HAL
archives-ouvertes.fr

Sound-based control with two microphones

Aly Magassouba, Nancy Bertin, François Chaumette

To cite this version:

Aly Magassouba, Nancy Bertin, François Chaumette. Sound-based control with two microphones. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'15, Sep 2015, Hamburg, Germany. pp.5568-5573, 2015. <hal-01185841>

HAL Id: hal-01185841

https://hal.inria.fr/hal-01185841

Submitted on 21 Aug 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Sound-based control with two microphones

Aly Magassouba[1], Nancy Bertin[2] and François Chaumette[3]

***Abstract*— This paper presents a novel approach to robot audition by performing robotic tasks with auditory cues. Unlike many previous works, we propose a control scheme, that does not require any explicit sound source localization. This approach is capable of controlling all the three degrees of freedom in a plane from two microphones. Built upon the sensor-based control framework, this approach relies on implicit sound source direction obtained from the time difference of arrival (TDOA). We introduce an analytical modelling of auditory cues considering a robot equipped with a pair of microphones and multiple sound sources, from which a control scheme is designed. A stability analysis is provided as well. The results obtained in simulation show the feasibility and the suitability of this method even in reverberant area.**

I. INTRODUCTION

In robotics, the use of aural perception has received recently a growing interest but still remains marginal in comparison to other senses like vision. Audio sensing can be a source of robotic interaction, for instance in positioning tasks, as well as complementary to other perception senses. Most existing works about aural perception are based on the relative localization of a defined system with respect to a sound source. The control is generally performed outside the localization system. In this context, a robot auditory system performing planar localization is proposed in [8] with the use of three microphones. By increasing the number of microphones in a non planar shape, a 3D localization system with cross-correlation method is presented in [17]. These two methods consider a static sound source and the localization is performed once before the positioning task. A dynamic localization method is developed in [14] that uses a tracking based on a model of the sound source. Multi-source localization and tracking based on particle filtering are dealt with in [16]. More recent approaches have gone towards binaural localization with application on humanoid robots [7]. All these approaches mainly propose and exploit different techniques of explicit localization adapted to a specific context. The path of improvement taken by these methods is to focus on the preliminary explicit source localization. Especially, these works tend to robustify the localization process in presence of several sound sources and reverberation. Indeed the localization process is ill-conditioned and depends strongly on reverberation. To overcome this issue, the number of microphones is generally increased to more than three and angular information redundancy is exploited to get a more robust sound source position. In binaural field, the localization is generally limited to direction of the azimuthal and elevation angles [9] without any knowledge of the distance to the sound sources. [15] proposes a stochastic approach to estimate the planar distance to a single source. However this method requires a motion from the sensors or the sound source to build an estimate of the sound source position after many iterations.

In contrast, the approach presented in this paper builds on a different point of view by introducing a sensor-based framework instead of explicit localization. This line of work considers the hearing sense as a direct and real-time input of closed loop control schemes. Unlike most previous works, this approach does not require any explicit source localization: we focus on developing a modelling based on sound features. By skipping the localization step and processing auditory cues directly in a control loop this approach tends to be more robust to reverberation while using a limited number of microphones. A simple approach of this type has been proposed in [10] as an audio servo system. The control scheme considers a linear relationship between the angular source position and the stereo cues retrieved by a pair of microphones. This method is designed to control only one degree-of-freedom (dummy head rotation) to the direction of the sound source. Similarly, this audio servo is also used for robot playing music instrument like Theremin [1]. But this type of work is dedicated to the Theremin case, and is hardly extensible to more complex tasks.

We propose a more generic modelling capable of controlling all the three DOF of a robot in a plane with only two microphones. This work draws inspiration from the wide body of knowledge available in the field of vision-based control [4], [6], [11]. We based this modelling on the measurement of the time difference of arrival (TDOA) between the two microphones. While TDOA itself is not new in robot aural perception, the idea is to define a robotic task with respect to (w.r.t) angular information: it can thus be related to the field of beaconing or bearing-only homing. To the best of our knowledge, no other prior work has tried to solve the problem of sound-based control in this way.

The main contributions of this paper are: (**i**) an analytical modelling linking the sound features to the control input in Section II; (**ii**) the definition and the analysis of the robotic tasks that can be achieved from the number of sound sources in Section III; (**iii**) the validation of the proposed method by simulation in Section IV.

[1]Université Rennes I - IRISA, Campus de Beaulieu, 35042 Rennes cedex, France aly.magassouba@irisa.fr
[2]CNRS - IRISA, Campus de Beaulieu, 35042 Rennes cedex, France nancy.bertin@irisa.fr
[3]Inria - IRISA, Campus de Beaulieu, 35042 Rennes cedex, France francois.chaumette@inria.fr

II. AN AUDIO-BASED CONTROL MODELLING

A. Sound features

The sensor-based control performed in this work is based on the TDOA between a pair of microphones. In acoustics, the TDOA corresponds to the time difference τ in sound propagation between two sensors. There exists several methods to estimate this value, that have been classified and compared in [2]. Among them the generalized cross-correlation with phase transform GCC-PHAT method is the most popular and has been used in this work. This TDOA refers implicitly to a sound source direction and can be expressed under the far field assumption as:

$$\tau = A\cos\alpha \tag{1}$$

where $A = d/c$ in which c is the sound celerity, d the distance between the two microphones and α the sound direction (see Fig. 1) also known as direction of arrival (DOA). Eq. (1) implies that $\alpha \in [0 \;\; \pi]$ since $\tau(\alpha) = \tau(-\alpha)$. This leads to an ambiguity in the direction of the sound source generally faced in the binaural field. In the following development, we consider a half plane working space, that is to say the sound sources are always on the same side (front or back) of the microphones.

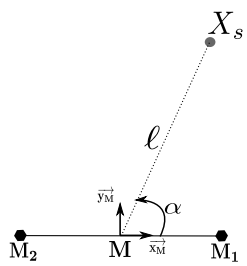


Fig. 1: Planar system modelling

B. Sensor-based framework

Considering a set of measurements $\mathbf{s}(t)$ obtained from a sensor, the goal of the servo is to minimize an error $\|\mathbf{e}(t)\|$ characterized by

$$\mathbf{e}(t) = \mathbf{s}(t) - \mathbf{s}^* \tag{2}$$

where $\mathbf{s}^*$ denotes the set of measurements at the desired state. Once $\mathbf{s}$ is selected, the time variation of this set is related to the sensors velocity by

$$\dot{\mathbf{s}} = \mathbf{L_s}\mathbf{v_s} \tag{3}$$

in which $\mathbf{L_s} \in \mathbb{R}^{k\times n}$ is a matrix sized by k the number of measurements and n the number of controlled DOF. When $\mathbf{v_s} = (\boldsymbol{v_s}, \boldsymbol{\omega_s}) = (v_x, v_y, v_z, \omega_x, \omega_y, \omega_z)$ where $\boldsymbol{v_s}$ denotes the spatial linear and $\boldsymbol{\omega_s}$ the angular velocity, then $k = 6$ and $\mathbf{L_s}$ is named the interaction matrix of $\mathbf{s}$ [4]. From then on, the control scheme can be designed with for instance a purpose of exponential decoupled decrease of the error. The time variation of the expected error follows

$$\dot{\mathbf{e}} = -\lambda\mathbf{e} \tag{4}$$

with $\lambda > 0$ a gain that tunes the time to convergence. Then, by combining (3) and (4) we obtain

$$\mathbf{v_s} = -\lambda\widehat{\mathbf{L_s^+}}\mathbf{e} \tag{5}$$

where $\mathbf{L_s^+} \in \mathbb{R}^{n\times k}$ is the Moore-Penrose pseudo-inverse of the interaction matrix. The pseudo-inverse is used when the inverse is not defined, that is when $n \neq k$ or $n = k$ and $|\mathbf{L_s}| = 0$. Otherwise $\mathbf{L_s^+} = \mathbf{L_s^{-1}}$. Additionally, an approximation $\widehat{\mathbf{L_s^+}}$ is considered since it is impossible to know perfectly either $\mathbf{L_s}$ or $\mathbf{L_s^+}$ in practice [4].

The global asymptotic stability of the control law (5), following Lyapunov definition, is ensured as soon as $\mathbf{L_s}\widehat{\mathbf{L_s^+}} > 0$, if $k \leq n$. Otherwise only the local asymptotic stability can be ensured in the neighbourhood of $\mathbf{s}^*$ under the condition that $\widehat{\mathbf{L_s^+}}\mathbf{L_s} > 0$ [4].

C. Geometric configuration

We consider a robot instrumented with a pair of microphones $\mathbf{M_1}$ and $\mathbf{M_2}$ in an area free from obstacle. A continuously emitting static sound source $\mathbf{X_s}(x_s, y_s)$ is then placed at an unknown position. Fig. 1 illustrates this configuration, where a Cartesian frame $\mathcal{F}_{\mathbf{M}}(\overrightarrow{x_M}, \overrightarrow{y_M})$ is attached such that its origin $\mathbf{M}$ is at the mid-point between $\mathbf{M_1}$ and $\mathbf{M_2}$. Let ℓ be the distance between $\mathbf{M}$ and $\mathbf{X_s}$ so that the sound source is characterized by the following relationships:

$$\begin{cases} x_s = \ell\cos\alpha \\ y_s = \ell\sin\alpha \end{cases} \text{and} \begin{cases} \alpha = \text{atan2}\,(y_s, x_s) \\ \ell = \sqrt{x_s^2 + y_s^2} \end{cases} \tag{6}$$

This configuration is endowed with 3 DOF, translations along $\overrightarrow{x_M}$ and $\overrightarrow{y_M}$ axis, and rotation around $\overrightarrow{z_M}$ that are parametrized by a position and an orientation.

D. Sound features modelling

The time variation of the angle α is defined by:

$$\dot{\alpha} = (\dot{y_s}x_s - \dot{x_s}y_s)/\ell^2. \tag{7}$$

Eq. (7) can also be expressed as a matrix relationship:

$$\dot{\alpha} = \mathbf{L}_\alpha\mathbf{v_s} \tag{8}$$

where $\mathbf{L}_\alpha$ is the interaction matrix like $\mathbf{L_s}$ in visual servoing. From the basic kinematic equation:

$$\dot{\mathbf{X_s}} = -\boldsymbol{v_s} - \boldsymbol{\omega_s} \times \mathbf{X_s} \Leftrightarrow \begin{cases} \dot{x_s} = -v_x - \omega_y z_s + \omega_z y_s \\ \dot{y_s} = -v_y - \omega_z x_s + \omega_x z_s \\ \dot{z_s} = -v_z - \omega_x y_s + \omega_y x_s \end{cases} \tag{9}$$

which relates the velocity of a 3-D point $\mathbf{X_s}$ to the sensor spatial velocity $\mathbf{v_s}$, we obtain

$$\dot{\alpha} = (-v_y x_s + v_x y_s - \omega_z(y_s^2 + x_s^2))/(x_s^2 + y_s^2). \tag{10}$$

Eventually referring to (8), the interaction matrix related to α is identified as:

$$\mathbf{L}_\alpha = \begin{bmatrix} \frac{y_s}{\ell^2} & -\frac{x_s}{\ell^2} & 0 & 0 & 0 & -1 \end{bmatrix} \tag{11}$$

In connection with the geometric configuration, we consider only the non-zero terms of this matrix so that $k = 3$ and $\dot{\alpha} = \mathbf{J}_\alpha \mathbf{u}$ where:

$$\mathbf{J}_\alpha = \begin{bmatrix} \frac{y_s}{\ell^2} & -\frac{x_s}{\ell^2} & -1 \end{bmatrix} \text{ and } \mathbf{u} = \begin{bmatrix} v_x & v_y & \omega_z \end{bmatrix}^\top. \quad (12)$$

In this case $\mathbf{J}_\alpha$ is the feature Jacobian and $\mathbf{u}$ the velocity control input. Moreover, although the sound source position $\mathbf{X_s}(x_s, y_s)$ is unknown, (6) allows to reshape the analytical value of the Jacobian matrix as:

$$\mathbf{J}_\alpha = \begin{bmatrix} \frac{\sin\alpha}{\ell} & -\frac{\cos\alpha}{\ell} & -1 \end{bmatrix}. \quad (13)$$

This result is then extended to the feature Jacobian matrix related to τ (see (1)) since the TDOA is the direct available measurement. So, taking the time derivative of (1) we obtain:

$$\dot{\tau} = -A \sin\alpha \, \dot{\alpha}. \quad (14)$$

The feature Jacobian matrix $\mathbf{J}_\tau$ is then defined by

$$\mathbf{J}_\tau = -A \sin\alpha \mathbf{J}_\alpha. \quad (15)$$

By replacing (13) in (15), $\mathbf{J}_\tau$ is eventually equal to

$$\mathbf{J}_\tau = \begin{bmatrix} -\frac{A^2-\tau^2}{A\ell} & \frac{\tau\sqrt{A^2-\tau^2}}{A\ell} & \sqrt{A^2-\tau^2} \end{bmatrix}. \quad (16)$$

Similarly to classic binaural localization, the distance ℓ that appears in $\mathbf{J}_\alpha$ and $\mathbf{J}_\tau$ is unknown. This result is also similar to classical image-based visual servoing where the depth of a point, which is unknown, appears in the interaction matrix of an image point. However, as studied in Section IV-C, the robustness of the controller allows approximating the Jacobian matrix with $\widehat{\ell}$ without deteriorating the system behaviour [6].

III. POSE ANALYSIS

From the feature Jacobian defined by (16), it is possible to achieve positioning tasks so that α reaches a particular desired value α^*. By using only one feature from a single sound source, it can be intuitively expected that several sensor poses exist so that $\alpha = \alpha^*$. Actually only one DOF can be controlled by using one sound source. However, it is possible to extend the approach to control more DOF by simply increasing the number of sound sources. In the context of bearing localization, this fact has been exposed in [3] by an observability analysis. The latter work provides an empirical and numerical proof of the behaviour of such a system considering up to 3 beacons. Since the observability cannot always be proved, the results of this analysis have been mostly obtained through simulations. Hence, we develop in this section a pose analysis supported by geometrical and analytical proof based on virtual links [5].

Using (3) with a motion $\mathbf{v}^*_{\mathbf{s_i}}$ so that the sensor output $\mathbf{s_i}$ remains constant for a given configuration, leads to:

$$\mathbf{L}_{\mathbf{s_i}} \mathbf{v}^*_{\mathbf{s_i}} = 0. \quad (17)$$

This concept is generalized by the vectorial subspace $\mathbf{S}_n{}^* = \text{Ker } \mathbf{L_s}$, where each column characterizes a motion of $\mathbf{v}^*_{\mathbf{s_i}}$ type. Likewise, all compatible and independent configurations where $\mathbf{s} = \mathbf{s}^*$ build up a virtual linkage between the sensor and the sound source. With m the rank of the interaction matrix, N the rank of $\mathbf{S}_n{}^*$ is equal to $n - m$. This rank N defines the class of the virtual link.

A. Case of one sound source

When considering only one sound source, $\mathbf{J}_\alpha$ given by (13) is a rank one matrix. Consequently $\mathbf{S}_1{}^* \in \mathbb{R}^{3\times 2}$ implies a virtual link of class 2:

$$\mathbf{S}_1{}^* = \begin{bmatrix} \cos\alpha & \ell\sin\alpha \\ \sin\alpha & -\ell\cos\alpha \\ 0 & 1 \end{bmatrix}. \quad (18)$$

The first column of $\mathbf{S}_1{}^*$ induces a translational motion along the sound source direction, that is along $\overrightarrow{\mathbf{MX_s}}$. The second column describes a rotation around $\overrightarrow{z_M}$ axis combined with a translation. This last motion can be geometrically represented by a circle of radius ℓ centred on $\mathbf{X_s}$ (see Fig. 2a). Thus, any linear combination of these two motions implies infinite poses to complete the task $\alpha = \alpha^*$, from which we set up a first Lemma.

Lemma *1:* For each random position of the microphones defined by $\mathbf{M}$, there exists one orientation $\theta_\mathbf{M}$ of the microphones such that $\alpha = \alpha^*$.

In terms of Lyapunov stability of the control law (5), the condition $\mathbf{J_s}\widehat{\mathbf{J_s^+}} > 0$ is ensured as soon as $\widehat{\ell}$ is set to a positive value, which is of course the case in practice, since:

$$\mathbf{J_s}\widehat{\mathbf{J_s^+}} = \frac{\widehat{\ell}(\widehat{\ell}\ell + 1)}{\ell(\widehat{\ell}^2 + 1)}. \quad (19)$$

B. Case of two sound sources

When considering two sound sources $\mathbf{X_{s_1}}$ and $\mathbf{X_{s_2}}$, the feature Jacobian matrix is obtained by merely stacking (13) for each sound source:

$$\mathbf{J_\alpha} = \begin{bmatrix} \frac{\sin\alpha_1}{\ell_1} & -\frac{\cos\alpha_1}{\ell_1} & -1 \\ \frac{\sin\alpha_2}{\ell_2} & -\frac{\cos\alpha_2}{\ell_2} & -1 \end{bmatrix}. \quad (20)$$

This time $\mathbf{S}_2{}^* \in \mathbb{R}^{3\times 1}$ implies a class 1 virtual link:

$$\mathbf{S}_2{}^* = \begin{bmatrix} \ell_1\cos\alpha_2 - \ell_2\cos\alpha_1 \\ \ell_1\sin\alpha_2 - \ell_2\sin\alpha_1 \\ \sin(\alpha_1 - \alpha_2) \end{bmatrix}. \quad (21)$$

Geometrically $\mathbf{S}_2{}^*$ refers to the circumscribed arc of circle characterized by $\mathbf{X_{s_1}}$, $\mathbf{X_{s_2}}$ and $\mathbf{M}$, as it can be proved by the inscribed angle theorem. Let λ be $\angle\mathbf{X_{s_1}}\mathbf{M}^*\mathbf{X_{s_2}}$ (i.e $\lambda = \alpha_2^* - \alpha_1^*$) from a specified pose as it appears in Fig. 2b. This theorem exposes that given $\mathbf{X_{s_1}}\mathbf{X_{s_2}}$, the set of points $\mathbf{M}$ in the plane for which the angle $\angle\mathbf{X_{s_1}}\mathbf{M}\mathbf{X_{s_2}}$ is equal to λ is an arc belonging to the circumscribed circle of $\mathbf{M}^*\mathbf{X_{s_1}}\mathbf{X_{s_2}}$. Then by considering a random point $\mathbf{M}$ on this arc, from Lemma 1, it is guaranteed to find an orientation $\theta_\mathbf{M}$ such that $\alpha_1 = \alpha_1^*$ (or $\alpha_2 = \alpha_2^*$). Knowing that λ remains constant, $\alpha_2 = \alpha_2^*$ (or $\alpha_1 = \alpha_1^*$) is guaranteed from $\lambda = \alpha_2^* - \alpha_1^*$ This result sets up the following Lemma:

Lemma *2:* For each random position $\mathbf{M}$ on the arc of the circumscribed circle defined by $\mathbf{X_{s_1}}$, $\mathbf{X_{s_2}}$ and $\mathbf{M}^*$, there exists one orientation $\theta_\mathbf{M}$ of the microphones so that $\alpha_1 = \alpha_1^*$ and $\alpha_2 = \alpha_2^*$.

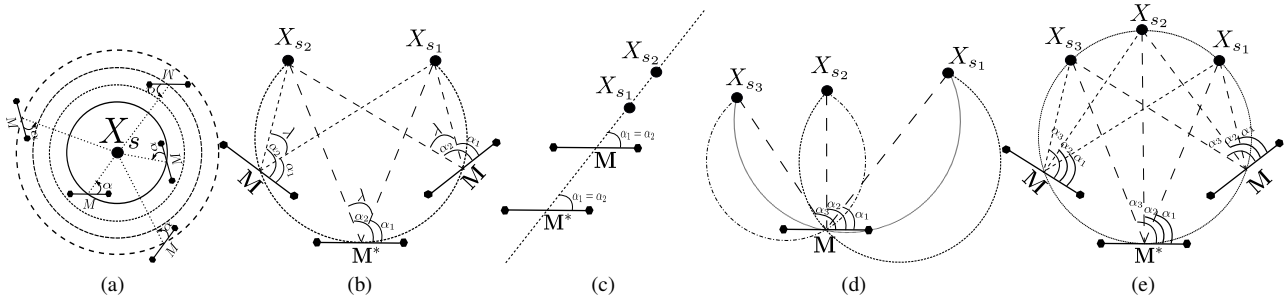


Fig. 2: Microphones admissible set of poses with (a) one source, (b) two sources, (c) two sources in singular configuration, (d) three sources and (e) three sources in singular configuration

This Lemma is valid for any sources configuration but the situation where $\mathbf{M}^*$, $\mathbf{X_{s_1}}$ and $\mathbf{X_{s_2}}$ are aligned. In this particular case where $\alpha_1^* = \alpha_2^*$ we get:

$$\mathbf{S}_2{}^* = \begin{bmatrix}\cos\alpha_1 & \sin\alpha_1 & 0\end{bmatrix}^\top. \tag{22}$$

As expected, we obtain a trajectory similar to the translational motion of $\mathbf{S}_1{}^*$ in (18), since only one angular data is available (see Fig. 2c). For the stability, the condition $\mathbf{J_s}\widehat{\mathbf{J_s^+}} > 0$ is of course ensured when $\widehat{\ell_i} = \ell_i$ since we have $\mathbf{J_s}\widehat{\mathbf{J_s^+}} = \mathbb{I}_2$ in that case.

C. Case of three sound sources

When considering three sound sources $\mathbf{X_{s_1}}$, $\mathbf{X_{s_2}}$ and $\mathbf{X_{s_3}}$ the feature Jacobian is generally full rank 3,

$$\mathbf{J}_{\boldsymbol{\alpha}} = \begin{bmatrix} \frac{\sin\alpha_1}{\ell_1} & -\frac{\cos\alpha_1}{\ell_1} & -1 \\ \frac{\sin\alpha_2}{\ell_2} & -\frac{\cos\alpha_2}{\ell_2} & -1 \\ \frac{\sin\alpha_3}{\ell_3} & -\frac{\cos\alpha_3}{\ell_3} & -1 \end{bmatrix} \tag{23}$$

In most cases, this configuration implies that there exists only one pose where $\mathbf{s} = \mathbf{s}^*$. However, depending on the specified desired position or the sound sources configuration, a set of poses for which $\mathbf{s} = \mathbf{s}^*$ might exists. Indeed, inspired by the case of two sound sources, a set can be expected if $\mathbf{X_{s_1}X_{s_2}X_{s_3}}$ and $\mathbf{M}$ design a concyclic quadrilateral. Wherever $\mathbf{M}$ belongs to the corresponding circumscribed circle, an orientation $\theta_{\mathbf{M}}$ exists such that $\mathbf{s} = \mathbf{s}^*$ (see Fig. 2d) and $\mathbf{S}_3{}^*$ is designed by (21). Likewise, another set is obtained if the three sound sources are aligned with $\mathbf{M}$ in which $\mathbf{S}_3{}^*$ is defined by (22).

Geometrically, these results are obtained by considering the three circumscribed arcs of circle defined by $\mathbf{X_{s_1}X_{s_2}M}$, $\mathbf{X_{s_2}X_{s_3}M}$ and $\mathbf{X_{s_1}X_{s_3}M}$. According to Lemma 2, $\mathbf{M}$ must belong to these three arcs to ensure $\alpha_1 = \alpha_1^*$, $\alpha_2 = \alpha_2^*$ and $\alpha_3 = \alpha_3^*$. As shown in Fig. 2d, these arcs intersect in only one position when considering a nominal configuration, otherwise this intersection corresponds to the circumscribed circle (see Fig.2e).

Lemma *3:* In presence of three sound sources in a non-concyclic or non-degenerate configuration with $\mathbf{M}^*$, there exists only one pose $\mathbf{M}$ of the microphones so that $\alpha_1 = \alpha_1^*$, $\alpha_2 = \alpha_2^*$ and $\alpha_3 = \alpha_3^*$.

Apart from degenerate cases including aligned or juxtaposed sound sources, the condition $\mathbf{J_s}\widehat{\mathbf{J_s^{-1}}} > 0$ is theoretically ensured (with $\widehat{\ell_i} = \ell_i$) and the system stable since $\mathbf{J_s}\widehat{\mathbf{J_s^{-1}}} = \mathbb{I}_3$ if $\mathbf{M}$ does not belong to the circumscribed circle.

Indeed, crossing this singular circle during the task challenges the stability of the control law since $\mathbf{J}_{\boldsymbol{\alpha}}$ is not full rank at this pose. This is confirmed by computing the determinant of $\mathbf{J}_{\boldsymbol{\alpha}}$:

$$|\mathbf{J}_{\boldsymbol{\alpha}}| = \frac{\sin(\alpha_1-\alpha_2)}{\ell_1\ell_2} + \frac{\sin(\alpha_2-\alpha_3)}{\ell_2\ell_3} + \frac{\sin(\alpha_3-\alpha_1)}{\ell_3\ell_1}. \tag{24}$$

On the circumscribed circle, from the inscribed angle theorem and the sinus law injected in (24), the determinant of the Jacobian matrix becomes

$$|\mathbf{J}_{\boldsymbol{\alpha}}| = \frac{\sin(\alpha_3-\alpha_1)(\ell_2 - \frac{\|X_{s_2}X_{s_1}\|}{\|X_{s_3}X_{s_1}\|}\ell_3 - \frac{\|X_{s_3}X_{s_2}\|}{\|X_{s_3}X_{s_1}\|}\ell_1)}{\ell_1\ell_2\ell_3}. \tag{25}$$

Furthermore, using the property of the cyclic quadrilateral defining the product of the diagonals as equal to the sum of the products of opposite sides,

$$\ell_2\|X_{s_3}X_{s_1}\| = \ell_1\|X_{s_3}X_{s_2}\| + \ell_3\|X_{s_2}X_{s_1}\|, \tag{26}$$

leads to $|\mathbf{J}_{\boldsymbol{\alpha}}| = 0$. A parallel can be emphasized with IBVS region of instability, where the circumscribed circle is the planar equivalent of the singular cylinder [13]. In all other configurations, it is possible to show that $|\mathbf{J}_{\boldsymbol{\alpha}}|$ is not null.

D. Using more than three sound sources

Using more than three sound sources can be beneficial by giving more robustness to the control law with redundant data and allows avoiding the singularity of the circumscribed circle. Those sound sources imply a single pose such that $\mathbf{s} = \mathbf{s}^*$. The demonstration developed for three sound sources remains applicable. Unlike the previous case, the singularity circle is cancelled by a fourth sound source positioned in a non-concyclic configuration with the three first. On the other hand only the local asymptotic stability is verified under the condition $\widehat{\mathbf{J_s^+}}\mathbf{J_s} > 0$ (see Section II-B), which is ensured when $\widehat{\ell_i} = \ell_i$ since $\widehat{\mathbf{J_s^+}}\mathbf{J_s} = \mathbb{I}_3$.

To conclude, with three sound sources in a non-singular configuration or with at least four sources in any other case, all the three robot DOF can be controlled. Hence a

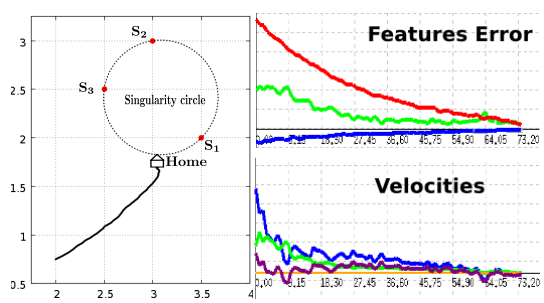
homing task can be performed without an exact knowledge of the position of the sound sources (an approximation $\widehat{\ell}$ is sufficient). This is a clear advantage of our approach with respect to classic techniques of sound source localization which require more microphones.

IV. SIMULATION AND RESULTS

A. Simulation settings

The developed method was tested through simulations by using Roomsimove [2], a room acoustics simulator. From this tool, we designed a room of 4.55m $\times 3.55$m $\times 2.5$m in which three sound sources $\mathbf{S_1}$, $\mathbf{S_2}$ and $\mathbf{S_3}$ were positioned as they appear in Fig. 3a. The estimation of the TDOA has been performed with BSSLocate [2] with a non-linear GCC-PHAT function. First, each of the sound sources emitted in succession so that it can be identified uniquely. Then the task was performed using the TDOA τ_i estimated from the global signal (simultaneously received from all sound sources). In the two simulations described in section IV-B, we considered the real interaction matrix $\widehat{\mathbf{L_s^{-1}}} = \mathbf{L_s^{-1}}$ assuming that each distance ℓ_i is known. In practice this is not the case and different strategies of approximation of the interaction matrix are discussed in Section IV-C.

B. Results



(a) (b)

Fig. 3: A homing task in reverberant area

1) Typical example: We simulated first a homing task with an initial pose defined by $\mathbf{M}(2, 0.75)$ and $\theta_{\mathbf{M}} = -5°$ and a home pose defined by $\mathbf{M}^*(3.1, 1.75)$ and $\theta_{\mathbf{M}^*} = 10°$. That home pose corresponds to $\boldsymbol{\alpha}^*$=$(31°, 85°, 118°)$ and was located outside of the circumscribed circle. In addition, a moderate reverberation was added with a sound absorption rate $\gamma = 0.7$ (e.g reverberation time RT60 $\approx$ 75ms) and with a background white noise that guarantees a signal-to-noise ratio (SNR) of 20 db. The result of this simulation is shown on Fig. 3. Despite the uncertainties on the TDOA measurement caused by the noise and reverberation, the system converged successfully to the home pose with an exponential decrease of the error. Moreover, since the singularity circle was not crossed during the task, the stability of the control law was ensured. As for simulation data, despite the jittery behaviour in the velocities curves caused by TDOA estimation, the trajectory remains smooth.

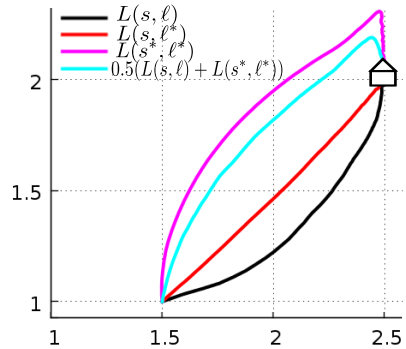


Fig. 5: Homing trajectories with different interaction matrix approximation

2) Robustness to reverberation: A second scenario was conducted including 99 trials on different steps of reverberation with uniformly distributed starting pose. With the front/back ambiguity implied by the use of only two microphones, $\theta_{\mathbf{M}}$ was chosen so that the sound sources were on the same side of the microphones at the initial pose. Thus the orientation $\theta_{\mathbf{M}}$ varied from $-30°$ to $30°$ depending on the starting pose. With a home pose at $\mathbf{M}(2.25, 1)$ and $\theta_{\mathbf{M}^*} = -5°$, the results summed up in Fig. 4 show that with no reverberation the system always converge to that pose. This result confirms the relevance and the suitability of the auditory modelling. However, we distinguished two failure areas in presence of reverberation. The first one was located on the alignment of $\mathbf{S_2}$ and $\mathbf{S_3}$ and mainly occurs when $\gamma = 0.7$. In this area, the main difficulty is to identify each measurement. With variation in the estimation and outlier (sound echoes) values, a wrong TDOA could be associated to a sound source. Likewise, a feature loss was observed on the right extremity of the test area since one of the TDOA was not correctly estimated. Consequently the system could fail to converge in those two areas especially with high reverberation.

C. Approximation of the interaction matrix

Several ways exist to compute $\widehat{\mathbf{L_s}}$ and these different strategies affect the behaviour and the stability of the system as shown in [12]. Since the distances ℓ_i are unknown in practice, a first strategy consists in using ℓ_i^*, the distances at the desired pose, instead. As illustrated by Fig. 5, $\mathbf{L_{s(s,\ell^*)}}$ gives a straighter and shorter trajectory. In addition, the desired measurement s^* can be used in the interaction matrix. In that case $\mathbf{L_{s(s^*,\ell^*)}}$ leads to a longer and unexpected trajectory but still converges to the home pose. These two solutions solve the problem of the unknown distance to the sound source. Moreover $\mathbf{L_{s(s^*,\ell^*)}}$ is constant and need to be computed only once. As a result the control law is not ill conditioned any more when crossing the circumscribed circle. A last strategy considers the combination of the real and the constant interaction matrix leading to $\widehat{\mathbf{L_s}} = (\mathbf{L_{s(s,\ell)}} + \mathbf{L_{s(s^*,\ell^*)}})/2$. The corresponding trajectory is improved compared to the constant interaction matrix, with the influence of $\mathbf{L_{s(s,\ell)}}$.

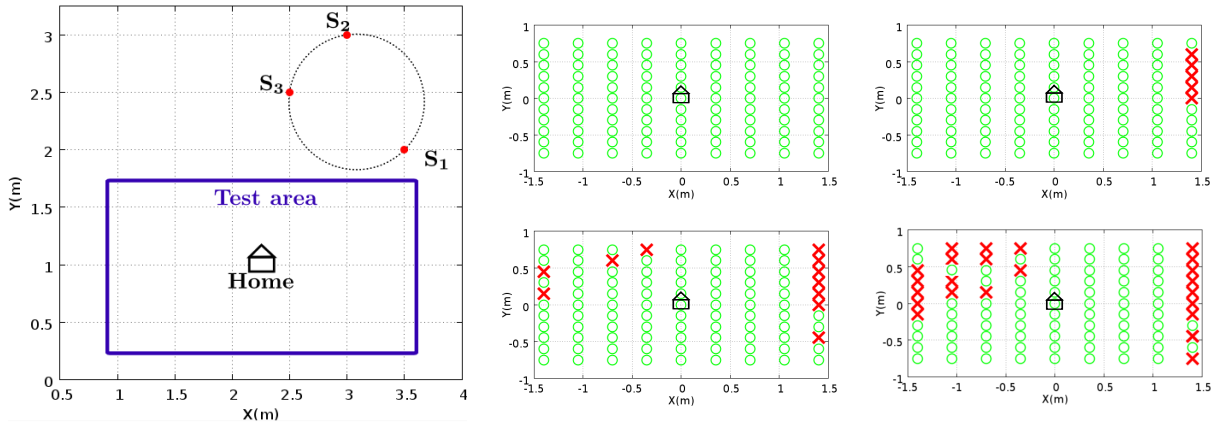


Fig. 4: Homing task success in function of reverberation and initial pose: success(**O**) and failure (**X**) with from top-left to bottom-right with wall absorption rate $\gamma = \{1,\ 0.9,\ 0.8,\ 0.7\}$.

V. CONCLUSION

We presented a novel sound-based approach for controlling robot motions, without any explicit sound source localization. Several positioning tasks can be achieved controlling up to three degrees of freedom in a plane with only two microphones, while classic localization techniques need more microphones or an exact knowledge about the environment to complete the same type of task. The theoretical developments established in the sensor-based control framework showed the capability of positioning from auditory cues: a single and stable pose of the microphones can be ensured with at least three sound sources as validated by the simulations. This result opens several perspectives in binaural field, even if the scattering effect of the head between the microphones should be taken into account in the modelling. For our future work, we would like to extend this approach to different auditory cues modelling for instance the interaural time difference and the interaural phase difference. Likewise a simple auditory feature like the sound intensity could be modelled to control the distance to the sound source. Experimental application of this framework on mobile robots is also intended in a near future.

REFERENCES

[1] A Alford, S Northrup, K Kawamura, KW Chan, and J Barile. A music playing robot. In *Proc. of the Conf. on Field and Service Robots*, pp 29–31, 1999.

[2] C. Blandin, A. Ozerov, and E. Vincent. Multi-source tdoa estimation in reverberant audio using angular spectra and clustering. *Signal Processing*, 92(8):1950–1960, 2012.

[3] P. Bonnifait and G. Garcia. Design and experimental validation of an odometric and goniometric localization system for outdoor robot vehicles. *IEEE Tr. on robotics and automation*, 14(4):541–548, 1998.

[4] F. Chaumette and S. Hutchinson. Visual servo control. i. basic approaches. *IEEE Robotics and Automation Magazine*, 13(4):82–90, 2006.

[5] F. Chaumette, P. Rives, and B. Espiau. Classification and realization of the different vision-based tasks. *Visual Servoing*, 7:199–228, 1993.

[6] P. Corke. Mobile robot navigation as a planar visual servoing problem. In *Robotics Research*, pp 361–372. Springer, 2003.

[7] J. Hornstein, M. Lopes, J. Santos-Victor, and F. Lacerda. Sound localization for humanoid robots-building audio-motor maps based on the hrtf. In *IEEE Int. Conf. on Intelligent Robots and Systems*, pp 1170–1176. 2006.

[8] J. Huang, T. Supaongprapa, I. Terakura, F. Wang, N. Ohnishi, and N. Sugie. A model-based sound localization system and its application to robot navigation. *Robotics and Autonomous Systems*, 27(4):199–209, 1999.

[9] F. Keyrouz, Y. Naous, and K. Diepold. A new method for binaural 3-d localization based on hrtfs. In *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, volume 5, pp V–V. 2006.

[10] M. Kumon, T. Sugawara, K. Miike, I. Mizumoto, and Z. Iwai. Adaptive audio servo for multirate robot systems. In *IEEE Int. Conf. on Intelligent Robots and Systems*, volume 1, pp 182–187. 2003.

[11] M. Liu, C. Pradalier, Q. Chen, and R. Siegwart. A bearing-only 2d/3d-homing method under a visual servoing framework. In *IEEE Int. Conf. on Robotics and Automation*, pp 4062–4067. 2010.

[12] E. Malis, Y. Mezouar, and P. Rives. Robustness of image-based visual servoing with a calibrated camera in the presence of uncertainties in the three-dimensional structure. *IEEE Tr. on Robotics*, 26(1):112–120, 2010.

[13] H. Michel, P. Rives, et al. Singularities in the determination of the situation of a robot effector from the perspective view of 3 points. *INRIA Research Report, Tech. Rep. 1850*, 1993.

[14] J C. Murray, H. Erwin, and S. Wermter. Robotics sound-source localization and tracking using interaural time difference and cross-correlation. In *Proc. of NeuroBotics Workshop*, pp 89–97, 2004.

[15] A. Portello, P. Danes, and S. Argentieri. Active binaural localization of intermittent moving sources in the presence of false measurements. In *IEEE Int. Conf. on Intelligent Robots and Systems*, pp 3294–3299. 2012.

[16] J-M. Valin, F. Michaud, and J. Rouat. Robust localization and tracking of simultaneous moving sound sources using beamforming and particle filtering. *Robotics and Autonomous Systems*, 55(3):216–228, 2007.

[17] J-M. Valin, F. Michaud, J. Rouat, and D. Létourneau. Robust sound source localization using a microphone array on a mobile robot. In *IEEE Int. Conf. on Intelligent Robots and Systems*, volume 2, pp 1228–1233. 2003.