Sound source identification in time domain exploiting the Clustering Inverse Beamforming

C. Colangeli[1], K. Janssens[1], P. Chiariotti[2], P. Castellini[2]
[1] Siemens Industry Software NV
Interleuvenlaan 68, B-3001, Leuven, Belgium
[2] Università Politecnica delle Marche
Via Brecce Bianche 1, I-60131, Ancona, Italy

e-mail: **claudio.colangeli@siemens.com**

Abstract
Acoustic imaging techniques allow the user to see what he/she is listening to. This paper aims at proposing an inverse procedure that allows for retrieving the evolution of the noise source identified in a beamforming map. Such approach overcomes the limit of frequency domain strategies, and opens up different application fields such as auralization, coherence analyses, etc… The source localization step is performed in frequency domain with the goal of accurately identifying the source coordinates. The corresponding time signals are subsequently obtained by convolving in time domain the microphones data with multiple input – multiple output impulse responses corresponding to the back-propagating functions identifying the receiver-source link. The formulation of the algorithm is presented in this paper and its main strengths and limitations are discussed. Applications are shown in simulated and real experiments.

1 Introduction

The proposed technique leverages on the localization step performed in frequency domain for reducing the calculation plane to the few points that host the actual sources. In this way, the further inverse source identification problem, wherein the corresponding time signals are obtained, is no longer underdetermined, since the number of sources active in the field is reasonably lower than the number of microphones available in the array. The above mentioned Impulse Responses are computed by inverting, in frequency domain, the matrix containing the Noise Transfer Functions between the sources' and the microphones' locations (see references [1] and [2]). The obtained inverse Noise Transfer Functions are then inversely Fourier transformed. This set of Impulse Responses intrinsically take into account the mutual interaction between the sources, granting an optimal separation of the corresponding signals as long as all the main sources active in the field are correctly localized and included in the computation. The omission of important contributions could in fact badly compromise the correct identification of the other sources (see also [3]). To prevent this for happening, the sound source localization results must allow large dynamic ranges and accurate calculation of the sources locations. The former aspect is important for correctly identifying the weakest sources as well as the strongest ones; the latter is required in order to properly select the Noise Transfer Functions to be adopted in the time domain source identification step. In order to meet these requirements, the Clustering Inverse Beamforming algorithm ([4], [5] and [6]) is exploited for the sound source localization task, since this method guarantees indeed accurate localization results with large dynamic range also in complex scenarios where multiple correlated and uncorrelated sources are active at the same time. In section 2 are pointed out the main theoretical features and in section 3 some fundamental aspects are investigated on a virtual experiment. Section 4 is dedicated to the validation of the method on experimental data. In particular will be treated in detail the cases of correlated sources (multiple sources and source plus reflections) and uncorrelated sources. Section 5 is dedicated to the conclusions.

2 Theory

The proposed method requires initial source localization in frequency domain and consequent inverse source identification in time domain. Hereafter a theoretical description of the two steps is reported.

2.1 Inverse sources identification in time domain

The inverse load identification step is allows for retrieving the time signals of the main noise sources active in the acoustic scene observed with the microphones array. As already mentioned the main idea is convolving the microphones signals with inverse impulse responses. Those are calculated by inverse Fourier transforming (symbol: $F^{-1}\{...\}$) a set of MIMO estimated inverse Noise Transfer Functions (NTF) modeled in frequency domain. Those are obtained by inverting per frequency line the direct radiation model (A in (1)) including the candidate N source locations and the M microphones locations. Since the matrix A is in general not square, a pseudo-inversion is required. Moreover the system may be ill-conditioned and require regularization. For deepening this aspect the interested reader may refer to [7]. This process allows obtaining the inverse impulse response $h_{n,m}(t)$ between each of the mth microphones locations and each of the nth sources locations.

$$F^{-1}\{A^{H}{}_{n,m}\} = h_{n,m}(t) \tag{1}$$

A is the radiation matrix whose elements describe the radiation model adopted. In case of free-field propagation, each element of the A matrix can be expressed as: $A_{mn} = e^{-ikr_{mn}} / 4\pi r_{mn}$; where the subscripts m and n represent the m^{th} microphone (over M) and n^{th} calculation point (over N) respectively and r_{mn} represents the distance between the geometrical positions of these two points.
It is very important to notice that by doing so the contribution of all the candidate sources is considered together at the same time. This, in one hand, ensures the correct identification of correlated as well as uncorrelated source and provides the best source separation possible because it includes in the model the cross-talk between the sources. In the other hand that means also that all the main acoustic sources should be taken into account. If this is not the case a wrong estimation of all the remaining source signals may most likely occour.
Once the set of inverse impulse responses is available, each source signal can be retrieved adopting the equation (2) which is valid for source n out of the N active in the acustic scene.

$$q_n(t) = \sum_{m=1}^{M} h_{n,m}(t) \otimes p_m(t) \tag{2}$$

In order to ensure that all the main sources are taken into account in the computation, an efficient sound source localization strategy is required. In this paper the Clustering Inverse Beamforming is proposed for accomplishing this task. This method in fact: ensures high accuracy in localization almost independently from the frequency range; it moreover allows to correctly identifying correlated sources as well as uncorrelated sources with a large dynamic range. The latter is needed for localizing the weakest sources as well as the strongest ones.

2.2 Clustering Inverse Beamforming

The Clustering Inverse Beamforming (CIB) approach is a processing strategy that combines different clusters of microphones belonging to the same array in order to enhance the solution of an inverse beamformer. Indeed, the solution obtained by an inverse beamforming algorithm is strongly dependent on the combination of input available for retrieving the equivalent source distribution at the calculation points. By changing this set of input (i.e. considering different sub-sets of microphones in the inverse beamforming calculation), numerical instabilities, as well as phenomena not directly related to the source

distributions (e.g. cavity modes), will be interpreted differently, while the actual noise sources will be identified similarly.
The approach exploits the Generalized Inverse Beamforming (GIB) method (see [8] , [9] and [10]). The concept can be extended to different algorithms.
In GIB the source distribution is recovered by first decomposing the acoustic field in eigenmodes and then by solving, for each eigenmode, the general radiation problem reported in (3)

$$a_i = A^{-1} v_i = \mathrm{Y}(A, v_i) \quad (3)$$

The following symbolism is adopted in (3):

- v_i : i^{th} eigenmode obtained from the Eigenvalue Decomposition of the microphone Cross Spectral Matrix ($C_M = ESE^{\dagger} \Rightarrow v_i = E\sqrt{S}$);
- a_i : reconstructed source distribution;
- Y : operator taking into account regularisation strategy needed to cope with ill-the conditioning of the inverse problem and an iterative process for optimising the solution.

The Clustering approach consists in performing, per eigenmode, the inverse beamforming iterative process N_c times on N_c different clusters of microphones. Considering a subset – cluster - of microphones among those constituting the whole array means selecting only certain rows of the radiation matrix *A*, and thus changing the mathematical formulation of the problem. However, the associated physical event remains obviously the same. The regularisation strategies and the iterative solution of the inverse problem (refer to [10] for a deeper insight about regularisation strategies and sensitivity analysis in GIB) will perform differently depending on the radiation matrix considered. In this way, any numerical instability that gives rise to ghost sources will vary, while the actual sources will be constantly identified. The set of solutions obtained in this way ($a_c^{(i)}, c = 1 : N_c$) for each i^{th} main eigenmode, are statistically processed in order to obtain a so-called *clustering mask matrix* ($\gamma^{(i)}$). The *clustering mask matrix* is obtained by vector multiplying (Hadamard product) two matrices: the *normalised mean matrix* ($\bar{a}^{(i)}$) and the *normalised occurrences matrix* ($\hat{a}^{(i)}$).
The first one is obtained by averaging, per eigenmode, the set of solutions $a_c^{(i)} = \mathrm{Y}(\{A\}_{c,N}, v_i)$ obtained processing each cluster of microphones over the *N* target points. The *normalised occurrences matrix* $\hat{a}^{(i)}$ somehow quantifies the consistency of the pattern produced at the calculation plane every clustering solution. Eventually each element $\gamma^{(i)}(k)$ of the *clustering mask matrix* is calculated as

$$\gamma^{(i)}(k) = \bar{a}^{(i)}(k)\hat{a}^{(i)}(k) \quad (4)$$

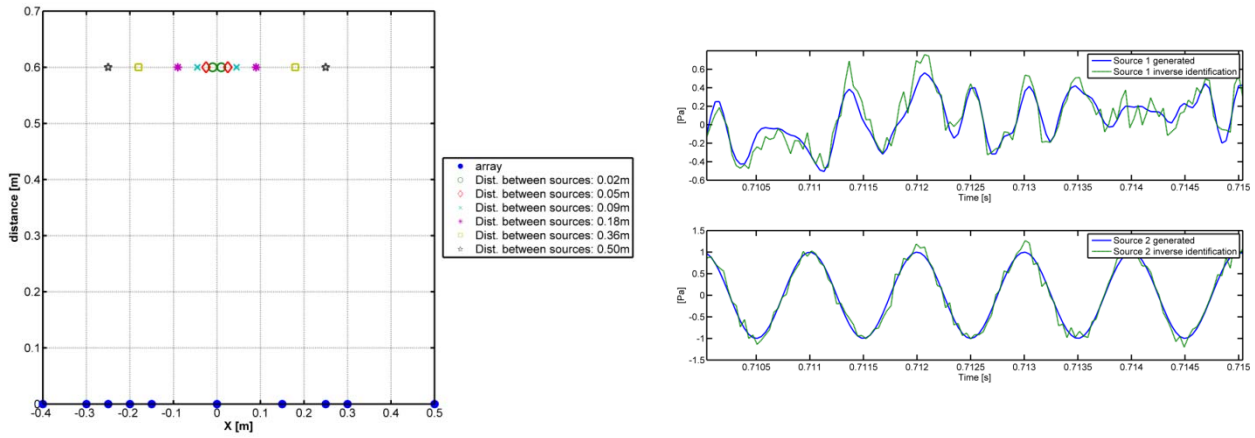
and therefore the Clustering Inverse Beamforming (CIB) solution can be obtained as

$$a^{(i)}(k) = f(\mathrm{Y}(A, v_i), \gamma^{(i)}) \quad (5)$$

For a more detailed mathematical formulation of the method the interested reader can refer to [5] and [6].
In this particular application, the CIB is needed for accurately identifying the sources' coordinates and for investigating the level of correlation between the retrieved sources. The quantification step will be deputed, instead, to the inverse source identification described in section 2.1. For accomplishing the localization task, equation (4), which is the one that will be exploited from now on, provides a very powerful tool. In fact, since the clustering mask matrix is a function that varies from 0 to 1 assigning to each candidate scanned grid point the level of confidence of being the location of one of the actual main sources responsible of the observed acoustic field, the local maxima of such function give the most likely source distribution that it is possible to retrieve with the available data recorded in the far field (for a more deep description of the theoretical aspects see reference [6]). Moreover the separation of the acoustic field in uncorrelated source distributions is automatically obtained.

3 Preliminary analysis on simulated data

Before applying the inverse source identification method described in section 2.1 on real data, a preliminary study on a simulated scenario has been carried out with the goal of understanding the influence of two potential sources of inaccuracy: the presence of different SNR conditions and the cross-talk between the two sources due to their closeness. Such study has been conducted by considering the beamforming problem depicted in Figure 1(a). Two sources (one 1 kHz sine tone and one white noise band-pass filtered between 0.1-2 kHz) are placed 0.6 m far from a randomly distributed linear array of 10 microphones. The distance between the two sources has been varied between 0.02 m and 0.5 m. The SNR between the microphones clean signal and the added Gaussian background noise has been varied between 25 and 50 dB.



(a) *(b)*

Figure 1: (a) simulated scenarios with a randomly distributed linear array (10 microphones) and sources placed at different distances. (b) example of identification in presence of severe SNR conditions.

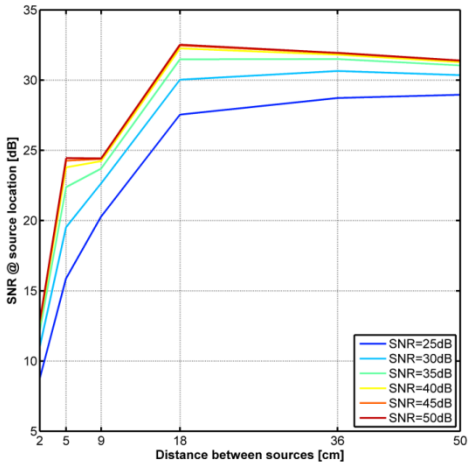


Figure 2 – Influence of the SNR at microphones location (measurement noise) on the SNR obtained in the retrieved sources signals. Results reported as a function of the distance between the sources.

Figure 1(b) shows one example of results in which the effect of background noise is visible. Analyzing Figure 2 it is possible to observe that the presence on measurement noise in the microphones is not the only cause of inaccouracy in the retrieved signals. Cross-talk between the sources always occours and its influence becomes dramatic when the sources become closely spaced (distance < 0.18 m).

4 Validation on experimental data

In this section the method is applied on real test cases. Two different measurement campaigns will be presented in which the strengths and limitation of the technique in presence of correlated and uncorrelated sources are respectively evaluated.

4.1 Indentification of correlated sources

With this validation case two conditions will be tested: the case in which two correlated sources are generated simultaneously by two different devices and the case in which a second correlated source takes place due to the presence of a reflecting surface. In the first case, although correlated, the two signals' signature can be different since they are generated by two different physical devices, while in the other case the main difference between the two signals is the phase shift occourring due to the different travelled path.

4.1.1 Measurement setup

The tests have been performed in a semi-anechoic room, adopting a microphone array of 36 microphones distributed over a pattern of three concentric circles (LMS HDCam36). Two high frequency referenced sources in the range 2 kHz – 20 kHz have been used. Two perpendicular reflecting walls have been used for producing reflections.

(a) (b) (c)

Figure 3: (a) Measurement setup for correlated sources identification. (b) Scenario A, two correlated sources. (c) Scenario B, one source and a reflective wall.

Figure 3 shows the setup. Microphone array: LMS HDCam36. Sources: LMS High Frequency Q Sources (prototype). The origin of the axis coincides with the center of the array which lies in the xy plane. The microphone array points towards –z direction. In all the cases shown in this paper the two sources are correlated random noises in the bandwidth 2 kHz – 20 kHz.

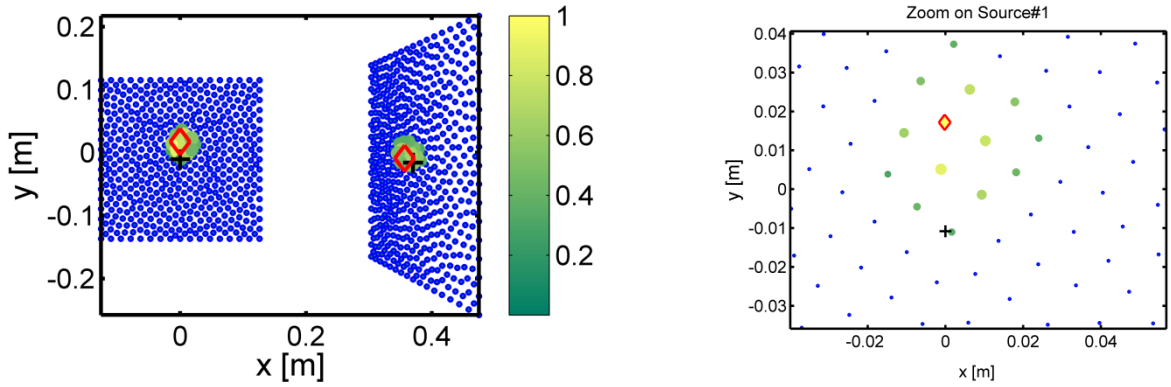
Scenario	Source#1	Source#2	Vertical wall	Horizontal wall
A	√	√		
B	√		√	(absorbing material)

Table 1: Tested scenarios for correlated sources identification.

As shown in Figure 3(c) absorbing material has been added (in scenario B) for selectively avoiding (damping as much as possible) reflections produced by the horizontal wall. In the case of scenario A the two walls have been removed.

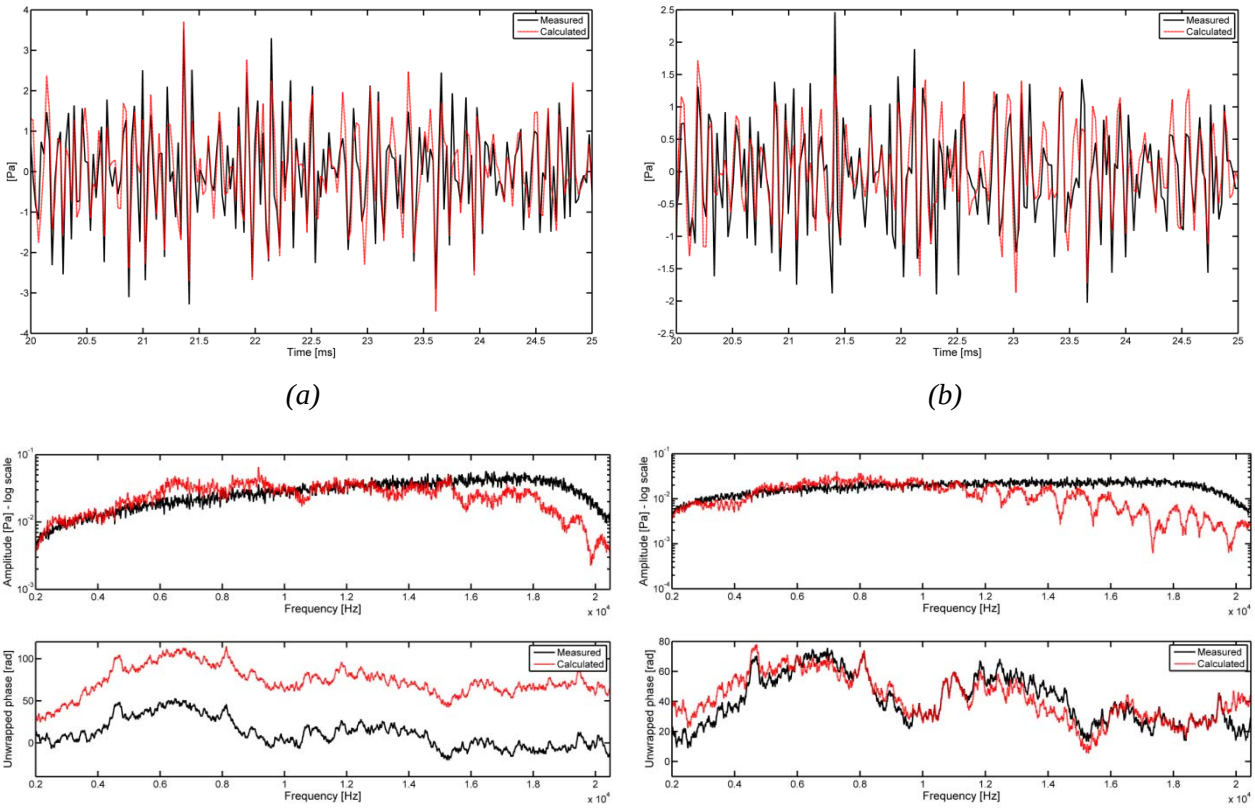
4.1.2 Results on two differently generated correlated sources (Scenario A)

Figure 4(a) shows the results of the localization step for the scenario depicted in Figure 3(b). At this stage of the process the sources are identified among a set of candidate elementary sources (in blu in the picture). The clustering mask matrix presents local maxima in the proximity of the ideal position of the sources as visible in the zoom for the localization of Source#1 in Figure 4(b). The red diamond marker indicates the local maximum.



(a) *(b)*

Figure 4: Source localization in Scenario A. (a) Localization by means of CIB. (b) Zoom on Source#1: the local maximum is selected. Frequency range: 3 – 3.1 kHz.



(a) *(b)*

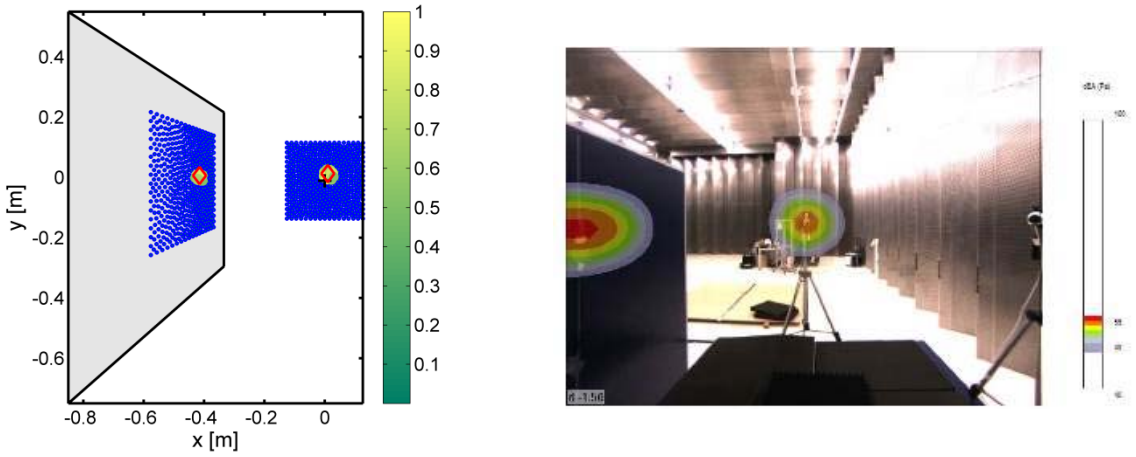
(c) *(d)*

Figure 5: Scenario A. (a) and (b) Source#1 and Source#2 compared with the source's reference signal. (c) and (d) Comparison of the spectra.

This allows selecting the coordinates of the candidate sources to be identified and at the same time turns the previous undetermined inverse acoustic problem into a well determined one since the number of candidate sources becomes sufficiently lower than the number of available sensors.
Figure 5 compare the obtained signals (in red) with the reference ones (in black) both in time and in frequency domain. The results obtained in this scenario testify that the method is able to retrieve the two sources active in the acoustic field. The accuracy drops for frequencies greater than 10 kHz.

4.1.3 Results on main source and its reflection (Scenario B)

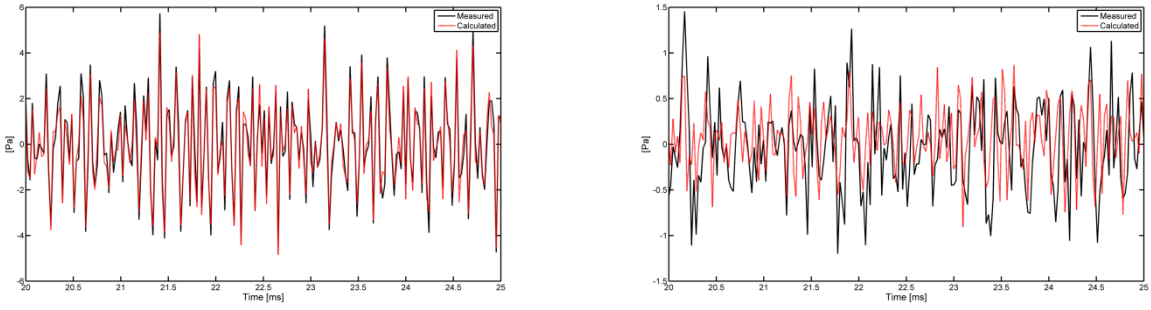
Figure 6(a) shows the results of the localization step for the scenario depicted in Figure 3(c). The blue dots represent the scanned grid of target poits. The red diamond markers represent the coordinates of the retrieved sources that will be identified.



(a) *(b)*

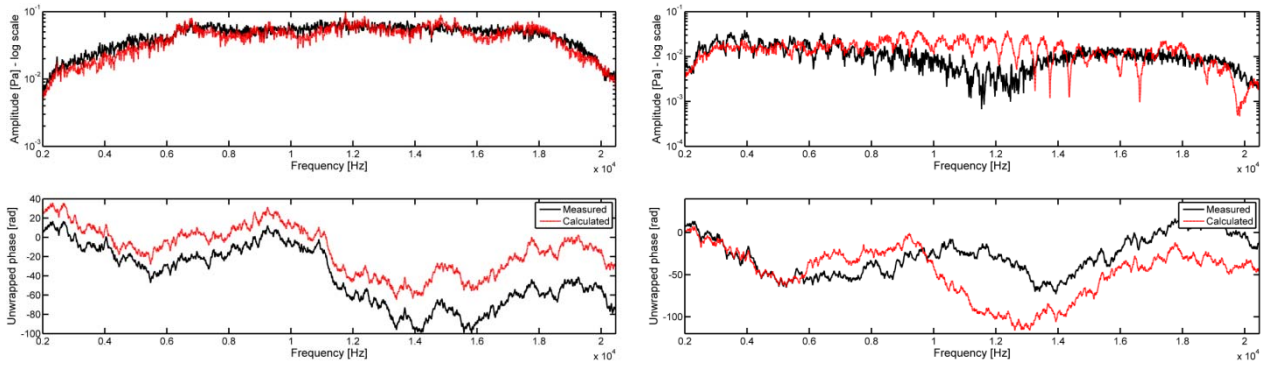
Figure 6: Scenario B. (a) CIB for accurate source localization. (b) Conventional beamforming analysis for placing reference microphone in correspondence of the reflection. Frequency range: 3 – 3.1 kHz.

In this case the source localized on the vertical wall is the reflection of Source#1. In order to obtain the reference signal for comparison, a microphone has been placed in correspondence of the theoretical position of such reflection. Such position has been calculated based on geometrical considerations and validated by the beamforming analysis shown in Figure 6(b).
The result of the inverse source identification step is reported in Figure 7 for this scenario.



(a) *(b)*

Figure 7: Scenario B: Comparison with the reference signals in time domain.



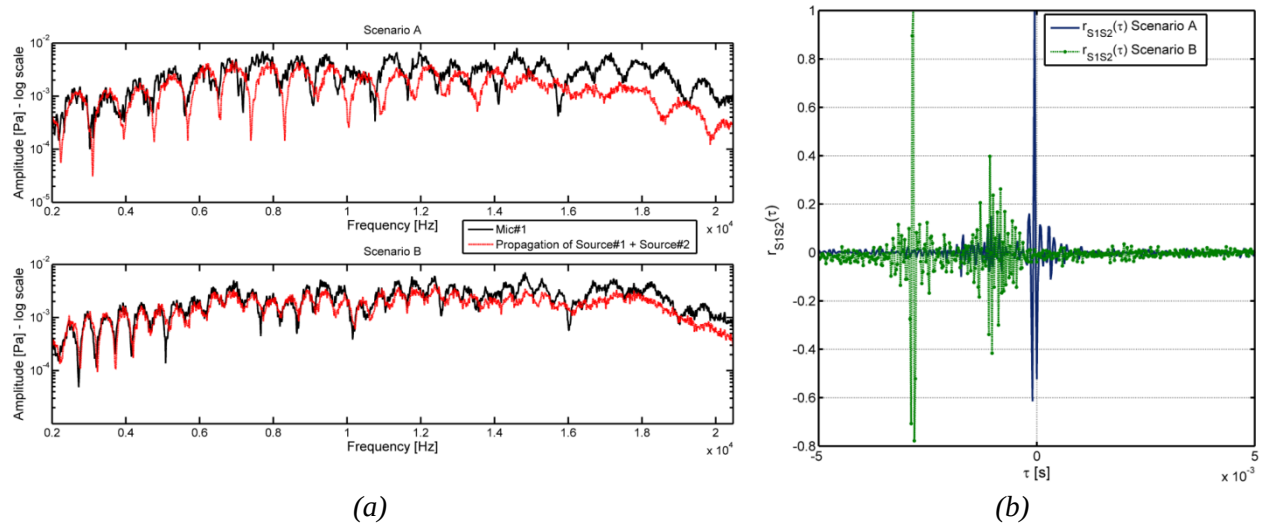
(a) *(b)*

Figure 8: Scenario B: comparison with the reference signals in frequency domain.

In the sub-case (b) of Figure 7 and Figure 8 the match is not optimal, but it is mainly due to the inaccuracy of the used reference signal for the reflection on the vertical wall.

4.1.4 Comparison of the two cases

Reconstruction of source distribution, and their corresponding signals, from far-field data alone is ambiguous since different source distributions can generate an identical far field. From this point of view, the retrieved source distribution and source signals must be interpreted in a most likelihood sense. This considered, the retrieved sources are supposed to be able to reconstruct the acoustic far field. In order to prove this assumption, the two inverse-identified sources have been propagated towards the location of one of the microphones of the array. The spectra of the obtained signal and the microphone's one are compared in Figure 9(a).



(a) *(b)*

Figure 9: Scenario A and B. (a) Mic#1 (array) spectrum vs. propagation of the retrieved sources towards the microphone location. (b) Cross-correlation function between the two retrieved signals.

The comparison between the results obtained for scenario A and for scenario B allows to certify also that the proposed method does automatically take into account the phase relationship (time delay or "distance") between the sources.

This aspect has been investigated by retrieving the distance between the sources by means of cross-correlation. The two cross-correlation functions are reported in Figure 9(b). Table 2 reports that:

- For scenario A, since the Source#1 and Source#2 are acting simultaneously (with no phase shift), as expected, despite the distance between the two sources is 0.927 m, the two retrieved signals present a phase shift quantifiable in ~0.03 m.
- In the case of scenario B, the second source is the reflection of Source#1 on the vertical wall. Therefore between the two sources there must be a phase shift corresponding to the distance covered by the sound emitted by Source#1 before impinging on the wall. The value retrieved in this case is indeed 1.006 m which is very well comparable with the theoretical value (1.088 m) calculated from the geometry.

					d_{12} [m]		
Scenario	**x [m]**	**y [m]**	**z [m]**	**Reflection**	**Geometry**	**Reference**	**Calculated**
A	0.7	-0.002	-0.863		0.927	< 0.03	< 0.03
B	-0.85	-0.014	-0.86	√	1.088	1.022	1.006

Table 2: Source#1 and: Source#2 in Scenario A; its reflection on the wall in scenario B. Time delays and covered distances calculated by cross-correlation and considering speed of sound: 342 m/s.

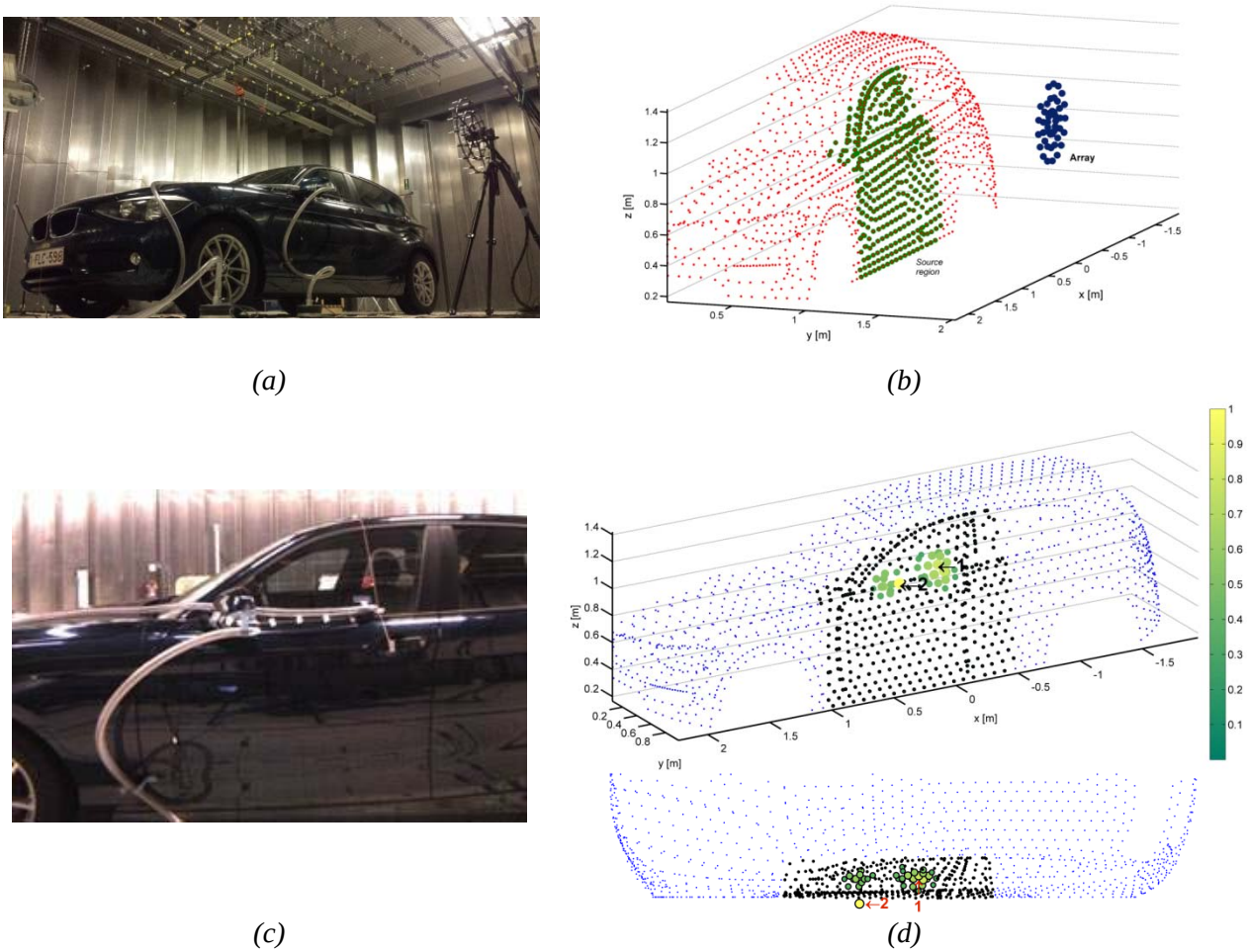
4.2 Indentification of uncorrelated sources

The following test is used to study the case of two uncorrelated sources simultaneously acting in an acoustic field. The main difference w.r.t. the previous case is that the two signals do not have any deterministic phase relationship.

4.2.1 Measurement setup

Two calibrated volume velocity sources have been positioned as in Figure 10(c) on a vehicle and measured by using an array of 45 microphones. The two sources (LMS Mid-High Frequency Q Sources) emit two uncorrelated white noises filtered in the band 200 Hz – 10000 Hz. One of the two sources is placed at the tip of the left side mirror of the car. The other source is placed 40 cm distant from the previous one, on the left side window, in the proximity of the B-pillar as visible in Figure 10(c). The array (LMS HDCam45) is placed 1m far from the vehicle (Figure 10: Identification of uncorrelated sources. (a) and (b) Setup and geometry of the problem. (c) Tested configuration. (d) CIB localization results. Frequency range: 2 – 2.1 kHz.

(b)).



(a) *(b)*

(c) *(d)*

Figure 10: Identification of uncorrelated sources. (a) and (b) Setup and geometry of the problem. (c) Tested configuration. (d) CIB localization results. Frequency range: 2 – 2.1 kHz.

4.2.2 Results

Figure 10(d) shows the localization of the two sources adoptiong the Clustering Inverse Beamforming and in particular assigning to the two sources the coordinates corresponding to the local maxima of the overall clustering mask matrix. Notice that in the case of the source placed on the tip of the rear mirror, labeled with the number 2, the local maximum of the clustering mask matrix correctly occours in the corresponding point representative of this component.

The Figure 11 (a)-(b) and (c)-(d) pairs report the comparison between the calculated source signals and the corresponding reference signals both in time and in frequency domain. In this case the sources' volume accelerations are compared (direct output of the calibrated sources). It is observed an overestimation of the low frequency content in the case of source 2.

Figure 11 (e) and (f) report the clustering mask matricies corresponding to the two uncorrelated source distributions obtained by means of CIB.

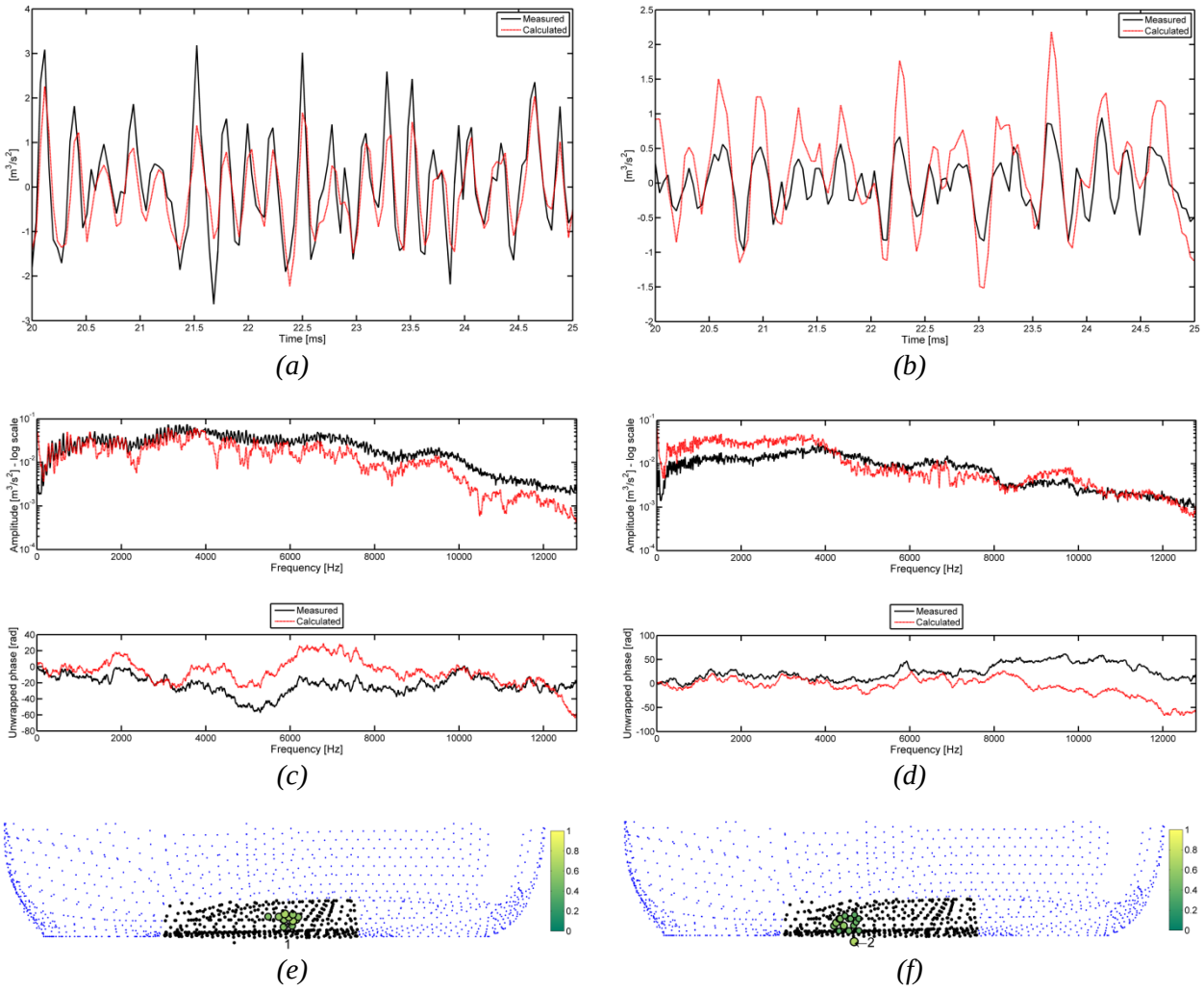


Figure 11: inverse identification of the sources. (a) and (b): comparison of measured and calculated time signals (volume acceleration). (c) and (d): comparison of the corresponding spectra. (e) and (f) localization and separation of the two uncorrelated source distributions (frequency range: 2 – 2.1 kHz).

Figure 12 shows the spectrum of the propagation of the sources towards the location of one of the microphones compared with the spectrum of the signal of the microphone in the same location. The good match confirms that the retrieved source distribution is indeed able to generate the measured far-field.

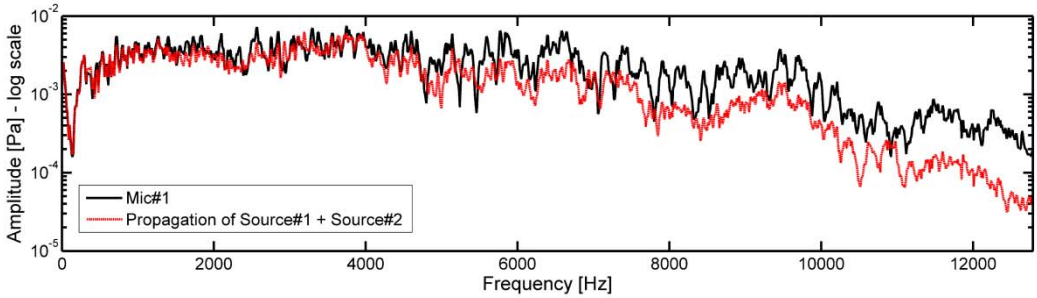


Figure 12: Mic#1 (array) spectrum vs. propagation of the retrieved sources at microphone location.

5 Conclusions

A procedure that allows for retrieving the time evolution of the noise source identified in a beamforming map has been developed and applied on virtual and real experiments.

The idea of turning the initial undetermined inverse acoustic problem into a well determined one by means of a preliminary sound source localization step in frequency domain has been formalized and successfully applied. In this stage the Clustering Inverse Beamforming has proven to be suited for accomplishing the task of accurately localize the sources. In fact the concept of clustering mask matrix allows assigning the correct coordinate of the source with high spatial accuracy; the large dynamic range allows identifying strong sources as well as the weakest ones. Moreover the same technique allows assessing wether the retrieved sources are correlated or uncorrelated with each other.
Once the localization task is completed, the identified sources are synthesized and their time-domain signals become available.

Taking into account that the reconstruction of the sources' corresponding signals from far-field data alone must be interpreted in a most likelihood sense because different source distributions can generate an identical far field, the results presented have shown a promising correspondence between the synthesized sources and the corresponding reference signals. Successful results have been obtained both in presence of correlated sources, where also the time delay between the signals is correctly estimated, and uncorrelated ones.

The method presented also limitations mainly due to the abovementioned ambiguity. The performed preliminary analysis on a virtual experiment shows in fact that this ambiguity is translated in presence of cross-talk between the retrieved sources and/or the presence of a consequent noise disturbance in the retrieved signals. Nevertheless the same analysis ensured also the robustness of the proposed approach in presence of severe SNR conditions and/or complex acoustic fields.

The exploitation of the ideas described and implemented in this paper enables the user to obtain a realistic estimation of the time evolution of the main acoustic sources under investigation by means of far-field measurements only. This can be a unique advantage in many applications such as aero-acoustics, condition monitoring, etc…

Acknowledgements

This work was supported by the Marie Curie ITN project ENHANCED (joined Experimental and Numerical methods for HumAN CEntered interior noise Design). The project has received funding from the European Union Seventh Framework Programme under grant agreement n° 606800. The whole consortium is gratefully acknowledged.

References

[1] J.W. Verheij, *Inverse and reciprocity methods for machinery noise source characterization and sound path quantification, part 1: sources.* International Journal of Acoustics and Vibration Vol. 2 [1], p. 11-20, (1997).

[2] K. Janssens, F. Bianciardi, *Time-domain ASQ method for pass-by noise engineering,* Proceedings NOISE-CON 2013, Denver, Colorado ,USA, 26-28 August (2013).

[3] F. Bianciardi, K. Janssens, L. Britte, *Critical assessment of OPA: effect of coherent path inputs and SVD truncation,* 20th International Congress of Sound & Vibration. Bangkok (Thailand), 7-11 July, (2013).

[4] C. Colangeli, P. Chiariotti, K. Janssens, P. Castellini, *A microphone clustering approach for improved generalized inverse beamforming formulation*, Proceeding of the NOVEM 2015 conference, Dubrovnik (Croatia), 2015, April 13-16, Dubrovnik (2015).

[5] C. Colangeli, P. Chiariotti, K. Janssens, P. Castellini, *Microphone clustering inverse beamforming*, Submitted to journal (2015).

[6] C. Colangeli, P. Chiariotti, G. Battista, P. Castellini, K. Janssens, *Clustering inverse beamforming for interior sound source localization: application to a car cabin mock-up,* Proceeding of the Berlin Beamforming Conference 2016, Berlin (Germany), 2016, February 29-March 1, Berlin (2016).

[7] P.C. Hansen, *REGULARIZATION TOOLS: A Matlab package for analysis and solution of discrete ill-posed problems*, Numerical Algorithms Vol. 6, 1-35, (1994).

[8] T. Suzuki, *Generalized inverse beamforming algorithm resolving coherent/incoherent, distributed and multipole sources*, Proceedings of the 14th AIAA/CEAS Aeroacoustics Conference, Vancouver (British Columbia Canada), (2008).

[9] P. Zavala, W. De Roeck, K. Janssens, J.R.F. Arruda, P. Sas, W. Desmet, *Generalized inverse beamforming with optimized regularization strategy*, Mechanical Systems and Signal Processing, Vol. 25, 928-939, (2011).

[10] C. Colangeli, P. Chiariotti, K. Janssens, *Uncorrelated noise sources separation using inverse beamforming,* Proceedings of the XXXIII IMAC Conference. Orlando (Florida US), (2015).