ACOUSTIC RECONSTRUCTION OF THE GEOMETRY OF AN ENVIRONMENT THROUGH ACQUISITION OF A CONTROLLED EMISSION

D. Aprea, F. Antonacci, A. Sarti, S.Tubaro

Dipartimento di Elettronica ed Informazione - Politecnico di Milano
Piazza Leonardo da Vinci, 32, 20133 Milano - Italy

ABSTRACT

This paper presents a novel solution for the reconstruction of an environment from the acquisition of a controlled emission, in particular for what concerns the estimation of the position of acoustic reflectors. The solution proposed in this paper makes use of a loudspeaker rotating on a circular pattern and emitting a controlled noise and a microphone located at the center of the circle. A likelihood map is built by means of a template matching between the signal acquired at the microphone and a template signal obtained by simulating the propagation of the signal to all the potential obstacles locations. The position of the reflectors is inferred by the analysis of the likelihood map. Simulations assess the best working conditions for the device in terms of test signal duration, radius of the circular trajectory and Signal to Noise Ratio. A simulation has been carried out also in the presence of multiple and interacting reflectors to show the feasibility of the approach also in more complex scenarios.

1. INTRODUCTION AND RELATED WORK

The problem of acoustic scene reconstruction is an increasing field of research. The goal is to assess, by means of the joint emission and acquisition of an acoustic test signal, the configuration of obstacles (e.g. reflectors) in the environment. Historically, one of the first research fields interested in the problem of scene reconstruction was that of underwater inspection: the water turbidity prevents in fact the use of traditional cameras to assess the configuration of reflectors. In [1] Castellani et al. propose to use *acoustic cameras*, that are bi-dimensional arrays of microphones. Acoustic cameras aim at reconstructing location, principal dimensions and possibly shape of obstacles in the environment. In particular the authors propose to use multiple acoustic cameras to attain a three dimensional reconstruction. In [2] the authors fuse the information coming from video and acoustic cameras to obtain a high resolution image of the environment.

Underwater acoustics inspection is interesting from a historical point of view, however the problems met in this field are quite different from sound propagation in air. In the last few years spherical arrays [3], [4] have seen an increasing interest in sound processing due to their resolution together with their compactness. For this reason spherical arrays are good candidates for scene reconstruction. In [5] and [6] the authors propose the adoption of spherical arrays to infer the temporal sequence of reflections in the environment (in that case a concert hall) together with the three dimensional directions of arrival.

This work acknowledged the financial support of the SCENIC project, under the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET-Open grant number: 226007

The solutions shortly described above present, however, the disadvantage of requiring a large set of microphones. In order to overcome this issue, one may think to use a single microphone and to accomplish several measures in a sequential fashion, the microphone being moved from one measurement to the other on a specific path. In this paper we will elaborate on this device: we will present a solution which uses a microphone and an omnidirectional loudspeaker rotating on a circular pattern in a continuous fashion. The rotation of the loudspeaker induces a time-dependent impulse response between the microphone and the loudspeaker and makes it possible to discern reflections coming from objects located at different positions. In particular, we build a likelihood map of the reflector position based on the template matching process. Much information can be extracted from the likelihood map. In this paper we will estimate the position of non-absorptive obstacles. We will characterize the estimator through theoretical and experimental approaches. Moreover, we will show what happens when two or more obstacles are present. The rest of the paper is organized as follows: Section 2 describes the theory and the mathematical aspects of the solution. Section 3 presents some experimental results to show the feasibility of the technique. Finally Section 4 summarizes the paper and presents future work on the topic.

2. THE PROPOSED SOLUTION

We start this section with the data model used throughout the rest of the paper. Thereafter we will derive the likelihood map. In the next sections we will refer explicitly to the case of 2D environments. However, the following considerations apply also (with little changes) to 3D coordinates.

2.1 Data Model

Consider the geometry presented in Figure 1. A microphone located at point $\mathbf{o}$ (which is also the center of our reference frame) captures the signal $s_i(t)$ while a loudspeaker moves on the circular trajectory $\mathbf{p}_l(t)$ and emits the controlled noise $s_l(t)$. Let us call with $h(\tau,t)$ the time-dependent environment impulse response between $\mathbf{p}_l(t)$ and $\mathbf{o}$. $s_i(t)$ is the result of a time-dependent filtering between $s_l(t)$ and $h(\tau,t)$:

$$s_i(t) = \int_0^\infty s_l(t) \cdot h(\tau,t) d\tau + n(t), \qquad (1)$$

where $n(t)$ is the additive environmental noise. We will start our treatment with the case of a single reflector, denoted in Figure 1 with the symbol $\mathbf{w}$. Later, we will generalize to the case of multiple reflectors. The perpendicular to the reflector through $\mathbf{o}$ hits the reflector in $\mathbf{p}_0(\rho_0,\theta_0)$. $d_{\mathbf{p}_0}(t)$ is the distance from $\mathbf{p}_l(t)$ to $\mathbf{w}$ and then from $\mathbf{w}$ to $\mathbf{o}$. We assume

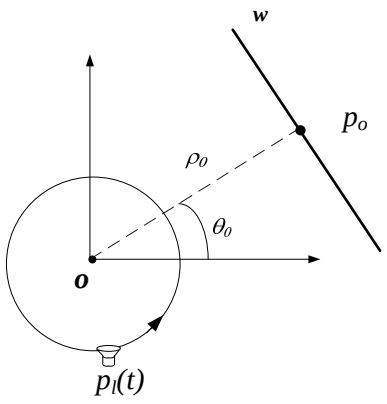

Figure 1: Description of the device and notation used throughout the rest of the paper. $\mathbf{o}$ is the position of the microphone and the center of our reference frame, $mathbf p_0$ is the intersection of the perpendicular line to the reflector $\mathbf{w}$ with the reflector itself. ρ_0 and θ_0 are the polar coordinates of $\mathbf{p}_0$. The point $\mathbf{p}_l(t)$ is the position of the emitter.

that the theory of optical acoustics is valid, therefore the reflective path must honour the Snell's law. The term $\alpha_{\mathbf{p}_0}$ is the corresponding attenuation. The time dependent impulse response of the reflective path is:

$$h(\tau,t) = \alpha_{\mathbf{p}_0}\delta(\tau - \frac{d_{\mathbf{p}_0}(t)}{c}) \,. \tag{2}$$

We observe that (2) introduces an approximation, since we are considering the distance $d_{\mathbf{p}_0}(t)$ at the acquisition time, while the distance at the time at which the signal was emitted should appear instead. This approximation, however, does not imply significant errors for our purposes.

In order to simplify the treatment, we assume that $|\mathbf{p_0}| \gg |\mathbf{p}_l(t)| \; \forall t$, which means that the reflector is far more distant from the microphone than the loudspeaker. In this context we have that:

$$d_{\mathbf{p}}(t) \approx 2|\mathbf{p}| - \frac{\mathbf{p}_l(t)\cdot\mathbf{p}}{|\mathbf{p}|} \,, \tag{3}$$

Our goal is to compare the signal $s_i(t)$ with a template signal $s_{\mathbf{p}}(t)$ built under the hypothesis that the reflector is placed in $\mathbf{p}$. More specifically, we find the estimate $\hat{p}$ as the position that maximizes the coherence of $s_i(t)$ with $s_{\mathbf{p}}(t)$. The template signal $s_{\mathbf{p}}(t)$ is built by assuming that the signal $s_l(t)$ is reflected from an obstacle in $\mathbf{p}$ and acquired at $\mathbf{o}$. The template signal $s_{\mathbf{p}}(t)$ has a non stationary nature due to the time-dependent impulse response from the loudspeaker to the microphone. Considering that the obstacle is a perfect reflector (i.e. optical acoustics is valid) and ignoring the attenuation term, we get that the impulse response that relates $s_l(t)$ with $s_{\mathbf{p}}(t)$ is:

$$h_{\mathbf{p}}(\tau,t) = \delta(\tau - \frac{d_{\mathbf{p}}(t)}{c}) \,, \tag{4}$$

where the meaning of the terms is analogous to (2). Thereafter, the template signal for point $\mathbf{p}$ is:

$$s_{\mathbf{p}}(t) = \int_0^\infty s_l(\tau)\cdot h_{\mathbf{p}}(\tau,t)d\tau. \tag{5}$$

In a time-discrete implementation of (5), the signal $s_{\mathbf{p}}(t)$ is computed as a time-dependent delay of the signal $s_l(t)$.

From equations (4) and (3) we can observe that all the points in the same direction of $\mathbf{p}$ are characterized by an impulse response which is a delayed version of $h_{\mathbf{p}}(\tau,t)$. In fact if we consider a point $\mathbf{p}'$ that shares the same direction as $\mathbf{p}$ but it is placed at a different distance (i.e. $\mathbf{p}' = a\mathbf{p}$), it can be demonstrated that the signal $s_{a\mathbf{p}}(t)$ is obtained as a time-delay of $s_{\mathbf{p}}(t)$:

$$s_{a\mathbf{p}}(t) = s_{\mathbf{p}}(t - a') \,, \tag{6}$$

where $a' = 2(a-1)|\mathbf{p}|/c$. This observation is important since it enables us to build templates for different angles and then to obtain templates for the same angles but different distances just as a time-shift of them.

2.2 Template matching

We define the likelihood map as the correlation between $s_{\mathbf{p}}(t)$ and $s_i(t)$:

$$m(\mathbf{p}) = \frac{1}{T}\int_0^\infty s_{\mathbf{p}}(t)\cdot s_i(t)dt \,, \tag{7}$$

where T is the duration of $s_l(t)$. If we substitute (1) into (7), we use the definition of convolution and we take the expectation we obtain that:

$$E[m(\mathbf{p})] = \int_0^\infty\int_0^\infty \alpha_{p_0}(h_{\mathbf{p}}(\tau,t) * r_{s_l}(\tau))h(\tau,t)dtd\tau \,, \tag{8}$$

where $r_{s_l}(t)$ is the autocorrelation of $s_l(t)$. In order to make more explicit the importance of equation (8), we consider the specific case of $\mathbf{p}_l(t)$ covering a circular trajectory. In this case (3) assumes the form:

$$d_{\mathbf{p}}(t) = 2|\mathbf{p}| - \frac{\mathbf{p}_x}{|\mathbf{p}|}R\cos(\omega t) - \frac{\mathbf{p}_y}{|\mathbf{p}|}R\sin(\omega t) \,. \tag{9}$$

Let us assume the signal $s_l(t)$ to be white noise between 0 and T. The signal is low-pass filtered to keep into account for the transfer function of electronic devices. We remark, however, that the distortion brought by an uncorrect estimation of the transfer function is negligible. With these assumptions we obtain that

$$m(\mathbf{p}) = \frac{1}{T}\int_0^T s_l(t - \frac{d_{\mathbf{p}}(t)}{c})[s_l(t - \frac{d_{\mathbf{p}_0}(t)}{c}) + n(t)]dt \,. \tag{10}$$

We remark that the template matching in (10) computed the likelihood of the presence of an image source in point $\mathbf{p}$. When multiple reflectors are present, the same reasoning holds: template signals are built for each point in space. The resulting acoustic map is the sum of the acoustic maps of each individual source, due to the linearity of the cross-correlation operator involved in (10). In the next paragraph we will see how the estimation of the obstacles location can be attained from (10).

2.3 Assessment of obstacles location

By taking the expectation of (10) we obtain that:

$$E[m(\mathbf{p})] = \frac{1}{T}\int_0^T \alpha_{p_0} r_g(\frac{d_{\mathbf{p}}(t)}{c} - \frac{d_{\mathbf{p}_0}(t)}{c})dt \,. \tag{11}$$

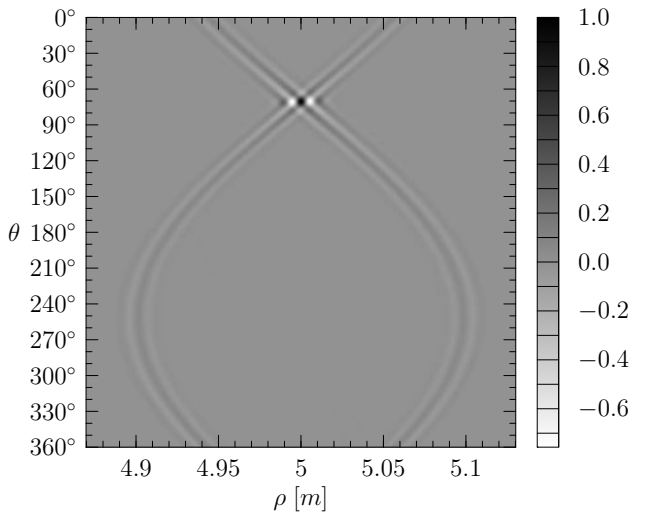

Figure 2: Zoom on $E[m(\mathbf{p})]$ when $R = 0.1$ m and a reflector is placed at $\rho = 5$ m from the microphone and at an azimuth angle of $\theta = 70$ degrees. Map is plotted in polar coordinates

The term $\frac{d_{\mathbf{p}}(t)}{c} - \frac{d_{\mathbf{p}_0}(t)}{c}$ is identically zero for $\mathbf{p} = \mathbf{p}_0$, while it is generally different from zero if $\mathbf{p} \neq \mathbf{p}_0$. It is easy to observe that $E[m(\mathbf{p})]$ assumes the maximum value if $\mathbf{p} = \mathbf{p}_0$. This is independent from the particular choice of the test signal $s_l(t)$. Figure 2 shows a zoom in on $E[m(\mathbf{p})]$ when $R = 0.1$ m, the reflector is placed 5 m from the microphone and at an angle of 70 degrees. The signal $s_l(t)$ is a band pass signal between 10kHz and 20kHz. The map is plotted in polar coordinates. We can observe a very sharp peak for $E[m(\mathbf{p} = \mathbf{p}_0)]$. However, we note that $\mathbf{p}_0$ is centered on an "X-shaped" curve in the (θ, ρ) axes, which means that other points different from $\mathbf{p}_0$ partially match with $s_i(t)$. The shape of this curve is characteristic of the loudspeaker trajectory. Finally, we infer the position of the reflector through:

$$\hat{\mathbf{p}}_0 = \arg\max_{\mathbf{p}} m(\mathbf{p}) \,. \tag{12}$$

We observe that the integrand function in (10) is non-gaussian [7]. However, due to the integral summation in (10), when T is sufficiently long, we can invoke the central limit theorem and state that $m(\mathbf{p})$ is gaussian. In a scenario of multiple reflectors the likelihood map presents multiple local maxima, each corresponding to the position of a reflector. When reflectors are mutually visible, we will observe also local maxima corresponding to signal coming from multiple reflections. A simulation in Section 3 concerns this case.

2.4 Robustness of the estimator against additive noise

In order to validate the estimator, it is necessary to assess its robustness against additive environmental noise. In particular, we will consider the ratio between the signal and noise components in $m(\mathbf{p}_0)$. We will refer to this measure as *PSNR* (Peak-SNR). *PSNR* is useful because it reveals the ability of the estimator in (12) to distinguish between peaks related to obstacles and noise. More specifically, we relate *PSNR* with the noise and loudspeaker powers, denoted by σ_n^2 and P_g. After some passages and assuming that the signal duration is longer than the de-correlation time of $s_l(t)$ we get:

$$PSNR \approx \alpha_{\mathbf{p}_0}^2 \frac{P_g}{\sigma_n^2} = \alpha_{\mathbf{p}_0}^2 SNR \,, \tag{13}$$

In Figure 3 we plot the contour lines of *PSNR* (values expressed in *dB*) predicted according to (13) for various SNR's and distances of the obstacle. Equation (13) is important not only from a theoretical viewpoint but also from an operating one. Consider the scenario in which we have to measure an environment with the proposed technique. From an initial assessment we can estimate the noise level σ_n^2 and the principal dimensions of the room that determine the attenuation term $\alpha_{\mathbf{p}_0}^2$. Therefore by using (13), or by inspection of Figure 3, we estimate the power P_g of the signal to be emitted by the loudspeaker.

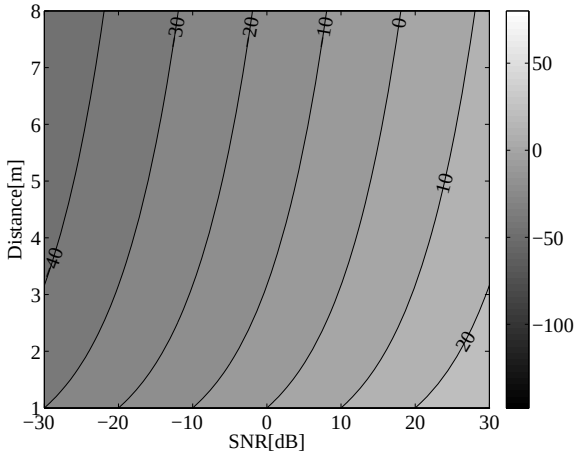

Figure 3: Contour lines of *PSNR* for various distances of the obstacle and SNR's. We observe that PSNR improves as the SNR increases and/or the obstacle gets closer to the rotating device.

3. EXPERIMENTAL RESULTS

This Section is divided into three parts. First, we will simulate the effect of additive noise at different work conditions (i.e. rotation speeds and radii of the circular trajectory). The second experiment demonstrates the effectiveness of the algorithm while varying the level of the additive noise. The last experiment shows an example of $m(\mathbf{p})$ in a multiple reflectors scenario.

According to the particular scenario we have considered, we have used either the image source or the beam tracing to simulate the acquisition of the reflected signal. Both methodologies enable us to accurately simulate the reflective paths in far and near fields. As motivated in Section 2, the proposed algorithm assumes that the image source is in the far field. As a consequence, we expect that in some situations our algorithm fails. It is worth to notice that beam tracing and image sources do not account for other propagation phenomena such as diffraction and diffusion. Therefore, some distortion to the acoustic map may appear in a real scenario. Nonetheless, from some preliminary experiments (a short demo is available at [8]) we have observed a good match between the predicted and observed acoustic maps.

3.1 Variation of device parameters

In order to estimate the effectiveness of the algorithm with respect to the parameters of the device, we simulated through image source [9] the acquisition of a signal in a simple scenario: a reflector is placed in a dry room at $\rho = 5$ m and $\theta = 130°$. We verified the estimation error for different durations of the test signal (between 0.06 s and 0.6 s), and radii of the circular trajectory (between 0.02m and 0.4m). In particular, the term ω in (9) is related to the signal duration: the signal $s_l(t)$ lasts for a single rotation of the device. For each point in the grid (T,R) the average RMS angular error is computed over $N = 30$ realizations. SNR has been kept fixed at 40dB. Figure 4 shows the RMS of the angle estimation error in the range $R = [0.02\text{ m}, 0.2\text{ m}]$ and $T = [0.06\text{s}, 0.6\text{s}]$. We can notice that for R higher than 0.1 m the localization capabilities of the device greatly improve. This fact can be easily explained if we consider that, as R increases, the variability of the delay $d_{\mathbf{p}}(t)/c$ increases too, thus making the template signals $s_p(t)$ "much more different" for different positions. Figure 5 is similar to Figure 4 but is built in the interval

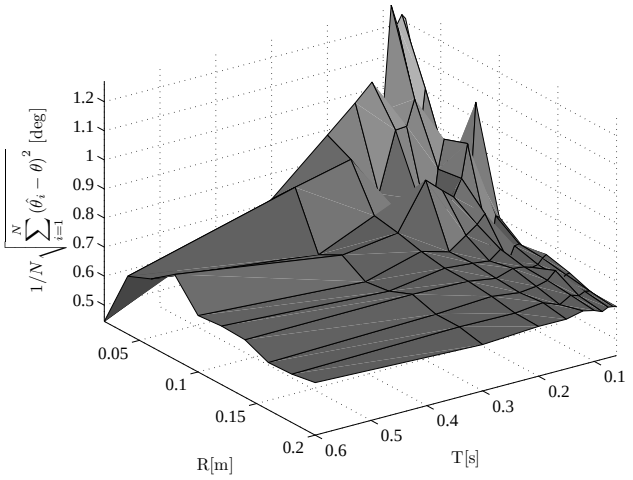

Figure 4: RMS of the angle estimation error in the range $R = [0.02\text{ m}, 0.2\text{ m}]$ and $T = [0.06\text{ s}, 0.6\text{ s}]$. We observe that increasing the radius R of the device and/or the duration T of the signal we improve the effectiveness of the algorithm.

$R = [0.22\text{ m}, 0.4\text{ m}]$. We can observe a sudden decrease of the localization capabilities for $R > 0.24$ m, independently from the signal duration. This is not surprising, since for these radii the condition $|\mathbf{p}_0| \gg R$ does not hold anymore and the template signal $s_{\mathbf{p}}(t)$ built according to the approximation in (3) significatively differs from the signal $s_i(t)$. We conclude that a trade-off between resolution capabilities (which involves the use of larger R) and far-field approximation (use of smaller R) is necessary to obtain efficient localization of obstacles. As far as the signal duration is concerned, we can observe from Figure 4 that we attain better performances if we use longer signals.

3.2 Robustness against additive noise

In the second simulation we have verified with experimental results the robustness of the estimator in (12): we have simulated the presence of a reflector at a distance variable in the range $[0.5m, 8m]$. The angle of the reflector is $\theta = 130°$.

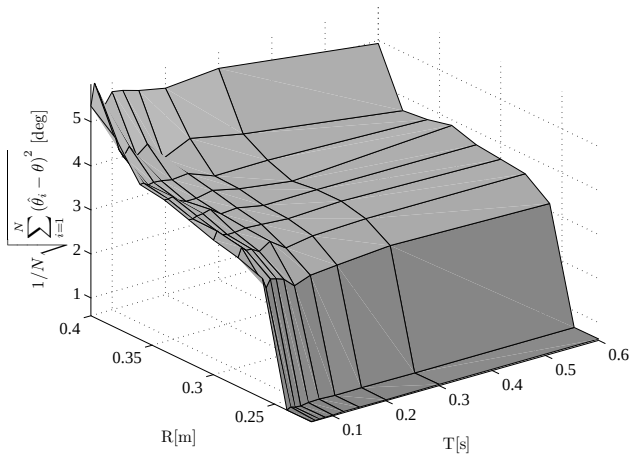

Figure 5: RMS angle estimation error in the range $R = [0.22\text{ m}, 0.4\text{ m}]$ and $T = [0.06s\text{ s}, 0.6\text{ s}]$. We observe that for radii greater than 0.25 m (which corresponds to a ratio $R/\rho_0 \geq 0.05$) a threshold behavior of the algorithm since we do not meet the far field hypothesis.

The radius of the trajetory is $R = 0.1$ m and $T = 0.6s$ s($\omega = 100\,rpm$) and we have made the SNR variable in the interval $[-18dB, 30dB]$. Figure 6 shows the RMS distance estimation error over $N = 10$ realization of the experiment. As in other

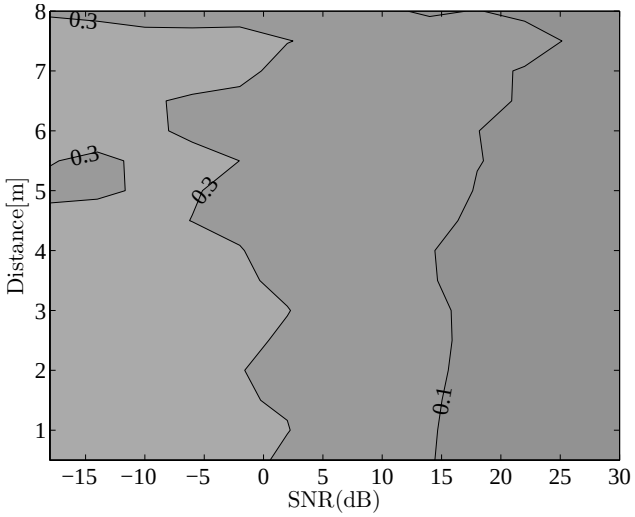

Figure 6: Contour lines of the RMS distance estimation error (meters) for variable SNR and obstacle distance. We observe a good match between the RMS distance estimation error and the contour lines of PSNR, shown in Figure 3.

estimators (i.e. Generalized Cross Correlation), we observe a threshold behavior. The plot of the RMS angle estimation error exhibits the same behavior and therefore is omitted. From a joint analysis of Figures 3 and 6 we observe a rough correspondence between the contour lines of *PSNR* in Figure 3 and the RMS distance estimation error. This analogy is useful under a practical viewpoint: in fact once we are given the range of distances of the obstacles and the desired precision, we determine the correct signal level to be used simply by analysis of (13).

3.3 A simulation with multiple reflectors

We now simulate the presence of multiple mutually visible reflectors. The environment, together with the reference frame, is plotted in Figure 7: two reflectors that extend from $y = -5$ m to $y = 5$ m are placed at $x = -5$ m and $x = 5$ m. Their reflection coefficient is 0.9. The environment impulse response has been simulated using fast beam tracing [10]. The parameters of the device in this experiment are: $R = 0.1$ m, $T = 0.6$ s $\omega = 100$ *rpm* and $SNR = 20dB$. Figure 8 shows $m(\mathbf{p})$ in polar coordinates. We observe that the "X-shaped" curves visible in Figure 2 are not evident in Figure 8. This is due to the fact that the distance range on which we focus in 2 (0.2 m)is much smaller than the distance range in Figure 8. Even if they are not evident, "X-shaped" patterns are present also in Figure 8. We observe the presence of multiple peaks, related to single and higher order reflections. The text superimposed on the image symbolically describes the sequence of reflections that generated the relative events on the likelihood map. Even if in this experiment we have used a reflection coefficient close to 1, the attenuation effect makes the secondary reflections much more dimmed than the primary ones.

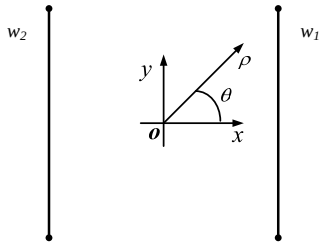

Figure 7: Environment for the experiment with multiple reflectors. w_1 and w_2 represent the obstacles (walls). The rotating device is placed between the two walls. Since the two walls are mutually visible we expect to observe, together with first order ones, higher order reflections.

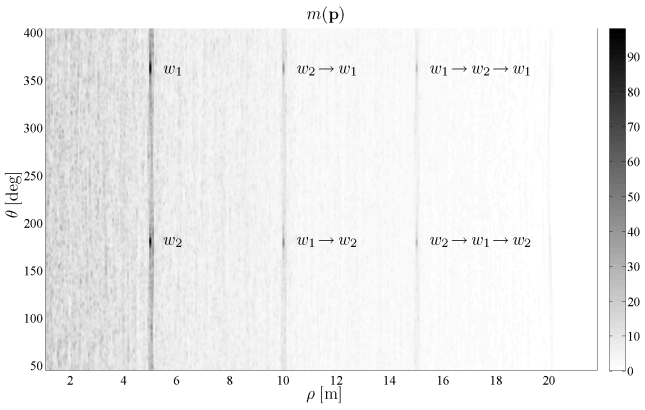

Figure 8: Likelihood map for the experiment with multiple reflectors. We observe the presence of first and higher order reflections. The label close to each peak in the acoustic maps denote the reflective path that links source and receiver through the walls $\mathbf{w}_1$ and $\mathbf{w}_2$.

4. CONCLUSIONS AND FUTURE WORK

In this paper we have presented a novel technique to assess the position of reflectors by means of a single microphone and a probing signal emitted by a loudspeaker which moves on a circular trajectory. The esteem is found as the maxima of a likelihood map built with template matching. In a multiple reflector scenario higher order reflections are visible from the likelihood map. Experimental results demonstrate the robustness of the estimator with respect to the SNR.

We are now working on a generalization of the work presented in this manuscript. More specifically, we are generalizing the construction of the likelihood map to the case of arbitrary trajectories. Furthermore we will remove the hypothesis of a continuous trajectory. A first demonstration of the video is available at [8].

REFERENCES

[1] U.Castellani, A.Fusiello, V.Murino, L.Papaleo, E.Puppo, and M.Pittore, "A complete system for on-line 3D modelling from acoustic images," in *Proc. of Signal Proc. Image Comm*, 2005, vol. 20, pp. 832–852.

[2] S. Negahdaripour, H. Sekkati, and H. Pirsiavash, "Opti-acoustic stereo imaging, system calibration and 3-d reconstruction," in *Proceedings of IEEE Conference Computer Vision and Pattern Recognition CVPR '07.*, June 2007, pp. 1–8.

[3] J.Meyer and G.Elko, "Spherical harmonic modal beamforming for an augmented circular microphone array," in *Proceedings of IEEE ICASSP 2008*, 2008.

[4] Z. Li and R. Duraiswami, "Flexible and optimal design of spherical microphone arrays for beamforming," *IEEE Transactions on Audio, Speech, and Language Processing*, Feb. 2007.

[5] A.O'Donovan, R. Duraiswami, and D. Zotkin, "Imaging concert hall acoustics using visual and audio cameras," in *Proceedings of IEEE ICASSP 2008*, 2008.

[6] A.O'Donovan, R.Duraiswami, and Jan Neumann, "Microphone arrays as generalized cameras for integrated audio visual processing," in *Proc. IEEE International Conference Computer Vision and Pattern Recognition (CVPR-07)*, 2007, vol. 1, pp. 1–8.

[7] V.V.Khlobystov and V.K.Zadiraka, "Distribution density of scalar product of Gaussian vectors," *Cybernetics and Systems Analysis*, vol. 8, no. 3, pp. 477–481, 1972.

[8] "Demonstration of the rotating device," Apr. 2009, `http://www-dsp.elet.polimi.it/ispg/SCENIC/index.php?option=com_remository&Itemid=13&func=fileinfo&id=8`.

[9] J. Borish, "Extension of the image model to arbitrary polyhedra," *J. of the Acoustical Society of America*, vol. 75, no. 6, 1984.

[10] M. Foco, P. Polotti, A. Sarti, and S. Tubaro, "Sound spatialization based on fast beam tracing in the dual space," in *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFx-03), London, Great Britain*, Sep. 2003.