

Acoustic source localization using a polyhedral microphone array and an improved generalized cross-correlation technique

Thomas Padois^{a,b}, Franck Sgard^b, Olivier Doutres^a, Alain Berry^c

^a*Department of Mechanical Engineering, École de Technologie Supérieure (ÉTS),
Montréal, (Qc), Canada, H3C 1K3*

^b*IRSST, Montréal, (Qc), Canada, H3A 3C2*

^c*Department of Mechanical Engineering, Université de Sherbrooke, Sherbrooke (Qc),
Canada, J1K 2R1*

Abstract

Millions of workers are exposed to excessive noise levels each day. Acoustic solutions have to be developed to protect workers from hearing loss. The first step of an acoustic diagnosis is the source localization which can be performed with a microphone array. Spherical microphone arrays can be used to detect the acoustic source positions in a workplace. In this study, a spherical microphone array, with polyhedral discretization, is proposed and compared with a spherical array with a slightly different geometry. The generalized cross-correlation technique is used to detect the source positions. Moreover, two criteria are introduced to improve the noise source map. The first is based on the geometric properties of the microphone array and the scan zone whereas the second is based on the energy of the spatial likelihood function. Numerical data are used to provide a systematic comparison of both geometries and criteria. Finally, experiments in a reverberant room reveal that the polyhedral microphone array associated with both criteria provide the best noise source map.

Keywords: acoustic source localization, spherical microphone array, generalized cross-correlation, reverberant environment

1. Introduction

Each year, millions of workers are exposed to excessive noise levels. In Québec, between 400,000 and 500,000 of the 2,5 million workers are exposed

4 daily to noise levels likely to generate hearing impairment. Hearing loss is by
5 far the most recorded occupational disease [1] and its total cost to the society
6 is the highest [2]. Therefore noise control solutions have to be implemented to
7 continue prevention efforts. To reduce noise levels at the workstation, direct
8 noise sources and their reflections by the environment have to be detected.

9 Microphone array techniques have become widespread for noise source
10 localization [3, 4, 5, 6]. These techniques are based on the calculation of the
11 propagation delay between a pair of microphones in the time domain or the
12 phase shift in the frequency domain. An accurate assessment of a source
13 location requires several microphones. Commonly, microphone array mea-
14 surements lead to an acoustic image or noise source map coded with colors
15 to highlight the noise source locations [7]. Depending on the situation, the
16 microphone array processing, the number of microphones and the array geom-
17 etry may differ. Detecting aeroacoustic sources, speakers in a meeting room
18 or noise sources in industrial halls does not involve the same constraints and
19 challenges.

20 For transportation applications, like cars, trains or aircrafts, there is
21 no ambiguity on the source location with respect to the microphone array.
22 Therefore a planar acoustic array geometry is preferred. Today, the most
23 common planar geometry corresponds to several spiral arms which provide an
24 acoustic image with moderate side lobes and operate over a broad frequency
25 range [8]. The microphone array technique commonly used is the conven-
26 tional beamforming in the frequency domain based on the cross-spectral ma-
27 trix of the microphone signals [9]. Due to the poor resolution of the acoustic
28 image provided by beamforming, deconvolution techniques or inverse prob-
29 lems are used to clean up the noise map [10, 11, 12, 13].

30 Microphone array techniques are also used in speech processing to detect
31 speakers in a meeting room [14, 15]. In this case, the microphone array
32 involves few microphones and is located either at the center of the room on
33 the table or on a wall [16, 17]. The source is the human voice and the goal
34 is to track the speakers. The main constraints are a high background noise
35 and a reverberant environment. The common microphone array technique
36 used is the beamforming, also called the steered response power based on the
37 Generalized Cross-Correlation (GCC) of the microphone signals [18].

38 In workplaces such as an industrial hall, the dimensions of the rooms
39 are very large and the reverberation time may be high (more than one sec-
40 ond). The sound sources are usually broadband and may come from any
41 direction therefore spherical microphone arrays are preferred. A 32-spherical

microphone array associated to eigenbeam beamforming processing is used to localize a main source and its reflections [19]. The obtained noise source maps allowed for a correct source localization. However, the processing had to be done at each frequency which is time consuming for broadband sources, moreover the number of microphones is high. Time-domain algorithms based on the GCC can be used to create a noise source map showing the main sources and potential reflections from the walls, floor or ceiling. However, the noise source map obtained may not be accurate due to the presence of large side lobes or high background noise and inverse problems have to be used to improve the resolution [20]. Noël *et al.* [21] developed a time domain technique associated with an inverse problem to detect source positions in complex industrial environments. Despite a correct source localization, the computation time of the inverse problem is very high. Moreover, their microphone array geometry is not well adapted for omnidirectional localization due to the absence of microphones at the top and bottom of the array.

The objective of the paper is to propose a fast signal processing technique with few microphones to localize acoustic sources in a reverberant environment. First, instead of using post-processing techniques to improve the noise source maps through the solution of associated inverse problems, the focus is rather put on generating a clean initial noise source map. Each microphone pair leads to a Spatial Likelihood Function (SLF) [22]. The noise source map is then obtained by summing up all the SLFs. Two criteria, based on the SLF energy or geometry of the microphone array and scan zone, are proposed to discard the microphone pairs which do not bring useful information. Then a polyhedral microphone array geometry with a small number microphones is proposed. The 14-microphone array is compared to the spherical microphone array developed by Noël *et al.* [21]. The main originality of this work is that the proposed technique does not operate on the *a posteriori* noise source map but acts before generating the noise source map by removing SLFs with useless information. Moreover, by removing useless SLFs and using few microphones, the computation time and complexity is not increased.

Section 2 presents the source localization method and criteria used to compute the noise source map. Then, numerical data are used to assess the proposed microphone array geometry and criteria in Section 3. Finally, the performances of the microphone array geometries and criteria are confirmed by experiments carried out in a fully reverberant room (section 4).

2. Source localization method

2.1. Acoustic signal model

The objective of this section is to present the theoretical formulation of the source localization technique. The acoustic pressure source signal $s(\mathbf{r}_s, t)$ is generated by an omnidirectional point source at location \mathbf{r}_s and recorded by a set of M free-field microphones at locations \mathbf{r}_m . The acoustic pressure signal x_m recorded by the microphone m is given by

$$x_m(t) = \alpha_m(\mathbf{r}_s) s(\mathbf{r}_s, t - \Delta t_{ms}) + v_m(t), \quad (1)$$

where t represents time, $\alpha_m(\mathbf{r}_s)$ is the geometrical attenuation due to the propagation between the source and the microphone and $v_m(t)$ is an uncorrelated additive noise due to background or sensor noise. The term Δt_{ms} corresponds to the propagation delay between the source and the microphone, called Time of Flight (ToF), and is defined by the Euclidean distance

$$\Delta t_{ms} = \frac{1}{c_0} \|\mathbf{r}_m - \mathbf{r}_s\|_2, \quad (2)$$

where c_0 is the sound speed and $\|\cdot\|_p$ is the l_p -norm of a vector or a matrix. In a real scenario, the microphone signals are sampled at a frequency f_s and the processing is performed on discrete signals. The acoustic pressure microphone signal is then divided into segments of N points and the source localization technique presented in the following section may be performed frame by frame.

2.2. Conventional Beamforming

The main idea behind most source localization techniques is to compensate the time delay between each microphone. For a given sound speed, the ToF between the microphones and a set of L virtual sources at location \mathbf{r}_l is computed. Each microphone signal is delayed by the corresponding ToF, which corresponds to steer the microphone array in the direction of the virtual source. When the sum of the delayed microphone signals is maximized the source is localized. This technique is known as delay-and-sum beamformer and its output $y_l(t)$ can be expressed for the virtual source l location as

$$y_l(t) = \sum_{m=1}^M x_m(t + \Delta t_{ml}), \quad (3)$$

106 where Δt_{ml} is the time delay which steers the beamformer into a virtual
 107 source location. The beamformer output power Y^{BF} for a virtual source
 108 location l is given by

$$Y_l^{BF} = E\{y_l^2(t)\} = \sum_{m=1}^M \sum_{n=1}^M R_{x_m, x_n}(\Delta t_{ml} - \Delta t_{nl}), \quad (4)$$

109 where $E\{\cdot\}$ is the mathematical expectation and R_{x_m, x_n} the cross-correlation
 110 function between the signals x_m and x_n defined by

$$R_{x_m, x_n}(\tau) = E\{x_m(t)x_n(t + \tau)\}, \quad (5)$$

111 where τ is the time lag. The beamformer output power is obtained after the
 112 summation process over all microphones. The estimated source position \mathbf{r}_e
 113 is provided by the peak value of the beamformer among all virtual source
 114 locations

$$\mathbf{r}_e = \arg \max_l (Y_l^{BF}). \quad (6)$$

115 The set of virtual source locations is called a scan zone and may be in one, two
 116 or three dimensions. In this study, the scan zone is a plane and the result ob-
 117 tained is a two dimensional image called noise source map. The noise source
 118 map is composed of a main lobe (source position), side lobes and spurious
 119 lobes due to the microphone array geometry. Commonly, the computation
 120 of the beamformer output is performed over the $M_p = M(M - 1)/2$ micro-
 121 phone pairs due to the symmetry of the cross-correlation matrix (redundant
 122 information). Moreover, the auto-correlation terms (auto-correlation of mi-
 123 crophone signals) are removed because they do not bring any information
 124 to the noise source map. These terms are known as "DC" component or
 125 self-noise. Usually, the beamformer output is computed thanks to the Gen-
 126 eralized Cross-Correlation (GCC) [18], which is a weighted cross-correlation
 127 function, therefore the noise source map obtained is called GCC in the fol-
 128 lowing.

129 2.3. Truncated cross-correlation

130 To obtain the beamformer output power, the time delay ($\Delta t_{ml} - \Delta t_{nl}$)
 131 between a virtual source location and a microphone pair has to be computed.
 132 Then, the cross-correlation of the microphone pair is estimated. Finally the
 133 value of the cross-correlation corresponding to the time delay is picked up and
 134 added to the source map. For a large number of virtual source locations and

number of microphones the computation time may be intractable. Rather than searching the time delay for each virtual source location, one by one, the time delays between all the virtual source locations and a microphone pair (m, n) are computed and stored into a time lag vector $\tau^{m,n}$

$$\tau^{m,n} = [|\Delta t_{m1} - \Delta t_{n1}|, \dots, |\Delta t_{ml} - \Delta t_{nl}|], \text{ with } l = 1, \dots, L, \quad (7)$$

with $|\cdot|$ the absolute value. This vector corresponds to the time delay between all the virtual source positions and the microphone pair (m, n) . Then, the peak value of the time lag vector is looked up. This time lag corresponds to the largest time delay possible for the considered microphone pair and all the virtual source positions. Therefore, the time delays of the cross-correlation exceeding this maximum time lag are removed because they cannot be reached. A truncated cross-correlation function referred to as \hat{R}_{x_m, x_n} is introduced,

$$\text{if } |\tau| < \max(\tau^{m,n}), \quad \hat{R}_{x_m, x_n}(\tau) = R_{x_m, x_n}(\tau), \quad (8)$$

\hat{R}_{x_m, x_n} is a vector corresponding to the N_t allowed time delays. Now, the goal is to match each truncated cross-correlation value (with size $[1 \times N_t]$) to the time lag vector (with size $[1 \times L]$). However, the size of the time lag vector and the truncated cross-correlation are different. Therefore, the truncated cross-correlation vector is interpolated over the time lag vector which means that for each virtual source location a truncated cross-correlation value is assigned. The result is called the Spatial Likelihood Function (SLF) [22] for a microphone pair. Theoretically, the maximum of SLF is a hyperboloid in three dimensions and a hyperbola in two dimensions. However, in real situations, the hyperbolas are spread out and blurred and rather represent hyperbolic areas [16], called hyperbolas in the following. These hyperbolic areas can be narrow or wide according to the microphone positions. Finally, the processing is repeated for all the microphone pairs. The number of SLF is thus equal to the number of microphone pairs. To create the noise source map, all the SLFs are summed up. The peak values of the noise source map, which are the locations where the hyperbolas intersect, correspond to the source positions.

2.4. Examples of Spatial Likelihood Function

The SLFs are investigated for three microphone pairs called vertical, diagonal and collinear configurations (see Figure 1.d-f, considered microphones

are plotted as red circles). The source generates a broadband signal. For each case, the cross-correlation is shown in (Figure 1.a-c). The SLF is shown in three dimensions and a slice is cut in the plane of the source (scan zone).

The vertical case corresponds to microphones at the same position but with a different z position. The source location is at equal distance from both microphones therefore the time delay is zero and the resulting SLF is a line as expected. In the case of the diagonal configuration, the time delay is not zero. This leads to a hyperboloid in three dimensions and a hyperbola in the scan zone plane. In the case of several microphone pairs, the intersection between the hyperbolas yields the source location. In the collinear configuration, the first microphone is in front of the second microphone, i.e. the first microphone is between the source and the second microphone on the same axis. In this case, the time delay is always negative because the acoustic waves always reach the first microphone before the second (for a source located in the scan plane). Therefore for the considered scan zone, the time delays between the scan points and the microphone pair are almost similar (very low variance). The resulting SLF is a cone in three dimensions and a large main lobe in two dimensions due to the intersection of the scan zone with the cone.

The SLF for each configuration are shown in the scan zone Figure 2.a-c). The final noise source map is obtained by summing all the SLFs provided by the microphone pairs, Figure 2.d-f). Owing to the last comment, some microphone pairs do not provide accurate information about the source location, such as the collinear configuration (Figure 2.e-f). The sum of SFLs with a narrow hyperbola area provides a more accurate localization (Figure 2.d) than the SFLs with a large main lobe. Criteria are discussed in the following to reject the SLF which provides inaccurate source localization.

2.5. Geometrical criterion (GC)

A Geometrical Criterion (GC) based on the vectors formed by a scan point and a microphone pair to improve the resolution of the noise source map obtained with GCC is proposed. Consider a scan point coordinate \mathbf{r}_l and a microphone pair defined by points \mathbf{r}_m and \mathbf{r}_n . Two vectors from the scan point to the microphones can then be defined. If the l_2 -norm of the normalized cross product of these two vectors is small, the angle θ between the two vectors is small and an inaccurate SLF for the localization, similar

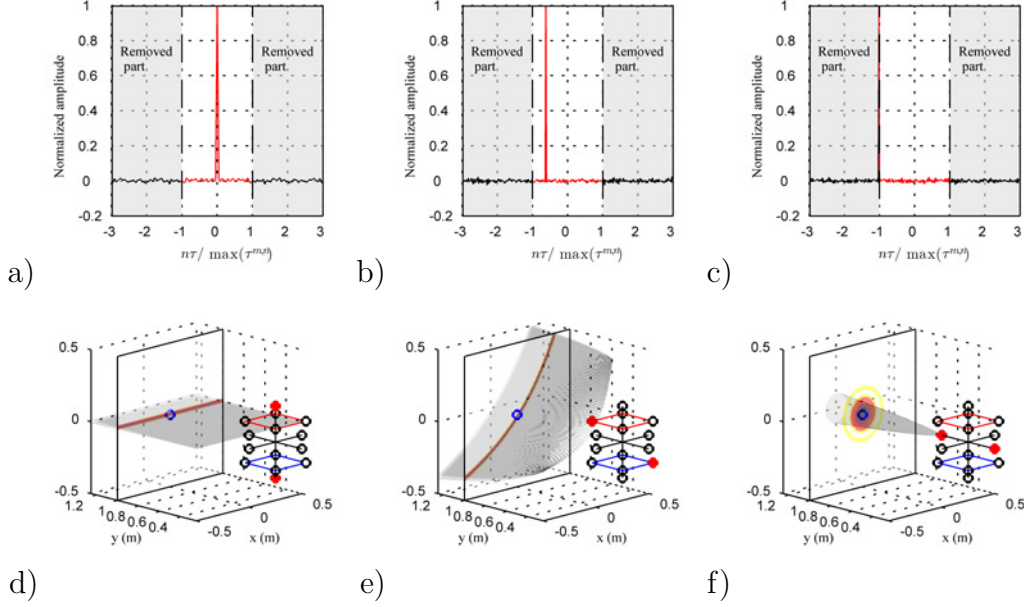


Figure 1: Cross-correlation between a microphone pair signal (top) and SLF (bottom) for three microphone pair configurations, a-d) vertical, b-e) diagonal and c-f) collinear. The microphone array center is at $y = 0$ m and $z = 0$ m. The filled red dots represent the microphone pair of interest. The black circles are the other microphone positions. The scan zone is the black square at $y = 1$ m and the blue circle is the real source position. The shaded surface is the hyperboloid.

to the collinear configuration, may result,

$$\frac{\|\overrightarrow{\mathbf{r}_m - \mathbf{r}_l} \wedge \overrightarrow{\mathbf{r}_n - \mathbf{r}_l}\|_2}{\|\overrightarrow{\mathbf{r}_m - \mathbf{r}_l}\|_2 \|\overrightarrow{\mathbf{r}_n - \mathbf{r}_l}\|_2} = \sin(\theta). \quad (9)$$

The GC criterion is presented as a spatial weighting function $W_{mn}(\mathbf{r}_l)$ multiplying the SLF (Eq. 8). When the angle θ is lower than β_{GC} , the value of the SLF (corresponding to a scan point location) is set to zero

$$W_{mn}(\mathbf{r}_l) = \begin{cases} 1 & \text{if } \theta > \beta_{GC} \\ 0 & \text{otherwise.} \end{cases} \quad (10)$$

The GC criterion depends on a scan point and a microphone pair, therefore some scan points of a SLF or the whole SLF may be discarded. In the following, this method is called GCC-GC.

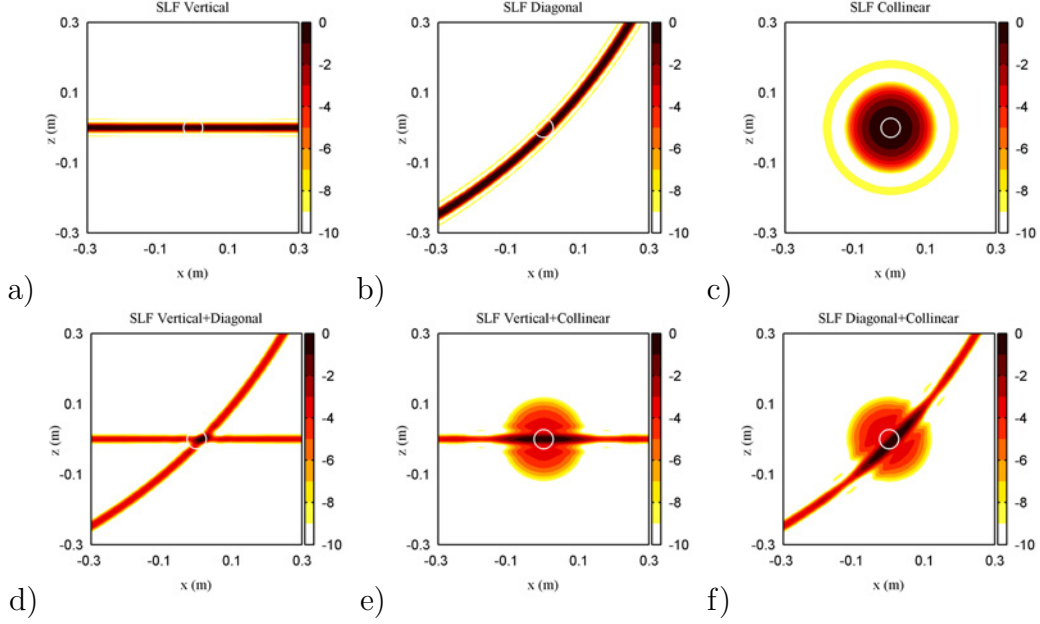


Figure 2: Spatial Likelihood Function (SLF) for three microphone pairs configurations, a) vertical, b) diagonal and c) collinear. Sum of the (SLF) d) vertical + diagonal, e) vertical + collinear and f) diagonal + collinear. The white circle is the source position.

2.6. Energy criterion (EC)

An alternative Energy Criterion (EC) is proposed and called GCC-EC. As shown in figure 2.a-c the pattern of the SLF depends on the microphone locations. The noise source map is obtained by summing all SLFs, thus a SLF with large main lobe does not provide accurate information about the source position but contains a high energy (due to the size of the main lobe). Therefore SLFs with a high energy may be removed. The energy of the SLF can be obtained by the l_2 -norm operation (root mean square value of the SLF for all the scan points). Therefore, removing the SLF given by a microphone pair with energy higher than a threshold β_{EC} should improve the resolution of the final noise source map. The main difference between the criteria is the computation. With the GCC-GC the value of the angle θ is computed for all the scan points and microphone pairs whereas the GCC-EC is computed

221 for each SLF,

$$\hat{R}_{x_m, x_n}(\tau) = \begin{cases} \hat{R}_{x_m, x_n}(\tau) & \text{if } 10 \log_{10} \left(\frac{\|\hat{R}_{x_m, x_n}(\tau)\|_2}{\min \|\hat{R}_{x_m, x_n}(\tau)\|_2} \right) < \beta_{EC} \\ 0 & \text{otherwise.} \end{cases} \quad (11)$$

222 3. Numerical source localization

223 3.1. Numerical set-up

224 The performances of the GCC associated to the two criteria are first in-
 225 vestigated from synthetic data. The acoustic signal is generated by a broad-
 226 band omnidirectional point source in free-field conditions. The acoustic point
 227 source is located at $(x = 0, z = 0)$ at 1 m from the microphone array center
 228 at the same z position. The geometry of the microphone array is discussed
 229 in Section 3.2. The acoustic time signal is sampled at $f_s = 44$ kHz. The scan
 230 zone, where the source is searched, is a square with side equal to 0.6 m. The
 231 distance between the microphone array and the scan zone is 1 m. It was ver-
 232 ified through numerical simulations that a small error on this distance (less
 233 than 10%) leads to a correct source position with a similar noise source map.
 234 A larger error provides a correct source position detection but with higher
 235 spurious lobes levels. The scan zone is discretized with 101 scan points along
 236 each direction which leads to a total number of scan points $L = 10,201$. The
 237 cross-correlation is computed using the inverse Fourier transform of the cross-
 238 spectrum between microphone pair signals. Commonly the cross-spectrum is
 239 whitened by the PHase Transform (PHAT) filter, which removes its magni-
 240 tude. This technique enhances the cross-correlation estimation because only
 241 the phase information is used to compute the inverse Fourier transform [18].

242 3.2. Microphone array geometries

243 In the context of acoustic imaging various microphone array geometries
 244 can be used depending on the configuration (planar, circular...). In an in-
 245 dustrial hall, the acoustic energy may come from all the directions, therefore
 246 spherical microphone arrays are preferred. The distribution of the micro-
 247 phones on the sphere has to be correctly chosen. Polyhedral discretization
 248 of a sphere has been discussed theoretically in [23]. In the present study, a
 249 polyhedral geometry is proposed with 14 microphones arranged on the ver-
 250 tices of the compound of a cube and an octahedron. This geometry fits on
 251 a sphere of radius of 0.25 m (diameter $d = 0.5$ m). It is compared with the

252 microphone array developed by Noël *et al.* [21] that involves 15 microphones
 253 arranged on a sphere. Their microphone array is composed of 3 parallel circles with 5 microphones at different z positions. The radius of the sphere is
 254 also 0.25 m. Noël *et al.* [21] chose a small diameter to keep the microphone
 255 array compact. Moreover they used few microphones to keep their inverse
 256 problem tractable from a computational point of view. Both microphone
 257 array geometries are depicted in Figure 3. The noise source maps obtained
 258 with the two microphone array geometries, called CUBE (for the polyhedron)
 259 and ISIT as proposed in [21], are compared in the following.

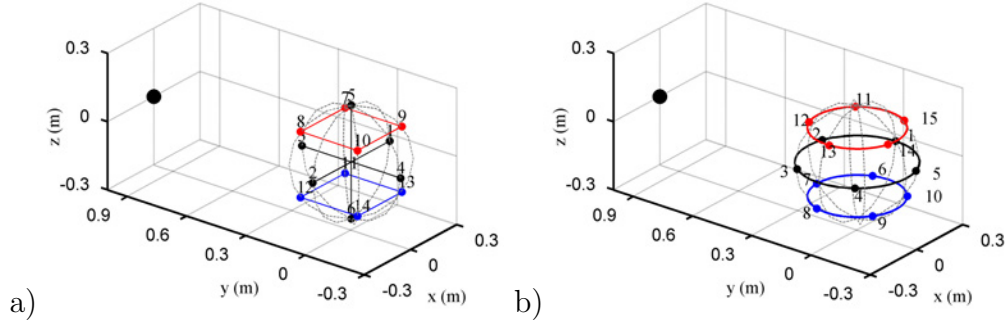


Figure 3: a) Spherical 14-microphone array (called CUBE). b) Spherical 15-microphone array (called ISIT). The large black dot is the source position. The dashed gray lines represent circles of radius 0.25 m.

260

261 3.3. Noise source maps

262 First, the noise source maps provided by both geometries and GCC are
 263 compared. The SLFs generated by each microphone pair are summed up to
 264 produce the noise source map. The amplitudes of the noise source map are
 265 normalized by the peak value and shown in dB (Figure 4), darker colors cor-
 266 responding to higher amplitudes. The source location is correctly detected
 267 with both microphone array geometries. The CUBE array provides a sym-
 268 metric map in both directions (due to the symmetry of the array) whereas
 269 the noise source map given by the ISIT array provides a large lobe in the
 270 vertical direction and an asymmetry in the horizontal direction. The verti-
 271 cal lobe of the ISIT array (z -axis) is larger than the horizontal due to the
 272 absence of microphones on top of the sphere and the asymmetry is due to
 273 the non-symmetric geometry along the x -axis. To get a better insight into

274 the noise source maps, slices in the horizontal and vertical directions are dis-
 275 played in Figure 4.c-d. The horizontal size of the main lobe is not clearly
 276 improved by the CUBE array because both arrays have the same size in this
 277 direction. An improvement is clearly obtained in the vertical direction where
 the main lobe is narrower and the side lobes level is smaller. Using the cri-

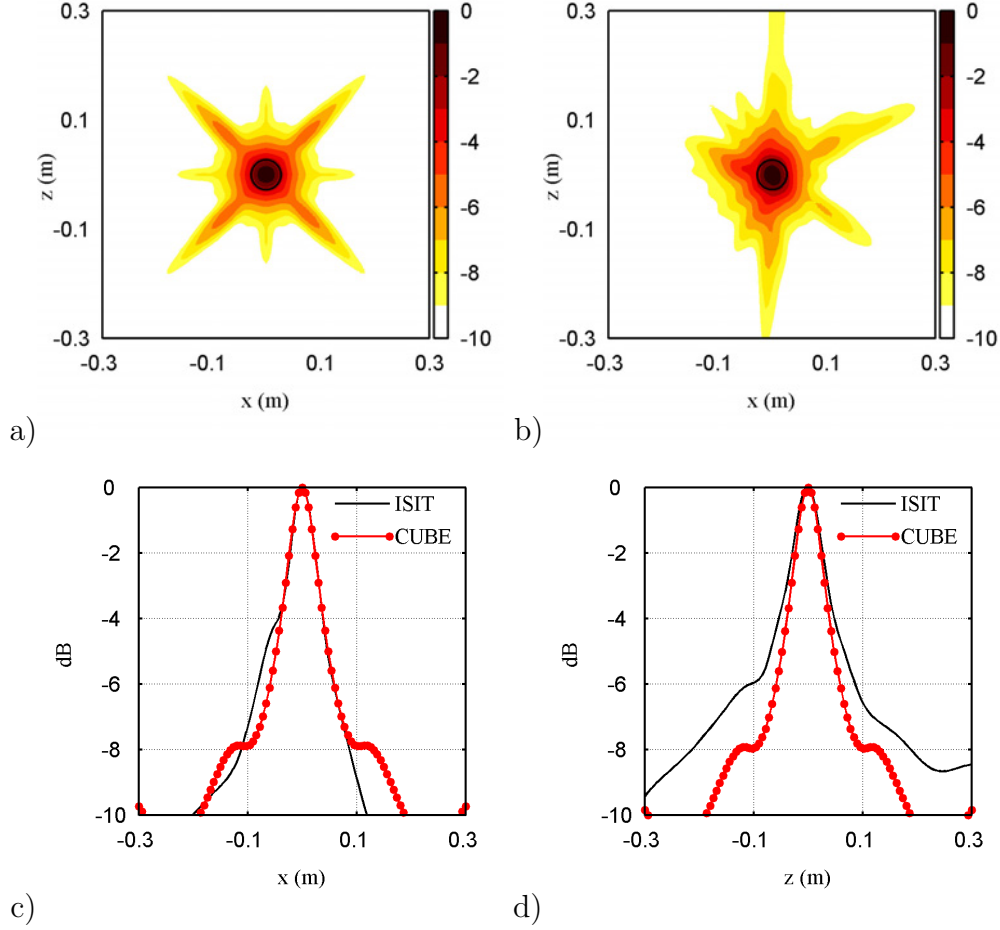


Figure 4: Noise source maps with a) GCC and the CUBE array and b) GCC and the ISIT array. The black circle is the source position. Slices of the noise source maps c) at $z = 0$ m and d) at $x = 0$ m.

278
 279 teria should improve the noise source map. However, both criteria require
 280 defining a tuning factor (β_{GC} or β_{EC}). To get the best value of the tuning

factor, the width of the main lobe in both directions is picked up for several values of the tuning factor. The main lobe width is defined by the length of the main lobe at -3 dB. The number of scan points is increased up to $L = 361,201$ (one point per millimeter) in order to accurately estimate the main lobe width. β_{EC} varies from 0 dB up to 8 dB and β_{GC} from 0° up to 23° . The main lobe width versus tuning factor is displayed in Figure 5 for both microphone array geometries. With $\beta_{EC} = 8$ dB and $\beta_{GC} = 0^\circ$, the noise source maps obtained correspond to the GCC source map (without criteria). In this case, the main lobe width along the x -axis is similar for both microphone array geometries. However, the main lobe width along the z -axis is larger with the ISIT microphone array as shown in Figure 4.d. Note that the main lobe width is similar in both directions with the CUBE array due to the symmetry of the geometry. With both arrays, if $\beta_{EC} > 3$ dB the main lobe width is not improved by the criterion. For lower values, the main lobe width is enhanced until it reaches a minimum for the CUBE and ISIT (z -axis). With the CUBE array, the smallest main lobe width is reached when β_{EC} is in the range $[0.6; 1]$ dB. With the ISIT array, the main lobe width along the x -axis decreases as β_{EC} decreases but at the expense of a larger main lobe width along the z -axis (larger than without criterion). For $\beta_{EC} < 1$ dB, the main lobe width along the z -axis is large because too many SLFs are removed and only SLFs with a vertical line are kept (which explains the small main lobe width along the x -axis). Therefore a trade-off has to be made. In order to compare the noise source maps provided by both microphone array geometries with the GCC-EC, a value of $\beta_{EC} = 1$ dB is chosen. This allows for the best main lobe width with the CUBE array and a small main lobe width along the z -axis with the ISIT array is achieved. As for the GCC-EC, when $\beta_{GC} < 10^\circ$, the main lobe width is not clearly improved for both microphone array geometries. With the CUBE array, the optimal main lobe width is reached when β_{GC} is in the range $[18^\circ; 20^\circ]$. With the ISIT array, the trend of the main lobe width is similar to GCC-EC, the x main lobe width decreases with β_{GC} but the z main lobe width increases. The optimal values chosen for both criteria are $\beta_{GC} = 20^\circ$ and $\beta_{GC} = 21^\circ$ for the CUBE and ISIT array, respectively. The noise source map obtained with both microphone array geometries and above optimal values of the criteria, are plotted in Figure 6. The CUBE array provides a narrower, more symmetric main lobe than the ISIT array for which the main lobe is blurred with side lobes in the vertical direction. The best noise source map is given by the CUBE array with GCC-EC which exhibits a narrow main lobe and

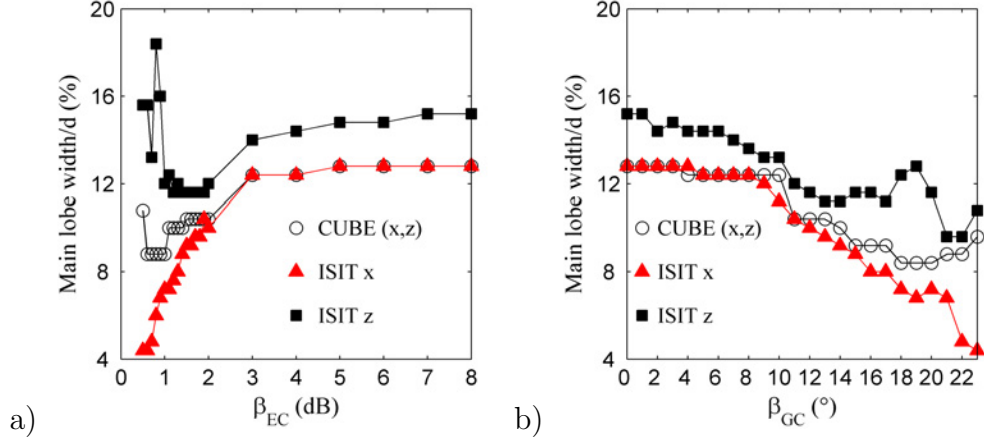


Figure 5: Main lobe width in both directions (x -axis and z -axis) versus tuning factor, a) GCC-EC and b) GCC-GC. The main lobe width is divided by the diameter of the spherical microphone array and expressed in %.

low side lobes levels. The number of microphone pairs (which corresponds to the number of SFLs) used to generate a noise source map with both criteria is investigated in Figure 7. The β_{EC} criterion corresponds to the energy of a given SLF normalized by the SLF with the lowest energy. The number of microphone pairs considered as a function of β_{EC} is displayed in Figure 7.a for both microphone array geometries. The number of microphone pairs is different for both microphone arrays, therefore the number of microphone pairs is expressed in percent, 100% meaning that all the SFLs are used to compute the noise source map. The trend of both curves is similar which means that the number of discarded SLFs is approximately the same for both array geometries. For instance, if the threshold is chosen equal to 1 dB, only 30% of the SLFs are kept and all the others SLFs are rejected from the computation.

The β_{GC} criterion depends on the angle θ formed by the vectors joining a scan point and a microphone pair; this angle is therefore computed for all possible values. The size of the spatial weighting function matrix is $[M_p \times L]$, which represents a large set of values. To compare the number of microphone pairs considered as a function of β_{GC} for both arrays, the average value is computed for all the microphone pairs (Figure 7).b. These values are similar for both array geometries. For instance, if the value of $\beta_{GC} = 20^\circ$ is chosen,

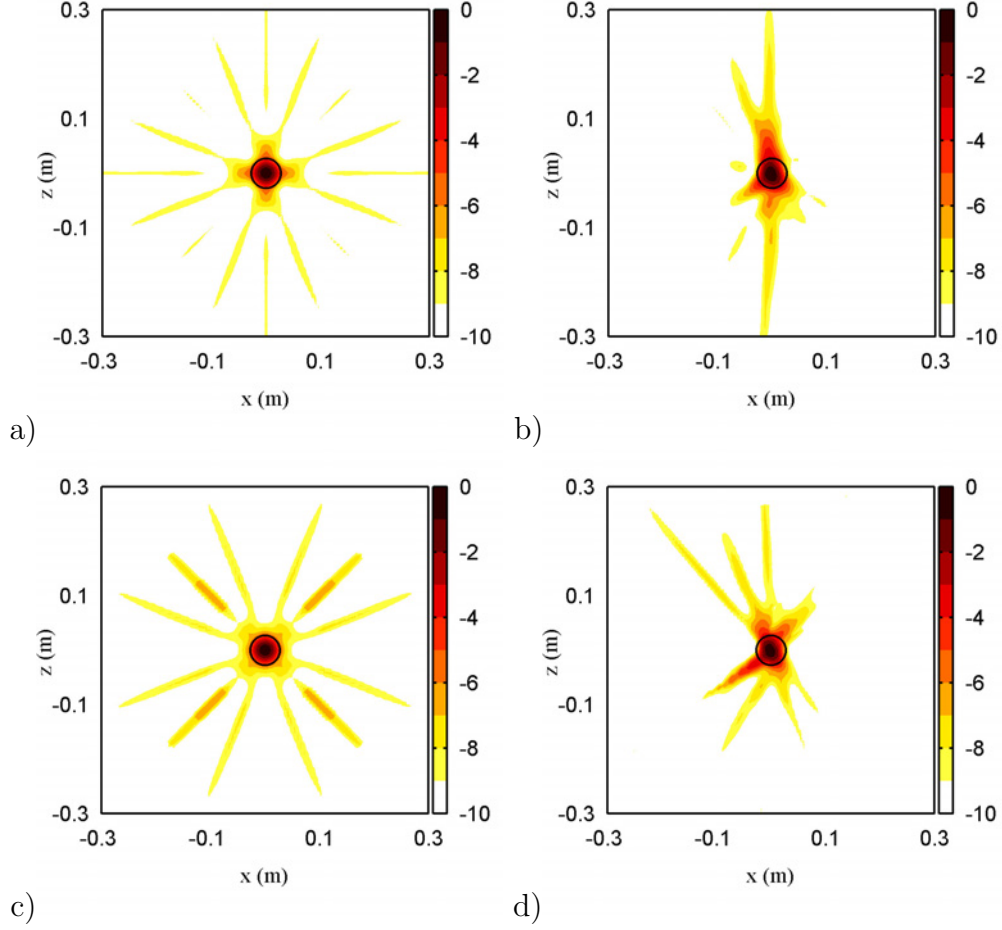


Figure 6: Noise source map with a) CUBE array and GCC-EC, b) ISIT array and GCC-EC, c) CUBE array and GCC-GC and d) ISIT array and GCC-GC. The black circle is the source position.

339 70% of microphone pairs are removed on average from the computation of
 340 the noise source map. Vectors formed by a scan point and microphone pairs
 341 leading to a low value of β_{GC} are rejected.

342 In summary, both criteria reject the same number of microphone pairs
 343 (with the β_{EC} and β_{GC} values considered). However in the case of a single
 344 source in front of the microphone array the best noise source map is given
 345 by the GCC-EC and the CUBE array.

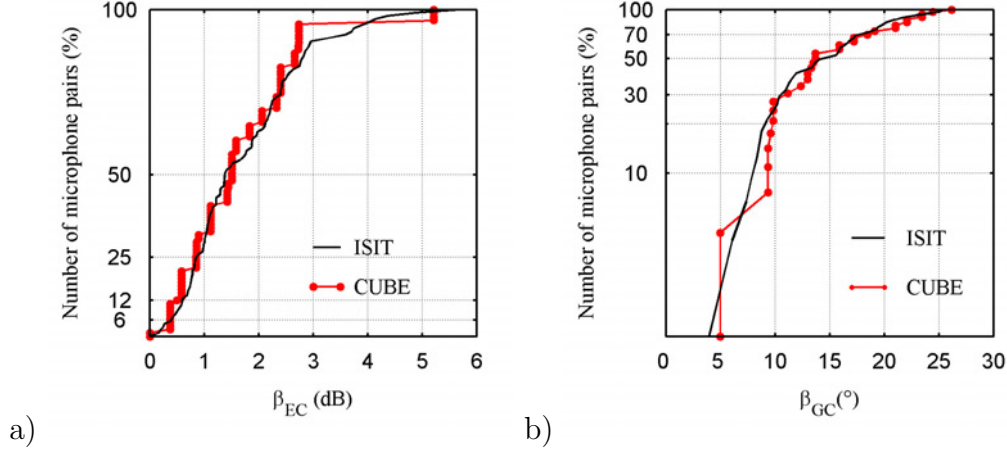


Figure 7: Number of microphone pairs considered in the computation of the noise source maps as a function of the criteria values a) GCC-EC. b) GCC-GC.

3.4. Influence of background noise

The previous noise source maps have been generated in the case of microphone signals without additive noise (or background noise). Now, an uncorrelated white noise with zero mean is added to each microphone signal (see Eq. 1). In the case of a workplace, the background noise can be electrical or due to undesirable noise sources. The electrical noise depends on the quality of the acquisition system and is considered of low amplitude as compared to the main sources. However, the amplitude of the acoustic background noise such as the heat and ventilation air conditioning can be high. To determine the amount of noise added to the microphone signals, the Signal to Noise Ratio (SNR) criterion is used. The previous noise source maps have been generated without noise, therefore the SNR is equal to $+\infty$. An SNR= 0 dB is chosen to generate the noise source maps which correspond to a source and a noise with the same level (see Figure 8). Without criteria, the noise source maps obtained with both microphone arrays are very similar. With the CUBE array, the side lobes in the horizontal and vertical directions are more extended but the main lobe size is slightly narrower. With the ISIT array, the noise source map keeps the same pattern with side lobes slightly more extended. Therefore, even with background noise added to the microphone signals, the source can be located and the noise source maps are almost the same as without background noise. Using the criteria,

the noise source maps are improved. With the ISIT array, the noise source maps obtained are almost the same as Figures 6.b-d. With the CUBE array the noise source maps are slightly different but the spurious lobes are still removed. Finally, even with a high background noise (SNR= 0 dB), the source can be located and both criteria still improve the noise source maps. In the following, the background noise amplitude is set to zero to only investigate the performances of both arrays and criteria.

3.5. Influence of source-array distance

The numerical set-up of Section 3.1 is used to illustrate how the localization results are affected by the distance between the source and the array. The source-array distance is set to 1 m for generating the acoustic field, but different source-array distances are used for generating the noise source maps ([0.7; 1; 1.3] m). The noise source maps obtained, with the CUBE array, are shown in Figure 9.

As expected, the best result is obtained when the source-array distance is well estimated. However, even with biased source-array distances, the source position is located. The main difference is the number and level of spurious lobes. Figure 10 shows the noise source maps for a smaller bias of the source-array distances ([0.95; 1; 1.05] m). If the source-array distance is estimated with a small error, the noise source maps are very similar. Therefore, if the source-array distance can be experimentally estimated, with a small error, the source position will be correctly estimated.

Commonly, the source-array distance is an input into the microphone array processing and can be quite accurately estimated experimentally. If the source-array distance cannot be estimated, the alternative way is to use a plane wave formulation.

3.6. Three acoustic sources with different source levels

In this section the performances of the microphone arrays associated with the criteria are compared in the case of three sources with different levels [0; -3; -6] dB. The source locations are $x = [-0.25; 0; +0.25]$ m and $z = 0$ (horizontal case) and $x = 0$ and $z = [-0.25; 0; +0.25]$ m (vertical case). The acoustic signals are uncorrelated white noises. The noise source maps are computed with GCC, GCC-EC and GCC-GC for both microphone array geometries.

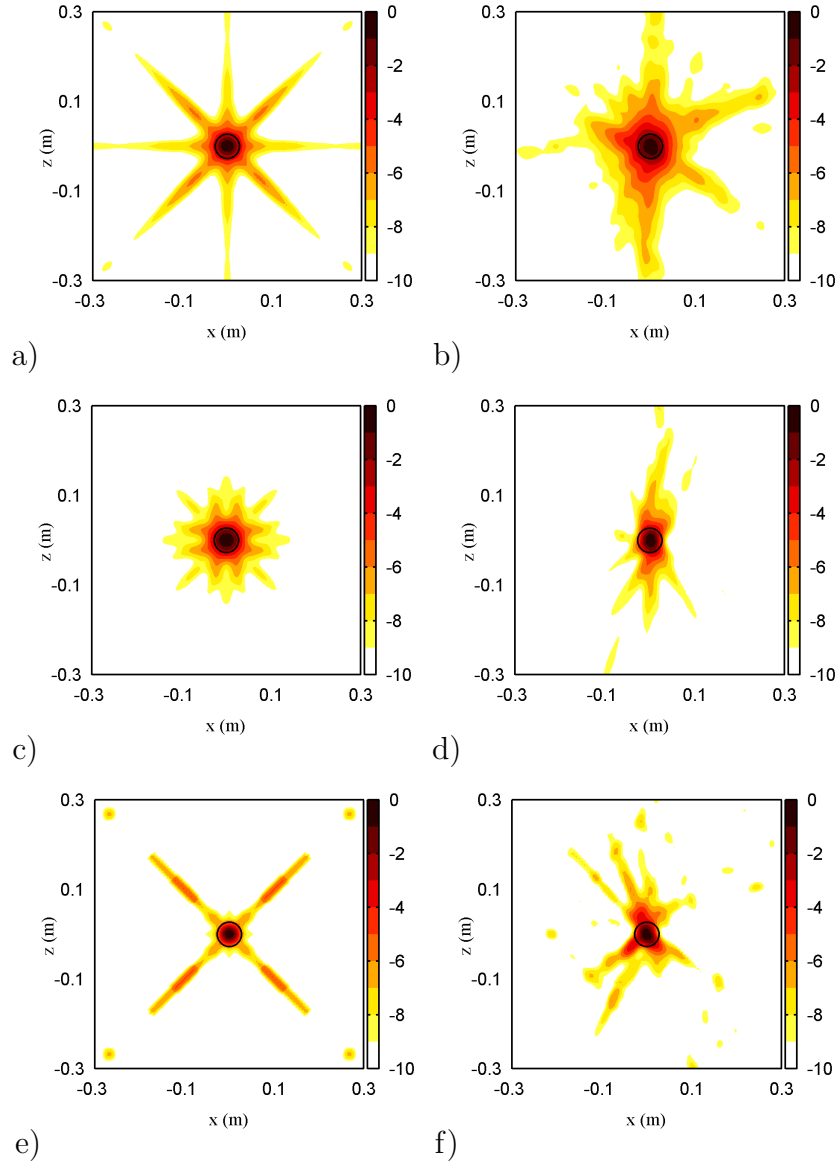


Figure 8: Noise source maps in the case of a Signal to Noise Ratio (SNR) equal to 0 dB . a) CUBE array with GCC, b) ISIT array with GCC, c) CUBE array with GCC-EC, d) ISIT array with GCC-EC, e) CUBE array with GCC-GC and f) ISIT array with GCC-GC. The black circles are the source positions.

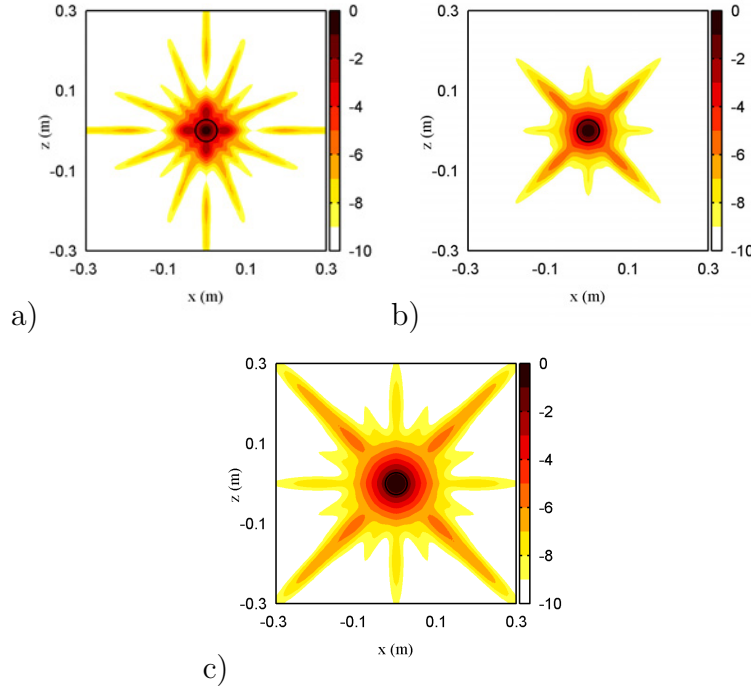


Figure 9: Noise source maps for three source-array distances with CUBE array and GCC, a) 0.7 m, b) 1 m and c) 1.3 m. The black circles are the source positions.

403 The noise source maps are shown in Figure 11 for the horizontal source
 404 configuration. In all the cases, the source positions are correctly detected.
 405 However with GCC, the detection of the source with the weakest level is
 406 difficult due to the spurious lobes. The criteria allow for a better detection
 407 of the source positions. In this configuration the ISIT microphone array
 408 provides a noise source map with few spurious lobes due to its lower side
 409 lobes level along the x -axis. To get a better insight into the noise source
 410 maps, slices along the x -direction are presented in Figure 12. Both criteria
 411 improve the main lobe width and reduce the background noise of source
 412 maps. For both microphone array geometries, the GCC-EC provides the
 413 best results with low side lobes level and narrow lobes. Moreover, the GCC-
 414 EC better estimates the noise source levels. The GCC-GC is less efficient
 415 with the CUBE array and under-estimates the weakest source level with the
 416 ISIT array.

417 As discussed in Section 3.3, the vertical main lobe width of the ISIT

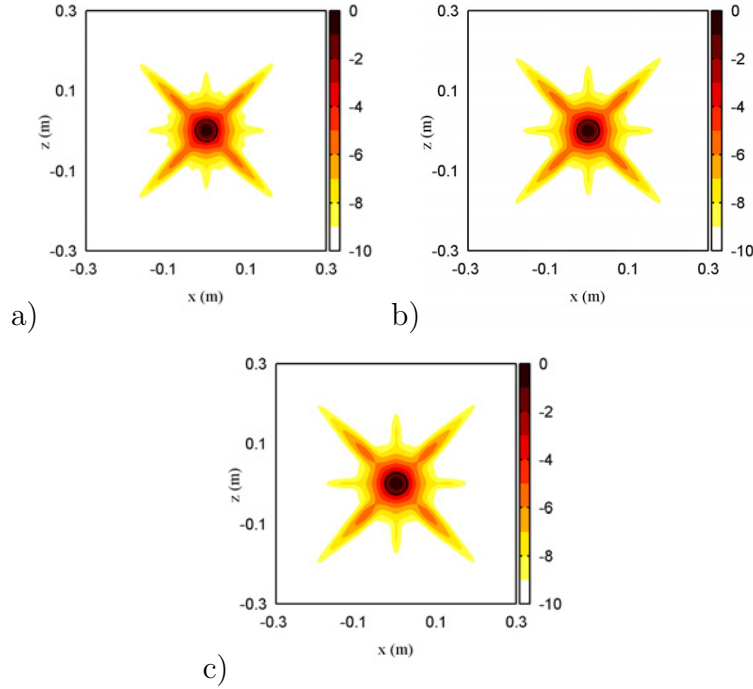


Figure 10: Noise source maps for three source-array distances with CUBE array and GCC, a) 0.95 m, b) 1 m and c) 1.05 m. The black circles are the source positions.

array is poor due to the smaller z -dimension of the microphone array, thus
inaccurate noise source maps may be expected in the vertical axis. The
three acoustic sources, previously set horizontally, are now along the z -axis.
The noise source maps are shown in Figure 13 for both microphone array
geometries and criteria. Without any criterion, the noise source maps exhibit
a high side lobe level and the detection of the weakest source is difficult. Both
criteria improve the noise source maps by removing side lobes and narrowing
the main lobe. However, the noise source map based on GCC-GC and ISIT
still exhibits a very poor vertical main lobe width whereas this artifact is not
present in the noise source map based on the GCC-EC and GCC-GC with
the CUBE array. The slices along the z -direction are presented in Figure 14.
When GCC-GC is used, the ISIT array shows a large side lobe. The best main
lobe width is provided by the GCC-EC which leads to the best estimation of
the source level.

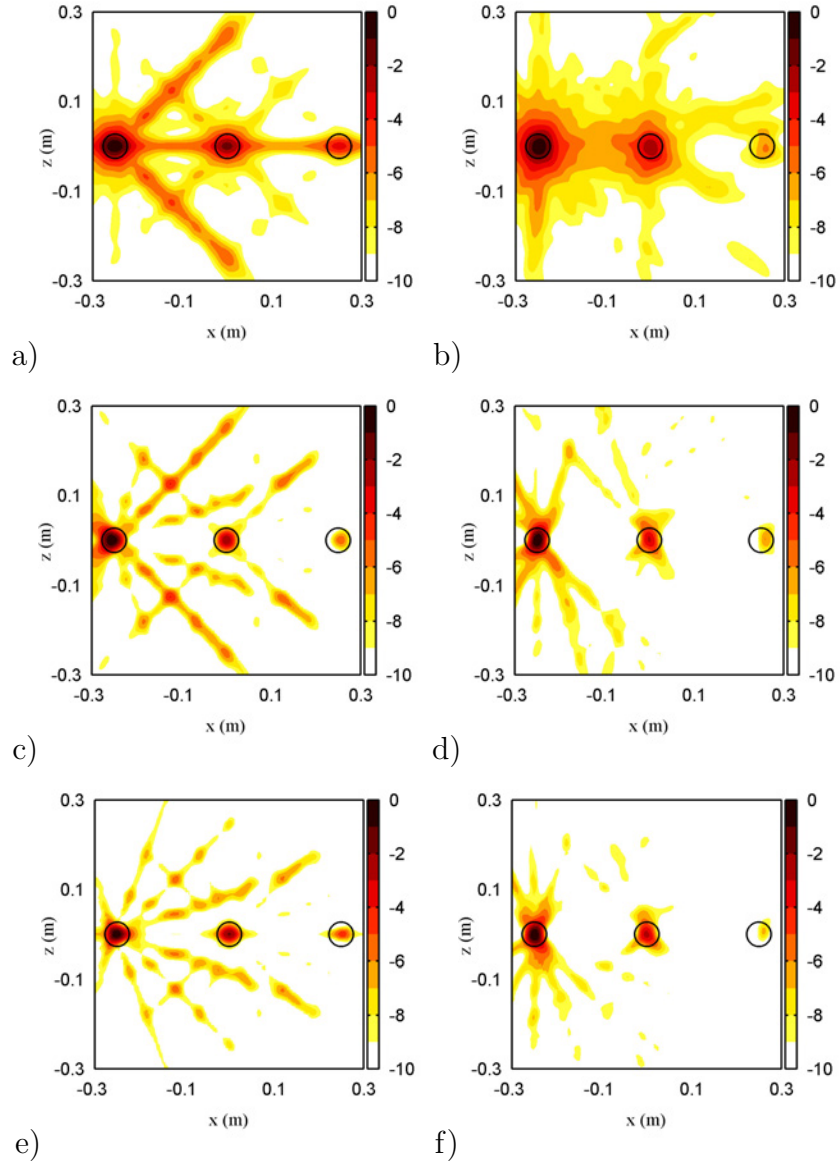


Figure 11: Noise source maps in the case of three horizontal acoustic sources with different sound levels ($[0; -3; -6]$ dB). a) CUBE array with GCC, b) ISIT array with GCC, c) CUBE array with GCC-EC, d) ISIT array with GCC-EC, e) CUBE array with GCC-GC and f) ISIT array with GCC-GC. The black circles are the source positions.

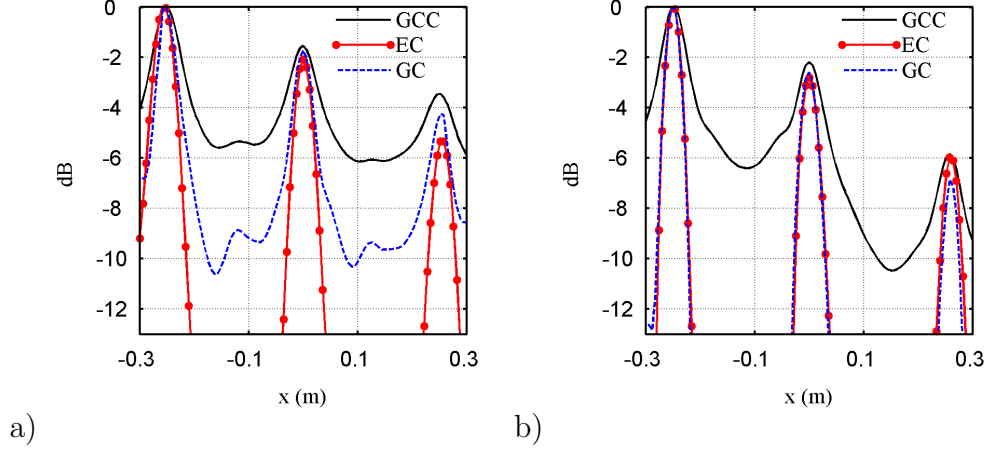


Figure 12: Slices of the noise source maps in the case of three horizontal sources at $z = 0$ m, a) CUBE array and b) ISIT array.

4. Experimental source localization

4.1. Experimental set-up

In order to confirm the above findings on array geometries (CUBE and ISIT) and criteria (GCC-EC and GCC-GC), experiments were conducted in a reverberant room. The reverberation time based on the TR60 is about 3 s at 1 kHz which leads to a reverberation radius of 0.5 m.

To set-up the microphone arrays, two similar frames have been designed. Each frame was composed of a sphere of radius 3.81 cm supported by a tripod. Holes were drilled in the sphere according to each microphone array geometry pattern. Rods with 20 cm length were inserted into the holes and the microphones were mounted at the end of the rods (see Figure 15) to obtain an array radius of 0.25 m. Brüel&Kjaer microphones type 4935 were used and the signals were recorded using a Brüel&Kjaer 3038B front end and Brüel&Kjaer Pulse software. The acoustic signals were sampled at 32,768 Hz during 15 seconds. The source signals were uncorrelated white noises, with the same amplitude, generated by a NI PXI-4461 card controlled with Labview. The signals were amplified by a BSWA audio amplifier SWA 100 and emitted by two loudspeakers. The distance between the loudspeakers and the center of the microphone array was 1.7 m, this distance is larger than the reverberation radius therefore the microphone array is in the reflected

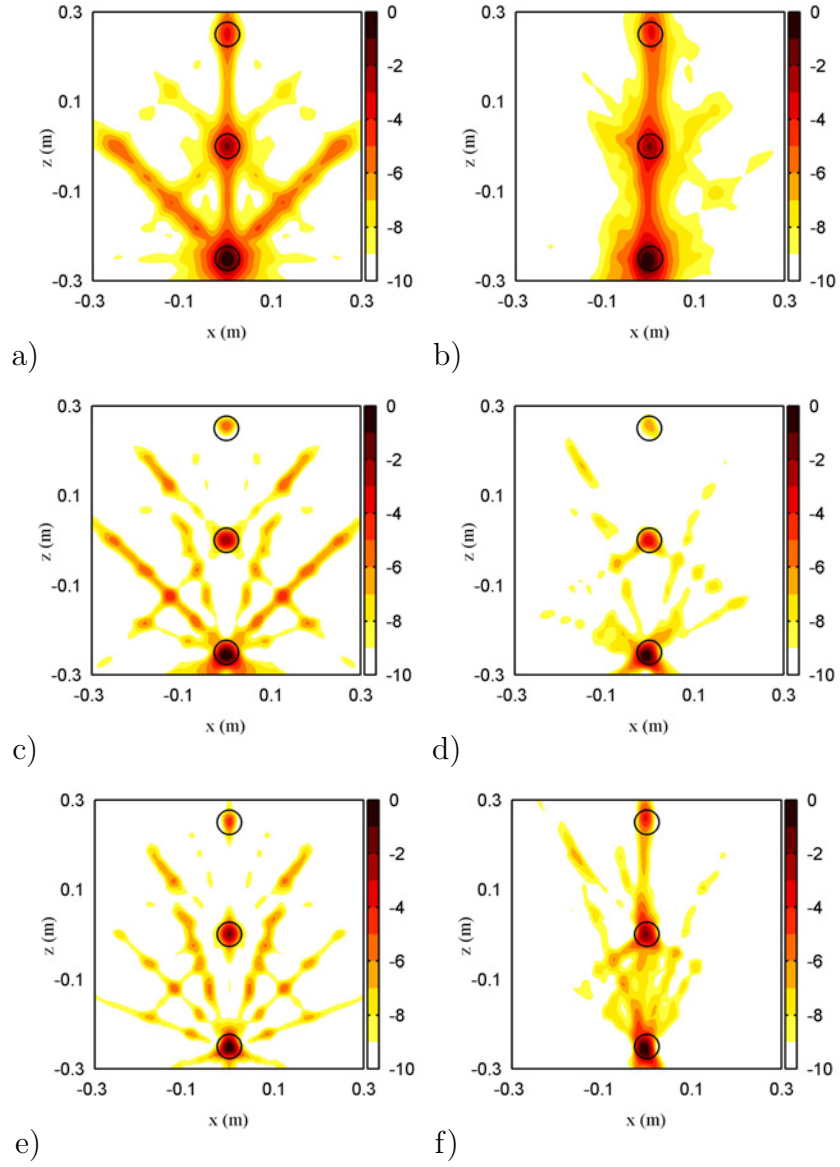


Figure 13: Noise source maps in the case of three vertical acoustic sources with different sound levels ($[0; -3; -6]$ dB). a) CUBE array with GCC, b) ISIT array with GCC, c) CUBE array with GCC-EC, d) ISIT array with GCC-EC, e) CUBE array with GCC-GC and f) ISIT array with GCC-GC. The black circles are the source positions.

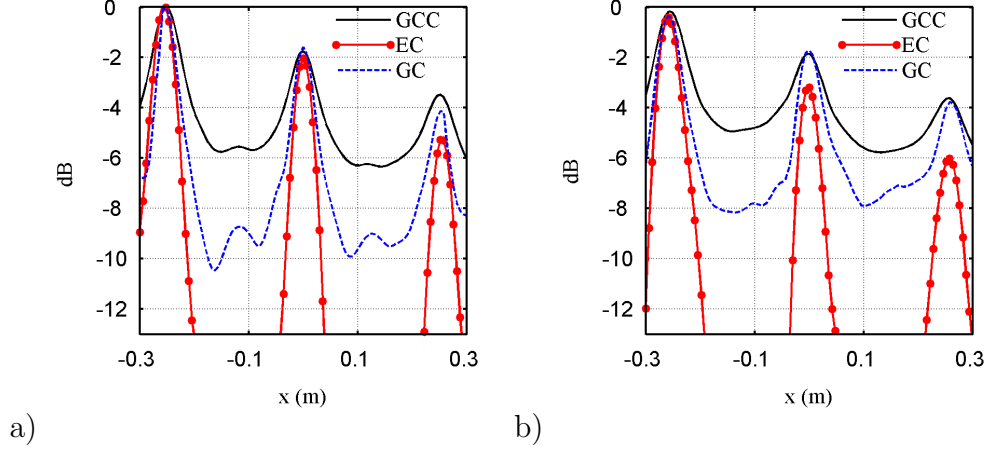


Figure 14: Slices of the noise source maps in the case of three vertical sources at $x = 0$ m, a) CUBE array and b) ISIT array.

field. The two Boston CR67 loudspeakers, composed of a 133 mm woofer at the bottom and a 2 mm tweeter at the top, were separated by 0.55 m. The microphone array was located at the same distance from the loudspeakers as the back wall and the side wall was closer (see Figure 15). The center of the microphone array and the loudspeakers were set at 1.3 m above the ground. In this configuration, the microphone array records the direct acoustic field and the multiple reflections from the ground and walls. The microphone signals were filtered by a band-pass Butterworth second order filter between 500 Hz and 8000 Hz. The scan zone where the sources are searched was a $1.2 \text{ m} \times 1.2 \text{ m}$ square at 1.7 m away from the microphone array (including the loudspeakers). The scan zone was discretized with 121×121 points (14,641 scan points, i.e. one scan point by centimeter). For both microphone arrays, the GCC with and without criteria was applied.

The noise source maps are shown in Figure 16 for both microphone array geometries and detection algorithms. The GCC with the ISIT array exhibits two large spots at source positions with a vertical main lobe width larger than the horizontal main lobe width. The noise source map is enhanced when the criteria are applied. The GCC-GC and GCC-EC remove the spurious lobes and decrease the size of the main lobes. With the CUBE array, the noise source map, obtained with GCC, exhibits two main spots with a smaller main lobe width than the ISIT array in the vertical direction. The GCC-GC

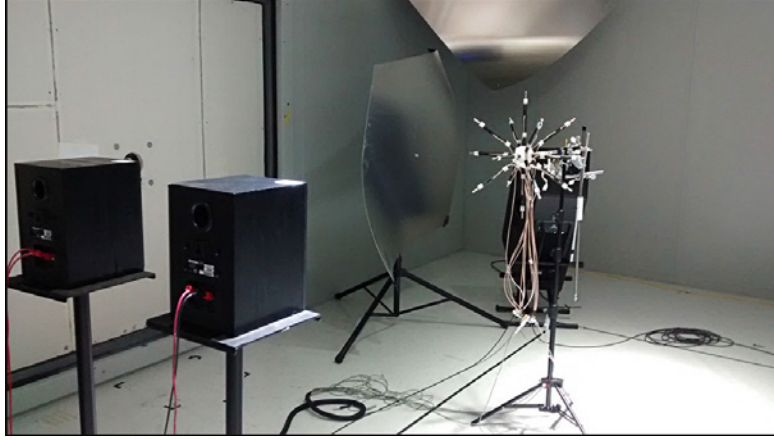


Figure 15: Photograph of the experimental set-up.

473 and GCC-EC remove the spurious lobes and allow the best detection of the
 474 two sources as compared to the GCC with the ISIT array.

475 Slices of the noise source maps along the axis $z = 0$ for both microphone
 476 arrays are shown in Figure 17. With the CUBE array, the left and right
 477 loudspeakers are localized at $x = -0.32$ m and $x = 0.23$ m which lead to a
 478 separation distance of 0.55 m. With the ISIT, the left and right loudspeakers
 479 are localized at $x = -0.27$ m and $x = 0.27$ m which lead to a separation
 480 distance of 0.54 m. Therefore, both microphone arrays are able to correctly
 481 estimate the source separation distance which was set to 0.55 m. The esti-
 482 mated position of the left (or right loudspeaker) is different with both arrays
 483 due to the difficulty to set-up each microphone array (CUBE or ISIT) ex-
 484 actly at the same position in front of the two loudspeakers in the reverberant
 485 chamber. Moreover, the difference between the estimated positions of the
 486 left loudspeaker with both arrays is equal to 5 cm which is lower than the
 487 diameter of the woofer (0.13 m). For both microphone arrays, the GCC-EC
 488 and GCC-GC decrease the size of the main lobes and allow clearly detecting
 489 the loudspeakers.

490 4.2. Computation time

491 This section is dedicated to the computation time of the different meth-
 492 ods. In the context of source localization in a workplace, all the problematic
 493 workstations have to be tested. In order to disturb as little as possible the

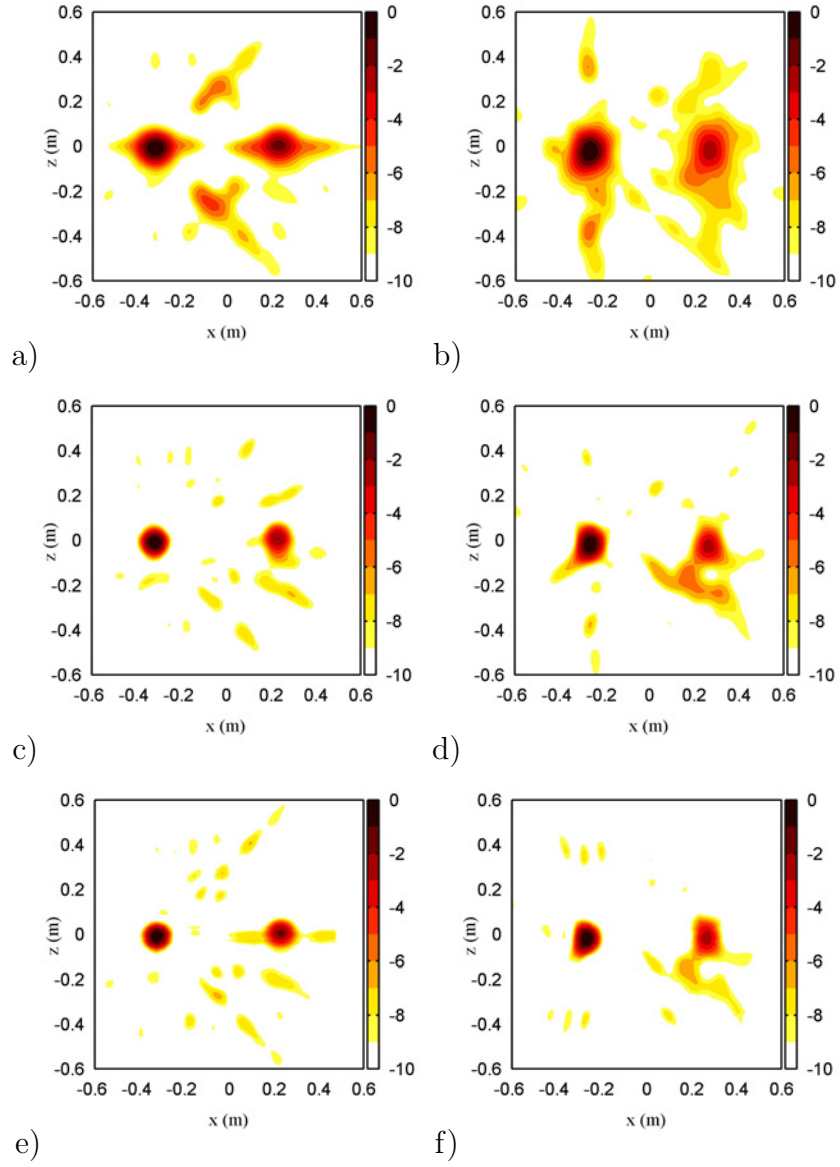


Figure 16: Noise source maps in the case of two loudspeakers. a) CUBE array with GCC, b) ISIT array with GCC, c) CUBE array with GCC-EC, d) ISIT array with GCC-EC, e) CUBE array with GCC-GC and f) ISIT array with GCC-GC.

494 workers and to avoid slowing down the production, the computation time

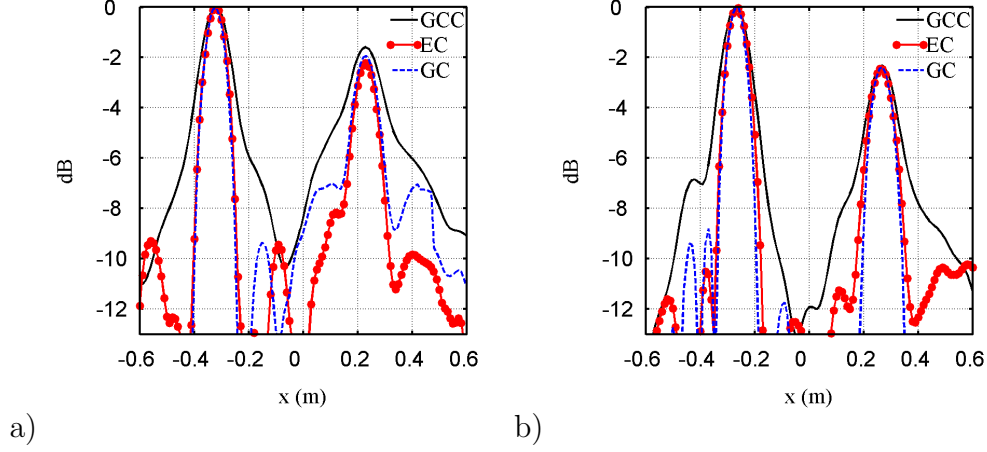


Figure 17: Slices of the noise source maps in the case of the loudspeakers at $z = 0$ m, a) CUBE array and b) ISIT array.

495 of the method should be low. The GCC, GCC-EC and GCC-GC methods
 496 were implemented with custom-made Matlab R2012b routines and the compu-
 497 tation time was provided by the tic-toc function. A personal computer
 498 was used with an Intel processor (Xeon(R) @ 2.4 GHz) and 16 Go of RAM.
 499 The computation time was displayed as a function of the number of scan
 500 points from 100 points (scan zone 10×10 points) up to 10^6 points (scan zone
 501 1000×1000 points). The computation time of the GCC is shown in Fig-
 502 ure 18.a for three numbers of microphone pairs $M_p = 105$ (15 microphones);
 503 $M_p = 210$ (21 microphones) and $M_p = 435$ (30 microphones) ($M_p = 105$
 504 corresponds to the ISIT array, the CUBE array has $M_p = 91$ pairs). With
 505 a number of scan points $L < 10,000$, the computation time is steady and
 506 equal to 0.1 second with $M_p = 105$. When $L > 10,000$, the computation
 507 time increases linearly with L (as expected due to Eq.4). However, even
 508 with a large number of scan points ($L=100,000$, scan zone 100×1000), the
 509 computation time is still reasonable approximately 1 second, which is com-
 510 patible with almost real-time post-processing of the microphone data. If the
 511 number of microphone pairs is multiplied by two or four, the computation
 512 time is twice or four times higher. Now, the computation time of the GCC
 513 associated with the criteria is compared to the case of $M_p = 105$ (ISIT array),
 514 considered as the baseline (Figure 18.b). With $L < 10,000$, the application
 515 of the GCC-EC does not increase the GCC computation time. The GCC-GC

slightly increases the computation time but practically, it is negligible. For GCC-GC, the computation of the cross product is more time consuming than the 2-norm operation (for GCC-EC). When $L > 10,000$, the GCC-EC and GCC-GC computation times increase linearly as for the GCC. With $L = 10^6$, the computation time of the GCC, GCC-EC and GCC-GC are 5.8, 10.3 and 25.6 seconds. Therefore even with a densely resolved scan zone (1000×1000), the computation time is still reasonable especially with the GCC-EC which is able to improve the noise source map with a moderate increase of the computation time.

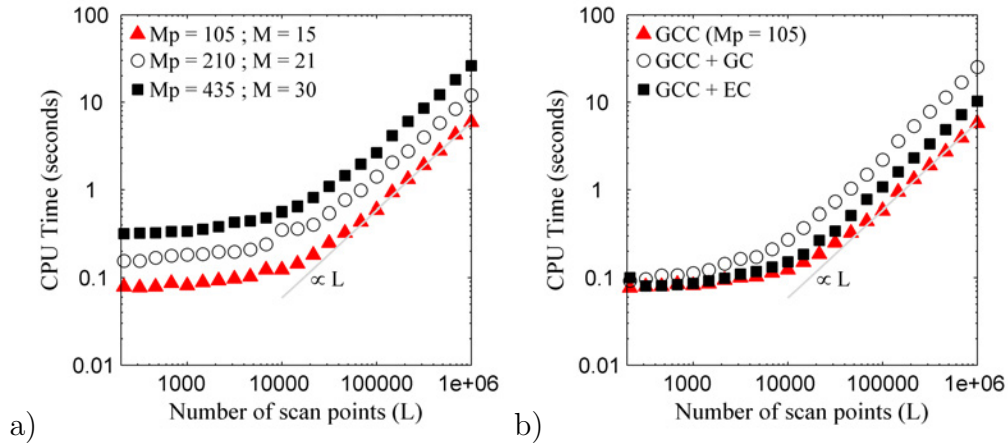


Figure 18: a) Computation time of the GCC method with several numbers of microphone pair. b) Computation time of the GCC and the two criteria (CPU Intel(R) Xeon(R) @ 2.4 GHz, RAM 16 Go, using tic-toc function Matlab R2012b).

524

525 5. Conclusion

In this study, the performance of spherical microphone arrays, combined with the Generalized Cross-Correlation (GCC) technique, to detect broadband point sound sources was investigated. The objective of the paper was to propose a fast signal processing technique to generate a clean noise source map with few microphones. A geometry based on the compound of a cube with an octahedron (called CUBE) and comprising 14 microphones has been proposed. This geometry has been compared with another 15 spherical microphone array used previously in the literature and called ISIT by the authors.

534 For both geometries, the microphones were distributed on a sphere with a ra-
 535 dius of 0.25 m. An Energy Criterion (GCC-EC) and a Geometrical Criterion
 536 (GCC-GC) were proposed to select a subset of relevant microphone pairs
 537 in Cross-Correlation estimations, and improve the noise source map. The
 538 noise source maps provided by both microphone arrays and criteria were as-
 539 sessed using numerical data in free-field and experimental data obtained in
 540 a reverberant room.

541 The case of a single source, three horizontal sources and three vertical
 542 sources with different sound levels were studied through numerical simula-
 543 tions. The best noise source map in terms of main lobe width and side lobes
 544 is provided by the CUBE array with the GCC-EC. Even with background
 545 noise, both criteria are efficient and lead to clean noise source maps.

546 Then, the performances of both microphone arrays were compared in the
 547 case of experimental data. Two loudspeakers were set in a reverberant room.
 548 The CUBE array associated to the GCC-EC clearly detected the two source
 549 positions with the best resolution.

550 Finally, it has been shown that the smallest GCC computation time is
 551 achieved for the CUBE array (due to it is low number of microphone pairs).
 552 Moreover, the GCC-EC is faster than the GCC-GC.

553 This study demonstrated that the CUBE array is more efficient than the
 554 ISIT array and that the GCC-EC allows for improving the noise source map
 555 with a weak increase of the computation time.

556 Usually in acoustic imaging, inverse problems can be used to achieve a
 557 high resolution. The input of the inverse problem is the noise source map and
 558 the computation time may be long due to the iteration process. The next step
 559 of this work is to study how the improved noise source map obtained with
 560 the CUBE array combined with the GCC-EC criterion affects the resolution
 561 of inverse problems.

- 562 [1] P. Duguay, A. Boucher and M.-A. Busque. *Statistiques sur les maladies*
 563 *professionnelles au Québec, Colloque IRSST Maladie professionnelles :*
 564 *portrait, défis et perspectives*, Novembre 2014.
- 565 [2] M. Lebeau. *Maladies professionnelles : impact économique au Québec,*
 566 *Colloque IRSST Maladie professionnelles : portrait, défis et perspec-*
 567 *tives*, Novembre 2014.
- 568 [3] D. H. Johnson and D. E. Dudgeon. *Array Signal Processing: Concepts*

- 569 *and Techniques*, Prentice Hall, (ISBN-10: 0130485136), (ISBN-13: 978-
570 0130485137), 1993.
- 571 [4] M. Brandstein and D. Ward. *Microphone Arrays: Signal Processing*
572 *Techniques and Applications*, Springer, (ISBN-10: 3540419535), (ISBN-
573 13: 978-3540419532), 2001.
- 574 [5] J. Benesty, J. Chen and Y. Huang. *Microphone Array Signal Processing*,
575 Springer, (ISBN-10: 3540786112), (ISBN-13: 978-3540786115), 2008.
- 576 [6] M. R. Bai, J.-G. Ih and J. Benesty. *Acoustic Array Systems: The-*
577 *ory, Implementation, and Application*, Wiley-IEEE Press, (ISBN-10:
578 0470827238), (ISBN-13: 978-0470827239), 2013.
- 579 [7] J. Lanslots, F. Deblauwe and K. Janssens. *Selecting sound source lo-*
580 *calization techniques for industrial applications*, Sound and Vibration,
581 Vol.**44**(**6**), pp 6-10, 2010.
- 582 [8] J. R. Underbrink. *Circularly symmetric, zero redundancy, planar array*
583 *having broad frequency range applications*, US Patent No 6,205,224 B1,
584 2001.
- 585 [9] U. Michel. *History of acoustic beamforming*, 1st Berlin Beamforming
586 Conference (BeBec), Berlin, 2006.
- 587 [10] T.F. Brooks and W.M. Humphreys. *A deconvolution approach for the*
588 *mapping of acoustic sources (DAMAS) determined from phased micro-*
589 *phone arrays*, Journal of Sound and Vibration, Vol.**294**(**4-5**), pp 856-
590 879, 2006 (doi.org/10.1016/j.jsv.2005.12.046).
- 591 [11] P. Sijtsma. *Clean based on spatial source coherence*, Interna-
592 tional Journal of Aeroacoustics, Vol.**6**(**4**), pp 357-374, 2007
593 (doi:10.1260/147547207783359459).
- 594 [12] T. Suzuki. *L1 generalized inverse beam-forming algorithm resolv-*
595 *ing coherent/incoherent, distributed and multipole sources*, Jour-
596 nal of Sound and Vibration, Vol.**330**(**4**), pp 5835-5851, 2011
597 (doi.org/10.1016/j.jsv.2011.05.021).
- 598 [13] T. Padois, P-A. Gauthier and A. Berry. *Inverse problem with beamform-*
599 *ing regularization matrix applied to sound source localization in closed*

- 600 *wind-tunnel using microphone array*, Journal of Sound and Vibration,
601 Vol.**330**(25), pp 6858-6868, 2014 (doi.org/10.1016/j.jsv.2014.07.028).
- 602 [14] J.P. Dmochowski, J. Benesty and S. Affes. *A generalized steered response*
603 *power method for computationally viable source localization*, IEEE
604 Transactions on Audio, Speech, and Language Processing, Vol.**15**(8),
605 pp 2510-2526, 2007 (doi.10.1109/TASL.2007.906694).
- 606 [15] H. Do and H. Silverman. *SRP-PHAT methods of locating simultaneous*
607 *multiple talkers using a frame of microphone array data*, Proceedings
608 of 2010 IEEE International Conference on Acoustics Speech and Signal
609 Processing, Dallas, TX, USA, pp 125-128, 14-19 March 2010.
- 610 [16] J. Velasco, D. Pizarro and J. Macias-Guarasa. *Source localiza-*
611 *tion with acoustic sensor arrays using generative model based fit-*
612 *ting with sparse constraints*, Sensors, Vol.**12**, pp 13781-13812, 2012
613 (doi:10.3390/s121013781).
- 614 [17] A. Brutti, M. Omologo and P. Svaizer. *Multiple source localization based*
615 *on acoustic map de-emphasis*, EURASIP Journal on Audio, Speech, and
616 Music Processing, Vol.**2010**, pp 1-17, 2010 (doi:10.1155/2010/147495).
- 617 [18] C. Knapp and G. C. Carter. *The generalized correlation method*
618 *for estimation of time delay*, Transactions on Acoustics, Speech
619 and Signal Processing, IEEE, Vol.**24**(4), pp 320-327, 1976
620 (doi.10.1109/TASSP.1976.1162830).
- 621 [19] H. Sun, E. Mabande, K. Kowalczyk and W. Kellermann. *Localization of*
622 *distinct reflections in rooms using spherical microphone array eigenbeam*
623 *processing*, Journal of the Acoustical Society of America, Vol.**131**(4), pp
624 2828-2840, 2012 (doi.org/10.1121/1.3688476).
- 625 [20] T. Padois, F. Sgard, O. Doutres and A. Berry. *Comparison of acoustic*
626 *source localization methods in time domain using sparsity constraints*,
627 Proceedings of Internoise, San Francisco, CA, USA, 9-12 August, 2015.
- 628 [21] C. Noël, V. Planeau and D. Habault. *A new temporal method*
629 *for the identification of source directions in a reverberant hall*,
630 Journal of Sound and Vibration, Vol.**296**(3), pp 518-538, 2006
631 (doi.org/10.1016/j.jsv.2005.12.056).

- 632 [22] P. Aarabi. *The fusion of distributed microphone arrays for sound local-*
633 *ization*, EURASIP Journal on Applied Signal Processing, Vol.**2003**, pp
634 338-347, 2003 (doi:10.1155/S1110865703212014).
- 635 [23] J.-S. Hu, C.-M. Tsai, C.-Y. Chan and Y.-J. Chang. *Geometrical arrange-*
636 *ment of microphone array for accuracy enhancement in sound source lo-*
637 *calization*, Proceedings of 8th Asian Control Conference (ASCC), Kaoh-
638 siung, Taiwan, 15-18 May, 2011.