

# On the use of geometric and harmonic means with the generalized cross-correlation in the time domain to improve noise source maps

Thomas Padois,<sup>1,2,a)</sup> Olivier Doutres,<sup>2</sup> Franck Sgard,<sup>1</sup> and Alain Berry<sup>3</sup>

<sup>1</sup>*Institut de Recherche Robert-Sauvé en Santé et en Sécurité du Travail, Montréal, Québec H3A 3C2, Canada*

<sup>2</sup>*Department of Mechanical Engineering, École de Technologie Supérieure, Montréal, Québec H3C 1K3, Canada*

<sup>3</sup>*Group of Acoustics of the University of Sherbrooke, Department of Mechanical Engineering, Université de Sherbrooke, Sherbrooke, Québec J1K 2R1, Canada*  
*Thomas.Padois@etsmtl.ca, olivier.doutres@etsmtl.ca, franck.sgard@irsst.qc.ca, alain.berry@usherbrooke.ca*

**Abstract:** Microphone array techniques are an efficient tool to detect acoustic source positions. The delay and sum beamforming is the standard method. In the time domain, the generalized cross-correlation can be used to compute the noise source map. This technique is based on the arithmetic mean of the spatial likelihood functions. In this study, the classical arithmetic mean is replaced by the more standard generalized mean. The noise source maps provide by the arithmetic, geometric and harmonic means are compared in the case of numerical and experimental data obtained in a reverberant room. The geometric and harmonic means provide the best noise source maps with no side lobes and a better source level estimation.

© 2016 Acoustical Society of America

[DHC]

Date Received: January 29, 2016 Date Accepted: May 10, 2016

## 1. Introduction

Microphone array techniques are a common tool for localizing sound sources in several domains such as aeroacoustics,<sup>1,2</sup> speech<sup>3,4</sup> or occupational health and safety.<sup>5,6</sup> With these techniques, the acoustic field is recorded at each microphone position and the signals are processed in order to build a noise source map which highlights the source positions. The standard technique is the delay and sum (DAS) beamforming which can be performed in the frequency domain via the cross-spectral matrix or in the time domain via the cross-correlation of the microphone signals.<sup>7</sup> The DAS beamforming is known to be less efficient at low frequencies, therefore several techniques based on inverse methods have been developed to improve its performances.<sup>5,8–10</sup> This study focuses on the DAS beamforming obtained with the generalized cross-correlation in the time domain.<sup>11</sup> In DAS beamforming, spatial likelihood functions (SLFs) are classically averaged using an arithmetic mean to form the noise source map. Similarly to R. P Dougherty's approach that operates in the frequency domain (functional beamforming),<sup>12</sup> this work investigates an alternative way of processing the SLFs, namely, the use of the geometrical or harmonic means in order to minimize the computation complexity and increase the accuracy of the localization. The performances of the proposed methods are compared in terms of source position detection and source level estimation both with numerical and experimental data.

## 2. Acoustic source localization based on generalized cross-correlation

### 2.1 Acoustic signal

An acoustic point source located at position  $\mathbf{r}_s$  generates a broadband signal  $s(\mathbf{r}_s, t)$  (where  $t$  is time) which is recorded by a set of  $M$  distributed microphones at positions  $\mathbf{r}_m$  (with  $m = 1, \dots, M$ ). The signal  $x_m(t)$  at the microphone  $m$  can be written as

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

where  $\alpha_m(\mathbf{r}_s)$  is the geometrical attenuation due to the propagation between the source and microphone  $m$  and  $v_m(t)$  is an uncorrelated additive background or sensor noise.

<sup>a)</sup>Author to whom correspondence should be addressed.

The time of flight (ToF)  $\Delta t_m$  is the propagation delay between the source and the microphone  $m$  and is defined by the Euclidean distance

$$\Delta t_m = \frac{1}{c_0} \|\mathbf{r}_m - \mathbf{r}_s\|_2, \tag{2}$$

where  $c_0$  is the sound speed and  $\|\cdot\|_p$  is the  $l_p$ -norm of a vector or a matrix. Since the location information of the acoustic source lies in the ToF, the next step is to estimate the time delay between signals measured by a microphone pair.

### 2.2 Generalized cross-correlation

The time delay between two microphone signals  $x_m$  and  $x_n$  can be estimated with the cross-correlation function  $R_{x_m x_n}(\tau)$ . Commonly, the cross-correlation function between two signals is obtained by the inverse fast Fourier transform of the cross-spectrum  $C_{x_m x_n}$  which allows prefiltering operation

$$R_{x_m x_n}(\tau) = \sum_{k=0}^{N_f-1} W(k) C_{x_m x_n}(k) \exp\left(j2\pi \frac{k}{N_f} \tau\right), \tag{3}$$

where  $W(k)$  is the prefilter and  $k$  is the frequency index. The common technique used to prefilter the cross-spectrum is the phase transform (PHAT) which allows removing its magnitude in order to only retain the phase information. This approach to compute the cross-correlation function is named the generalized cross-correlation (GCC) or GCC-PHAT when the PHAT prefiltering operation is used (in the following the term PHAT is omitted for simplicity). If the estimated GCC between two microphone signals is interpolated over spatial locations  $\mathbf{r}_l$  where the source is searched, the resulting map is typically a hyperboloid in three dimensions and hyperbola in two dimensions, this map is called SLF. However, the source position cannot be accurately detected with a single hyperboloid or hyperbola. If three microphones are used, three hyperbolas (in two dimensions) can be defined. If the corresponding SLFs are summed, the maximum value corresponds to the intersection of the hyperbolas and provides the source position. Commonly, the output signal  $y^{AM}(\mathbf{r}_l)$  is obtained by calculating the arithmetic mean (AM) of the SLFs determined for  $M_p$  microphone pairs as

$$y^{AM}(\mathbf{r}_l) = \frac{1}{M_p} \sum_{p=1}^{M_p} R_p(\tau_{r_l}), \tag{4}$$

where  $R_p$  is the cross-correlation function of microphone pair  $p$  and  $\tau_{r_l} = \Delta t_{ml} - \Delta t_{nl}$  is the time lag for the spatial source location  $l$ .

### 2.3 Modified generalized cross-correlation

In this study, alternative means are used to compute the output signal. First, the generalized mean or power mean  $g_n$  (with exponent  $n$ ) of a set of positive real numbers  $z_p$  with  $p = 1, \dots, P$  is introduced,

$$g_n = \left( \frac{1}{P} \sum_{p=1}^P z_p^n \right)^{1/n}. \tag{5}$$

If  $n = 1$  the generalized mean corresponds to the AM. If  $n = 0$ , the generalized mean is equal to the geometrical mean (GM)

$$g_0 = \left( \prod_{p=1}^P z_p \right)^{1/P}. \tag{6}$$

If  $n = -1$  the harmonic mean (HM) is obtained and can be expressed as

$$g_{-1} = P \left( \sum_{p=1}^P \frac{1}{z_p} \right)^{-1}. \tag{7}$$

In the case of the GCC and  $M_p$  microphone pairs, the GM and HM are expressed as

$$y^{GM}(\mathbf{r}_l) = \left( \prod_{p=1}^{M_p} |R_p(\tau_{r_l})| \right)^{1/M_p}, \tag{8}$$

$$y^{HM}(\mathbf{r}_l) = M_p \left( \sum_{p=1}^{M_p} \frac{1}{|R_p(\tau_{r_l})|} \right)^{-1} \quad (9)$$

The set of numbers has to be positive with the GM and HM.<sup>13</sup> However the cross-correlation functions may contain negative values, therefore  $|\cdot|$  denotes the absolute value. It can be shown that when the exponent  $n \rightarrow +\infty$  (respectively,  $n \rightarrow -\infty$ ) the mean corresponds to the maximum (respectively, minimum) of the set values. The following inequality holds

$$y^{AM} > y^{GM} > y^{HM} \quad (10)$$

On the basis of Eq. (10), differences in the resulting noise source map may be expected. The noise source maps obtained with the different means are compared in Sec. 3.

### 3. Comparison of simulated noise source maps obtained with the different means

To highlight the influence of each mean in perfectly controlled conditions, numerical data are used. The goal is to compare the noise source maps provided by each mean (AM, GM, and HM) in terms of source position detection and level estimation. Four acoustic point sources located at  $x = [-0.6; -0.2; 0.2; 0.6]$  m,  $y = 1$  m, and  $z = 0$  m generate uncorrelated white noise; the associated source levels are  $[0; -3; -6; -9]$  dB, respectively. The acoustic source signal is recorded by a six-microphone array. Therefore, the number of microphone pairs is 15. Each microphone is located at the face center of a cube of side length 0.5 m [see Fig. 1(a)]. The center of the microphone array is located at the origin of the coordinate system. The source positions are searched in a square scanning zone with side equal to 0.8 m in the  $x - z$  plane at  $y = 1$  m. The square is discretized with 41 points in each direction which leads to a total number of 1681 scan points. The cross-correlation function of the microphone pair signals is obtained with the GCC. The different means described in Sec. 2.3 are used to compute the noise source map. Therefore the only difference between the noise source maps is the combination of the SLFs. The noise source map obtained classically with the AM is shown in Fig. 1(b). It should be recalled that the AM does not require taking the absolute value of the cross-correlation functions. In this case, it is difficult to detect the four source positions due to the presence of several side lobes which merge. Moreover, due to the merging of the side lobes, the level of each source is not

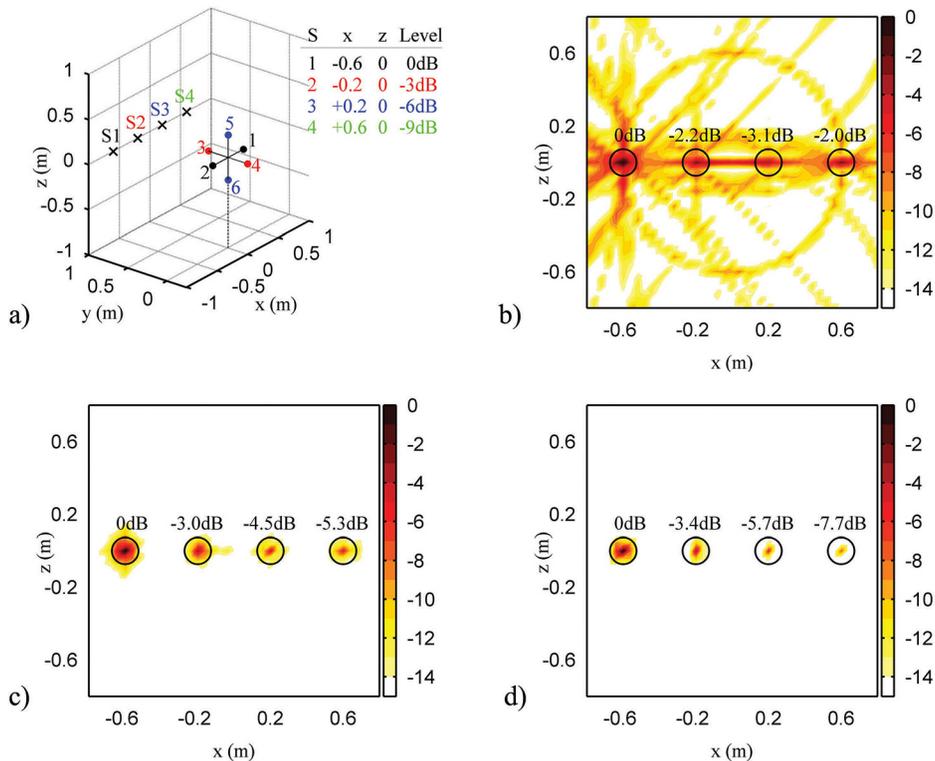


Fig. 1. (Color online) (a) Numerical set-up with the four sources (black crosses) and the six-microphone array (dots with number). Noise source maps with (a) AM, (b) GM, and (c) HM. The black circles are the source positions. The source level is given above the circles. The color bar is in dB.

correctly estimated. The level of the two lowest sources is over-estimated by 3 dB for the third source and 7 dB for the fourth source. Therefore, the GCC with the AM is not able to correctly detect the source positions and estimate the source levels in this case. Figure 1(c) presents the noise source map obtained with the GM. In this case, the side lobes are removed, and it is possible to clearly detect the source positions. Moreover, the source level estimation with the GM is better than the AM. However the best estimation of the source level is reached with the HM which leads to an error less than 0.5 dB for the second and third sources and a 1.3 dB error for the fourth source. Therefore the GM and HM clearly improve the noise source map obtained with the GCC in terms of source positions and level estimation without an increase in the computation complexity.

#### 4. Comparison of experimental noise source maps obtained with the different means

In order to confirm the above findings, experiments were conducted in the reverberant room of the Infrastructure Commune en Acoustique pour la Recherche laboratory (École de Technologie Supérieure-Institut de recherche Robert-Sauvé en santé et en sécurité du travail, Montréal). The reverberation time TR60 is larger than 2.6 s in the frequency range between 100 Hz and 3 kHz and the peak value is 3.3 s at 500 Hz which leads to reverberation radii of 0.52 and 0.46 m, respectively. To set-up the microphone array, a frame was composed of a sphere of radius 3.81 cm supported by a tripod. Holes were drilled in the sphere according to the microphone array geometry. Rods with 20 cm length were inserted into the holes and the microphones were mounted at the end of the rods to obtain an array radius of 0.25 m. Brüel & Kjær (Nærum, Denmark) microphones type 4935 were used and the signals were recorded using a Brüel & Kjær 3038B front end and Brüel & Kjær Pulse software. The acoustic signals were sampled at 32 768 Hz during 15 s. The source signals were uncorrelated white noise, with the same amplitude, generated by a NI PXI-4461 (Austin, TX, USA) card controlled with Labview. The signals were amplified by a BSWA audio amplifier SWA 100 (Beijing, China) 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 reverberant field. The two loudspeakers were separated by 55 cm. The microphone array was located at the same distance from the loudspeakers as the back wall and the side wall was closer. The centers of the microphone array and of 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 scan zone where the sources are searched was a 1.2 m  $\times$  1.2 m vertical square 1.7 m from the microphone array (the scan zone therefore contains the loudspeakers). The scan zone was discretized with 31  $\times$  31 points. The noise source map provided by the GCC with the AM is shown in Fig. 2(a). The noise source map exhibits two spots at the source positions, therefore the GCC with the AM is able to detect the loudspeakers. However, side lobes and spurious lobes are present which prevent an efficient localization. Figures 2(b) and 2(c) present the noise source maps obtained with the GM and HM. The best noise source map is provided by the HM which narrows the main lobes and removes side lobes.

Despite an accurate detection and a better source separation, the noise source maps obtained with the GM and HM still exhibit some spurious lobes. Without the PHAT prefiltering, the cross-correlation values are generally positive. If the PHAT prefiltering is used the cross-correlation function is whitened and the values may oscillate around zero. When using numerical data, the cross-correlation function is always positive. However, the cross-correlation function of the experimental data exhibits values

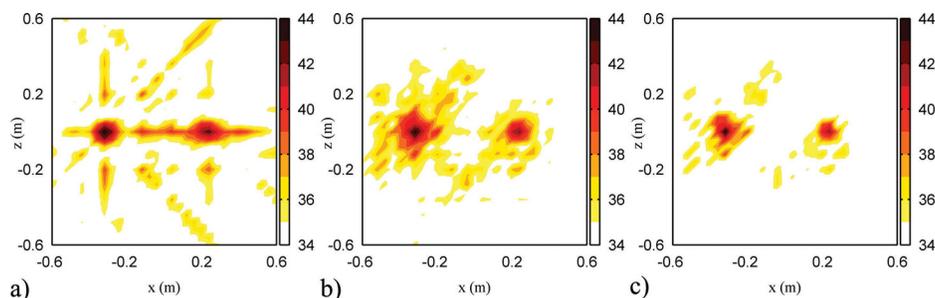


Fig. 2. (Color online) Experimental noise source maps of two loudspeakers in front of a six-microphone array obtained with (a) AM, (b) GM, and (c) HM. The experiments have been conducted in a reverberant room. The color bar is in dB.

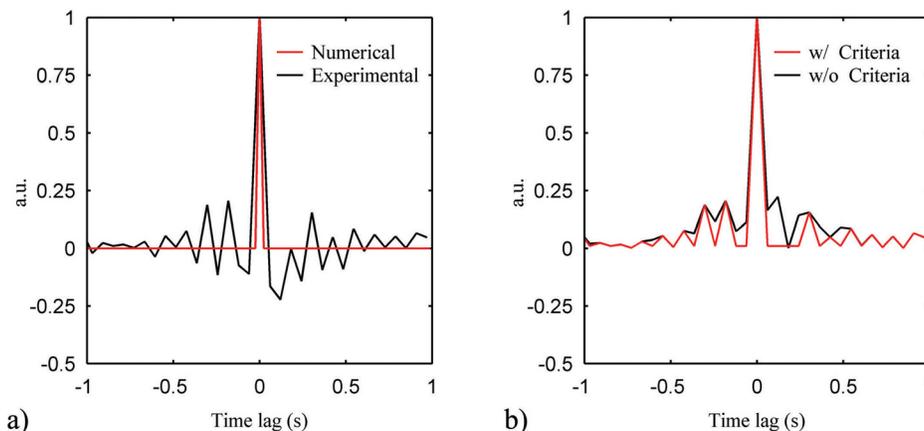


Fig. 3. (Color online) (a) Example of cross-correlation functions with numerical and experimental data. (b) Absolute value of the cross-correlation function with and without the threshold criteria.

oscillating around zero. An example of numerical and experimental cross-correlation functions is given in Fig. 3(a). If the AM is used, these negative values are beneficial. Indeed, when the noise source map is computed the negative and positive values of the SLFs average out and therefore the spurious lobe amplitudes are reduced. However, since the GM and HM operate on positive values [see Eq. (8) and Eq. (9)], taking the absolute value of the cross-correlation functions introduces additional spurious lobes.

To improve the noise source maps provided by the GM and HM, a selection criterion is proposed. The goal is to replace the negative values of the cross-correlation function by a positive threshold value. This threshold value has to be as low as possible and could be zero but introducing zero in a product could lead to the removal of useful information. Therefore, the threshold value is based on a low percentage of the maximum of the cross-correlation function. The following threshold value is proposed

$$\text{if } R_p(\tau_{rl}) < 0 \text{ then } R_p(\tau_{rl}) = 1\% \max(R_p(\tau_{rl})). \tag{11}$$

This threshold value avoids absolute zero and replaces negative values by very small positive values, which can then be considered as background noise [see Fig. 3(b)]. Smaller values such as 0.01% do not improve further and larger values (above 5%) introduce spurious lobes.

The noise source maps obtained with the threshold criteria for the three means are shown in Fig. 4. As expected, removing the negative values of the cross-correlation function does not improve the noise source map obtained with AM; the side lobe levels are higher which may introduce false sources. With the GM and HM, the noise source maps exhibit small spots at both source positions. Almost all the side and spurious lobes are removed. The best detection is provided by the HM which presents only two tiny spots at the source positions.

### 5. Conclusion

This paper deals with the use of different means to compute noise source maps based on the GCC. Commonly, the AM of the spatial likelihood functions is used to produce the noise source map. In this study, the AM has been replaced by the GM and HM.

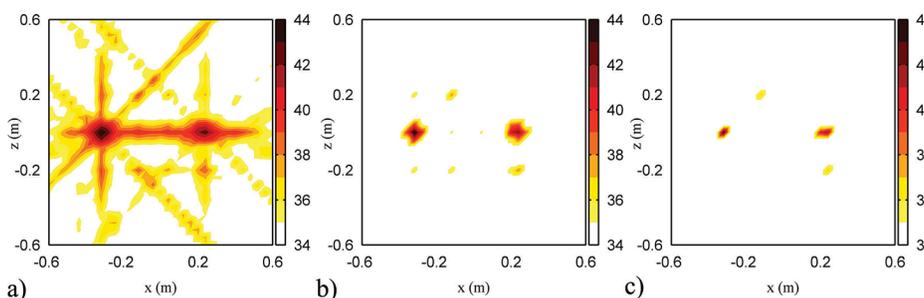


Fig. 4. (Color online) Experimental noise source maps of two loudspeakers in front of a six-microphone array obtained with (a) AM, (b) GM, and (c) HM. The experiments have been conducted in a reverberant room. The negative values of the GCC are removed from the computation using the threshold criterion in Eq. (11). The color bar is in dB.

In the numerical case of four sources with different sound levels in front of a six-microphone array, the classic AM does not allow for a clear detection of the source positions and levels unlike GM and HM which remove the side lobes and narrow the main lobes thereby improving the source localization. The best result is provided by the HM which slightly underestimates the sound level of the four sources.

When operating on experimental data obtained in a reverberant room, the GM and HM provide a better noise source map than the AM. However, some spurious lobes are always present.

A threshold criterion was introduced to remove the negative values of the GCC which are due to background noise. The threshold criterion associated with the GM or HM allows for cleaning the noise source map. The best localization is obtained with the HM where only two tiny spots are present at the source positions.

Finally, with few microphones, the proposed method allows for an enhanced detection of source positions. Moreover, the sum operation used in the classic AM being replaced by a product (GM) or by a combination of divisions and sum, the computational cost of the signal processing is not increased.

### References and links

- <sup>1</sup>C. Camier, T. Padois, J. Provencher, P.-A. Gauthier, A. Berry, J.-F. Blais, M. Patenaude-Dufour, and R. Lapointe, "Fly-over source localization on civil aircraft," in *19th AIAA/CEAS Aeroacoustics Conference*, Berlin, Germany (May 27-29, 2013), pp. 1–11.
- <sup>2</sup>Q. Wei, S. Zhong, and X. Huang, "Experimental evaluation of flow-induced noise in level flight of the pigeon (*Columba livia*)," *J. Acoust. Soc. Am.* **134**(1), EL57–EL63 (2013).
- <sup>3</sup>J. P. Dmochowski, J. Benesty, and S. Affes, "A generalized steered response power method for computationally viable source localization," *IEEE Trans. Audio Speech Language Processing* **15**(8), 2510–2526 (2007).
- <sup>4</sup>H. Do and H. Silverman, "SRP-PHAT methods of locating simultaneous multiple talkers using a frame of microphone array data," in *Proceedings of 2010 IEEE International Conference on Acoustics Speech and Signal Processing*, Dallas, TX (March 14-19, 2010), pp. 125–128.
- <sup>5</sup>T. Padois, F. Sgard, O. Doutres, and A. Berry, "Comparison of acoustic source localization methods in time domain using sparsity constraints," in *Proceedings of Internoise*, San Francisco, CA (August 9-12, 2015), pp. 1–10.
- <sup>6</sup>C. Noël, V. Planeau, and D. Habault, "A new temporal method for the identification of source directions in a reverberant hall," *J. Sound Vib.* **296**(3), 518–538 (2006).
- <sup>7</sup>D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques* (Prentice Hall, Upper Saddle River, NJ, 1993), Chap. 4, pp. 111–198.
- <sup>8</sup>T. F. Brooks and W. M. Humphreys, "A deconvolution approach for the mapping of acoustic sources (DAMAS) determined from phased microphone arrays," *J. Sound Vib.* **294**(4-5), 856–879 (2006).
- <sup>9</sup>T. Suzuki, "L1 generalized inverse beam-forming algorithm resolving coherent/incoherent, distributed and multipole sources," *J. Sound Vib.* **330**(4), 5835–5851 (2011).
- <sup>10</sup>T. Padois, P.-A. Gauthier, and A. Berry, "Inverse problem with beamforming regularization matrix applied to sound source localization in closed wind-tunnel using microphone array," *J. Sound Vib.* **333**(25), 6858–6868 (2014).
- <sup>11</sup>C. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust. Speech Signal Processing* **24**(4), 320–327 (1976).
- <sup>12</sup>R. P. Dougherty, "Functional beamforming," in *Berlin Beamforming Conference (BeBeC)*, Berlin, Germany (May 19-20, 2014), pp. 1–25.
- <sup>13</sup>F. W. J. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark, *NIST Handbook of Mathematical Functions* (Cambridge University Press, Cambridge, UK, 2010), Chap. 1, pp. 3.