# ENHANCED MULTIDIMENSIONAL SPATIAL FUNCTIONS FOR UNAMBIGUOUS LOCALIZATION OF MULTIPLE SPARSE ACOUSTIC SOURCES

*Francesco Nesta and Maurizio Omologo*

Center of Information Technology, Fondazione Bruno Kessler - Irst
via Sommarive 18, 38123 Trento, Italy
email:{nesta|omologo}@fbk.eu

**ABSTRACT**

The Steered Response Power with PHAT transform (SRP-PHAT) or Global Coherence Field (GCF), has become a standard method for acoustic source localization, thanks to their simplicity, computational inexpensiveness and robustness against mid-high reverberation. However, originally formulated for the single source localization case, it does not apply satisfactorily to the multiple source case. In this paper, we analyze the structure of the spatial function and reshape it according to a generic multidimensional metric. We show that traditional functions are based on the $L_1$ norm which is prone to generate ambiguous locations with high likelihood (i.e. *ghosts*). A more generic multidimensional kernel based on higher norms and on a partitioned representation of the cross-power spectrum is introduced, which better exploits the source sparseness in the discrete time-frequency domain.

Evaluation results over simulated data show that the new spatial functions considerably improve the detection of multiple competing sources in both spatial and multidimensional TDOA domains.

***Index Terms—*** multiple speaker localization, kernel methods, TDOA estimation, multidimensional signal processing, sparse sources

## 1. INTRODUCTION

The problem of multiple acoustic source localization has been tackled for more than 20 years but still arises a high interest in the related community. The most used methods are based on the spatial coherence observed at multiple microphone pairs, i.e. the widely known Steered Response Power with PHAT transform (SRP-PHAT) [1] or Global Coherence Field (GCF)[2]. Although those methods are commonly applied for their simplicity and robustness against strong reverberation, they lack of reliability when multiple sources are competing at the same time. In particular, in the multidimensional case, they suffer of presence of ambiguous location with high likelihood (i.e. ghosts) which can only be mitigated increasing the microphone pair redundancy or using an additional postprocessing step [3]. In this paper, we analyze the reason for such a limitation and show that it is intrinsically related to the weak multidimensional structure of the coherence function, which is based on the $L_1$ norm. It is shown that when a more discriminative metric $L_{\gamma,\beta}$ is adopted, ghosts are drastically reduced, with a consequent increase of localization performance. Moreover, using a partitioned structure for the cross-power spectrum observations, an extended likelihood function is adopted in order to reduce the interference across the sources and further improve the detection capability.

Note, the proposed spatial likelihood functions are tightly related to the recently proposed Generalized State Coherence Transform (GSCT) [4]. However, the assumptions underlying their formulation are different. The GSCT was originally formulated for estimating multiple TDOAs from mixing parameters retrieved through the Independent Component Analysis (ICA). In contrast, here we focus on the source sparseness in time-frequency domain and show that the proposed functions may estimate multiple source locations directly from the cross-power spectrum. Then, they can easily substitute standard spatial likelihood functions in preexisting localization frameworks, without the need of any ICA stage.

## 2. MODELS FOR MULTICHANNEL OBSERVATIONS IN FREQUENCY-DOMAIN

Assuming to observe $N$ sources by an array of $M$ elements we denote with $s_n(t)$ the time-domain signal generated by the *n-th* source and with $x_m(t)$ the signal sampled at the *m-th* microphone. By means of a short-time Fourier transform (STFT), applied to frames of $N_{bins}$ samples, each signal can be transformed from the time-domain to a discrete time-frequency representation. Therefore, let $S_n(k,l)$ and $X_m(k,l)$ be the *l-th* STFT frame coefficients obtained for the *k-th* frequency bin. Indicating the source STFT vector with $\mathbf{S}(k,l) = [S_1(k,l) \cdots S_N(k,l)]^T$, the vector of the transformed observed mixtures $\mathbf{X}(k,l) = [X_1(k,l) \cdots X_M(k,l)]^T$ can be modeled as $\mathbf{X}(k,l) = \mathbf{H}(k)\mathbf{S}(k,l)$, where $\mathbf{H}(k)$ is the $M \times N$ mixing matrix corresponding to the transfer function between sources and microphones at *k-th* frequency bin.

In anechoic environments, the mixing matrix depends only on the relative position between source and microphones and can be modeled as $\mathbf{H}(k) = [\alpha_{mn} e^{-j2\pi f_k T_{mn}}]_{m,n}$, where $T_{mn}$ is the propagation time delay from the *n-th* source to the *m-th* microphone for *k-th* frequency bin, while $\alpha_{mn}$ is the corresponding magnitude of the frequency response between the *n-th* source and the *m-th* microphone. Note, $f_k$ is a real frequency which denotes the center of the corresponding *k-th* frequency bin, with $0 \leq k < N_{bins}$. Assuming ideal sparseness of the source signals in the STFT domain, i.e. only a single source has not negligible energy in each discrete time-frequency point, the observation $\mathbf{X}(k,l)$ can be modeled as $\mathbf{X}(k,l) = [\alpha_{1\widetilde{n}(k,l)} e^{-j2\pi f_k T_{1\widetilde{n}(k,l)}}, \cdots, \alpha_{M\widetilde{n}(k,l)} e^{-j2\pi f_k T_{M\widetilde{n}(k,l)}}]^T \times$
$\times S_{\widetilde{n}(k,l)}(k,l)$, where $\widetilde{n}(k,l)$ is the index of the most dominant source. The estimation of the correct source mixing parameters cannot be achieved without knowledge of the original signal $S_{\widetilde{n}(k,l)}(k,l)$. However, in anechoic environments, the normalized cross-power spectrum between the signals $X_{a_p}(k,l)$ and $X_{b_p}(k,l)$, acquired by the p-*th* microphone pair, is approximated as

$$r^p_{kl} = \frac{X_{a_p}(k,l) X^*_{b_p}(k,l)}{|X_{a_p}(k,l) X^*_{b_p}(k,l)|} \simeq e^{-j2\pi f_k \Delta t^p_{\widetilde{n}(k,l)}}, \qquad (1)$$

i.e. it is a complex-valued transformation of the time-difference of arrivals (TDOA) $\Delta t^p_{\widetilde{n}(k,l)} = T_{a_p\widetilde{n}(k,l)} - T_{b_p\widetilde{n}(k,l)}$ of the sound propagating from the $\widetilde{n}(k,l) - th$ source to the given microphone pair.

## 3. SOURCE LOCALIZATION THROUGH GLOBAL COHERENCE FIELD

If multiple microphone pairs are available, a function of global coherence can be used in order to estimate the likelihood of presence of a source in a given spatial location. We indicate with $\mathbf{\Delta T}^{\Pi}(\mathbf{q})$ the vector of TDOAs $[\Delta T^1(\mathbf{q}); \Delta T^2(\mathbf{q}), \cdots, \Delta T^P(\mathbf{q})]$ related to a source located at the Cartesian coordinates $\mathbf{q} = (x, y, z)$ observed by a given subset $\Pi$ of $P$ microphone pairs $[(1,2);(1;3);\cdots]$. A typical coherence function is realized through the GCF, which is defined as

$$f_1^{\Pi}(\mathbf{q}) = \sum_l \sum_{p\in\Pi} \mathrm{IDFT}(r^p_{kl}, \overline{\Delta T}^p(\mathbf{q})), \tag{2}$$

where $\overline{\Delta T}^p(\mathbf{q})$ is the multiple integer of the sampling period nearest to $\Delta T^p(\mathbf{q})$ and $\mathrm{IDFT}(r^p_{kl}, \overline{\Delta T}^p)$ is the inverse discrete Fourier transform of the normalized cross-power spectrum of the *l-th* frame, evaluated in $\overline{\Delta T}^p(\mathbf{q})$. According to the Hermitian symmetry of the DFT and neglecting the DC components, eq. (2) can be approximated as

$$f_1^{\Pi}(\mathbf{q}) \simeq \frac{2}{N_{bins}} \sum_l \sum_{p\in\Pi} \sum_{k=1}^{\frac{N_{bins}}{2}-1} \mathrm{Re}[r^p_{kl} e^{j2\pi f_k \overline{\Delta T}^p(\mathbf{q})}] \tag{3}$$

which, after some mathematical manipulation [4], can be equivalently rewritten as

$$f_1^{\Pi}(\mathbf{q}) \simeq \frac{2}{N_{bins}} \sum_l \sum_{p\in\Pi} \sum_{k=1}^{\frac{N_{bins}}{2}-1} \left(1 - \frac{|r^p_{kl} - e^{-j2\pi f_k \overline{\Delta T}^p(\mathbf{q})}|^2}{2}\right). \tag{4}$$

The term $e^{-j2\pi f_k \overline{\Delta T}^p(\mathbf{q})}$ approximates the anechoic propagation model in the direction of the pair $p$ of a source located in $\mathbf{q}$. Therefore, each term in the summation is a measure of similarity between each single observed cross spectrum and the ideal models. Under the aforementioned time-frequency sparseness assumption in each frequency $k$ and frame $l$, a different source is predominant. Therefore, we should expect (4) be maximized in all the source locations (on condition that the microphone geometry provides a sufficient spatial resolution). It can be noted that the coherence over all the dimensions is obtained by a simple summation over the $P$ directions. In other terms, the overall likelihood is based on a weak multidimensional metric of the acoustic propagation coherence. To better understand this aspect, (4) can be reformulated as

$$f_1^{\Pi}(\mathbf{q}) \simeq \frac{2}{N_{bins}} \sum_l \sum_{k=1}^{\frac{N_{bins}}{2}-1} g(L[\mathbf{e}_{kl}]_\gamma), \quad \gamma = 1, \tag{5}$$

where $\mathbf{e}_{kl}$ is the error vector $\mathbf{e}_{kl} = [e^1_{kl}; \cdots; e^P_{kl}]$ with $e^p_{kl} = \frac{|r^p_{kl} - e^{-j2\pi f_k \overline{\Delta T}^p(\mathbf{q})}|}{2}$ and $g(x) = P - x^2$. Therefore, the coherence is computed as a function of the $L_1$ norm of the distance between the observed normalized cross-power spectrum and the ideal anechoic propagation model. It is known that the $L_1$ norm is a weak metric in the multidimensional space, i.e. measure discrepancies across the dimensions do not influence considerably the total norm value. Consequently, high coherence in the propagation directions related to each microphone pair increases the overall value of (5) in all the multidimensional locations lying along such directions.

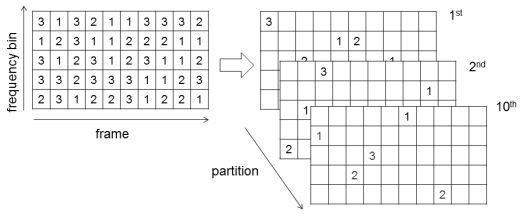



**Fig. 1**. Graphical representation of the partitioning of the observations $r^p_{kl}$ obtained in different time frames and frequency-bins.

## 4. ENHANCED MULTIDIMENSIONAL EXTENSIONS OF SPATIAL COHERENCE FUNCTION

In order to reduce the likelihood in ghost locations the $L_1$ norm in eq. (5) can be substituted with a better multidimensional metric. It is known that higher norms have better discrimination capability in the multidimensional space. In general, the $L_1$ metric can be substituted with an extended norm we call $L_{\gamma,\beta}$ norm, which is defined as

$$L_{\gamma,\beta}[\mathbf{e}_{kl}] = ||\mathbf{e}_{kl}||\gamma \times [1 + \beta \times \mathrm{std}(\mathbf{e}_{kl})], \tag{6}$$

where $\mathrm{std}(\mathbf{e}_{kl})$ is the standard deviation of the error vector with respect to the $P$ dimensions. The $L_{\gamma\beta}$ is a modification of the standard $L_\gamma$. With high values of $\gamma$ (i.e. $\gamma = \infty$) and $\beta > 0$, $L_{\gamma,\beta}$ considerably increases the discrimination of the observation vectors in the multidimensional space, even under high spatial aliasing [5].

As a second extension, the structure of the likelihood function (5) can be improved in order to account for the sparse nature of the source signals in the time-frequency domain. It can be noted that, for a particular spatial location, the total likelihood is generated as a double summation over all the possible time-frequency points which is in contrast with the assumption of source sparseness.

To better account for this property, (5) is modified as follows

$$f_2^{\Pi}(\mathbf{q}) = \sum_{s=1}^{S} \max_{k,l\in Q(s)} g(L_{\gamma,\beta}[\mathbf{e}_{kl}]), \tag{7}$$

where the scaling $\frac{2}{N_{bins}}$ was removed since not useful for the analysis. The entire set of time-frequency observations is partitioned in $S$ subsets and $Q(s)$ indicates the set of time-frequency indexes belonging to the *s-th* partition. The partitioning should be determined according to the expected 2D distribution of the acoustic sources in the time-frequency domain. Here, we adopt a simple random partitioning, i.e. each partition includes time-frequency points randomly chosen from the entire set of observations, which has shown to work well with speech sources.

The meaning of (7) is clarified by figure 1 which provides a graphical representation of the random partitioning. We assume that three sources are sparsely distributed in the discrete time-frequency domain, representing with a different number a different source. We consider a set of 50 time-frequency observations partitioned in $S = 10$ subsets, where each partition contains only $C = 5$ time-frequency points. Here the cardinality $C$ is chosen to be small only to simplify the graphical representation. In general, it should be sufficiently larger than the expected number of sources, in order to ensure that each partition would include observation vectors representing all the sources. For each spatial point, the max over the $S$ partitions selects the observations related to the source leading to the highest likelihood, while discarding the contribute of the

remaining sources. This selection increases the rejection of the source interference in the global likelihood computation, without requiring to explicitly model the joint probability density function of all the source locations which would also require an iterative, or recursive, clustering of the observation vectors [6]. It can be shown that when $g(\cdot)$ is a kernel density derived from a statistical model for (1), applying the max over the $S$ partitions is equivalent to perform an approximated Maximum a Posteriori (MAP) clustering of the observations around the true source locations (for a related proof, see [5]).

## 5. KERNEL FUNCTION DEFINITION

Note that since the locations of multiple sources are derived from the maxima of a single likelihood map, it is also important how the function $g(\cdot)$ is defined, in order to reduce the interference across the sources and enable the correct estimation of multiple source locations. In general $g(\cdot)$ has to be a contrast function, rapidly approaching 0 with the increase of the error metric. While the function $g(x) = P - x^2$, intrinsically used in (4), is frequency independent and related to the steered power of a beamformer, a better $g(\cdot)$ can be derived using a statistical model for (1).

As long as the acoustic waves related to the propagation along the direct paths dominate the secondary reflections, the reverberation can be approximated as an uncorrelated additive random noise [7], and the observations in (1) can be approximated with the model $\mathrm{r}_{kl}^{p} = e^{-j2\pi f_k \tau^{p}_{\widetilde{n}(k,l)}}$, where $\tau^{p}_{\widetilde{n}(k,l)}$ is a random variable of given pdf with mode corresponding to the ideal TDOA of the $\widetilde{n}(k,l)$-*th* source observed at *p-th* microphone pair. Hence, the samples of (1) obtained varying $k$ and $l$ can be treated as a complex-valued transformation of samples of the hidden variables $\tau^{p}_{n}, \quad \forall n$. Under this model $g(\cdot)$ can be defined as an approximated spherical wrapped Gaussian kernel (GK)

$$g(\mathbf{e}_{kl}) = \frac{1}{2\pi f_k} e^{-\frac{\left(\frac{L[\mathbf{e}_{kl}]_{\gamma,\beta}}{2\pi f_k}\right)^2}{2h^2}}, \qquad (8)$$

where $h$ denotes the kernel bandwidth, related to the variance in the propagation time-delay (due to the reverberation)[4]. Note that here the Euclidean norm (originally used in [4]) has been replaced with the generic $L_{\gamma,\beta}$ norm.

We would like to remark that the proposed model requires that low frequencies are discarded [4](e.g., those under 200 Hz). In fact, using a constant bandwidth $h$ is only a convenient approximation, since in real-world environments the noise in the phase introduced by the reverberation is higher at lower frequencies, due to the higher correlation of the reverberant signal at the microphones [7]. However, for higher frequencies the variance can be considered constant, and the proposed kernel is sufficiently reliable to estimate the density of the entire time-delay distribution.

## 6. EVALUATION RESULTS

The proposed spatial function is validated considering the case of multiple sources recorded by two microphone pairs. Room impulse responses (RIRs) have been simulated for a mid-high reverberation, according to the geometry shown in figure 2. The mixtures were obtained by convolving each speech signal (sampled at $f_s$=16kHz), summing the contribute of the sources at each microphone, and perturbed by an AWGN of 40dB. Multiple mixtures were obtained modifying different simulation settings: $T_{60} = \{0.3s; 0.5s\}$,

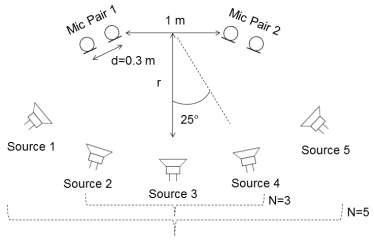
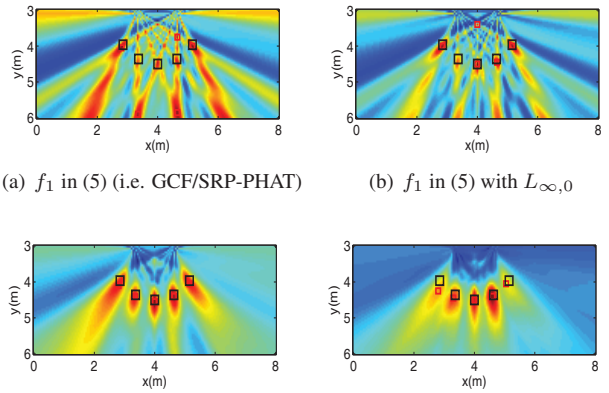



**Fig. 2**. Setup of the simulated scenario.




(a) $f_1$ in (5) (i.e. GCF/SRP-PHAT) (b) $f_1$ in (5) with $L_{\infty,0}$

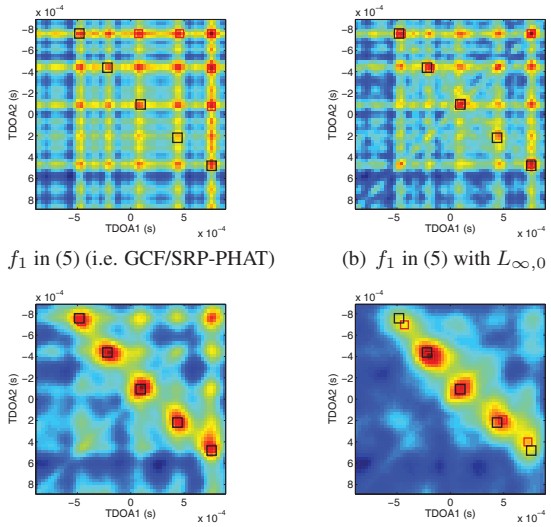
(c) $f_1$ in (5) with $L_{\infty,0}$ and GK (d) $f_2$ in (7) with $L_{\infty,0}$ and GK

**Fig. 3**. Spatial likelihood maps obtained by different functions. Black and red squares indicate true and detected locations (for shortage of space the likelihood is shown in a half-plane).

$r = \{1.5m; 2.5m\}$, $N = \{3; 5\}$. Time-domain mixtures were transformed to the discrete time-frequency representation by STFT applied to Hanning windows of 1024 samples, time-shifted of 256 samples. For each *p-th* microphone pair, the normalized cross-power spectrum was computed as in (1). The mixtures were segmented in non-overlapping blocks of 60 frames (about 1s) and the spatial functions were applied to each block, independently. For the computation of (7), the cardinality of the subset was set to $C = 16$. When the wrapped Gaussian kernel was adopted, the bandwidth was set to $h = \frac{d}{c*20}$ (where $c$ is the sound speed) and the total likelihood was computed discarding frequencies lower than 200 Hz.

In a first qualitative evaluation, we compared the resulting likelihoods obtained with different functions (indicated in figures 3 and 4) obtained simulating a mixtures of 1s for the case of $T_{60} = 0.5$s, $r = 1.5$m, $d = 0.3$m, $N = 5$. Likelihood maps were computed both in the spatial domain, by mapping each allowed Cartesian location in the corresponding bi-dimensional TDOAs, and directly in the bidimensional TDOA domain. For spatial likelihood maps, the function was evaluated on a grid having resolution of $0.05m$, while for the TDOA maps, the function was computed in the range $[-\frac{d}{c}; \frac{d}{c}]$ (seconds) with a resolution of $\frac{d}{c*60}$ (seconds). The functions (5) and (7) were directly evaluated in the frequency-domain, i.e. no approximation of the time-delays to a multiple integer of the sample period was necessary.

Figure 3 shows that with the proposed enhanced metrics the likelihood in ghost locations is drastically reduced. Moreover, the frequency-normalized Gaussian kernel further smooths the spatial representation of the observation vectors by mitigating the impact of outliers, introduced by the reverberation. In this specific case, the partitioning schema adopted in (7) does not seem to introduce any




(a) $f_1$ in (5) (i.e. GCF/SRP-PHAT) (b) $f_1$ in (5) with $L_{\infty,0}$

(c) $f_1$ in (5) with $L_{\infty,0}$ and GK (d) $f_2$ in (7) with $L_{\infty,0}$ and GK

**Fig. 4**. 2D TDOA likelihood maps obtained by different functions. Black and red squares indicate true and detected locations.

| Function | RMSE (m) | DR (%) |
|---|---|---|
| (a) $f_1$ in (5) $+\mathbf{L_{1,0}}$ (i.e. GCF) | 0.05 | 48.96 |
| (b) $f_1$ in (5)$+\mathbf{L_{\infty,0}}$ | 0.05 | 79.17 |
| (c) $f_1$ in (5)$+\mathbf{L_{\infty,1}}$ | 0.06 | 90.89 |
| (d) $f_1$ in (5)$+\mathbf{L_{\infty,0}}+\mathbf{GK}$ | 0.05 | **93.75** |
| (e) $f_1$ in (5)$+\mathbf{L_{\infty,1}}+\mathbf{GK}$ | 0.06 | 92.97 |
| (f) $f_2$ in (7)$+\mathbf{L_{\infty,0}}+\mathbf{GK}$ | 0.09 | 90.36 |
| (g) $f_2$ in (7)$+\mathbf{L_{\infty,1}}+\mathbf{GK}$ | 0.08 | 90.89 |

**Table 1**. Average accuracy and robustness for likelihood maps computed in the spatial domain.

further improvement in the map since many ghosts are likely to be generated outside the room bounds, where the function is not evaluated. On the other hand, when the spatial likelihood is computed in the multidimensional TDOA domains, which is less constrained, the reduction of ghosts becomes more evident with (7). We want to remark that the estimation of the correct multidimensional TDOAs is of crucial importance for applications where the microphone array geometry is not necessarily known, such as blind-beamforming, source separation and source detection.

In a second quantitative evaluation, we measured the average accuracy (distance between estimated and true locations) and robustness (number of detected locations) over 96 simulated mixtures (obtained varying the simulation parameters and taking different signal segments of about 1s). The evaluation was carried out computing the likelihood maps in both spatial and TDOA domains. In the former the accuracy was measured by the Root Mean Square Error (RMSE)

| Function | NTVE (%) | DR (%) |
|---|---|---|
| (a) $f_1$ in (5) $+\mathbf{L_{1,0}}$ (i.e. GCF) | 0.72 | 26.82 |
| (b) $f_1$ in (5)$+\mathbf{L_{\infty,0}}$ | 0.72 | 61.72 |
| (c) $f_1$ in (5)$+\mathbf{L_{\infty,1}}$ | 0.92 | 77.60 |
| (d) $f_1$ in (5)$+\mathbf{L_{\infty,0}}+\mathbf{GK}$ | 0.80 | 76.56 |
| (e) $f_1$ in (5)$+\mathbf{L_{\infty,1}}+\mathbf{GK}$ | 0.91 | 81.25 |
| (f) $f_2$ in (7)$+\mathbf{L_{\infty,0}}+\mathbf{GK}$ | 1.10 | 86.98 |
| (g) $f_2$ in (7)$+\mathbf{L_{\infty,1}}+\mathbf{GK}$ | 1.07 | **87.24** |

**Table 2**. Average accuracy and robustness for likelihood maps computed in the TDOA domain.

of the estimated locations (in meters), while in the latter by the Normalized TDOA Vector Error (NTVE) [4]. From each likelihood $N$ maxima were selected. In the spatial and TDOA domain, source locations were considered detected if RMSE$<$0.3 m and NTVE$<$5%, respectively. The average accuracy is computed only considering the detected locations. The detection rate (DR) is computed as the ratio between the detected sources and $N$.

Tables 1 and 2 report on the average performance obtained for different spatial functions. With (5) and without using GK, the metric $L_{\infty,1}$ is the least affected by ghosts and leads to a very high detection rate. With the Gaussian kernel, the detection rate further increases and the impact in the performance of different metrics is reduced. Note that the max selection over the partitions used in (7) actually reduces the number of observations used to compute the total likelihood map, generating more uncertainty but at the same time removing some ghosts. However, in the spatial domain many ghosts are already removed from the map, because are located outside the room bounds. Then the benefit of the structure of (7) may be overcome by its poor accuracy with a consequent reduction of the overall performance compared to (5). However, things totally change in the TDOA domain where more ghosts are generated and (7) considerably outperforms (5), achieving the highest detection performance.

## 7. CONCLUSIONS

In this paper, the structure of traditional spatial coherence function is analyzed. It is shown that in the case of multiple sources the functions based on the GCF/SRP-PHAT are more prone to generate wrong maxima (i.e. *ghosts*) since they are based on a measure of coherence which is weak in the multidimensional space. Exploiting the sparse representation of sources in time-frequency domain and moving to metrics more discriminative in the higher dimensional space, new spatial kernel functions less sensitive to *ghosts* are derived. Evaluation results over simulated data confirm that the proposed functions considerably improve the detection of multiple source locations with less ambiguity.

Future investigations will concern the use of better source and reverberation models and the integration of the functions in a general multiple source tracking framework. Furthermore, the combination of different likelihood functions seems to be another possible direction to improve both accuracy and detection rate.

## 8. REFERENCES

[1] M. Brandstein and D. Ward, Eds., *Microphone Arrays*. Springer, 2001.

[2] R. DeMori, Ed., *Spoken Dialogue with Computers*. London: Academic Press, 1998.

[3] A. Brutti, M. Omologo, and P. Svaizer, "Localization of multiple speakers based on a two step acoustic map analysis," in *ICASSP*, 2008.

[4] F. Nesta and M. Omologo, "Generalized state coherence transform for multidimensional tdoa estimation of multiple sources," *Audio, Speech, and Language Processing, IEEE Transactions on*, 2011.

[5] F. Nesta and A. Brutti, "Self-clustering non-euclidean kernels for improving the estimation of multidimensional TDOA of multiple sources," in *HSCMA*, 2011.

[6] H. Sawada, S. Araki, R. Mukai, and S. Makino, "Solving the permutation problem of frequency-domain bss when spatial aliasing occurs with wide sensor spacing," in *Proc. of ICASSP*, vol. 5, Toulouse, France, May 2006.

[7] T. Gustaffson, B. D. Rao, and M. Trivedi, "Source localization in reverberant environments: Modeling and statistical analysis," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 6, pp. 791–803, Nov. 2003.