Authors' accepted manuscript
Article published in *Applied Acoustics*, Volume 177, June 2021, 107914
The final publication is available at https://doi.org/10.1016/j.apacoust.2021.107914

Acoustic imaging with conventional frequency domain beamforming and generalized cross correlation: a comparison study

Thomas Padois[a], Jeoffrey Fischer[b], Con Doolan[b], Olivier Doutres[a]

[a] *Department of Mechanical Engineering, École de Technologie Supérieure (ÉTS), Montréal, (Qc), H3C 1K3 ,Canada*
[b] *School of Mechanical and Manufacturing Engineering, UNSW Sydney, NSW 2052, Australia*

Abstract

Nowadays, acoustic cameras have become a standard tool for visualizing sound sources. The main post-processing techniques for generating acoustic images are associated with beamforming in the time or frequency domain. While many advanced techniques have been developed for the latter in order to achieve high resolution, such as deconvolution or inverse methods, the former have been less investigated. Recent studies have shown that time-domain beamforming can provide a narrow main lobe with low side lobe levels in different situations. However, a comparison of both techniques in terms of acoustic imaging does not exist. This paper provides a detailed comparison of conventional frequency-domain beamforming (CBF) and time-domain beamforming based on the generalized cross-correlation technique (GCC). First, numerical data are used for assessing both techniques. Then, experimental tests are carried out in a hemi-anechoic room using loudspeakers or power tools. As expected, acoustic images provided by CBF and GCC are found to be very similar but with different computation times. In order to improve the acoustic image, advanced techniques are considered (Clean-SC and the GCC based on geometric mean). When dealing with low frequency, Clean-SC is not able to detect both sources while the GCC based on geometric mean provides an accurate source separation.

Keywords: Acoustic imaging, beamforming, time domain, frequency domain, generalized cross-correlation, deconvolution technique

1. Introduction

Nowadays, many acoustic manufacturers propose acoustic cameras for visualizing sound sources. These tools include a microphone array, a camera and

Email address: `Corresponding author : Thomas.Padois@etsmtl.ca` (Thomas Padois)

a data acquisition system. The digitized microphone signals are combined in order to generate an acoustic image while the camera captures an image of the environment under study. Then, both images are overlaid. The result is usually a black and white image of the environment with filled colored contours. The peaks of the contours exhibit the source positions and the colors provide the sound levels. Acoustic imaging techniques have been used for a wide range of applications from cars [1, 2, 3] to aircraft [4, 5, 6] or even snowmobiles [7]. If the microphone array is set close to the object under study, typically less than a wavelength away, the common technique is the near-field acoustic holography [8, 9]. However, if the source is far from the microphone array, the classic technique used is beamforming [10, 11]. In this work, the focus is made on the comparison of the frequency and time domain beamforming.

The standard beamforming technique for generating an acoustic image, also know as conventional beamforming (CBF) [12] is applied in the frequency-domain. First, the Cross Spectral Matrix (CSM), obtained from a Fast Fourier Transform of the microphone signals, is computed. The self noise of the microphones may be removed by setting to 0 the diagonal elements of the CSM [13]. Then, a virtual focusing plane containing the source is considered which allows for the computation of the steering vector [14, 15]. Finally, the CSM and the steering vector are combined in order to generate the acoustic image. The two main characteristics of an acoustic image are the main lobe, which indicates the source position, and the side lobes which represent spurious sources positions [16]. The CBF is known as a fast and robust technique, but presents some drawbacks when dealing with low or high frequencies. Indeed, the main lobe width enlarges when frequency decreases, while the side lobe amplitude increases with the frequency [17]. In order to tackle theses issues, improvements have been proposed such as optimized geometries [18, 19, 20, 21, 22, 23] or advanced post-processing techniques [24, 25, 26]. In terms of geometry, the common tradeoff seems to be now the spiral arrangements for planar array [27], while for the post-processing, deconvolution techniques or inverse problems become standard techniques. The growing interest in acoustic imaging with CBF has been recently highlighted by the publication of three review papers dealing with advanced post-processing techniques, array geometry or applications [24, 25, 26]. As mentioned by the Ref. [26], the major steps forward in acoustic imaging have been performed in the frequency-domain although some time-domain beamforming techniques exist.

The time-domain beamforming consists in delaying and summing the microphone signals (called D&S) [12, 28, 29]. Unlike the CBF which is a narrow-band technique (an acoustic image is obtained for each frequency bin), the D&S is a broadband technique. Indeed, the whole microphone signals are delayed, therefore no frequency band is selected. In 2004, Dougherty has investigated the influence of a pre-filter operation on microphone signals in order to select the n^{th} octave band [30]. He also introduced a formulation for diagonal removal and cross-shaped array for improving the acoustic image. However, the acoustic images provided by the time D&S and CBF were not compared. In 2006, Jaeckel has discussed the strength and weakness of D&S beamforming [31]. The strength of D&S beamforming were the applicability to non-stationary and strongly transient signals [32] as well as good resolution of the broadband signals. The weakness were the need for high sampling rate and the poor resolution at low frequency.

Advanced signal processing algorithms are mainly implemented in the frequency domain, few techniques exist in the time-domain [33, 34] and are inspired by the well known frequency-domain technique Clean-SC [35]. Recently, studies have taken the advantages of the generalized cross-correlation formulation

(GCC) to improve D&S beamforming [36, 37]. The GCC has been introduced by Knapp in 1976 [38]; this technique allows a pre-filtering operation by a weighting function before computing the cross-correlation. The most known weighting function is the PHAse Transform (PHAT) which divides the cross spectrum of a microphone pair by its absolute value [38]. When a broadband signal is considered, this weighting function narrows the main lobe of the cross-correlation and therefore improves the acoustic image. Other improvement techniques based on GCC have been proposed such as a spatial weighting [39], the use of the generalized mean [40] or enhanced weighting functions [41]. Each technique allows for narrowing the main lobe and decreasing the amplitude of the side lobes. Inverse method based on sparsity constraints have also been proposed to achieve acoustic imaging with high resolution [42, 43, 44].

In 2014, Hamid *et al.* have compared the performance of the time and frequency domain beamforming in terms of computation time [45]. However, from the author's knowledge no study has provided a detailed comparison of acoustic images given by both techniques. The CBF is considered as the standard technique and advanced techniques are mainly based on its formulation. The D&S beamforming is rarely used but seems to be more appropriated for some cases. Although, the CBF and D&S beamforming should provide similar results, their formulations are different as well as the advanced techniques. Therefore, a detailed comparison of both techniques would allow for selecting the more appropriated one.

The objective of this work is to propose a detailed comparison of both techniques in terms of acoustic images. First, the theoretical background is introduced in Section 2. Then, numerical data are used to compare the acoustic images provided by both techniques (Section 3). Finally, experimental data are considered in Section 4. The acoustic images are discussed in the simple case of two loudspeakers. Then power tools are considered, a leaf blower (aeroacoustic source) and a nail gun (impulsive source). A comparison of advanced post-processing is also discussed.

2. Theoretical background

2.1. Microphone signals

A source signal $s(\mathbf{r}_s, t)$ is generated by an omnidirectional acoustic point source located at $\mathbf{r}_s$ and recorded by a set of M microphones at locations $\mathbf{r}_m$. The signal x_m recorded by the microphone m is given by

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

where t represents time and $v_m(t)$ is an uncorrelated additive noise due to background or sensor noise. The term Δt_{ms} corresponds to the time of flight (ToF) between the source and the microphone and is defined by the Euclidean distance

$$\Delta t_{ms} = \frac{\mathbf{r}_{ms}}{c_0} = \frac{\|\mathbf{r}_m - \mathbf{r}_s\|_2}{c_0}, \tag{2}$$

where c_0 is the sound speed and $\|\cdot\|_p$ is the l_p-norm of a vector or a matrix.

The cross spectrum $C_{x_m x_n}(f)$ between the microphone signals x_m and x_n is given by

$$C_{x_m x_n}(f) = X_m(f) X_n(f)^H, \tag{3}$$

where X_m and X_n are the frequency domain microphone signals obtained with the Fast Fourier Transform (FFT) and the superscript $(\cdot)^H$ represents the Hermitian transpose.

With the CBF, the cross and auto spectra are gathered into the so-called Cross Spectral Matrix (CSM), denoted $\mathbf{C}$, and are usually averaged using the Welch's periodogram [46] for removing background noise.

With the GCC, the cross spectra are used to recover the cross-correlation function using the Inverse Fast Fourier Transform (iFFT).

For each technique, a scan zone where the source positions is sought has to be defined. In the case of a planar microphone array, the scan zone is usually a plane regularly discretized with L points at position r_l.

2.2. Frequency-domain beamforming: Conventional Beamforming (CBF)

In the frequency domain, the beamformer output power $Z_l(f)$ for a frequency bin is given by

$$Z_l(f) = \mathbf{h}_l(f)^H \mathbf{C}(f) \mathbf{h}_l(f), \tag{4}$$

where $\mathbf{h}_l(f)$ is known as the steering vector and is a normalized Green's function between the microphone located at $\mathbf{r}_m$ and the scan point at $\mathbf{r}_l$. In order to reduce the self-noise of the microphones, the diagonal elements of the CSM are set to 0.

There are several formulations of the steering vector in the literature [14], here one element of the steering-vector is

$$h_l(\mathbf{r}_m, f) = \frac{g_l(\mathbf{r}_m, f)}{\sqrt{M}\sqrt{\mathbf{g}_l(f)^H \mathbf{g}_l(f)}}, \tag{5}$$

which is referred as formulation IV in reference [14].

Note that the vector $\mathbf{g}_l(f)$ gathers elements from matrix $g_l(\mathbf{r}_m, f)$ that correspond to each microphone location $\mathbf{r}_m$. The vector $g_l(\mathbf{r}_m, f)$ is based on the free-field Green's function, thus frequency-domain beamforming makes the assumption of monopolar propagation,

$$g_l(\mathbf{r}_m, f) = \frac{r_{0l}}{r_{ml}} \exp(-jk(r_{ml} - r_{0l})), \tag{6}$$

where r_{0l} is the same as r_{ml} but using a reference microphone, usually the center microphone of the array and $k = 2\pi f / c_0$ represents the wavenumber. In the following, the acoustic images provided by the frequency domain beamforming are referred as CBF.

2.3. Time-domain beamforming: Generalized Cross-Correlation (GCC)

When the speed of sound is known, the ToF between a microphone and a scan point can be computed. Each microphone signal is delayed by the corresponding ToF, which corresponds to steer the microphone array into the direction of the scan point. 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 scan point l as

$$y_l(t) = \sum_{m=1}^{M} x_m(t + \Delta t_{ml}), \tag{7}$$

where Δt_{ml} is the Tof between microphone m and scan point l.

The beamformer output power $Y_l(t)$ is given by

$$\mathrm{Y}_l(t) = \mathrm{E}\{y_l^2(t)\} = \sum_{m=1}^{M} \sum_{n=1}^{M} \mathrm{R}_{x_m,x_n}(\Delta t_{ml} - \Delta t_{nl}), \tag{8}$$

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

$$\mathrm{R}_{x_m,x_n}(\tau) = \mathrm{E}\{x_m(t)x_n(t+\tau)\}, \tag{9}$$

where τ is a time lag.

The cross-correlation function R_{x_m,x_n} is typically recovered from the iFFT of the cross spectrum $C_{x_m x_n}$ (Eq. 3) between microphone signals

$$\mathrm{R}_{x_m,x_n}(\tau) = \sum_{f=0}^{N_f-1} W(f) C_{x_m x_n}(f) \exp\left(j2\pi \frac{f}{N_f} \tau\right), \tag{10}$$

where f is the frequency bin, N_f the number of frequency bins and $j = \sqrt{-1}$. A weighting function $W(f)$ can be used for improving the cross-correlation estimation which leads to the GCC [38, 47].

As the scan zone has a finite size, a peak time delay τ_{max} can be defined from all scan points which allows for selecting the values of the cross-correlation between $[-\tau_{max} : \tau_{max}]$. The peak time delay is given by

$$\tau_{max} = \max\left(\frac{1}{c_0}\left|\mathbf{r}_{ml} - \mathbf{r}_{nl}\right|\right), \tag{11}$$

with $|\cdot|$ the absolute value.

The projection of the cross-correlation function for a single microphone pair over the scan zone is called Spatial Likelihood Function [48] (SLF)

$$SLF = \mathrm{R}_{x_m,x_n}(-\tau_{max} : \tau_{max}). \tag{12}$$

Commonly, the computation of the beamformer output power Eq. 8 is performed over the $M_p = M(M-1)/2$ microphone pairs due to the symmetry of the cross-correlation matrix (redundant information) [49]. Moreover, the auto correlation terms (auto correlation of microphone signals) are removed because they do not bring any information to the acoustic image. These terms are known as "DC" component or self-noise [49]. Finally, the acoustic image is provided by the arithmetic mean of the SLF. In the following, the acoustic images provided by the time domain beamforming are referred as the GCC.

2.4. Detailed computation of CBF and GCC

For both acoustic imaging techniques, the first step is to gather the microphone signals into a matrix. Then, the scan zone is defined. With the CBF, the cross spectral matrix and the steering vector are computed. Then, the acoustic images are given by Eq. 4 or are averaged over the frequency band if required.

With the GCC, the microphone signals are filtered in the frequency band required. The SLF are computed for each microphone pair and averaged in order to get the acoustic image. In the following, both techniques (CBF and GCC) are compared with numerical and experimental data. The Figure 1 provides the detailed computation of both techniques.

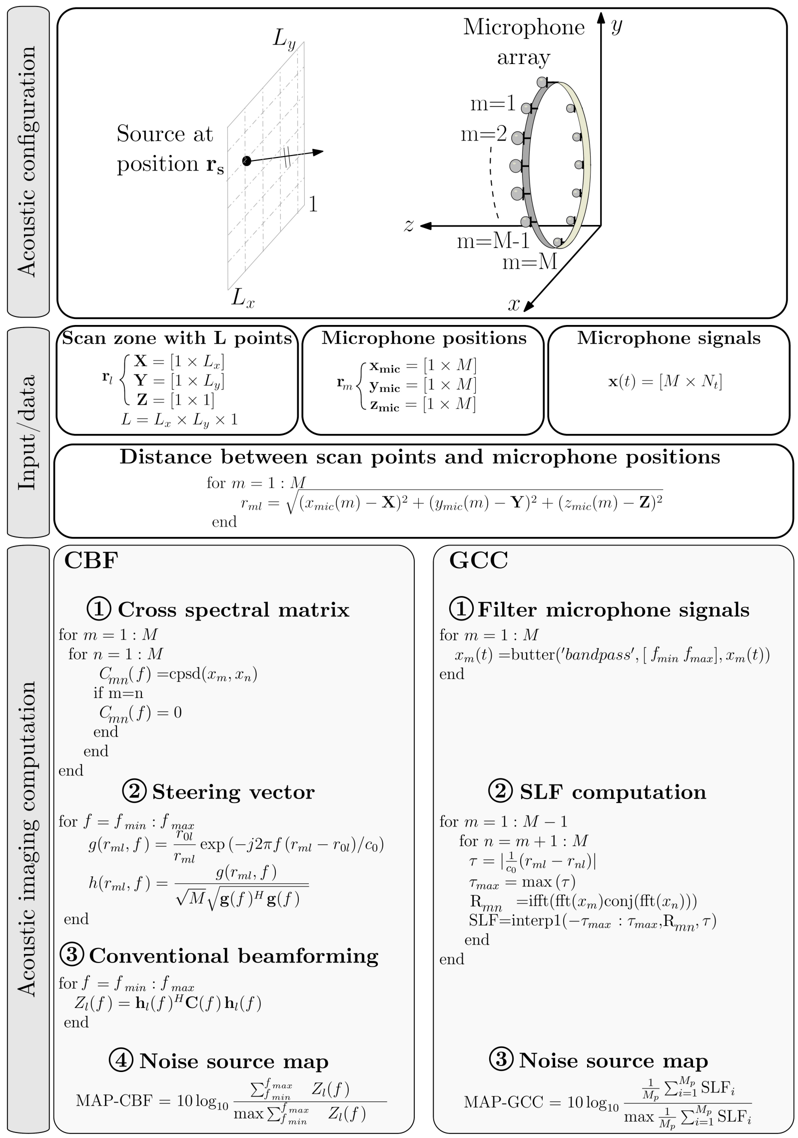


Figure 1: Detailed computation of the CBF and GCC.

2.5. Criteria for assessing an acoustic image

The result provided by both techniques is an acoustic image where a main lobe should be located at the source position. Unfortunately, in addition to the main lobe, the acoustic image also contains several side lobes (Figure 2.a). Moreover, the size of the main lobe and the amplitude of the side lobes are directly linked to the microphone array geometry, the source-array distance and the source characteristics [16, 17].

In the past, criteria have been developed to assess the acoustic image. The first criterion is the Main Lobe Width (MLW) at -3 dB which assesses the ability to separate two sources [16]. The MLW can be expressed in centimeters or degrees (Figure 2.b). In the case of a 2D acoustic image, this criterion is obtained by slicing the map along the main directions. If the distance between two sources is smaller than the MWL, the main lobes will be merged and the two sources can not be separated. Therefore, the smaller the MWL, the better the source localization. This criterion is also known as spatial resolution.

The second criterion is the Maximum Side lobe Level (MSL) which is given by the amplitude difference between the main lobe and the nearest side lobe. The MSL is expressed in decibel (dB). Again, in the case of a 2D acoustic image, this criterion is obtained by slicing the map along the main directions. If the side lobes amplitude is high, they could be interpreted as sources with lower amplitude. Therefore, the higher the MSL, the better the source localization. This criterion is also known as dynamic resolution.

An example of 2D acoustic image (for an arbitrary configuration with a single source) is given in Figure 2.a where the main lobe indicates the source position. This main lobe is surrounded by a side lobe. Secondary side lobes are also present. To make the difference with the main side lobe, these secondary side lobes are called spurious lobes in the following. To assess the acoustic image, slices along the lines $x = 0$, $y = 0$ and $x = y$ are shown in Figure 2.b. In this case, the MLW and MSL are independent of the slicing directions. However, both criteria do not take into account the spurious lobe which have a higher amplitude and are different for each direction. Therefore, although both criteria provide a useful information, they do not completely characterize the acoustic image especially the 2D aspect.

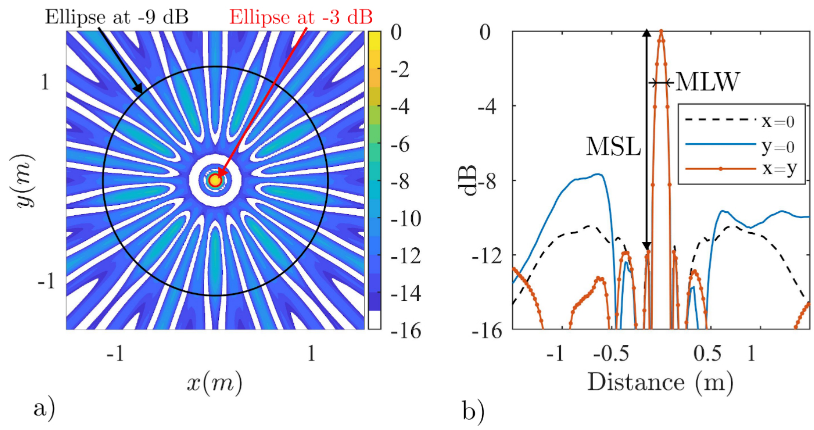


Figure 2: a) Example of acoustic image with ellipses at -3 dB (red) and -9 dB (black). b) Slices of the acoustic image along the lines $x = 0$, $y = 0$ and $x = y$. (color online)

To fully characterize the acoustic image, a criterion based on the surface of an ellipse is proposed. The goal is to surround the acoustic image values higher than a threshold. For instance, if the goal is to determine the spatial resolution, the threshold is set to -3 dB and the ellipse which surrounds all the higher values is searched. The main advantage is that the 2D aspect is taken into account. Again, the smaller the ellipse's surface, the better the source localization. If the threshold value is decreased, for instance up to -9 dB, the ellipse's surface will increase because the side lobes will be taken into account too. Finally, the ellipse's surface is also divided by the microphone array surface which provides a dimensionless criterion, called Ellipse Array Ratio (EAR). Low value of this criterion means a small surface ellipse and is expected for an efficient source localization. The EAR criterion has the advantage to provide an unique value to characterize either the main lobe or the side lobes no matter which direction is considered.

Many methods exist to define an ellipse surrounding a set of data. In this work the confidence ellipse is considered: first a threshold value has to be defined, then the acoustic image values below this threshold are discarded. The covariance matrix of this new data set is computed and the eigenvalues and vectors are searched. The minor and major axes of the ellipse are given by the eigenvalues and the orientation by the eigenvectors [50]. The 90% confidence ellipses are shown in Figure 2.a with thresholds equal to -3 dB and -9 dB. The value of 90% has been chosen to avoid influence of outliers. The first threshold allows for surrounding the main lobe only while the second takes also into account the side lobes. This value has been chosen because beamforming is not able to localize two sources with an amplitude difference larger than 9 dB (it is typically the side lobe level). Finally, the EAR criterion is given by

$$EAR = \frac{S_{ellipse}}{S_{array}} \tag{13}$$

where $S_{ellipse}$ and S_{array} are the ellipse and microphone array surfaces, respectively.

3. Numerical comparison of the CBF and GCC

3.1. Single source

The EAR criterion is now used to compare both techniques the CBF and GCC in the case of a single source in front of a 16-microphones circular array. The microphone array radius and the source-array distance range from 0.4 m to 1.4 m and from 0.2 m to 1.6 m respectively. The source signal is a white noise and the microphone signals are filtered in the 1000 Hz octave band. The frequency sampling is 44,100 Hz and 1 s of source signal is used. The scan zone is a square with 3 m sides and 40,401 points (201×201 grid). The EAR criterion values at -3 dB and -9 dB obtained with the CBF and GCC are shown in Figure 3 with respect to source-array distance and array radius. Both techniques provide similar trends. For instance, if the array radius is small, increasing the source-array distance increases the EAR criterion. Both techniques yield similar EAR criterion values which means that both techniques provide similar main lobe and side lobe levels.

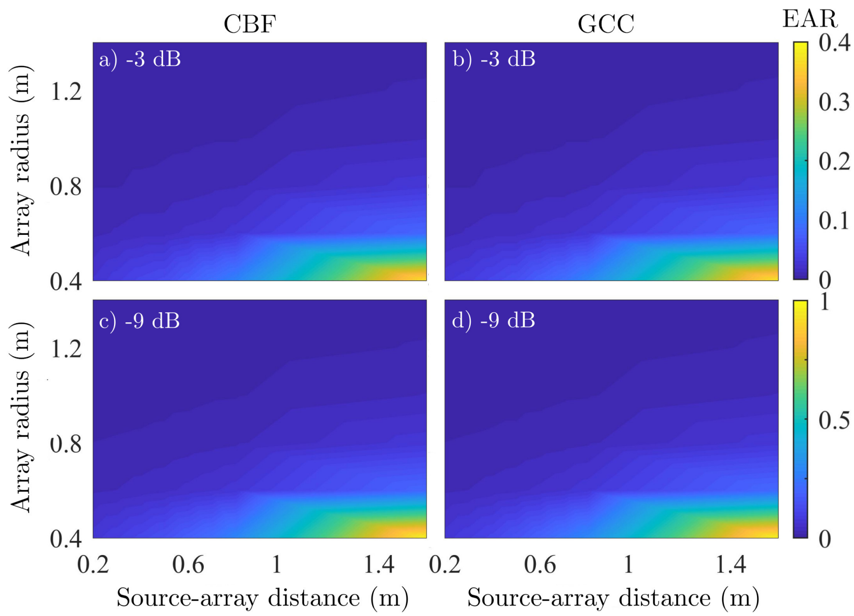


Figure 3: EAR criterion at -3 dB (a-b) and -9 dB (c-d) obtained with the CBF (left) and GCC (right). A 16-microphones circular array is used. The source signal is a white noise and the microphone signals are filtered in the 1000 Hz octave band. (color online)

As the EAR criterion is similar for both techniques at the 1000 Hz octave band, the acoustic images for octave bands ranging from 250 Hz to 4000 Hz are displayed in Figure 4 in order to visually assess the likely differences. Again, the acoustic images are very similar for both techniques, which provide the same EAR criterion for all the octave bands. The only difference is the spurious lobes below -13 dB which appear due to spatial aliasing [16]. For example, the

GCC acoustic image obtained for the 2000 Hz octave band (Figure 4.i) exhibits thinner spurious lobes than the CBF acoustic image (Figure 4.d). This first investigation allows for demonstrating that CBF and GCC similarly perform in terms of acoustic imaging in the case of a single source in front a microphone array.

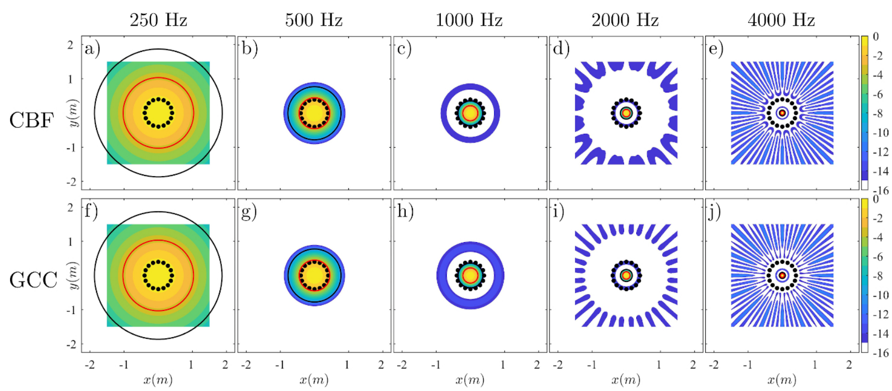


Figure 4: Acoustic images obtained with CBF (top) and GCC (bottom). The source signal is a white noise and the microphone signals are filtered in the 250 Hz (a-f), 500 Hz (b-g), 1000 Hz (c-h), 2000 Hz (d-i) and 4000 Hz (e-j) octave bands. A 16-microphones circular array is used. The red and black lines are the EAR criterion at -3 dB and -9 dB. The black dots denote the microphone positions. (color online)

3.2. Two sources with different amplitudes

In this section, the performance of both techniques is investigated in the case of two sources with different amplitudes. A 16-microphones circular array with 0.4 m radius is used. The source array distance is 1.2 m. The source amplitude difference is 6 dB and the source positions are $x = -0.3$ m and $x = 0.3$ m (with $y = 0$). The signal duration is 1 s and the octave bands 1000 Hz and 2000 Hz are considered. The acoustic images obtained and the slices along x-axis are presented in Figure 5. The EAR criterion is not considered here because the main interest is the amplitude difference. Both techniques correctly identify the two sources no matter the frequency. The side lobes amplitude below -12 dB are more important with the GCC but do not prevent an efficient source localization. The slices along the x-axis, Figure 5.c-f show that both techniques estimate similar source amplitude difference in both cases (octave bands 1000 Hz and 2000 Hz). In the octave band 1000 Hz, the source amplitude difference is slightly overestimated by both techniques, probably due to the presence of the side lobes. In the octave band 2000 Hz, the source amplitude difference is of 6.2 dB for both techniques which is close to the initial value. Therefore, both techniques provide similar source amplitude difference estimation.

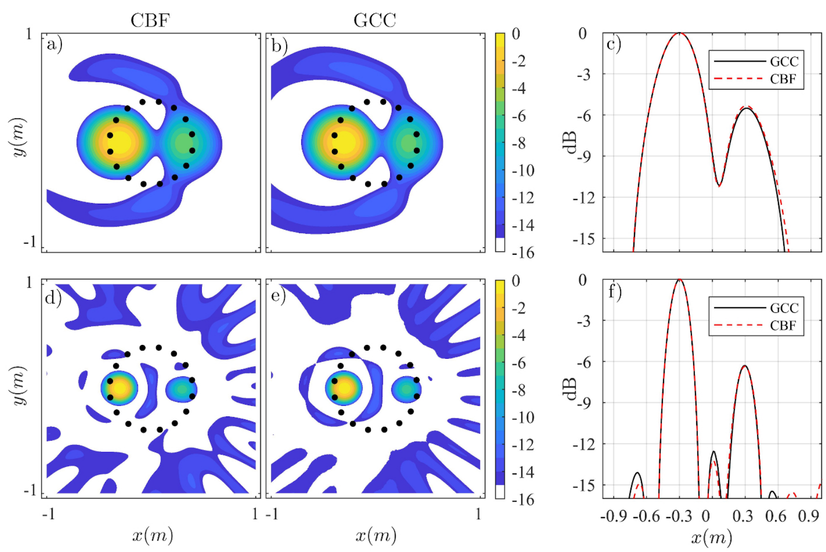
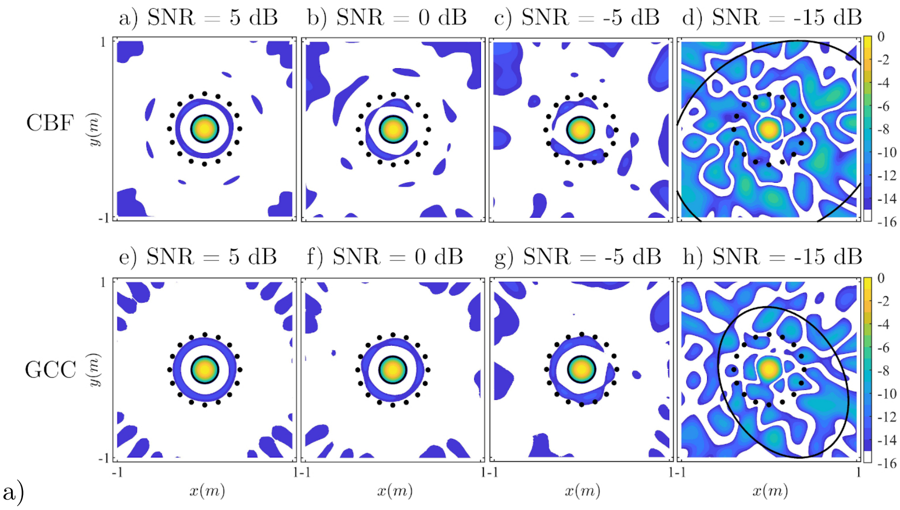


Figure 5: Acoustic images obtained with CBF (left) and GCC (center) for microphone signals filtered in the 1000 Hz (a-b) and 2000 Hz (d-e) octave bands. c-f) Slices along the line $x = 0$. A 16-microphones circular array is used. The black dots denote the microphone positions. (color online)

3.3. *Influence of the signal-to-noise ratio*

In this section, the performance of both techniques is investigated in the case of different Signal-to-Noise Ratio (SNR). A 16-microphones circular array with 0.4 m radius is used. The source array distance is 1.2 m. The signal duration is 1 s and the octave band 2000 Hz is considered. The SNR ranges from 5 dB to -15 dB. The acoustic images obtained are presented in Figure 6. With SNR=5 dB, the GCC acoustic image does not seem to be affected by the additive noise while some spurious lobes appear around the main lobe with the CBF. When the SNR decreases to -5 dB, the CBF acoustic image is contaminated by the presence of random spurious lobes. On the other hand, the GCC acoustic image is less affected. Decreasing the SNR to -15 dB, increases the spurious lobe amplitude for both techniques. However, the spurious lobes amplitude is larger with the CBF with values reaching -6 dB (as shown by the EAR criterion at -9 dB).



a)

Figure 6: Acoustic images obtained with CBF (top) and GCC (bottom). The SNR are 5 dB (a-e), 0 dB (b-f), -5 dB (c-g) and -15 dB(d-h). The source signal is a white noise and the microphone signals are filtered in the 2000 Hz octave band. A 16-microphones circular array is used. The black lines are the EAR criterion at -9 dB. The black dots denote the microphone positions. (color online)

3.4. Computation time

In this section, the computation time of both techniques is compared. The algorithms are coded with Matlab using classical functions, no optimization toolbox such as parallel computing are used. The codes, partially described in Figure 1, were run onto a personal laptop (Intel i7-7600 at 2.8 GHz, 16 Go Ram).

First the influence of the number of scan points on the computation time is investigated. The array is circular with 16 microphones and only one frequency bin is used to compute the CBF (in order to avoid frequency averaging). The number of scan points ranges from 10 to 10^6 and the computation time is displayed in Figure 7.a. In between 10 and 10^4, the GCC is four times faster than the CBF. When the number of scan point is equal to 4.10^5, the computation time is similar for both techniques. Above 4.10^5 points, the CBF becomes faster than GCC; however the time difference is only 1 sec at 10^6 points.

Now, the number of scan points is set to 4.10^4 (200×200 grid size) and the number of microphones ranges from 4 to 128. The computation time is shown in Figure 7.b. The computation times are similar when the number of microphones is equal to 10 but become largely different when the number of microphones increases. For instance, with 32 microphones the CBF is four times longer than the GCC. This fact can be easily explained by the computation process of both techniques. Indeed, the GCC requires only the microphone pairs ($M_p = (M \times (M-1))/2$), while the CBF requires the square of the number of microphones ($M^2 > M_p$).

Finally, the computation time of both techniques is compared when octave bands are considered (from 250 Hz to 8000 Hz). In this case, the number of microphones and scan points are 16 and 4.10^4, respectively. The frequency resolution (of the FFT) is an important parameter in this case because it determines the number of averages required in the CBF calculation. An arbitrary frequency resolution of 10 Hz is chosen and the result is displayed in Figure 7.c. As expected, the CBF computation time increases with the octave band because the number of frequency bins, required for the average, increases. The GCC computation time is the same for each octave band. Therefore, the GCC is much more efficient than the CBF when acoustic images have to be displayed for octave bands.

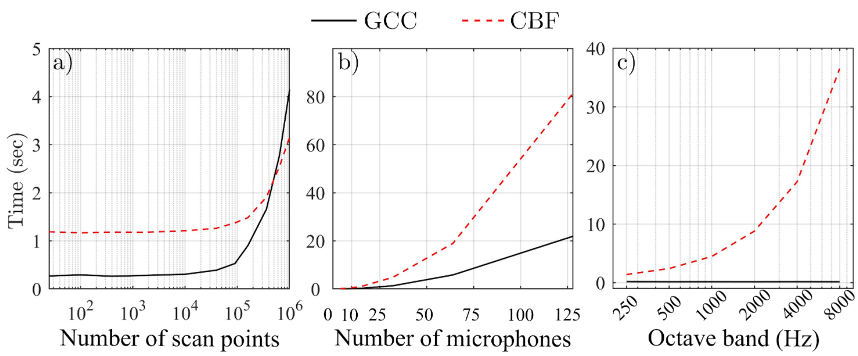


Figure 7: Computation times of the CBF (dashed red line) and the GCC (black line) versus a) the number of scan points, b) the number of microphones and c) the octave bands considered.

4. Experimental tests

The acoustic images provided by both techniques are now compared with experimental data obtained in a hemi-anechoic chamber. Various microphone array geometries and sources are considered. The aim is not to characterize the source behavior but to assess the performance of each technique for acoustic imaging. The colorbar is set to -10 dB to focus on acoustic imaging performance.

4.1. Two loudspeakers in front of a circular microphone array

During the first experiment, two loudspeakers (Eris E5 PreSonus), spaced by 50 cm were set in front of a 16-microphones circular array located at 1.2 m away. The loudspeakers signals were uncorrelated white noises with the same level and were generated with a PXI-4461 Sound and Vibration Module. The loudspeaker crossover frequency, which separates the signal of the tweeter (top) and the woofer (bottom), was of 3000 Hz. The acoustic signals were recorded with Brüel&Kjaer microphones type 4935 and 12-Ch input module type 3038B and sampled at 65,536 Hz. The microphone array aperture was 86 cm in both directions (see black dots in Figure 8.a). The scan zone, where the source positions were sought, is a square with side equal to 1 m and containing 160,801 points (401×401). The octave band 1000 Hz and 4000 Hz were considered.

The acoustic images obtained with the CBF and GCC are shown in Figure 8 for both octave bands investigated. Overall, the acoustic images are similar and detect the woofer when the 1000 Hz octave band is considered or the tweeter when the 4000 Hz octave band is considered. The main lobe width and the side lobes level are also equivalent. In this case, both techniques yield similar results.

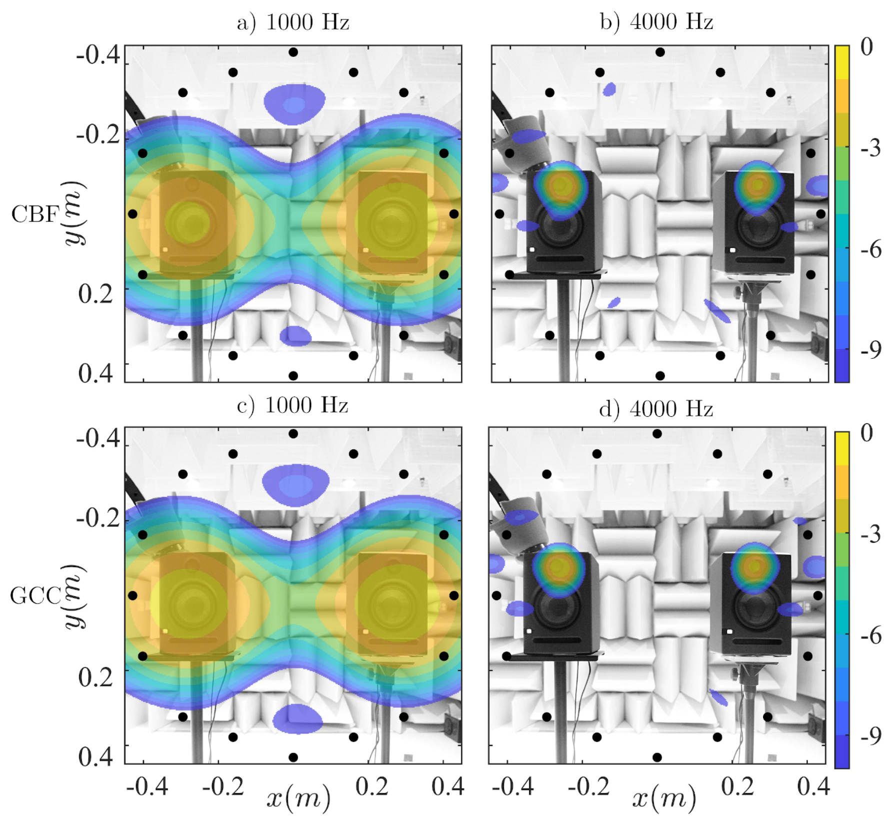


Figure 8: Acoustic images obtained with the CBF (top) and GCC (bottom). The source signals were white noises and the microphone signals were filtered in the 1000 Hz (a-c) and 4000 Hz (b-d) octave bands. A 16-microphones circular array in front of two loudspeaker with the same level was used. The black dots denote the microphone positions. (color online)

4.2. Leaf blower in front of a spiral-arm microphone array

The source considered now is a leaf blower which generates a stationary noise with a more complex frequency content. This hand-tool generates a tonal peak in mid frequencies (at 2631 Hz) and a broadband bump at higher frequencies (between 3400 and 4000 Hz). The array geometry is multi-arms logarithmic spiral shape with 41 microphones. First, the position of the source which generates the tonal peak at 2631 Hz is sought (9.a-b). Both techniques indicate that the position of this source is the air exhaust, the acoustic images are similar.

Then, the position of the source which generates the broadband bump is sought (9.c-d). Again, the acoustic images are similar and point to the engine as the main source with a secondary source at the air exhaust. Therefore in the case of a real source, both techniques provide similar results even if the noise radiated is tonal or broadband.

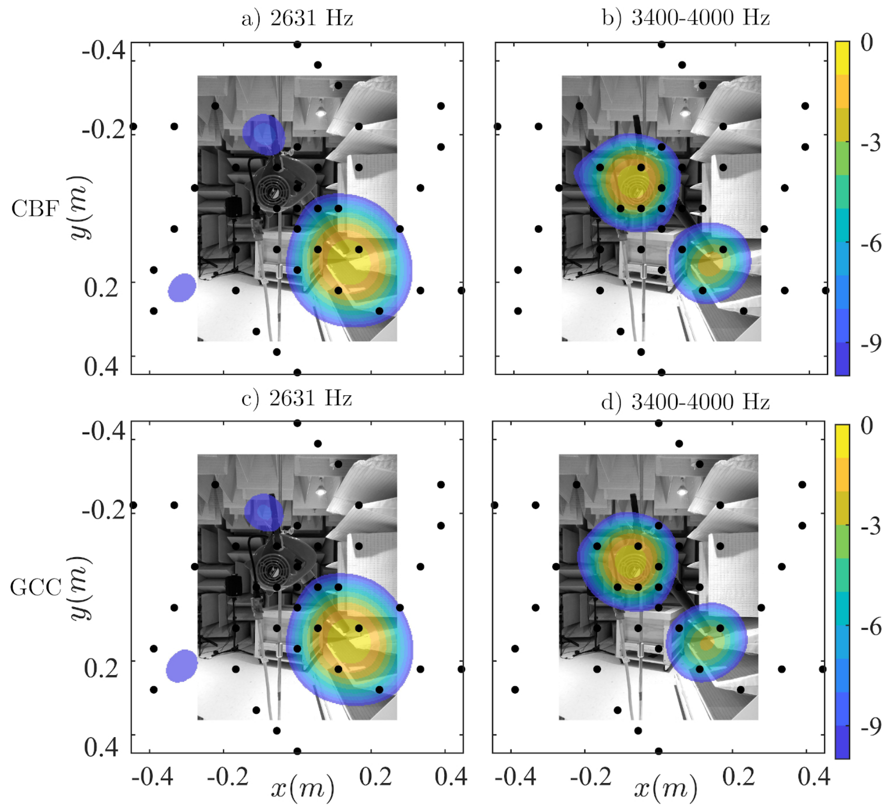


Figure 9: Acoustic images obtained with the CBF (top) and GCC (bottom). The microphone signals were filtered at 2631 Hz (a-c) and between 3400 Hz and 4000 Hz (b-d). A 41-microphones multi-arms logarithmic spiral array in front of a leaf blower was used. The black dots denote the microphone positions. (color online)

4.3. Nail gun in front of a spiral-arm microphone array

The source is now a pneumatic nail gun which typically generates impulsive noise. When the trigger nail gun is pulled up, the pressurized air is rushing in the chamber which moves down the piston and impacts the nail. Then, the piston goes up, the air is released by the exhaust at the top of the nail gun, and hits the high position. The noise generated by the exhaust is one of the main noise source of this nail gun. The noise frequency content being contained in the high frequency, the frequency band considered ranges from 2000 Hz to 5000 Hz [51]. The abilities of both techniques to localize the noise coming from the exhaust are compared in Figure 10. Both techniques provide similar results with a main lobe at the exhaust position.

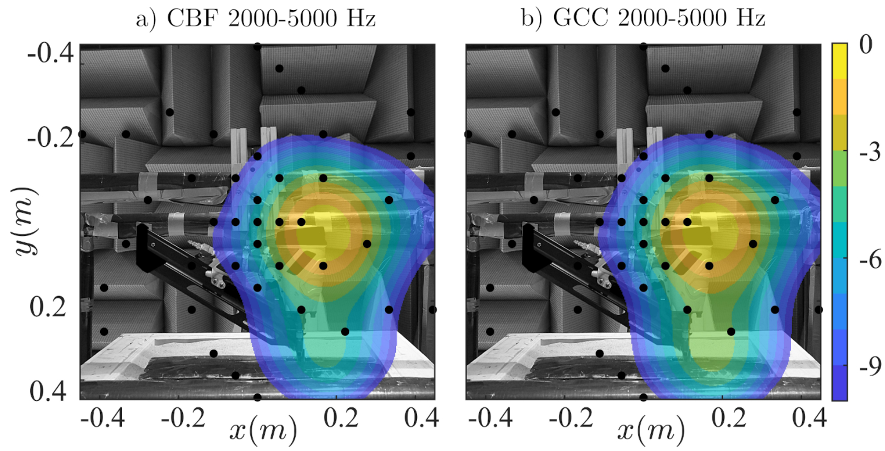


Figure 10: acoustic images obtained with the CBF (left) and GCC (right). The microphone signals were filtered between 2000 Hz and 5000 Hz. A 41-microphones spiral-arm array in front of a nail gun was considered. The black dots denote the microphone positions. (color online)

4.4. CBF and GCC with advanced techniques

It has been shown previously that CBF and GCC have a large main lobe for low frequency content or strong side lobes for higher frequencies (Figure 8). Many techniques have been developed for improving the acoustic image.

With the CBF, deconvolution methods have been proposed [24, 25, 26]. Although these methods improve the acoustic image by narrowing the main lobe width or removing side lobe, the computation time and complexity are usually greatly increased. A full comparison of all the deconvolution techniques with the GCC is not possible here because the number of proposed methods is too large. The deconvolution method selected in this study is CLEAN-SC [35] which is well known for its efficiency and short computation time.

This deconvolution technique is compared with the GCC based on the geometric mean with the weighting function ρ-PHAT-C. Although, other GCC improvements exist [52, 53], the latter has been chosen due to its performance with acoustic imaging [41]. Both improvements of the GCC are almost straightforward. Indeed, the geometric mean only modifies the average process of Eq. 10 by a product and the weighting function only divides the cross spectrum of the microphone signal by its absolute value at the power ρ and coherence of the microphone signals. These improvements do not greatly increase the computation time and complexity and are not based on iterative process. This technique is called GEO-ρ-PHAT-C.

Both techniques are compared in the case of two loudspeakers in front of a circular microphone array (Figure 11). First, the frequency band considered is the 500 Hz octave band. In this case, the loudspeakers are too close and the CBF and GCC are not able to separate them (Figure 11.a-c). The result is a main lobe in between the loudspeaker created by the merging of the main lobes. Although the frequency content is low, the GEO-$\rho = 1, 3$-PHAT-C is able to separate both loudspeakers and to identify the woofers as the main sources. The CLEAN-SC technique being based on the peak value of the CBF exhibits a source in between the loudspeakers. This technique is not able to separate both sources.

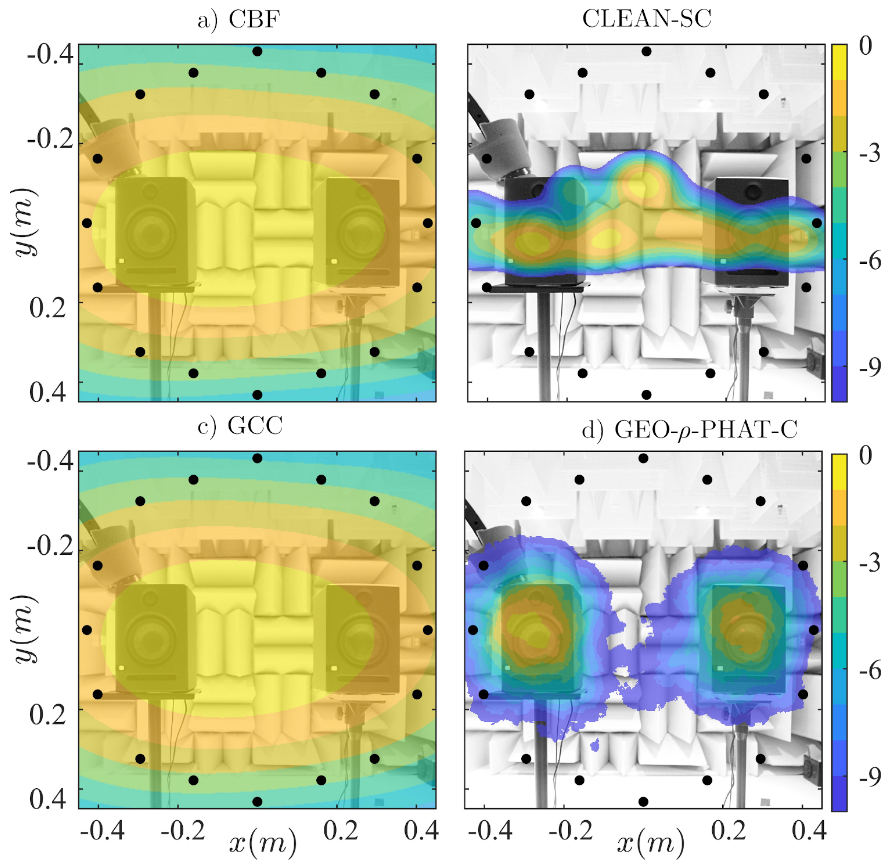


Figure 11: Acoustic images obtained a) CBF, b) CLEAN-SC, c) GCC and d) GEO-ρ-PHAT-C. A 16-microphones circular array in front of two loudspeaker was used. The microphone signals were filtered in the 500 Hz octave band. The black dots denote the microphone positions. (color online)

Then, the third octave band 3150 Hz is investigated. This frequency band is slightly larger than the cross over frequency between the woofer and tweeter. The CBF and GCC provide similar acoustic images with main lobes a little bit below the tweeter positions and strong side lobes (Figure 12). The GEO-$\rho = 1, 3$-PHAT-C perfectly detects both tweeters and removes all side lobes. The CLEAN-SC also removes the side lobes but identify the main sources such as the CBF i.e. a little bit below the tweeter positions.

To conclude, the GEO-ρ-PHAT-C provides similar results than CLEAN-SC but without using iterative process and without increasing the computation time and complexity.

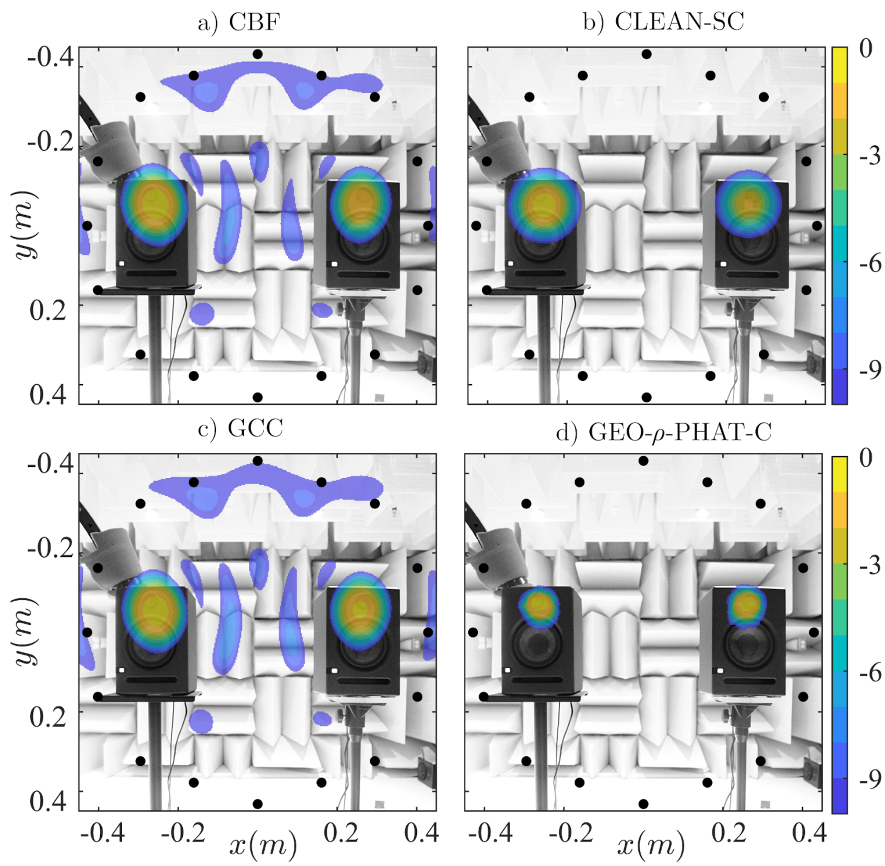


Figure 12: Acoustic images obtained a) CBF, b) CLEAN-SC, c) GCC and d) GEO-ρ-PHAT-C. A 16-microphones circular array in front of two loudspeaker was used. The microphone signals were filtered in the 3150 Hz third octave band. The black dots are the microphone positions. (color online)

5. Conclusion

Microphone arrays have become a standard tool to perform source localization or acoustic imaging. The post-processing of the microphone signals can be done in the time or frequency domain. In acoustic imaging, the standard technique is the conventional beamforming (CBF) based on the cross spectral matrix (frequency domain). However, acoustic images can be also obtained with the time domain technique such as the generalized cross-correlation (GCC). The objective of this work was to compare both techniques.

First, the computation of the CBF and GCC has been presented in details in the first section. Then, a criterion based on a covariance ellipse, denoted EAR, has been introduced to compare the acoustic images obtained numerically with each technique. This criterion allows for characterizing either the main lobe surface or side lobe influence. The CBF and GCC provided similar EAR criterion in the case of a single source. The abilities of both techniques for estimating a source amplitude difference or for localizing a single source in presence of an additive noise have also shown similar results. The main difference was the computation time which was faster with the GCC than the CBF (at the exception of very large grids).

Then, experimental data were used in order to compare both techniques, the sources were two loudspeakers, a leaf blower (stationary tonal and broadband noise) or a nail gun (impulsive noise). Again, the acoustic images provided by both techniques were similar. The main difference was when improved techniques were used. The deconvolution technique CLEAN-SC (usually used with the CBF) was compared with the GCC with the geometric mean associated to an improved weighting function. Both techniques were selected for their fast computation time. In the case of low frequency sources, the improved GCC outperformed the CBF, GCC and CLEAN-SC and was able to separate them.

For higher frequency source, the improved GCC enhanced the noise source map (as compared to GCC and CBF) in similar way than CLEAN-SC.

To conclude, the CBF and GCC provide similar performance, the main differences lies in the computation time and the use of improved techniques. A next step of this work would be a larger comparison of time and frequency domain techniques based on sparse representation.

References

References

[1] U. Michel and B. Barsikow, Localisation of sound sources on moving vehicles with microphone arrays. Proceedings of Euro-Noise, May 19–21 2003, Naples, Italy, 2003

[2] U. Michel, B. Barsikow, P. Böhning and M. Hellmig, Localisation of moving sound sources with phased microphone arrays. Inter-Noise, 22–25 August 2004, Prague, Czech Republic, 2004

[3] S.N.Y Gerges, W.D. Fonseca, W.D. and R.P. Dougherty, State of the art beamforming software and hardware for application. 16th International Congress on Sound and Vibration, July 5–9, Kraków, Poland (2009)

[4] P. Sijtsma and R. Stoker, R. Determination of absolute contributions of aircraft noise components using fly-over array measurements. 10th AIAA/CEAS Aeroacoustics Conference, May 10–12, Manchester, United Kingdom, AIAA Paper 2004-2958, 2004

[5] R. Merino-Martinez, M. Snellen and D.G Simons Functional beamforming applied to imaging of flyover noise on landing aircraft. J. Aircr. 53(6), 1830–1843, 2016

[6] C. Camier T. Padois, J. Provencher, P-A. Gauthier, A. Berry, J-F. Blais, M. Patenaude-Dufour and R. Lapointe R. Fly-over source localization on civil aircraft. 19th AIAA/CEAS Aeroacoustics Conference, May 27–29, Berlin, Germany, AIAA Paper 2013-2261, 2013

[7] T. Padois and A. Berry Application of acoustic imaging techniques on snowmobile pass-by noise. J. Acoust. Soc. Am. 141(2), EL134–EL139, 2017

[8] E. G. Williams, J. D. Maynard, and E. Skudrzyk. Sound source reconstructions using a microphone array. J. Acoust. Soc. Am. 68 (1), 340–344 (1980).

[9] E. G. Williams and J. D. Maynard. Holographic Imaging without the wavelength resolution limit. Phys. Rev. Lett. 45, 554–557 (1980).

[10] U. Michel, History of acoustic beamforming, 1st Berlin Beamforming Conference, 2006.

[11] T.J. Mueller, Aeroacoustic measurements, Springer-Verlag, 2002.

[12] D.H. Johnson and D.E. Dudgeon, Array Signal Processing, Concepts and Techniques. P T R Prentice Hall, Englewood Cliffs, 1993.

[13] T.F. Brooks and W.M. Humphreys, Effect of directional array size on the measurement of airframe noise components, 5th AIAA Aeroacoustics Conference, Bellevue, Washington, 1999.

[14] E. Sarradj, Three-dimensional acoustic source mapping with different beamforming steering vector formulations, Advances in Acoustics and Vibration, 1-12, 2012.

[15] J. Fischer and C. Doolan, Beamforming in a reverberant environment using numerical and experimental steering vector formulations, Mech. Syst. Signal Process. 91,10-22, 2017

[16] J.J. Christensen and J. Hald, Technical Review Beamforming, Brüel&Kjær Technical Review, 2004.

[17] A. Cigada, F. Ripamonti, M. Vanali, The delay & sum algorithm applied to microphone array measurement: Numerical analysis and experimental validation, Mech. Syst. Signal Process. 21 (6), 2645–2664, 2007.

[18] Z. Prime and C. Doolan, A comparison of popular beamforming arrays, Proceedings of ACOUSTICS 2013—Victor Harbor, Australia, 17-20 November, 2013.

[19] Z. Prime, C. Doolan and B. Zajamsek Beamforming array optimisation and phase averaged sound source mapping on a model wind turbine, InterNoise, Melbourne Australia, 16-19 November, 2014.

[20] E. Sarradj, A generic approach to synthesize optimal array microphone arrangements, 6th Berlin Beamforming Conference, 2016.

[21] W-H. Liao, Y. Mitsufuji, K. Osako and K. Ohkuri Microphone Array Geometry for Two Dimensional Broadband Sound Field Recording, 145th convention Audio Engineering Society, New York, USA, 17-20 October, 2018.

[22] F. Ramos do Amaral, J. C. Serrano Rico and M. A. Faraco de Medeiros Design of microphone phased arrays for acoustic beamforming, Journal of the Brazilian Society of Mechanical Sciences and Engineering, July, 2018.

[23] E. Arcondoulis and Y. Liu An iterative microphone removal method for acoustic beamforming array design, J. of Sound and vib, July, 442, 552-571, 2019.

[24] Q. Leclère, A. Pereira, C. Bailly, J. Antoni, and C Picard, A unified formal-

ism for acoustic imaging based on microphone array measurements, Int. J. Aeroacoust. 16(4–5), 431–456, 2017.

[25] R. Merino-Martinez, P. Sijtsma, M. Snellen, T. Ahlefeldt, J. Antoni, C. J. Bahr, D. Blacodon, D. Ernst, A. Finez, S. Funke, T.F. Geyer, S. Haxter, G. Herold, X. Huang, W.M. Humphreys, Q. Leclère, A. Malgoezar, U. Michel, T. Padois, A. Pereira, C. Picard, E. Sarradj, H. Siller, D.G. Simons and C. Spehr, A review of acoustic imaging methods using phased microphone arrays, CEAS Aeronautical Journal, 1–34, 2019.

[26] P. Chiariotti, M. Martarelli and P. Castellini, Acoustic beamforming for noise source localization - Reviews, methodology and applications, Mech. Syst. Signal Process., 120, 422–448, (2019).

[27] R.P. Dougherty and Y. Liu Spiral-shaped array for broadband imaging, US5,838,284; 1998.

[28] J. Fischer,V. Valeau and L.E. Brizzi, Beamforming of aeroacoustic sources in the time domain: An investigation of the intermittency of the noise radiated by a forward-facing step J. of Sound and vib, 383, 464-485, 2016.

[29] H. Kook, . B. Moebs, P Davies and J. S Bolton, An efficient procedure for visualizing thesound field radiated by vehicles duringstandardized passby tests J. of Sound and vib, 233(1), 137-156, 2000.

[30] R.P. Dougherty, Advanced time-domain beamforming techniques, 10th AIAA/CEAS Aeroacoustics Conference, American Institute of Aeronautics and Astronautics, 2004.

[31] O. Jaeckel, Strengths and weaknesses of calculating beamforming in the time domain, 1st Berlin Beamforming Conference, 2006.

[32] M. Bilodeau, N. Quaegebeur, O. Robin, P. O'Donoughue, P. Masson and A. Berry, Time domain imaging of extended transient noise sources using phase coherence J. Acoust. Soc. Am., 146(6), 4851–4859, 2019.

[33] D. Döblerand and R. Schröder, Contrast improvementand sound reconstruction of quiet sound sources using a high dynamic range algorithm 4th Berlin Beamforming Conference, 2012.

[34] R. Cousson, Q.Leclère, M.-A. Pallas and M. Bérengier, A time domain CLEAN approach for the identification ofacoustic moving sources, J. of Sound and vib, July, 443, 47-62, 2019.

[35] P. Sijtsma, Clean based on spatial source coherence Int. J. Aeroacoustics, 6(4), 357–374, 2007.

[36] T. Padois, S. Sgard, O. Doutres, and A. Berry, Acoustic source localization using a polyhedral microphone array and an improved generalized cross-correlation technique, J. of Sound and vib, 386(6), 82–99, 2017.

[37] T. Padois, O. Doutres, S. Sgard, and A. Berry, On the use of geometric and harmonic means with the generalized cross-correlation in the time domain to improve noise source maps, J. Acoust. Soc. Am., 140(1), EL56–EL61, 2016.

[38] C.H. Knapp and G.C. Carter, The generalized correlation method for estimation of time delay, IEEE Trans. Acoust. Speech, Signal Process., 24, 320-327, 1976

[39] N. Quaegebeur, T. Padois, P-A. Gauthier and P. Masson, Enhancement of time-domain acoustic imaging based on generalized cross-correlation and spatial weighting, Mech. Syst. Signal Process., 75, 512-524, 2015.

[40] T. Padois, Acoustic source localization based on the generalized cross-correlation and the generalized mean with few microphones, J. Acoust. Soc. Am., 15(2), EL393-EL398, 2018.

[41] T. Padois, O. Doutres and F. Sgard, On the use of modified phase transform weighting functions for acoustic imaging with the generalized cross correlation, J. Acoust. Soc. Am., 145(3), 1546-1555, 2019.

[42] J. Velasco, D. Pizarro and J. Macias-Guarasa, Source localization with acoustic sensor arrays using generative model based fitting with sparse constraints, Sensors, 12, 13781-13812, 2012.

[43] C. Noël, V. Planeau and D. Habault. A new temporal method for the identification of source directions in a reverberant hall, J. of Sound and vib, 296(3), 518-538, 2006.

[44] T. Padois, O. Doutres, F. Sgard and A. Berry, Time domain localization technique with sparsity constraint for imaging acoustic sources, Mech. Syst. Signal Process., 94, 85-93, 2017.

[45] U. Hamid, R. A. Qamar and K. Waqas Performance Comparison of Time-Domain and Frequency-Domain Beamforming Techniques for Sensor Array Processing Proceedings of 11th International Bhurban Conference on Applied Sciences & Technology (IBCAST) Islamabad, Pakistan, 14th-18th January, 2014

[46] A.V. Oppenheim, R.W. Schafer, J.R. Buck, Discrete-time signal processing, Prentice-hall Signal Processing Series, 1999.

[47] T. Padois, O. Doutres, F. Sgard and A. Berry, Optimization of a spherical microphone array geometry for localizing acoustic sources using the gener-

alized cross-correlation technique Mech. Syst. Signal Process., 132, 546-559, 2019.

[48] P. Aarabi. The fusion of distributed microphone arrays for sound localization, EURASIP Journal on Applied Signal Processing, 338-347, 2003.

[49] J.P. Dmochowski, J. Benesty and S. Affes. A generalized steered response power method for computationally viable source localization, IEEE Transactions on Audio, Speech, and Language Processing, 15(8), 2510-2526, 2007.

[50] R. A. Johnson and D. W. Wichern, Applied multivariate statistical analysis, 6th Ed., Upper Saddle River, NJ : Pearson Prentice Hall, Chap 2.3, 60-65, 2007.

[51] T. Padois, M-A. Gaudreau, P. Marcotte and F. Laville Identification of noise sources using a time domain beamforming on pneumatic, gas and electric nail guns Noise Control Engr. J. 67 (1), January-February 2019

[52] F. grondi and J. Glass SVD-PHAT: A fast sound source localization method IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May, 2019

[53] V. C. Raykar, R. Duraiswami, B. Yegnanarayana and S.R. Mahadeva Prasanna Tracking A Moving Speaker using Excitation Source Information Eurospeech, Geneva, 2003