SOUND SOURCE LOCALIZATION IN A REVERBERANT ROOM USING HARMONIC BASED MUSIC

Lachlan Birnie, Thushara D. Abhayapala, Hanchi Chen, Prasanga N. Samarasinghe
Contact:{lachlan.birnie, thushara.abhayapala, prasanga.samarasinghe}@anu.edu.au

College of Engineering and Computer Science
The Australian National University, Canberra, Australia

Summary

- **Goal:** Localize multiple sound source positions in a reverberant real-world environment.
- **Problem:** Acoustic reflections add confusion to source position.
- **Method:** Account for reflections by incorporating a harmonic coupling model of the room transfer function.
- **Results:** Improved robustness and position estimation.
- **Conclusions:** Reflections can be helpful when used carefully.

MUSIC Subspace Localization Method

Reverberant Room

- To a sound receiver, each acoustic reflection looks like a duplicated sound source.
- It is difficult to know which source is the original when we do not account for reflections.

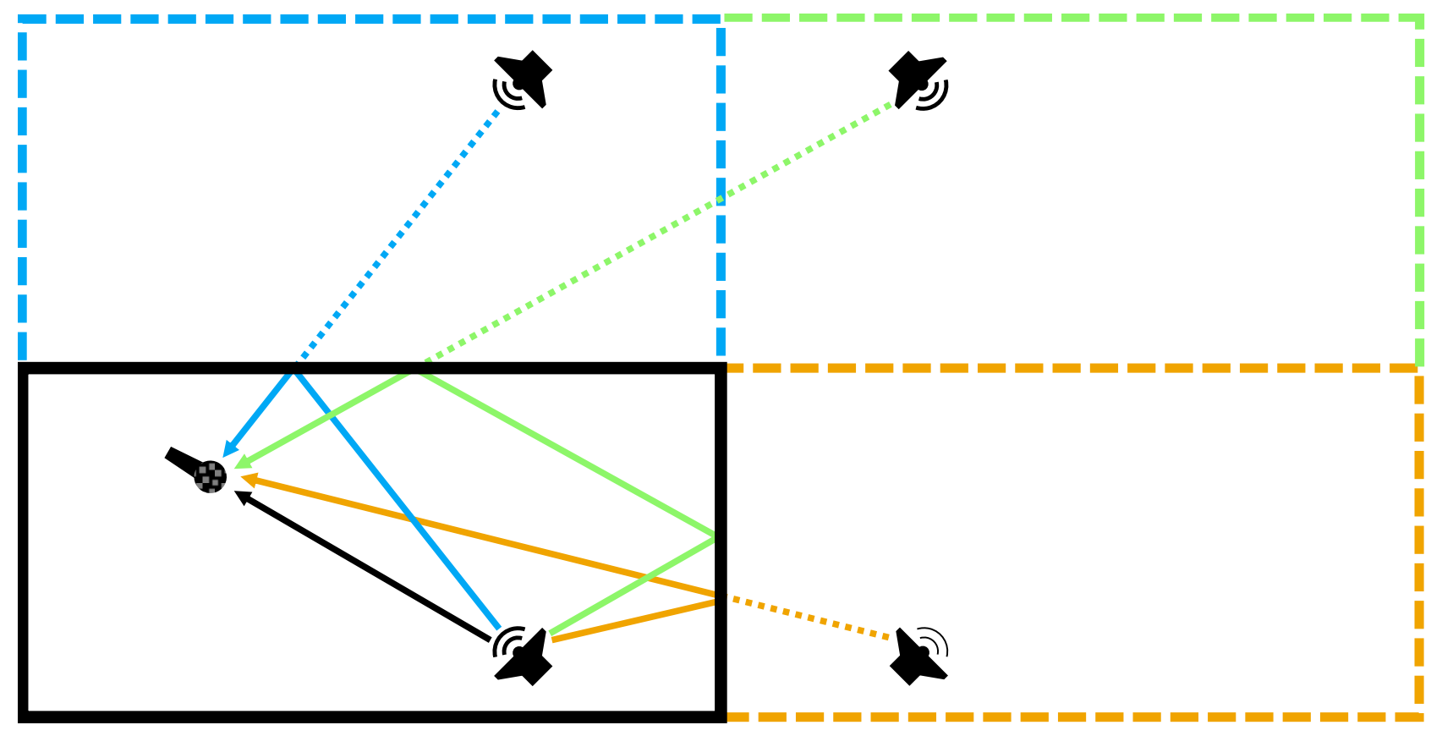

Fig.1 Acoustic reflections and their secondary sound sources

Direct Sound

- Sound incident to the receiver due to the sound source can be modeled with spherical harmonics.

$$P_{\mathrm{d}}(k,\boldsymbol{x_q}) = \sum_{\nu=0}^{V}\sum_{\mu=-\nu}^{\nu} \underbrace{\Big(ikh_\nu(k|\boldsymbol{y_\ell}|)Y^*_{\nu\mu}(\hat{\boldsymbol{y}}_{\boldsymbol{\ell}})\Big)}_{\text{direct sound } \boldsymbol{\Psi}(k)} j_\nu(k|\boldsymbol{x_q}|)Y_{\nu\mu}(\hat{\boldsymbol{x}}_{\boldsymbol{q}})$$

Measured Sound Field Model

- Measurements consider direct sound and noise.
- Reflections are observed as noise.

$$\underbrace{\boldsymbol{\gamma}(k)}_{\text{measured sound}} = \underbrace{\underbrace{\boldsymbol{\Psi}(k)}_{\text{direct sound}}}_{\text{steering vector } \boldsymbol{Y}} \times \underbrace{\boldsymbol{S}(k)}_{\text{source signals}} + \underbrace{\Big(\underbrace{\boldsymbol{n}(k)}_{\text{noise}} + \underbrace{\boldsymbol{n}_{\mathrm{r}}(k)}_{\text{reflections}}\Big)}_{\text{coherent noise}}$$

Noise Subspace

- Noise subspace is found from covariance of measured sound.

$$\begin{aligned}\boldsymbol{R_\gamma}(k) &\triangleq E\Big\{\boldsymbol{\gamma}(k)\boldsymbol{\gamma}(k)^{\mathrm{H}}\Big\}\\ &\approx \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\gamma}(k,t)\boldsymbol{\gamma}(k,t)^{\mathrm{H}}\\ &= \boldsymbol{Y}\boldsymbol{R}_{\mathrm{s}}(k)\boldsymbol{Y}^{\mathrm{H}} + \boldsymbol{R}_{\mathrm{n}}(k)\\ &= \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^{\mathrm{H}} = \big[\boldsymbol{U}_{\mathrm{s}} \;\; \underbrace{\boldsymbol{U}_{\mathrm{n}}}_{\text{noise subspace}}\big]\begin{bmatrix}\boldsymbol{\Sigma}_{\mathrm{s}} & 0\\ 0 & \boldsymbol{\Sigma}_{\mathrm{n}}\end{bmatrix}\begin{bmatrix}\boldsymbol{U}_{\mathrm{s}}^{\mathrm{H}}\\ \boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}\end{bmatrix}\end{aligned}$$

Harmonic MUSIC

- Source position is estimated with a MUSIC algorithm [1,2].

$$M(k,\boldsymbol{y}) = \frac{1}{||\boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}(k)\boldsymbol{Y}(\boldsymbol{y})||^2}$$

- Steering vector is orthogonal to the noise subspace when it points to a sound source.
- Sources appear as peaks in the MUSIC spectra plot.

Simulation

- $4 \times 6 \times 3$ m reverberant shoebox room by image source method.
- Source positions:

1) $(0.4\mathrm{m}, 60^\circ, 50^\circ)$
2) $(0.8\mathrm{m}, 120^\circ, 300^\circ)$
3) $(0.8\mathrm{m}, 140^\circ, 320^\circ)$
4) $(1.0\mathrm{m}, 60^\circ, 50^\circ)$

Source Localization Without Modeling Reflections

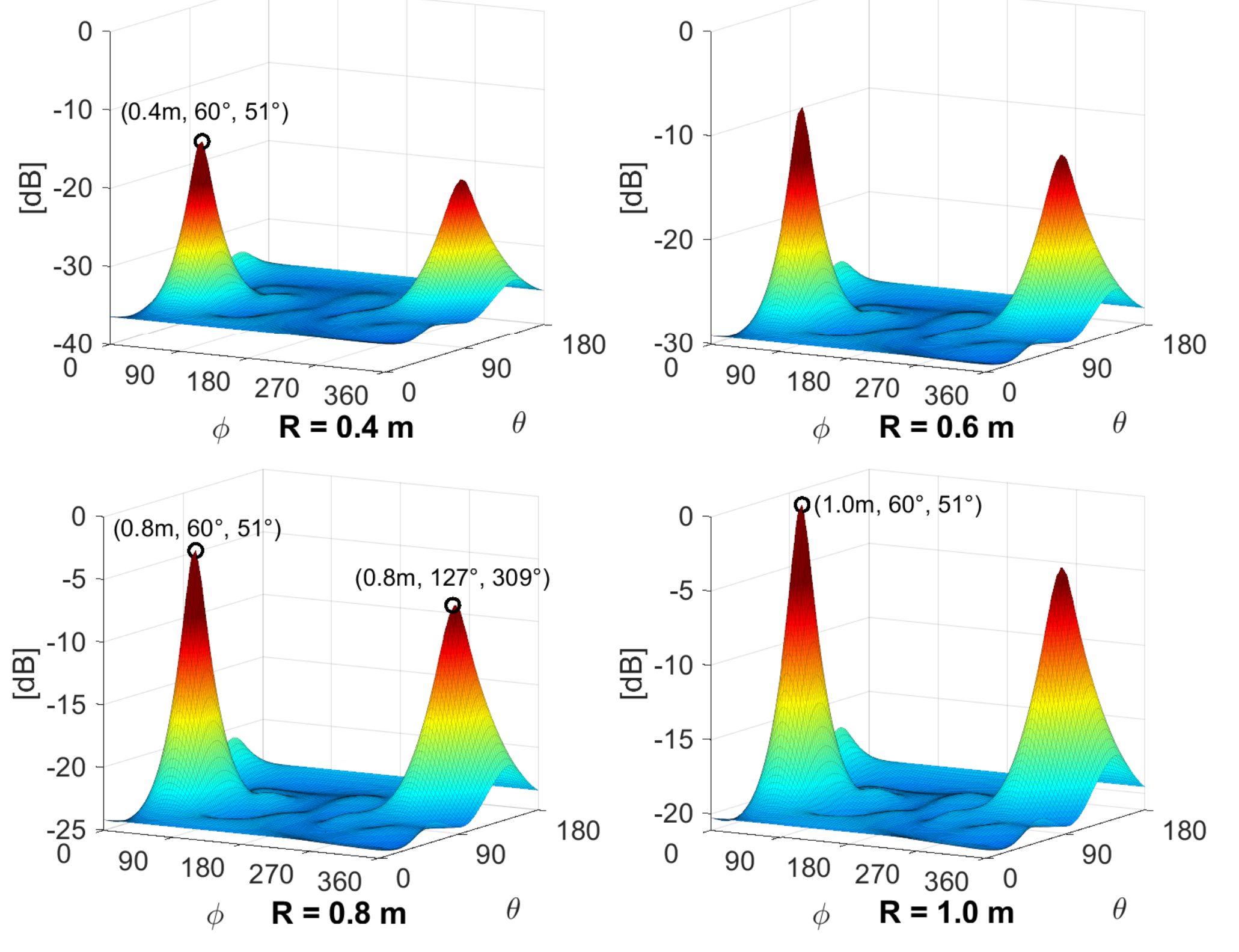

Fig.2 MUSIC spectra without modeling reflections

- Identifies two sound sources.
- Unable to uniquely distinguish nearby sources 2 & 3.
- Cannot radially separate same angular sources 1 & 4.

Proposed Localization Method

- We propose a localization method that models the direct sound and reflected sound components with spherical harmonics.
- Reflections are modeled and incorporated with harmonic room transfer function coupling coefficients [3].

Room Coupling Model

- Direct sound incident to the receive region is given by $\boldsymbol{\Psi}(k)$.
- Indirect sound emitted from the source region is given by $\boldsymbol{\beta}(k)$.
- Reflected sound is given by the room coupling coefficients $\boldsymbol{\alpha}(k)$.

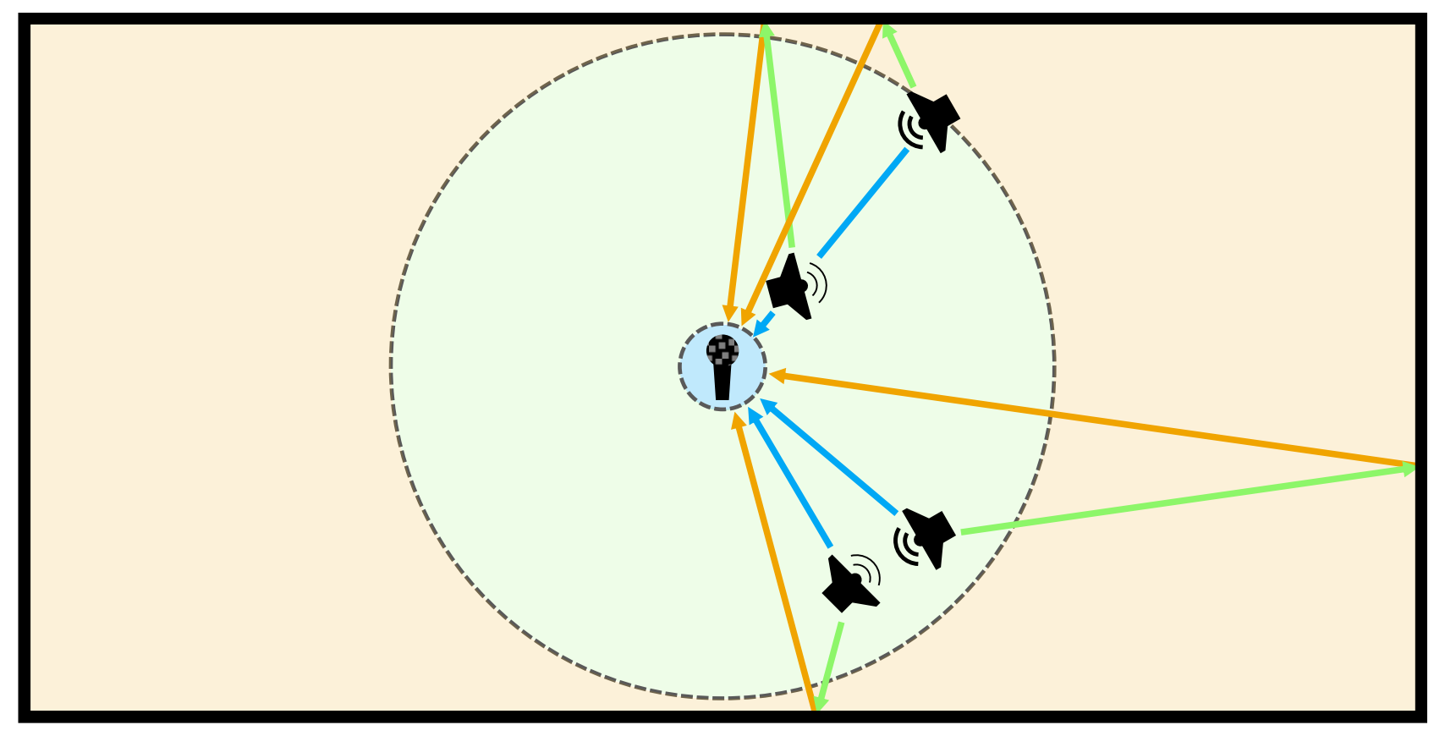

Fig.3 Acoustic regions

Parameterization of the Room Transfer Function

- RTF between regions of arbitrarily positioned or moving sources and receivers can be parameterized with spherical harmonics.
- Room characteristics are given by coupling coefficients $\boldsymbol{\alpha}(k)$ that are independent of source/receiver position.
- Coupling coefficients can be modeled or measured once for any arbitrary real-world environment [3,4].

$$\begin{aligned}P_{\mathrm{r}}(k,\boldsymbol{x_q}) = &\sum_{\nu=0}^{V}\sum_{\mu=-\nu}^{\nu} j_\nu(k|\boldsymbol{x_q}|)Y_{\nu\mu}(\hat{\boldsymbol{x}}_{\boldsymbol{q}})\\ &\times \underbrace{\sum_{n=0}^{N}\sum_{m=-n}^{n} \underbrace{\alpha_{\nu\mu}^{nm}(k)}_{\text{room coupling}} \underbrace{\Big(ikj_n(k|\boldsymbol{y_\ell}|)Y^*_{nm}(\hat{\boldsymbol{y}}_{\boldsymbol{\ell}})\Big)}_{\text{indirect sound } \boldsymbol{\beta}(k)}}_{\text{reflected sound}}\end{aligned}$$

Proposed Sound Field Model

- Measurements consider direct sound, reflected sound and noise.

$$\boldsymbol{\gamma}(k) = \underbrace{\Big(\underbrace{\boldsymbol{\Psi}(k)}_{\text{direct sound}} + \underbrace{\boldsymbol{\alpha}(k)}_{\text{coupling}} \times \underbrace{\boldsymbol{\beta}(k)}_{\text{indirect sound}}\Big)}_{\text{steering vector } \boldsymbol{A}(k)} \times \underbrace{\boldsymbol{S}(k)}_{\text{signals}} + \underbrace{\boldsymbol{n}(k)}_{\text{noise}}$$

Reflection Coupled MUSIC

$$M(k,\boldsymbol{y}) = \frac{1}{||\boldsymbol{U}_{\mathrm{n}}^{\mathrm{H}}(k)\boldsymbol{A}(k,\boldsymbol{y})||^2}$$

Source Localization With Modeled Reflections

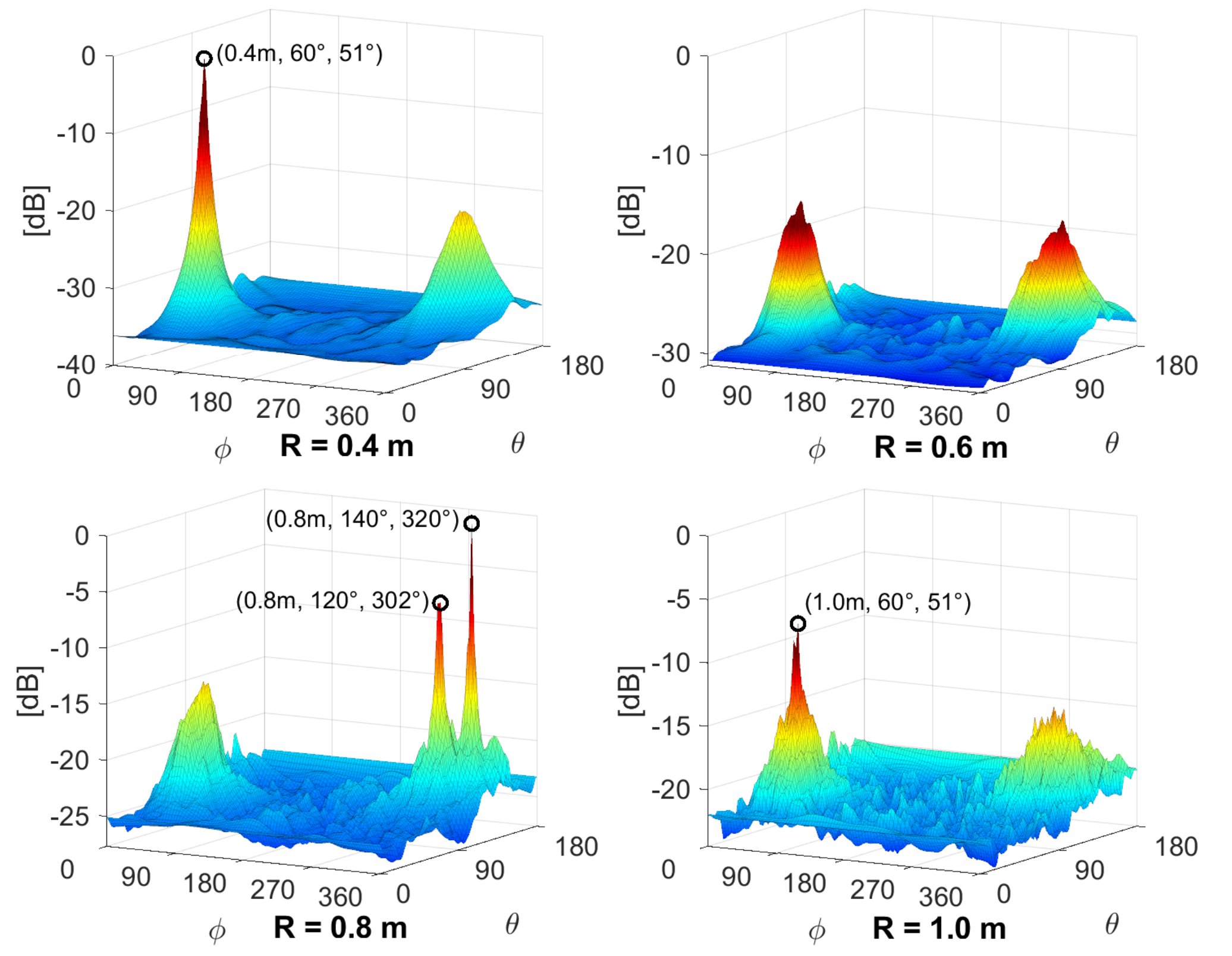

Fig.4 MUSIC spectra of proposed method

- Radial focusing confirms no sources at 0.6 m.
- Able to uniquely distinguish nearby sources 2 & 3.
- Can radially separate and identify same angular sources 1 & 4.

Robustness Against Reflection

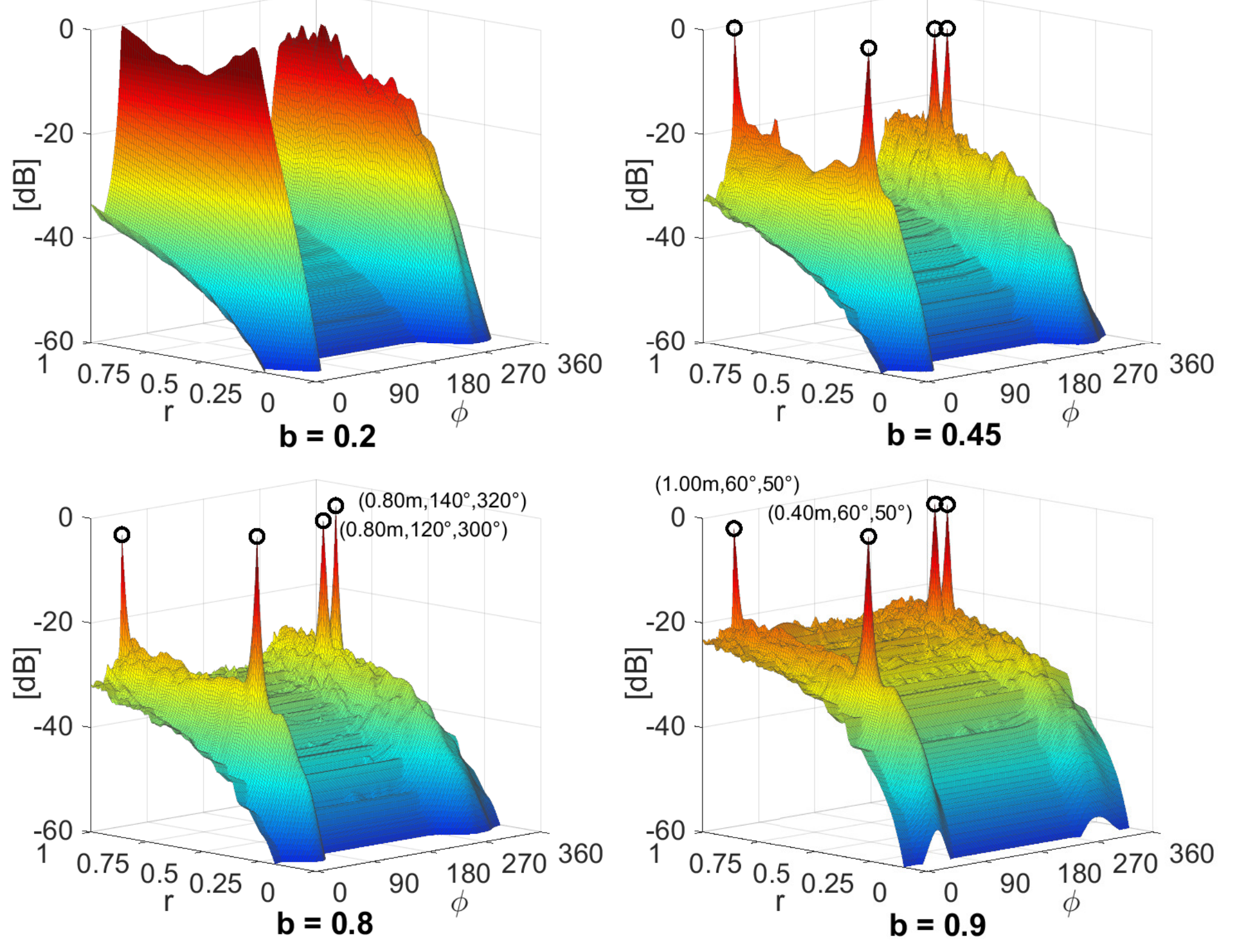

Fig.5 Proposed MUSIC spectra vs reflection coefficient $\boldsymbol{b}$

- Radial focusing improves with the reflection level.
- Sources are localized in highly reflective environments, $\boldsymbol{b} = 0.9$

[1] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, 1986.
[2] O. Nadiri and B. Rafaely, "Localization of multiple speakers under high reverberation using a spherical microphone array and the direct-path dominance test," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 22, no. 10, pp. 1494–1505, 2014.
[3] P. N. Samarasinghe, T. D. Abhayapala, M. Poletti, and T. Betlehem, "An efficient parameterization of the room transfer function," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 23, no. 12, pp. 2217–2227, 2015.
[4] P. N. Samarasinghe, T. D. Abhayapala, Y. Lu, H. Chen, and G. Dickins, "Spherical harmonics based generalized image source method for simulating room acoustics," *J. Acoust. Soc. Amer.*, vol. 144, no. 3, pp. 1381–1391, 2018.