BROADBAND MUSIC: OPPORTUNITIES AND CHALLENGES FOR MULTIPLE SOURCE LOCALIZATION

Jacek P. Dmochowski[*]*, Jacob Benesty, and Sofiène Affes*

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

ABSTRACT

It is well-known that the subspace MUltiple SIgnal Classification (MUSIC) method provides high-resolution spatial spectral estimates in narrowband signal environments. However, for broadband signals, such high-resolution methods still elude researchers. This paper proposes a broadband version of the MUSIC method using a parameterized version of the spatial correlation matrix. The proposed algorithm utilizes both inter-microphone amplitude and phase differences; as a result, simulation results show that the proposed method allows two broadband sources which are not resolvable by conventional steered beamforming to be accurately resolved.

1. INTRODUCTION

The literature concerning the localization of narrowband signals is quite developed – several now classical methods have been proposed, developed, and improved upon over the years [1]. Their study and understanding is facilitated by the fact that this family of narrowband localization methods lies within a central framework, namely the narrowband *spatial correlation matrix*, which conveys the correlation levels of the various signals across the array of sensors. While the earliest methods apply either fixed or adaptive linear weighting to the matrix (see [2] for a nice review of the conventional methods), more advanced "subspace" methods exploit the eigenstructure of the spatial correlation matrix to yield high resolution spatial spectral estimates – the pioneer of these subspace methods is termed the Multiple Emitter and Signal Parameter Classification (MUSIC) algorithm [3].

The localization of broadband signals is not as well understood – in general, the literature is quite scattered. It is hard to group the various methods into categories because of the disjoint that exists. Nevertheless, recently an attempt has been made to provide a centralized framework for broadband localization [4] – interestingly, this framework also involves a spatial correlation matrix; however, because of the nature of broadband signals, the matrix is nonlinearly parameterized leading to the *parameterized spatial correlation matrix*. This matrix was previously introduced in [5] under the name "steered covariance matrix."

In this paper, the case for establishing the parameterized spatial correlation matrix as the central structure in broadband localization is advanced by presenting a broadband formulation of the MUSIC algorithm. Indeed, the parametrization of the spatial correlation matrix leads to striking similarities between the second-order statistics of the classical narrowband model and that of the parameterized broadband model.

[*] This work was funded by the National Sciences and Engineering Research Council of Canada (NSERC)

2. SIGNAL MODEL

2.1. Propagation Model

Assume an array of M microphone elements, distributed in some fashion in three-dimensional space, whose outputs are denoted by $x_m(t)$, $m = 1, ..., M$, where t denotes time. The spherical coordinate system is used, where the range is denoted by r, elevation by ϕ, and azimuth by θ. For brevity, denote $\mathbf{r} = \begin{bmatrix} r & \phi & \theta \end{bmatrix}^T$ where ^{T} denotes transposition.

Consider a signal source located at $\mathbf{r}_\mathrm{s} = \begin{bmatrix} r_\mathrm{s} & \phi_\mathrm{s} & \theta_\mathrm{s} \end{bmatrix}^T$. Propagation of the signal to microphone m is modeled as:

$$x_m(t) = \alpha_m(\mathbf{r}_\mathrm{s})\, s\left[t - f_{1,m}(\mathbf{r}_\mathrm{s})\right] + v_m(t), \qquad (1)$$

where s is the desired signal, $v_m(t)$ is the additive noise at microphone m which includes any background or sensor noise, as well as reverberation, and microphone 1 is used as the phase reference. The function α_m models attenuation of the desired signal at microphone m due to propagation effects:

$$\alpha_m(\mathbf{r}_\mathrm{s}) \propto \frac{1}{d_{\mathrm{s},m}(\mathbf{r}_\mathrm{s})},$$

where $d_{\mathrm{s},m}(\mathbf{r}_\mathrm{s})$ is the distance from the source to microphone m, and the function $f_{i,j}$ relates the source location to the relative delay between microphones i and j:

$$f_{i,j}(\mathbf{r}_\mathrm{s}) = \frac{1}{c}\left[d_{\mathrm{s},j}(\mathbf{r}_\mathrm{s}) - d_{\mathrm{s},i}(\mathbf{r}_\mathrm{s})\right],$$

where c is the speed of sound.

2.2. Parameterized Spatial Correlation Matrix

The parameterized spatial correlation matrix is defined by:

$$\mathbf{R}_\mathbf{r} = E\left\{\mathbf{x}_\mathbf{r}(t)\,\mathbf{x}_\mathbf{r}^T(t)\right\}, \qquad (2)$$

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

$$\mathbf{x}_\mathbf{r}(t) = \begin{bmatrix} x_1[t] & x_2[t + f_{1,2}(\mathbf{r})] & \cdots & x_M[t + f_{1,M}(\mathbf{r})] \end{bmatrix}^T.$$

Before proceeding, assume that the desired signal s and additive noise v are mutually uncorrelated random processes. Substituting

(1) into (2) results in the following structure for the parameterized spatial correlation matrix:

$$[\mathbf{R_r}]_{i,j} = \alpha_i(\mathbf{r}_s)\,\alpha_j(\mathbf{r}_s)\,R_{s,s}[f_{i,j}(\mathbf{r}) - f_{i,j}(\mathbf{r}_s)] + R_{v_i,v_j}[f_{i,j}(\mathbf{r})] \quad (3)$$

where $[\mathbf{R_r}]_{i,j}$ refers to the ith row and jth column of $\mathbf{R_r}$ and

$$R_{x,y}(\tau) = E\{x(t)\,y(t+\tau)\}$$

is the cross-correlation function for two jointly wide-sense stationary and real random processes.

From (3), notice that the parametrization of the spatial correlation matrix spatially decorrelates the noise term $R_{v_i,v_j}[f_{i,j}(\mathbf{r})]$; if the additive noise is temporally white, and assuming that the lag $f_{i,j}(\mathbf{r}) \neq 0$ if $i \neq j$, the noise component of the parameterized spatial correlation matrix is simply:

$$\mathbf{R_r}|_{\text{noise}} = \sigma_v^2\mathbf{I}$$

where $\sigma_v^2 = R_{v_1,v_1}(0) = \sigma_{v_1}^2 = \cdots = R_{v_M,v_M}(0) = \sigma_{v_M}^2$. Thus, it is reasonable to assume that with the parametrization, the spatial correlation matrix of the noise is diagonal.

Secondly, when $\mathbf{r} = \mathbf{r}_s$, the signal component of the parameterized spatial correlation matrix takes the form of the following rank-one matrix:

$$\mathbf{R_{r_s}}|_{\text{signal}} = \sigma_s^2\boldsymbol{\alpha}(\mathbf{r}_s)\,\boldsymbol{\alpha}^T(\mathbf{r}_s),$$

where $\sigma_s^2 = R_{s,s}(0)$ and

$$\boldsymbol{\alpha}(\mathbf{r}_s) = \begin{bmatrix} \alpha_1(\mathbf{r}_s) & \alpha_2(\mathbf{r}_s) & \cdots & \alpha_M(\mathbf{r}_s) \end{bmatrix}^T.$$

If the parametrization is not matched to the location of the signal [i.e., $f_{i,j}(\mathbf{r}) \neq f_{i,j}(\mathbf{r}_s)$ if $i \neq j$)], the signal component is decorrelated. Assuming that the desired signal is a white process, the signal component takes the form of

$$\mathbf{R_r}|_{\text{signal}} = \sigma_s^2\text{diag}\left[\alpha_1^2(\mathbf{r}_s), \alpha_2^2(\mathbf{r}_s), \cdots, \alpha_M^2(\mathbf{r}_s)\right],$$

where diag$(\bullet)$ is a diagonal matrix with its nonzero entries denoted by the arguments. Putting all of this together, we arrive at

$$\mathbf{R_r} = \begin{cases} \sigma_s^2\,\boldsymbol{\alpha}(\mathbf{r}_s)\,\boldsymbol{\alpha}^T(\mathbf{r}_s) + \sigma_v^2\mathbf{I}, & \text{if } \mathbf{r} = \mathbf{r}_s \\ \sigma_s^2\text{diag}\left[\alpha_1^2(\mathbf{r}_s), \cdots, \alpha_M^2(\mathbf{r}_s)\right] + \sigma_v^2\mathbf{I}, & \text{otherwise} \end{cases}.$$

2.3. Multiple Broadband Sources

Consider the case of two broadband sources located at $\mathbf{r}_{s,1}$ and $\mathbf{r}_{s,2}$; for simplicity, the absence of noise is assumed. The output of the mth microphone is then

$$x_m(t) = \alpha_m(\mathbf{r}_{s,1})\,s_1[t - f_{1,m}(\mathbf{r}_{s,1})] + \alpha_m(\mathbf{r}_{s,2})\,s_2[t - f_{1,m}(\mathbf{r}_{s,2})],$$

where s_1 and s_2 are the source signals. Assuming that the two sources are uncorrelated, the entries of the parameterized spatial correlation matrix then follow as

$$[\mathbf{R_r}]_{i,j} = \alpha_i(\mathbf{r}_{s,1})\,\alpha_j(\mathbf{r}_{s,1})\,R_{s_1,s_1}[f_{i,j}(\mathbf{r}) - f_{i,j}(\mathbf{r}_{s,1})] + \alpha_i(\mathbf{r}_{s,2})\,\alpha_j(\mathbf{r}_{s,2})\,R_{s_2,s_2}[f_{i,j}(\mathbf{r}) - f_{i,j}(\mathbf{r}_{s,2})].$$

If s_1 and s_2 are white, we obtain

$$\mathbf{R_r} = \begin{cases} \sigma_{s_1}^2\,\boldsymbol{\alpha}(\mathbf{r}_{s,1})\,\boldsymbol{\alpha}^T(\mathbf{r}_{s,1}) + \sigma_{s_2}^2\text{diag}\left[\alpha_1^2(\mathbf{r}_{s,2}), \cdots, \alpha_M^2(\mathbf{r}_{s_2})\right], & \text{if } \mathbf{r} = \mathbf{r}_{s_1} \\ \sigma_{s_2}^2\,\boldsymbol{\alpha}(\mathbf{r}_{s,2})\,\boldsymbol{\alpha}^T(\mathbf{r}_{s,2}) + \sigma_{s_1}^2\text{diag}\left[\alpha_1^2(\mathbf{r}_{s,1}), \cdots, \alpha_M^2(\mathbf{r}_{s_1})\right], & \text{if } \mathbf{r} = \mathbf{r}_{s_2} \\ \sigma_{s_1}^2\text{diag}\left[\alpha_1^2(\mathbf{r}_{s_1}), \cdots, \alpha_M^2(\mathbf{r}_{s_1})\right] + \sigma_{s_2}^2\text{diag}\left[\alpha_1^2(\mathbf{r}_{s_2}), \cdots, \alpha_M^2(\mathbf{r}_{s_2})\right], & \text{otherwise} \end{cases}. \quad (4)$$

From (4), we see that the parametrization of the spatial correlation matrix results in the signal subspace of $\mathbf{R}_{\mathbf{r}_{s,i}}, i = 1, 2$, having dimension one – the parametrization "forces" the non-aligned source(s) to the noise subspace. Thus, we may proceed with the development assuming one signal source without loss of generality.

3. APPLYING MUSIC TO BROADBAND SIGNALS

3.1. Spectral Decomposition of Parameterized Spatial Correlation Matrix

Proceeding with the single-source scenario, consider the case when $\mathbf{r} = \mathbf{r}_s$; take any vector $\mathbf{u}$ such that

$$\boldsymbol{\alpha}^T(\mathbf{r}_s)\,\mathbf{u} = 0. \quad (5)$$

There are $M-1$ linearly independent vectors which fulfill (5); each of these vectors is therefore an eigenvector of $\mathbf{R_{r_s}}$ whose corresponding eigenvalue is σ_v^2. Because of this, these eigenvectors are termed the "noise eigenvectors" and are denoted by $\mathbf{u}_{n,\mathbf{r}_s,1}, \mathbf{u}_{n,\mathbf{r}_s,2}, \cdots, \mathbf{u}_{n,\mathbf{r}_s,M-1}$. The lone other eigenvector (denoted by $\mathbf{u}_{s,\mathbf{r}_s}$) must have an eigenvalue (denoted by $\lambda_{s,\mathbf{r}_s}$) greater than σ_v^2 and is thus termed the "signal eigenvector."

The spectral decomposition of $\mathbf{R_{r_s}}$ may thus be written as:

$$\mathbf{R_{r_s}} = \lambda_{s,\mathbf{r}_s}\mathbf{u}_{s,\mathbf{r}_s}\mathbf{u}_{s,\mathbf{r}_s}^T + \sigma_v^2\mathbf{U}_{n,\mathbf{r}_s}\mathbf{U}_{n,\mathbf{r}_s}^T,$$

where

$$\mathbf{U}_{n,\mathbf{r}_s} = \begin{bmatrix} \mathbf{u}_{n,\mathbf{r}_s,1} & \mathbf{u}_{n,\mathbf{r}_s,2} & \cdots & \mathbf{u}_{n,\mathbf{r}_s,M-1} \end{bmatrix},$$

with $\lambda_{s,\mathbf{r}_s} > \sigma_v^2$ and

$$\boldsymbol{\alpha}^T(\mathbf{r}_s)\,\mathbf{U}_{n,\mathbf{r}_s} = \mathbf{0}_{M-1}^T \quad (6)$$

where $\mathbf{0}_{M-1} = \begin{bmatrix} 0 & 0 & \cdots & 0 \end{bmatrix}^T$ is a vector of $M-1$ zeros.

Eq. (6) is the property that allows us to define the so-called "MUSIC spectrum" for the broadband case; at locations (parameters of spatial correlation matrix) near $\mathbf{r}_s$, the noise eigenvectors of the the parameterized spatial correlation matrix will be orthogonal to the vector of attenuation constants $\boldsymbol{\alpha}(\mathbf{r}_s)$. On the other hand, when the matrix is not time-aligned, this property fails to hold.

3.2. Broadband MUSIC

For each parameter $\mathbf{r}$ in the parameter space, we compute the signal eigenvector $\hat{\mathbf{u}}_{s,\mathbf{r}}$ as the principal eigenvector of $\hat{\mathbf{R}}_\mathbf{r}$, the estimated parameterized spatial correlation matrix. The remaining eigenvectors are designated as the noise eigenvectors $\hat{\mathbf{u}}_{n,\mathbf{r},i}, i = 1, 2, \cdots, M-1$ and comprise the noise subspace

$$\hat{\mathbf{U}}_{n,\mathbf{r}} = \begin{bmatrix} \hat{\mathbf{u}}_{n,\mathbf{r},1} & \hat{\mathbf{u}}_{n,\mathbf{r},2} & \cdots & \hat{\mathbf{u}}_{n,\mathbf{r},M-1} \end{bmatrix}.$$

For each $\mathbf{r}$, $\boldsymbol{\alpha}(\mathbf{r})$ is set to $\left[\begin{array}{ccc} \frac{1}{d_{\mathrm{s},1}(\mathbf{r})} & \frac{1}{d_{\mathrm{s},2}(\mathbf{r})} & \cdots \quad \frac{1}{d_{\mathrm{s},M}(\mathbf{r})} \end{array}\right]^T$ and then normalized such that $\| \boldsymbol{\alpha}(\mathbf{r})\| = 1$.

The broadband MUSIC spectrum at location $\mathbf{r}$ is formally defined as:

$$S_{\mathrm{MUSIC}}(\mathbf{r}) = \frac{1}{\boldsymbol{\alpha}^T(\mathbf{r})\hat{\mathbf{U}}_{\mathrm{n},\mathbf{r}}\hat{\mathbf{U}}_{\mathrm{n},\mathbf{r}}^T\boldsymbol{\alpha}(\mathbf{r})}.$$

Notice that the broadband MUSIC spectrum uses both inter-microphone amplitude *and* phase differences. The theoretical amplitude differences are given by $\boldsymbol{\alpha}(\mathbf{r})$; when $\mathbf{r} = \mathbf{r}_{\mathrm{s}}$, this vector is expected to be orthogonal to the noise eigenvectors in $\hat{\mathbf{U}}_{\mathrm{n},\mathbf{r}}$. The phase differences are exploited in the computation of the parameterized spatial correlation matrix itself: when the steered microphones become time aligned with respect to the source position, the MUSIC eigenstructure results.

In practice, we must frequently localize non-white sources. Thus, the cross-correlation functions are computed using the generalized cross-correlation (GCC) method with the phase transform (PHAT) [6]:

$$R_{x_i,x_j}^{\mathrm{PHAT}}(\tau) = \int_{-\infty}^{\infty} \frac{G_{x_i,x_j}(f)}{\left|G_{x_i,x_j}(f)\right|} e^{j2\pi f\tau} df,$$

where f denotes frequency. In the computation of $\mathbf{R}_{\mathbf{r}}$, $R_{x_i,x_j}^{\mathrm{PHAT}}(\tau)$ replaces $R_{x_i,x_j}(\tau)$, and the resulting algorithm is referred to as "MUSIC-PHAT" throughout the rest of this paper.

4. SIMULATION EVALUATION

To evaluate the ability of the proposed broadband MUSIC scheme to provide consistent high-resolution estimates of the locations of multiple broadband sources, a computer simulation using the image model method of [7] is performed. The simulations assume a uniform circular array of 10 omnidirectional microphones with an array radius of 12.7 cm, located at $(152.4, 228.6, 101.6)$ cm. The room dimensions are 304.8-by-457.2-by-381 cm. The reverberation level ranges from a 60 dB decay time (T_{60}) of 0 ms to 300 ms. The reverberation times are measured using the reverse-time integrated impulse response method of [8].

In order to show the spatial spectra as a function of one spatial dimension, the location space is assumed to be the circumference of a circle of radius 177.8 cm whose center is the center of the circular array. 360 candidate positions at increments of 1° degree around the circle are considered. Note that the proposed algorithm does not require knowledge of the range of the sources; however, for the simulations, this was assumed so that the resolution benefit may be plotted. In practice, the location set will be three-dimensional: (r, ϕ, θ). Two signal sources with equal power are present: one at $(213.21, 395.66, 101.6)$ cm or 70°, and the other at $(152.4, 406.4, 101.6)$ cm or 90°. In the first simulation scenario, the signals are band-limited realizations of a white Gaussian process; in the second, the signals are English speech.

The signal-to-noise ratio (SNR) is 20 dB, and the additive noise is a temporally and spatially white Gaussian process that is uncorrelated with the source signals. In order to attain precise spatial resolution, the sampling rate is chosen to be 48 kHz. The spatial spectra are computed once per 128 ms frame, and then averaged over all frames in the 30-second data set.

For the simulations with white signals, the proposed MUSIC method is compared to the popular steered response power (SRP) method [9], which corresponds to maximizing the power of a steered conventional beamformer. For simulations with speech signals, the proposed MUSIC-PHAT method is compared to the SRP-PHAT method [10].

Figure 1 shows the resulting spectra with two white sources. While the SRP spectrum is not able to resolve the two nearby sources, the broadband MUSIC spectrum shows two distinct peaks with only slight bias errors. Notice also the difference in spectral level between the background and the sources. As the level of reverberation increases, the MUSIC spectrum continues to accurately resolve the sources; furthermore, the reverberant field is also sharply depicted (i.e., smaller peaks in the background).

Figure 2 shows the spectra resulting from the two speech source scenario. In the anechoic and lightly reverberant environments (0 and 150 ms, respectively), the MUSIC-PHAT spectrum provides accurate resolution of the two simulated speakers, unlike SRP-PHAT which shows only a single peak. However, at $T_{60} = 300$ ms, the spectrum becomes noticeably corrupted with multiple spurious peaks of significant amplitude; these may be confused with sources. This is a result of the PHAT pre-filtering, which weights all frequencies equally in the cross-correlation computation: with a significant level of reverberation, frequencies containing much reverberation are boosted by this equal weighting. These boosted reflections may then be "perceived" by MUSIC as direct-path components as they may fulfill the orthogonality condition (6) at their corresponding locations. Further work is needed to achieve the desired correlation whitening without amplifying the unwanted reverberant components.

5. CONCLUSION

This paper has presented a broadband version of MUSIC which utilizes both amplitude and phase differences of the sound signals arriving at an array of microphones to localize multiple acoustic sources. It was shown that the algorithm provides greater spatial resolution than the commonly used SRP and SRP-PHAT methods.

6. REFERENCES

[1] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Processing Mag.*, vol. 13, pp. 67–94, July 1996.

[2] D. H. Johnson, "The application of spectral estimation methods to bearing estimation problems," *Proc. IEEE*, vol. 70, pp. 1018–1028, Sept. 1982.

[3] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. AP-34, pp. 276 – 280, Mar. 1986.

[4] J. Dmochowski, J. Benesty, and S. Affes, "Direction of arrival estimation using the parameterized spatial correlation matrix," *IEEE Trans. Audio, Speech and Language Proceessing*, vol. 15, pp. 1327 – 1339, May 2007.

[5] J. Krolik and D. Swingler, "Multiple broad-band source location using steered covariance matrices," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 1481–1494, Oct. 1989.

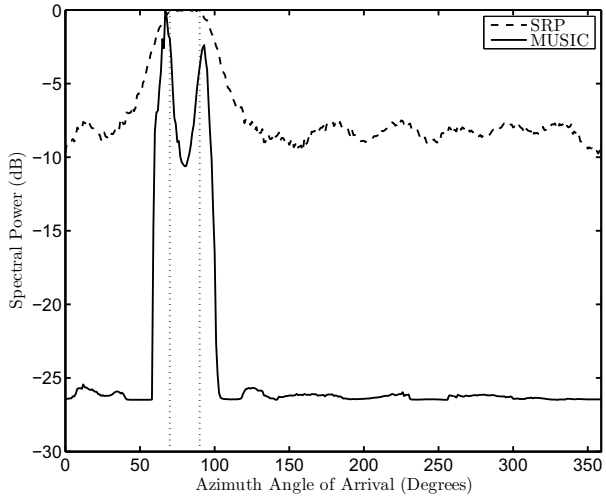
[6] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 24, pp. 320–327, Aug. 1976.

[7] 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.

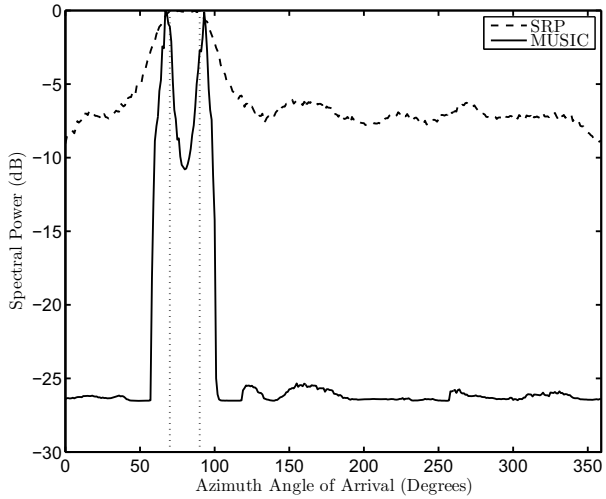
[8] M. R. Schroeder, "New method for measuring reverberation time," *J. Acoust. Soc. Am.*, vol. 37, pp. 409–412, 1965.

[9] 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.

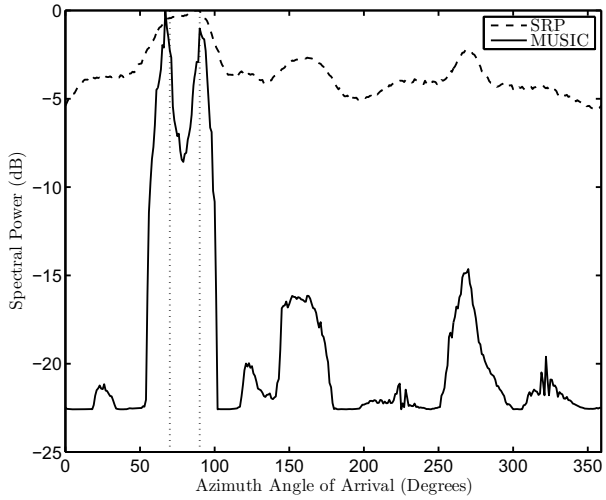
[10] J. Dibiase, "A High-Accuracy, Low-Latency Technique for Talker Localization in Reverberant Environments," PhD Thesis, Brown University, Providence RI, USA, May 2000.



(a) $T_{60} = 0$ ms

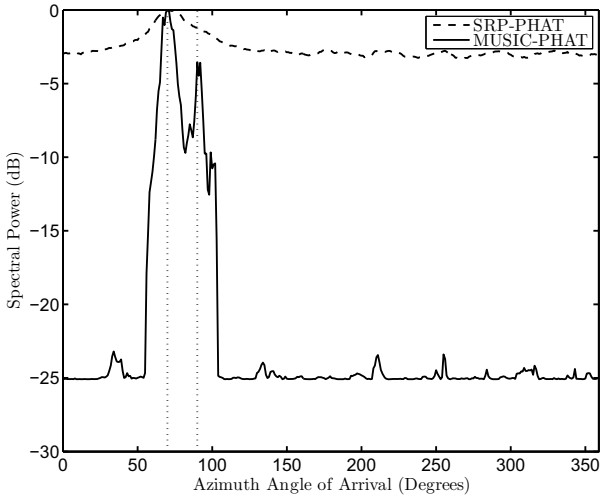


(b) $T_{60} = 150$ ms

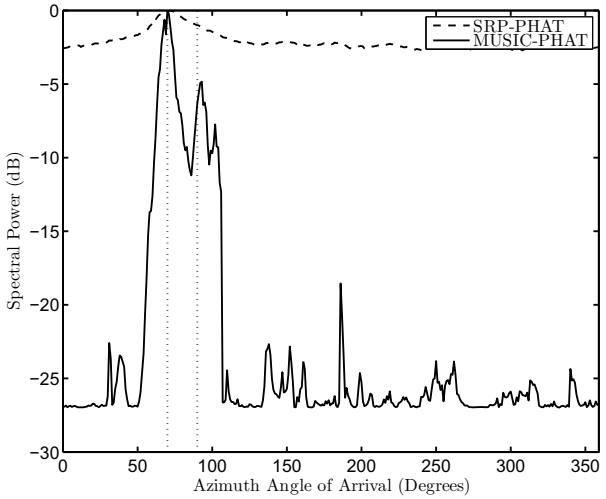


(c) $T_{60} = 300$ ms

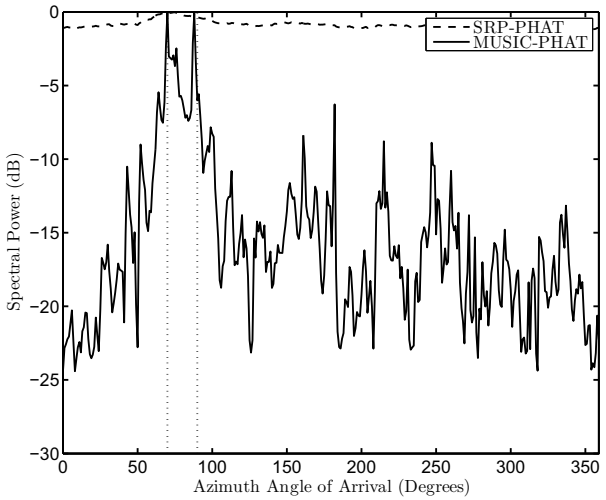
Figure 1: Spatial spectra for two nearby white sources.



(a) $T_{60} = 0$ ms



(b) $T_{60} = 150$ ms



(c) $T_{60} = 300$ ms

Figure 2: Spatial spectra for two nearby speech sources.