FAST STEERED RESPONSE POWER SOURCE LOCALIZATION USING INVERSE MAPPING OF RELATIVE DELAYS

Jacek Dmochowski, Jacob Benesty, and Sofiène Affes

Université du Québec, INRS-EMT
800 de la Gauchetière Ouest, Suite 6900
Montréal, Québec, H5A 1K6, Canada
{dmochow, benesty, affes} @emt.inrs.ca

ABSTRACT

The acoustic source localization problem has significance for modern intelligent communication systems and future human-computer interaction applications. Although steered-beamforming localization methods perform well in adverse conditions, the algorithms are inherently computationally burdensome due to their need to search the entire location space. This paper presents a computationally-reduced version of the steered response power (SRP) method. The proposed approach is rooted in the inverse mapping of relative delays to candidate source locations, which allows for the transformation of the iterative search from the multidimensional location space to the one-dimensional relative delay space. By subsetting the set of traversed relative delays to only those that experience a high-level of cross-correlation, the computational load is reduced by up to 90 % without incurring a loss in localization accuracy.

Index Terms— Microphone arrays, source localization.

1. INTRODUCTION

The ability to localize an acoustic source using measurements of the sound field across a spatial aperture plays an important role in modern applications including teleconferencing and automatic camera-steering. Moreover, as human-computer interaction becomes more common, our ability to effectively communicate with machines may rely on their ability to locate and track desired individuals.

Localization techniques rooted in steered beamforming [1] are arguably among the most robust methods when operating in reverberant conditions. Early works which paved the way for modern steered-beamforming algorithms include [2], [3], [4]. In the present-day, the steered response power (SRP) algorithm [5], [6] is the most popular steered-beamforming based localization algorithm. However, practical implementation of the algorithm is plagued by its undesirably high computational demands which stem from an inherent iterative search that is sequential in space. Previous works that address the computational load problem of SRP include [7] and [8]. Both of these approaches essentially limit the number of candidate locations using either location-space pruning or a time-difference-of-arrival (TDOA) based pre-processing. In this paper, a different approach to solving the complexity problem of SRP is proposed.

The proposed algorithm is rooted in the relationship between candidate source locations and the relative delays experienced at the array. While previous steered-beamformer algorithms utilize the mapping from source location to relative delay, there is also an inverse mapping from relative delay to a *set* of candidate locations which experience that particular delay. By utilizing this inverse mapping, the search may be performed across the one-dimensional relative delay space. Moreover, by limiting the set of traversed relative delays to only those that experience a high level of cross-correlation, the computational load of the resulting algorithm is drastically reduced. Experiments demonstrate the computational benefits of the proposed method which do not come at the cost of performance.

2. SIGNAL MODEL AND NOTATION

Assume an array of M microphone elements, distributed in some fashion in three-dimensional space, whose outputs are denoted by $x_m(t)$, $m = 0, 1, ..., M-1$, where t denotes time. The spherical coordinate system is used, where range is denoted by r, elevation by ϕ, and azimuth by θ. Consider a signal source located at $(r_\mathrm{s}, \phi_\mathrm{s}, \theta_\mathrm{s})$. Propagation of the signal to microphone m is modeled as:

$$x_m(t) = s\left[t - f_{0,m}\left(r_\mathrm{s}, \phi_\mathrm{s}, \theta_\mathrm{s}\right)\right] + v_m\left(t\right), \tag{1}$$

where x_m is the received microphone output (microphone 0 serves as the reference), t represents time, s is the desired signal, $v_m(t)$ is the additive noise at microphone m, and the function $f_{i,j}$ relates the source location to the relative delay between microphones i and j:

$$f_{i,j}\left(r_\mathrm{s}, \phi_\mathrm{s}, \theta_\mathrm{s}\right) = \frac{1}{c}\left[d_{\mathrm{s},j}\left(r_\mathrm{s}, \phi_\mathrm{s}, \theta_\mathrm{s}\right) - d_{\mathrm{s},i}\left(r_\mathrm{s}, \phi_\mathrm{s}, \theta_\mathrm{s}\right)\right], \tag{2}$$

where c is the speed of sound and $d_{\mathrm{s},i}$ is the distance between the sound source and microphone i. When the source is located in the far-field, the incoming wave front may be assumed to be planar, thus making $f_{i,j}$ independent of the range:

$$f_{i,j}\left(r_\mathrm{s}, \phi_\mathrm{s}, \theta_\mathrm{s}\right)|_{\mathrm{farfield}} = f_{i,j}\left(\phi_\mathrm{s}, \theta_\mathrm{s}\right) \approx \frac{1}{c}\,\boldsymbol{\zeta}_{\phi_\mathrm{s},\theta_\mathrm{s}}^T\left(\mathbf{p}_j - \mathbf{p}_i\right), \tag{3}$$

where $\boldsymbol{\zeta}_{\phi_\mathrm{s},\theta_\mathrm{s}} = \begin{bmatrix} \sin\phi_\mathrm{s}\cos\theta_\mathrm{s} & \sin\phi_\mathrm{s}\sin\theta_\mathrm{s} & \cos\phi_\mathrm{s} \end{bmatrix}^T$ is the unit direction vector of the source signal, and $\mathbf{p}_i$ is the location vector of microphone i. The received microphone signals are sampled with n denoting the time sample.

The set L denotes the location space (i.e., the set of possible source locations), which is three-dimensional in the near-field case and two-dimensional in the case of far-field propagation. In addition, P denotes the set of all unique (order-independent) pairs of microphones, indexed by (i, j) which refers to the microphone pair formed by microphones i and j, with $|P| = \binom{M}{2}$. For each microphone pairing, the set of physically realizable relative delays is given by $D_{i,j} = \left\{-\tau_{\max}^{i,j}, \ldots, -1, 0, 1, \ldots, \tau_{\max}^{i,j}\right\}$ where $\tau_{\max}^{i,j} = \mathrm{round}\left(\frac{f_\mathrm{sf}}{c}\|\mathbf{p}_i - \mathbf{p}_j\|\right)$ is the maximum physically realizable relative delay between microphones i and j, f_sf is the sampling frequency, and $\mathrm{round}\left(\bullet\right)$ denotes the rounding operation.

3. THE SRP AND SRP-PHAT ALGORITHMS

The output of a delay-and-sum beamformer steered to $(\rho, \varphi, \vartheta)$ is:

$$z_{\rho,\varphi,\vartheta}(n) = \sum_{m=0}^{M-1} w_m x_m \left[n + f_{0,m}(\rho, \varphi, \vartheta)\right]. \quad (4)$$

Assuming that $w_m = 1, m = 0, 1, \ldots, M-1$, the output power of the beamformer is then given by:

$$E\left\{z^2_{\rho,\varphi,\vartheta}(n)\right\} = \sum_{i=0}^{M-1}\sum_{j=0}^{M-1} R_{x_i,x_j}\left[f_{i,j}(\rho, \varphi, \vartheta)\right], \quad (5)$$

where $E\{\bullet\}$ denotes mathematical expectation and

$$R_{x_i,x_j}(\tau) = E\{x_i(n)\, x_j(n+\tau)\} \quad (6)$$

is the cross-correlation function for two jointly wide-sense stationary real random processes and $f_{i,j} = f_{0,j} - f_{0,i}$. The estimate of the source location (assuming a single source) is given by:

$$\left(\hat{r}_s, \hat{\phi}_s, \hat{\theta}_s\right) = \arg\max_{(\rho,\varphi,\vartheta)} \sum_{i=0}^{M-1}\sum_{j=0}^{M-1} R_{x_i,x_j}\left[f_{i,j}(\rho, \varphi, \vartheta)\right]. \quad (7)$$

The traditional SRP method implements the optimization of (7) in two distinct phases which are now described.

3.1. Computation of Cross-Correlations

In the first phase, the cross-correlation functions $R_{x_i,x_j}(\tau)$ are computed for all unique microphone pairs $(i,j) \in P$ and the set of all physically realizable relative delays pertaining to each microphone pairing $\tau \in D_{i,j}$. The computation of the cross-correlation functions is typically performed in the frequency-domain via the inverse fast Fourier transform (IFFT):

$$R^{\mathrm{g}}_{x_i,x_j}(\tau) = \sum_{k=0}^{N_\mathrm{f}-1} \psi_\mathrm{g}(k)\, G_{x_i,x_j}(k)\, e^{j2\pi\frac{k}{N_\mathrm{f}}\tau}, \quad (8)$$

where N_f is the FFT length, $G_{x_i,x_j}(k) = X_i(k)\, X_j^*(k)$ is the cross-spectrum between channels i and j, k is the discrete frequency index, $X_i(k)$ is the fast Fourier transform (FFT) of $x_i(n)$, $\psi_\mathrm{g}(k)$ is a pre-filter and $R^{\mathrm{g}}_{x_i,x_j}$ is termed the "generalized cross-correlation" (GCC) function [9]. A commonly used pre-filter is the phase transform (PHAT) weighting $\psi_{\mathrm{PHAT}}(k) = \frac{1}{\left|G_{x_i,x_j}(k)\right|}$ – the resulting algorithm is termed "SRP-PHAT." There are $\sum_{(i,j)\in P}|D_{i,j}|$ cross-correlations to compute.

3.2. SRP Search

The conventional SRP search process is outlined in Table I. An iterative search of the location space is performed. For each location $(\rho, \varphi, \vartheta) \in L$ and microphone pair $(i,j) \in P$, a lookup procedure translates $(\rho, \varphi, \vartheta)$ to a relative delay $\tau^{i,j}_{\rho,\varphi,\vartheta}$, which corresponds to the discrete relative delay experienced between microphones i and j if the source is located at $(\rho, \varphi, \vartheta)$. The steered response power at location $(\rho, \varphi, \vartheta)$ is then updated accordingly. The SRP spectrum (or "energy map") is known after the traversal of the last location, and the candidate location with the highest steered power becomes the estimated source location. The conventional SRP search consists of $|L|\,|P|$ look-up operations, and $|L|\,|P|$ updates.

Table 1: Conventional Search Algorithm.

initialization:
  for all $(\rho, \varphi, \vartheta) \in L$, $S^{\mathrm{SRP}}(\rho, \varphi, \vartheta) := 0$
search:
  for all $(\rho, \varphi, \vartheta) \in L$
    for all $(i,j) \in P$
      look up $\tau^{i,j}_{\rho,\varphi,\vartheta} = \mathrm{round}\left[f_{i,j}(\rho, \varphi, \vartheta)\right]$
      update $S^{\mathrm{SRP}}(\rho, \varphi, \vartheta) := S^{\mathrm{SRP}}(\rho, \varphi, \vartheta) + R^{\mathrm{g}}_{x_i,x_j}\left(\tau^{i,j}_{\rho,\varphi,\vartheta}\right)$
  $\left(\hat{r}_s, \hat{\phi}_s, \hat{\theta}_s\right) := \arg\max_{(\rho,\varphi,\vartheta)\in L} S^{\mathrm{SRP}}(\rho, \varphi, \vartheta)$

4. PROPOSED GENERALIZATION OF SRP

The generalization affects only the search portion of the SRP approach – the cross-correlation functions are computed as usual. At the heart of the generalization is the inverse mapping that maps relative delays to locations. We define this mapping by:

$$f^{-1}_{i,j}(\tau) = \{(\rho, \varphi, \vartheta) \in L \,|\, f_{i,j}(\rho, \varphi, \vartheta) = \tau\}. \quad (9)$$

The inverse mapping $f^{-1}_{i,j}$ maps a single relative delay (integer) to a discrete set of candidate locations. Since the inverse map is based only on array geometry, it may be computed offline and stored in memory. The memory requirements of this inverse look-up table are identical to those of the conventional forward look-up table that maps locations to relative delays.

The proposed search is outlined in Table II. Throughout the proposed search, instead of traversing the three-dimensional location space, the one-dimensional relative delay space is traversed. As each delay (lag) is traversed, all locations which are inverse mapped by that delay are "simultaneously" updated. This means that *the computation of the SRP energy map is no longer performed sequentially in space*. In other words, as the various relative delays are traversed, the energy map is being built-up at the corresponding inverse-mapped candidate locations. The more relative delays and microphones that we traverse, the more accurate the map.

In terms of reducing complexity, the key variable in the proposed implementation is $C_{i,j}$, a subset of $D_{i,j}$, which is the set of relative delays that is traversed in the proposed search process for microphone pair (i,j). In the proposed method, the set of traversed delays is restricted to a proper subset of all physically realizable relative delays. This subset includes the lags that produce high levels of cross-correlation. The traversal of the relative delay space is restricted to a subset that includes the lag that produces the peak in the cross-correlation for each microphone pair:

$$\hat{\tau}^{i,j} = \arg\max_{\tau} R^{\mathrm{g}}_{x_i,x_j}(\tau). \quad (10)$$

The subset of traversed relative delays is then given by:

$$C_{i,j} = \{\hat{\tau}^{i,j} - p \ldots, \hat{\tau}^{i,j} - 1, \hat{\tau}^{i,j}, \hat{\tau}^{i,j} + 1, \ldots, \hat{\tau}^{i,j} + p\} \cap D_{i,j},$$

which is a set of relative delays centered about $\hat{\tau}^{i,j}$. The parameter p determines how many adjacent lags are involved in the search process. The $\cap$ denotes intersection and the intersection with $D_{i,j}$ must be included to account for cases where $\hat{\tau}^{i,j}$ occurs near the edges of $D_{i,j}$. The parameter p determines both the reduction in computational load as well as the resulting source localization accuracy.

The traversal of relative delays is restricted to those that are deemed to potentially inverse map to a set that includes the true location. When $R^{\mathrm{g}}_{x_i,x_j}(\tau)$ is very small, we can be confident that τ does not inverse map to the source. Therefore, the updating of the locations which τ inverse maps to is omitted. By restricting the traversal

Table 2: Proposed Search Algorithm.

initialization:
  for all $(\rho, \varphi, \vartheta) \in L$, $S^{\mathrm{SRP}}(\rho, \varphi, \vartheta) := 0$
search:
  for all $(i, j) \in P$
    for all $\tau \in C_{i,j} \subseteq D_{i,j}$
      look up $f_{i,j}^{-1}(\tau)$
      for all $(\rho, \varphi, \vartheta) \in f_{i,j}^{-1}(\tau)$
        update $S^{\mathrm{SRP}}(\rho, \varphi, \vartheta) := S^{\mathrm{SRP}}(\rho, \varphi, \vartheta) + R^{\mathrm{g}}_{x_i x_j}(\tau)$
  $\left(\hat{r}_{\mathrm{s}}, \hat{\phi}_{\mathrm{s}}, \hat{\theta}_{\mathrm{s}}\right) := \arg\max_{(\rho,\varphi,\vartheta)} S^{\mathrm{SRP}}(\rho, \varphi, \vartheta)$

of the relative delay space, we are not wasting computational time updating locations far away from the peak of the energy map.

Notice that when $C_{i,j} = D_{i,j}$, the proposed search is the conventional (full) SRP search, just performed in different order. When $C_{i,j} = \left\{\hat{\tau}^{i,j}\right\}$, $\forall\,(i,j) \in P$, the proposed search involves only those locations which are inverse mapped by the optimal (peak) cross-correlation lag of at least one microphone pair. The search is scalable in the sense of the cardinality of $C_{i,j}$. The proposed SRP search consists of $\sum_{(i,j)\in P}\sum_{\tau\in C_{i,j}}\left|f_{i,j}^{-1}(\tau)\right|$ updates and $\sum_{(i,j)\in P}|C_{i,j}|$ look-up operations. The look-up operation involves inverse mapping a relative delay to an inverse set of locations.

5. SPATIAL DECOMPOSITION FORMULATION

Motivated by the proposed generalization, it is possible to expand the expression for the steered energy of a given location $(\rho, \varphi, \vartheta)$:

$$S(\rho, \varphi, \vartheta) = \sum_{(i,j)\in P}\sum_{\tau\in D_{i,j}} R^{\mathrm{g}}_{x_i,x_j}(\tau) \sum_{(r,\phi,\theta)\in f_{i,j}^{-1}(\tau)} \delta\left[(\rho, \varphi, \vartheta) - (r, \phi, \theta)\right],$$

where the basis functions are identified as functions whose value is 1 at locations belonging to the set $f_{i,j}^{-1}(\tau)$ and 0 elsewhere, and the weighting coefficients of the basis functions are $R^{(\mathrm{g})}_{x_i,x_j}(\tau)$. Since the second summation is over $D_{i,j}$, this refers to the full SRP map, with all relative delays used. The proposed generalization is indicated by simply switching $D_{i,j}$ to $C_{i,j}$:

$$S'(\rho, \varphi, \vartheta) = \sum_{(i,j)\in P}\sum_{\tau\in C_{i,j}} R^{\mathrm{g}}_{x_i,x_j}(\tau) \sum_{(r,\phi,\theta)\in f_{i,j}^{-1}(\tau)} \delta\left[(\rho, \varphi, \vartheta) - (r, \phi, \theta)\right],$$

Each basis function is identified by the microphone pair and lag which defines the corresponding inverse set $f_{i,j}^{-1}(\tau)$. A given array has $\sum_{(i,j)\in P}|D_{i,j}|$ basis functions. The summation over $C_{i,j}$ means that in $S'(\rho, \varphi, \vartheta)$, basis functions with low weighting $R^{\mathrm{g}}_{x_i,x_j}(\tau)$ are omitted in the representation.

6. EXPERIMENTAL EVALUATION

The proposed algorithm is evaluated in a computer simulation using the image model [10]. An open spherical array of $M = 13$ omnidirectional microphones and a radius of 7.62 cm is employed as the spatial aperture. The room dimensions in centimeters are (304.8, 457.2, 381.0). The center of the sphere is at (152.4, 228.6, 101.6). The speaker is situated at (254, 406.4, 203.2). The source is correctly assumed to be in the far-field, reducing the dimensionality of the location space to 2. Spatially uncorrelated additive noise with an SNR of 20 dB is added to the microphones. The simulated room has a 60 dB reverberation decay time (T_{60}) of 600 ms. The signal is two-minutes of English speech. The sampling rate and frame size are 48 kHz and 128 ms, respectively. The location space is given by the set of all azimuth/elevation pairs with a resolution of 1 degree and thus $|L| = 360 \times 180 = 64,800$.

Table 3: Performance and computational load with simulated data.

p	$\%_{\mathrm{nonanom.}}$	$e_{\theta,\mathrm{RMS}}$	$e_{\phi,\mathrm{RMS}}$	N_{updates}	N_{lookups}
0	82.25	2.34	1.65	172,400	78
1	92.47	2.40	1.98	514,990	234
2	94.38	2.34	1.84	849,470	390
3	94.83	2.15	1.62	1,165,400	545
4	93.26	2.39	1.76	1,464,500	699
5	93.15	2.43	1.75	1,755,100	852
6	93.71	2.43	1.88	2,045,000	1003
7	93.82	2.38	1.77	2,331,700	1153
8	93.26	2.34	1.76	2,610,600	1300
9	93.48	2.31	1.78	2,855,200	1443
10	92.70	2.37	1.76	3,076,700	1584
20	93.93	2.36	1.69	4,507,100	2734
30	94.61	2.35	1.67	4,982,500	3259
43	94.72	2.34	1.65	5,054,400	3354
full	94.72	2.34	1.65	5,054,400	5,054,400

Table 4: Performance and computational load with real data.

p	$\%_{\mathrm{nonanom.}}$	$e_{\theta,\mathrm{RMS}}$	N_{updates}	N_{lookups}
0	59.54	2.18	20,643	28
1	65.15	2.09	61,914	84
2	66.90	2.05	103,125	140
3	68.58	1.93	144,255	196
4	70.62	1.86	185,292	252
5	70.76	1.74	226,207	308
6	71.46	1.78	267,006	364
7	71.67	1.77	307,659	420
8	72.16	1.77	348,147	476
9	71.32	1.77	388,443	532
10	71.39	1.73	428,499	588
20	72.79	1.69	816,607	1148
40	73.42	1.62	1,536,124	2268
100	71.81	1.59	3,184,031	5628
full	72.23	1.56	4,987,892	4,987,892

The proposed and conventional SRP-PHAT algorithms are also evaluated with data obtained from the IDIAP Institute [11]. The array used is a uniform circular array with $M = 8$ omnidirectional microphones and a radius of 10 cm. Since the array is planar, the evaluation focuses on the localization accuracy of only the azimuth angle of arrival. The room dimensions are 8.2-by-3.6-by-2.4 m. The array rests on a centrally located table with dimensions 4.8-by-1.2 m. Throughout the recording process, the speaker moves to 16 locations in an L-shaped corner area of the room and utters a phrase. The microphones are sampled at 16 kHz. To ensure fine location resolution, the GCC measurements are interpolated by a factor of 20 before running the searches. A total of $|L| = 178,139$ locations are included in the search grid. The frame length is 64 ms.

Table 3 summarizes the localization accuracy and computational load of the proposed and conventional SRP searches for various values of p stemming from the simulated data. The performance is evaluated in terms of the percentage of nonanomalous estimates – those that differ from the true azimuth (and if applicable, elevation) angles by more than 5 degrees, and the root mean square (rms) error of the

nonanomalous estimates. Computational load is evaluated in terms of the number of lookups and updates required. The reductions in performance (percentage of nonanomalies) and computational load (number of updates) are plotted as a function of p in Fig. 1a.

The standard SRP search requires $|L|\,|P| = 5,054,400$ lookups and updates each, and yields a $\%_{\text{nonanom.}} = 94.72\%$ rate of nonanomalous estimates. At $p = 0$, the algorithm suffers only a 12.47 % reduction in the number of nonanomalous estimates, while reducing 96.59% of the number updates. At $p \geq 2$, the rate of nonanomalous estimates hovers around that of the full search, while the reduction in computational load decreases commensurately with p. To understand how it is possible to make such drastic cuts in computational load while maintaining optimal performance, Fig. 2 displays the SRP maps produced by the proposed search at $p = \{0, 2, 43\}$ for a sample frame. In these maps, energy is plotted as a function of the azimuth and elevation. White shades denote high levels of energy, the square denotes the actual source location, and the cross denotes the location chosen by the SRP search. The SRP map produced by the proposed algorithm with $p = 0$ consists of $|P| = 78$ basis functions (the light-shaded "ovals") which are concentrated around the actual source location. These basis functions contain enough information to nonanomalously localize the source. At $p = 2$, the energy map bears an even stronger resemblance to the full energy map which is represented by $p = 43$ (this value guarantees the inclusion of all relative delays).

Fig. 1b and Table 4 summarize the results utilizing real data. Once again, the proposed method attains the localization accuracy of the full search at surprisingly low values of p (i.e., $p = 8$). The wastefulness of the conventional SRP search and effectiveness of the inverse-mapping based spatial decomposition are evident.

7. CONCLUSION

This paper has presented a computationally viable paradigm for steered energy based source localization. It was shown that the SRP map may be decomposed into weighted basis functions, and that by including only the significant basis functions in the search, drastic reductions in computational load result. The proposed method may also be applied to other localization algorithms.

8. REFERENCES

[1] B. D. Van Veen and K. M. Buckley, "Beamforming: a versatile approach to spatial filtering," *IEEE ASSP Mag.*, vol. 5, pp. 4–24, Apr. 1988.

[2] W. Bangs and P. Schultheis, "Space-time processing for optimal parameter estimation," in *Signal Processing*, J. Griffiths, P. Stocklin, and C. V. Schooneveld, eds., pp. 577-590, Academic Press, 1973.

[3] G. Carter, "Variance bounds for passively locating an acoustic source with a symmetric line array," *J. Acoust. Soc. Am.*, vol. 62, pp. 922–926, Oct. 1977.

[4] W. Hahn and S. Tretter, "Optimum processing for delay-vector estimation in passive signal arrays," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 608–614, Sept. 1973.

[5] M. Omologo and P. G. Svaizer, "Use of the cross-power-spectrum phase in acoustic event localization," ITC-IRST Tech. Rep. 9303-13, Mar. 1993.

[6] J. Dibiase, "A High-Accuracy, Low-Latency Technique for Talker Localization in Reverberant Environments," PhD Thesis, Brown University, Providence RI, USA, May 2000.

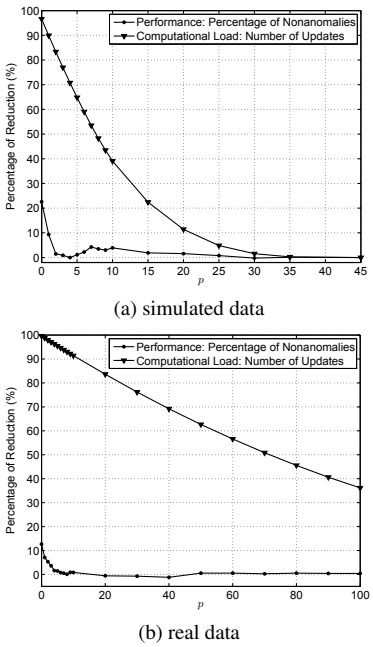
[7] D. N. Zotkin and R. Duraiswami, "Accelerated speech source localization via a hierarchical search of steered response power," *IEEE Trans. Speech and Audio Processing*, vol. 12, pp. 499–508, Sept. 2004.

[8] J. Peterson and C. Kyriakakis, "Hybrid algorithm for robust, real-time source localization in reverberant environments," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Processing*, 2005, vol. 4, pp. 1053–1056.

[9] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 24, pp. 320–327, Aug. 1976.

[10] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, pp. 943–950, Apr. 1979.

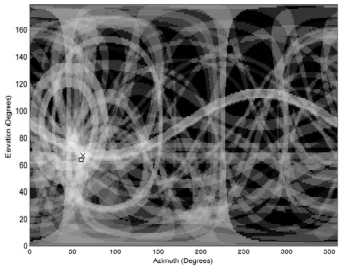
[11] G. Lathoud, J.M. Odobez, and D. Gatica-Perez, "AV16.3: an audio-visual corpus for speaker localization and tracking," in *Proc. MLMI'04 Workshop*, 2006.



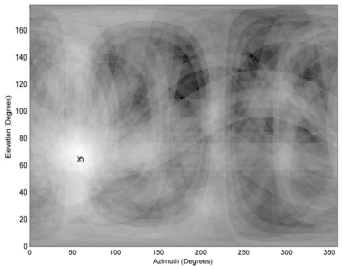
(a) simulated data

(b) real data

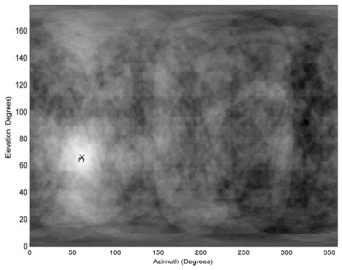
Fig. 1: Reductions in performance and computational load.



(a) $p = 0$



(b) $p = 2$



(c) $p = 43$ (full search)

Fig. 2: SRP energy maps as a function of p.