose and pseudo-inversion operations are denoted as $(\cdot)^T$ and $(\cdot)^\dagger$.

Organization: Sections II and III describe the fundamentals of ASL considering perfect knowledge of both the loudspeaker positions and the TOFs. Section III also describes an adaptation of the algorithm proposed in [3], whereas Section IV adjusts such algorithm to realistic problems in which the TOFs are unknown, requiring their estimation. Section V addresses the problems that may appear in the TOF estimation and proposes a way of circumventing them: the TOF-selection algorithm. Simulation results are shown in Section VI and the conclusions are drawn in Section VII.

II. Preliminary Definitions

The sound propagation model employed in this work assumes that the wave-sound travels from the loudspeaker to the microphone through multiple paths due to reverberation effects induced by surfaces of objects/obstacles present in the environment. Hence, the reverberated signal received at each microphone can be modeled as a filtered version of the emitted signal. It will be assumed that the signal which follows

W. A. Martins, L. O. Nunes, D. B. Haddad, L. W. P. Biscainho, M. V. S. Lima, and M. do V. M. da Costa are with Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brazil (e-mails: wallace.martins@ieee.org, lonnes@lps.ufrj.br, diegohaddad@gmail.com, wagner@lps.ufrj.br, markus.lima@ieee.org, mauricio.costa@lps.ufrj.br).

W. A. Martins is also with PPEEL/CEFET-RJ.

B. Lee is with the Networking and Mobility Lab, Hewlett-Packard Laboratories, Palo Alto, CA 94304 USA (e-mail:bowon.lee@hp.com).

the shortest path between loudspeaker and microphone, called line-of-sight (LOS), is present in all acquired signals. The time the wave-sound takes for traveling this path, called time-of-flight (TOF), is related to the distance between microphone and loudspeaker through the (assumed known) sound speed $c \in \mathbb{R}_+$.

Mathematically, by defining the 3-D spatial position of a single microphone as $\mathbf{m} \in \mathbb{R}^{3\times 1}$, and by also considering that the 3-D spatial position of the punctual sth sound source is $\mathbf{s}_s \in \mathbb{R}^{3\times 1}$, with $s \in \mathcal{S} \triangleq \{1, 2, \cdots, S\}$, in which $S \in \mathbb{N}$ denotes the number of sound sources, then one can compute the TOF, $t_s \in \mathbb{R}_+$, between the microphone and the sth loudspeaker as

$$t_s \triangleq \frac{\|\mathbf{m} - \mathbf{s}_s\|}{c}. \tag{1}$$

Observe that for ASL problems in which one can assume perfect knowledge about the loudspeaker positions (i.e. $\mathbf{s}_s$ is a known vector for all $s \in \mathcal{S}$), then the TOF t_s defines a sphere in the Euclidean space, whose center is at position $\mathbf{s}_s$ and whose radius is ct_s. All points on the surface of such a sphere (including $\mathbf{m}$) are possible candidates to the location of the microphone. Since there are infinitely many position candidates, the only way to estimate the microphone position employing TOFs is to combine the information from several loudspeakers, as described in Section III.

III. Ideal Determination of Sensor Positions

The authors in [3] proposed a method to solve the problem of automatic position calibration of multiple microphones and loudspeakers. The inputs for this algorithm are the signals emitted by each loudspeaker as well as the recorded signals received at each microphone. The outputs of the algorithm are, up to rotations and translations, the estimated positions of all agents of interest (nodes) of the acoustic environment, i.e. the locations of loudspeakers and microphones. The algorithm has access to all nodes, thus enabling centralized processing.

This section presents an adaptation of the algorithm proposed in [3] for solving exclusively ASL problems, assuming that the loudspeaker positions have been measured off-line and processing can be implemented locally at the sensors. Hence, all the related formulation will be developed for a single microphone without loss of generality, since it can be extended to multiple microphones by independently applying the algorithm to each microphone.

Assuming an ideal setup in which position and TOFs are known for every loudspeaker, the TOFs t_i and t_j between the microphone and two distinct loudspeakers respectively indexed by $i, j \in \mathcal{S}$ can be written as (see Eq. (1))

$$(\mathbf{s}_i - \mathbf{s}_j)^T \mathbf{m} = \frac{\|\mathbf{s}_i\|^2 - \|\mathbf{s}_j\|^2 - (t_i^2 - t_j^2)c^2}{2} \triangleq b_{i,j}. \tag{2}$$

Hence, for a given microphone, each pair of loudspeakers gives rise to an equation with three unknowns—the entries of vector $\mathbf{m}$. Considering S is the number of loudspeakers, then there are $\frac{S(S-1)}{2}$ distinct loudspeaker pairs, which eventually means one can stack $\frac{S(S-1)}{2}$ expressions like Eq. (2) as

$$\underbrace{\begin{bmatrix} (\mathbf{s}_1 - \mathbf{s}_2)^T \\ (\mathbf{s}_1 - \mathbf{s}_3)^T \\ \vdots \\ (\mathbf{s}_{S-1} - \mathbf{s}_S)^T \end{bmatrix}}_{\triangleq \mathbf{S} \in \mathbb{R}^{[S(S-1)/2]\times 3}} \mathbf{m} = \underbrace{\begin{bmatrix} b_{1,2} \\ b_{1,3} \\ \vdots \\ b_{S-1,S} \end{bmatrix}}_{\triangleq \mathbf{b} \in \mathbb{R}^{[S(S-1)/2]\times 1}} \Leftrightarrow \mathbf{S}\mathbf{m} = \mathbf{b}, \tag{3}$$

where both matrix $\mathbf{S}$ and vector $\mathbf{b}$ are known.

Note that if one has $S \geq 4$ (so[1] that the number of rows of matrix $\mathbf{S}$ is at least 6), and also $\mathbf{S}$ has full column rank (the number of columns of matrix $\mathbf{S}$ is 3), then one can exactly compute the microphone position using a pseudo-inverse $\mathbf{S}^\dagger \in \mathbb{R}^{3\times[S(S-1)/2]}$, as follows:

$$\mathbf{m} = \underbrace{(\mathbf{S}^T\mathbf{S})^{-1}\mathbf{S}^T}_{\triangleq \mathbf{S}^\dagger} \mathbf{b} = \mathbf{S}^\dagger \mathbf{b}. \tag{4}$$

Although the former exact computation of the microphone location is mathematically sound, it unfortunately takes advantage of the unrealistic assumption of perfect knowledge of the variables involved in the process. In practice, this is not true at all: an example of that is the uncertainty inherent to TOF estimations. Indeed, TOF t_s associated with any loudspeaker s should be replaced with its corresponding estimate $\hat{t}_s \in \mathbb{R}_+$. Section IV addresses this issue.

IV. Linear Estimation of Sensor Positions

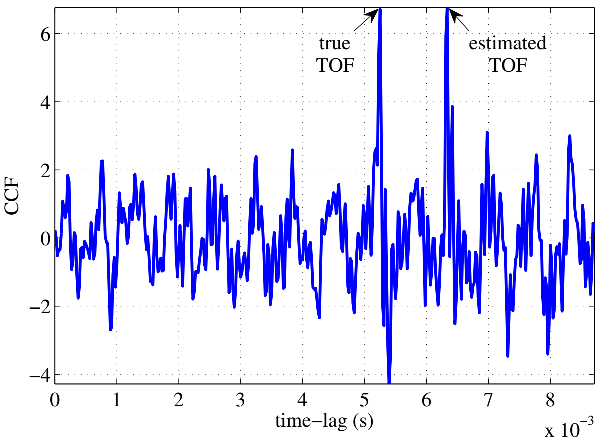
As mentioned before, given a group of loudspeakers, the localization of a single microphone must necessarily rely on estimates of their related TOFs. This paper assumes that, in addition to the signal received by the microphone, the emitted signal is available as well. Hence, the challenge is to estimate the LOS delay that the emitted signal suffered before reaching the microphone. Acoustic effects like noise and reverberation are the main challenges this task is bound to face.

There are many different ways to estimate TOFs associated with wave-sounds. This paper will address the conceptually and computationally simple, yet very effective, cross-correlation (CC) method, whose idea is to project the acquired signal into delayed versions of the emitted signal in order to find the best match between those signals. Indeed, in the absence of any signal degradation, the delay associated with the maximum peak of the resulting CC function (CCF) should indicate the desired TOF.

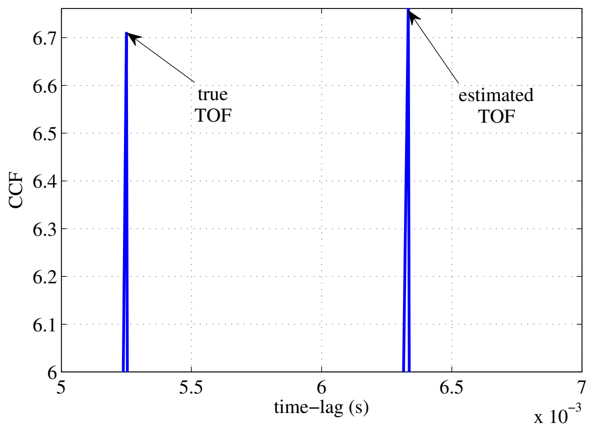
Mathematically, assume a discrete-time model whose related sampling frequency is $F \in \mathbb{R}_+$. Given the real-valued discrete-time signals $x_s[n]$ and $y[n]$, which denote the signal emitted by the sth loudspeaker and the signal acquired by the microphone, respectively, then their corresponding CCF $R_s[\cdot] : \mathbb{N} \longrightarrow \mathbb{R}$ is given as

$$R_s[k] \triangleq \sum_{n\in\mathbb{N}} y[n]x_s[n+k], \tag{5}$$

[1]Note that, if $S = 3$, then the last row of matrix $\mathbf{S}$ would be the difference between the first two rows, implying that $\mathbf{S}$ would not have full rank.



(a) CCF



(b) Zoom

Fig. 1. Typical example of CCF in which the largest peak is associated with reverberation. The true TOF actually corresponds to the second largest peak.

which can be efficiently computed in the frequency-domain. Hence, the TOF estimate that will substitute for t_s in Eq. (3) is

$$\hat{t}_s \triangleq \left(\arg\max_{k \in \mathbb{N}} \{R_s[k]\} \right) / F\,. \tag{6}$$

It is worth noting that $\hat{t}_{ms}$ is usually different from t_{ms} for, at least, three main reasons: (i) time sampling; (ii) measurement noise; and (iii) reverberation effects, which occur in most of the applications of interest. Ultimately, the latter becomes the most important source of errors in the sensor localization process (due to the deviations it produces in TOF estimations).

By assuming estimated instead of exact TOFs, Eq. (3) can be modified as follows:

$$\mathbf{S}\mathbf{m} \approx \begin{bmatrix} \hat{b}_{1,2} & \hat{b}_{1,3} & \cdots & \hat{b}_{S-1,S} \end{bmatrix}^T = \hat{\mathbf{b}}, \tag{7}$$

where, for each pair $(i, j) \in \mathcal{S}^2$ such that $i \neq j$, the element $\hat{b}_{i,j}$ is given by

$$\hat{b}_{i,j} \triangleq \frac{\|\mathbf{s}_i\|^2 - \|\mathbf{s}_j\|^2 - (\hat{t}_i^2 - \hat{t}_j^2)c^2}{2}. \tag{8}$$

The notation "$\approx$" which appears in Eq. (7) denotes the fact that in general vector $\hat{\mathbf{b}}$ might not be in the column space of matrix $\mathbf{S}$ due, among others things, to possible errors in the estimation of the related TOFs. Therefore, from a deterministic and linear point of view, one can consider a least-squares (LS) problem given by

$$\hat{\mathbf{m}}^{\mathrm{LS}} \triangleq \arg\min_{\mathbf{m} \in \mathbb{R}^{3\times 1}} \{\|\mathbf{S}\mathbf{m} - \hat{\mathbf{b}}\|\}, \tag{9}$$

whose solution is

$$\hat{\mathbf{m}}^{\mathrm{LS}} = \mathbf{S}^{\dagger}\hat{\mathbf{b}}. \tag{10}$$

Hence, the least-square solution is equal to the exact solution when no error is present. Unfortunately, as will be discussed in the next section, in reverberant environments this solution can be strongly impaired by large TOF errors.

V. MAIN CONTRIBUTION: TOF SELECTION

As shown in Eq. (6), TOFs are estimated by finding the largest peak of their corresponding CCFs. The underlying assumption is that the CCF exhibits its maximum value when the source and received signals are time aligned, which usually happens when one of them is advanced or delayed by a value corresponding to the true TOF t_s. However, in practical applications where moderate to severe reverberation is present, reflections of the source signal may add constructively with each other, thus generating other peaks in the CCF which can be larger than the peak associated with the LOS path. Fig. 1 depicts a practical example of CCF in which this phenomenon occurs. In this case, the CCF was computed using signals acquired in a real reverberant scenario. Even if the kind of CCF illustrated in Fig. 1 occurs for a very few loudspeakers, the induced TOF-estimation errors may be enough to completely deteriorate the estimate of the related microphone position. It is worth mentioning that this type of errors is not random in the sense that if the experiment is repeated using the same environment and microphone/loudspeaker positions, one would observe the same TOF-estimation error. Hence, in order to mitigate such errors, new algorithms must be conceived.

In this section, a TOF-selection scheme aiming at increasing the robustness of the TOF estimates against reverberation is presented. Such scheme requires the knowledge of source positions. Firstly, the TOF corresponding to each loudspeaker is computed by finding the time-lag associated with the largest peak of the CCF, as in Eq. (6). These TOF estimates are used to form vector $\hat{\mathbf{b}}$, as in Eq. (8), and then the initial microphone positions are estimated via the LS method yielding $\hat{\mathbf{m}}^{(0)\mathrm{LS}}$, as in Eq. (10). Then, denoting as $\bar{t}_s$ the exact TOFs that would be observed if the microphones were indeed at the estimated positions, i.e.,

$$\bar{t}_s \triangleq \frac{\|\hat{\mathbf{m}}^{(0)\mathrm{LS}} - \mathbf{s}_s\|}{c}, \tag{11}$$

the error between the TOF estimate $\hat{t}_s$ and $\bar{t}_s$ is

$$e_s \triangleq \bar{t}_s - \hat{t}_s. \tag{12}$$

Note that $\bar{t}_s$ depends indirectly on the $\hat{t}_s$. If all TOFs are correctly estimated, then $e_s \approx 0$, for all s. The idea of the algorithm is to explore this fact, in order to verify if one or more TOFs have been inconsistent[2] relative to the remaining TOFs, indicating that some TOF might have been wrongly selected. So, in order to check for inconsistencies, the average error $\bar{e} = \frac{1}{S}\sum_{s=1}^{S}|e_s|$ is computed. If the average error is greater than a threshold $\gamma \in \mathbb{R}_+$, an iterative search for better TOF estimates takes place.

At each iteration, a specific $\hat{t}_s$ is replaced with a new estimate and $\bar{e}$ is recomputed. The search for the new TOF combines two heuristics. The first heuristic constrains the corrected TOF to be associated with a CCF peak that necessarily precedes the previous (supposedly incorrect) CCF peak, since TOF-estimation errors induced by reverberation can only happen after the LOS TOF. The second heuristic selects the TOF associated with the loudspeaker whose secondary CCF peak has the largest amplitude ratio relative to the previously estimated CCF peak. The reason for this heuristic has roots in the observation that the amplitude of the correct CCF peak is frequently very close to that of the spurious peak (see Fig. 1). The method replaces older TOF estimates until $\bar{e}$ is smaller than γ. If at a given iteration no peak that satisfies the first constraint is found and $\bar{e}$ is still larger than γ, then one assumes the method failed, and the chosen sensor-location estimate is the one obtained with the standard LS method ($\hat{\mathbf{m}}^{(0)\mathrm{LS}}$). Algorithm 1 details the proposed procedure for TOF selection.

VI. Performance Evaluation

In this section, the proposed algorithm for sensor localization is evaluated. For this task, artificially generated signals are employed, allowing a high level of control over signal characteristics as well as precise knowledge of loudspeaker and microphone positions.

Algorithm 1 TOF selection.

Define a threshold $\gamma \in \mathbb{R}_+$
for all $s \in \mathcal{S}$ **do**
  $\hat{k}_s \leftarrow \arg\max_{k\in\mathbb{N}} \{R_s[k]\}$
  $\hat{t}_s \leftarrow R_s[\hat{k}_s]/F$
end for
Define $\mathbf{S}$ and $\mathbf{b}$ using Eq. (3)
$\hat{\mathbf{m}}^{(0)\mathrm{LS}} \leftarrow \mathbf{S}^{\dagger}\hat{\mathbf{b}}$
$\hat{\mathbf{m}}^{\mathrm{LS}} \leftarrow \hat{\mathbf{m}}^{(0)\mathrm{LS}}$
Compute e_s using Eq. (11) and Eq. (12) for all $s \in \mathcal{S}$
$\bar{e} \leftarrow \frac{1}{S}\sum_{s=1}^{S}|e_s|$
while $\bar{e} > \gamma$ **do**
  $s^* \leftarrow \emptyset$ and $r^* \leftarrow 0$
  for all $s \in \mathcal{S}$ **do**
    $\mathcal{P}_s \leftarrow \left\{k \in \mathbb{N} \,\middle|\, k < \hat{k}_s \text{ and } R_s[k] \text{ is a peak}\right\}$
    $k_s \leftarrow \arg\max_{k\in\mathcal{P}_s} R_s[k]$
    $r_s \leftarrow R_s[k]/R_s[\hat{k}_s]$
    if $r_s > r^*$ **then**
      $r^* \leftarrow r_s$
      $s^* \leftarrow s$
    end if
  end for
  if $s^* \neq \emptyset$ **then**
    $\hat{t}_{s^*} \leftarrow k_{s^*}/F$
    Update $\hat{\mathbf{b}}$ using Eq. (8)
    $\hat{\mathbf{m}}^{\mathrm{LS}} \leftarrow \mathbf{S}^{\dagger}\hat{\mathbf{b}}$
    Update e_{s^*} using Eq. (11) and Eq. (12)
    $\bar{e} \leftarrow \frac{1}{S}\sum_{s=1}^{S}|e_s|$
  else
    Abort procedure
    $\hat{\mathbf{m}}^{\mathrm{LS}} \leftarrow \hat{\mathbf{m}}^{(0)\mathrm{LS}}$
  end if
end while

A. Simulation Procedure

The simulation considered that the loudspeakers and microphones were inside a room with dimensions 5.2 m× 7.5 m × 2.6 m. A loudspeaker was positioned at the center of each wall of the room, totaling 6 loudspeakers. A set of 100 microphones were randomly positioned inside the room.

During the simulations, each loudspeaker emitted a white pseudo-Gaussian noise signal with 10-ms duration. A short silent interval was inserted between the emissions of sequenced loudspeakers, so that there is no time-overlap between signals originated from different loudspeakers arriving at a given microphone. The room acoustics were simulated using the standard and well-known image method [5].[3]

In the simulations, the 3-D position of each microphone was estimated using two methods: the standard least-squares (LS) solution and the LS plus TOF selection (SEL) scheme presented in Section V. Once both position estimates were computed, their associated errors were calculated as the Euclidean distance between the actual microphone position and the estimated one.

B. Performance in Reverberant Environment

The first evaluation strategy consists of verifying the performance improvement attained by the proposed algorithm under reverberant conditions. For this, the acoustic conditions of the room were set to different reverberation times (RT60), from 0 ms (anechoic) to 800 ms, in steps of 100 ms. For each reverberation time, 100 different microphone positions were simulated, and their respective estimation errors were averaged.

Fig. 2 compares the performances of the standard LS (solid line) and the proposed LS plus TOF selection (dashed line) algorithms. As can be gathered from the figure, the proposed algorithm improves the result for all reverberation times larger than 100 ms within the evaluated range. Indeed, the error

[2]In the sense that they convey information of a sensor position that is not the same as the one conveyed by the remaining TOFs.

[3]Implemented in the Audio Systems Array Processing Toolbox available at http://www.engr.uky.edu/~donohue/audio/Arrays/MAToolbox.htm.

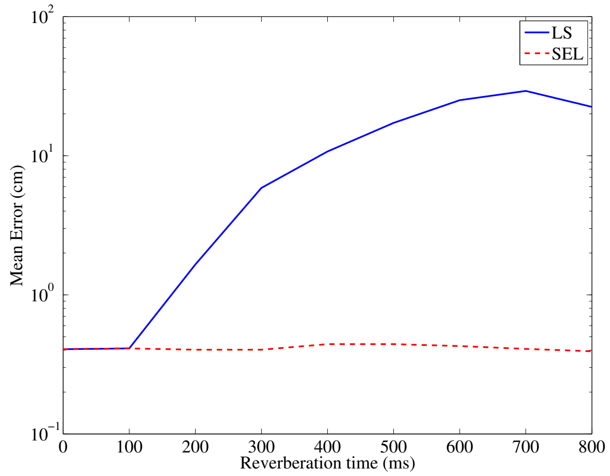


Fig. 2. Mean error (cm) *vs.* Reverberation time (ms).

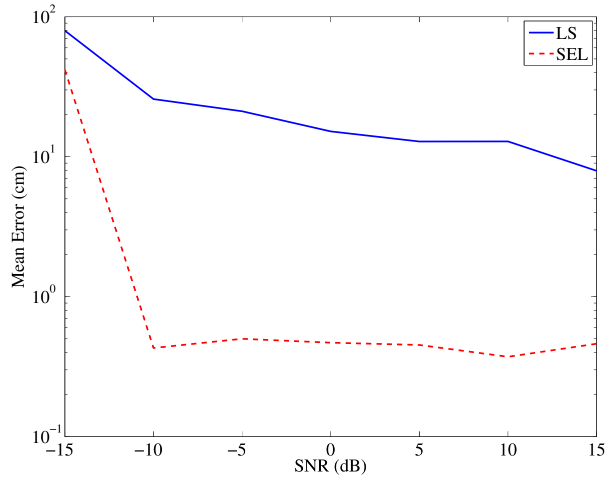


Fig. 3. Mean error (cm) *vs.* SNR (dB).

(around 0.4 cm) achieved by the proposed algorithm under reverberant conditions was similar to the error obtained by the standard LS algorithm for anechoic signals. This indicates that the proposed selection scheme was able to correct mistakenly chosen TOFs, mitigating the effects of reverberation, as desired. Note that, for the standard LS method, the sudden increase of the position-estimation error that occurred for RT60 larger than 100 ms takes place due to the high sensitivity of the position estimation with respect to TOF errors. Indeed, it was observed that TOF-estimation errors are not gradual; rather, once the reverberation time is large enough, spurious peaks in the CCF (see also Fig. 1) start to appear relatively "far" from the correct peak, thus generating large position-estimate deviations from the correct sensor location.

C. Performance for Noisy Signals

The second evaluation strategy looks into how the algorithm performs when noise is added to the received signals. For this, white pseudo-Gaussian noise is added to the signal received at every microphone, yielding signal-to-noise ratios (SNRs) varying from -15 to 15 dB in steps of 5 dB. In this experiment, reverberation time was set to 400 ms. Analogously to the previous simulation, for each SNR value 100 different microphone positions were randomly generated, and their mean estimation errors were computed.

Fig. 3 shows the mean error for each value of SNR. As can be gathered, the proposed solution is capable to correct the errors due to reverberation (mean error remains around 0.4 cm for SNRs larger than -10 dB) regardless the noise level, within the observed range.

VII. Concluding Remarks

This paper described a simple least-squares solution to the sensor localization problem that uses the signal emitted by several loudspeakers to find the 3-D position of a microphone. The main contribution presented was a new procedure for selection of TOF estimates, which relies explicitly on the fact that TOF-estimation errors usually derive from non-LOS paths. Additionally, the amplitude of the CCF peaks has proven to provide useful ancillary information with respect to good candidates to correct TOF estimates. As indicated by the simulation results, the proposed TOF selection algorithm is quite robust to reverberation and noise effects.

As future work, the proposed TOF selection scheme should be evaluated using real-world signals recorded under reverberant conditions. Furthermore, a careful study of the influence of TOF estimation errors as well as the impact of the number of loudspeakers on the microphone position estimates must be carried out.

Acknowledgements

This R&D project is a cooperation between Hewlett-Packard Brasil Ltda. and COPPE/UFRJ, being supported with resources of Informatics Law (no. 8.248, from 1991). L. W. P. Biscainho, Diego B. Haddad, M. V. S. Lima, W. A. Martins, and L. de O. Nunes would like to thank also CAPES, CNPq, and FAPERJ agencies for funding their research work.

References

[1] J. H. DiBiase, H. F. Silverman, and M. S. Brandstein, "Robust localization in reverberant rooms," in *Microphone Arrays: Signal Processing Techniques and Applications*, M. S. Brandstein and D. B. Ward, Eds. Berlin, Germany: Springer-Verlag, 2001, ch. 8, pp. 157–180.

[2] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*. Heidelberg, Germany: Springer, 2010.

[3] V. Raykar and R. Duraiswami, "Automatic position calibration of multiple microphones," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP'04)*, vol. 4. Montreal, Canada: IEEE, May 2004, pp. 69–72.

[4] N. Patwari, J. N. Ash, S. Kyperountas, A. O. H. III, R. L. Moses, and N. S. Correal, "Locating the nodes: Cooperative localization in wireless sensor networks," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 54–69, Jul. 2005.

[5] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *Journal of the Acoustical Society of America*, vol. 65, no. 3, pp. 943–950, April 1979.