Frank Schultz, Sascha Spors
Institute of Communications Engineering, University of Rostock, Germany
{frank.schultz;sascha.spors}@uni-rostock.de

Traditio et Innovatio

On the detection quality of early room reflection directions using compressive sensing on rigid spherical microphone array data

Introduction

We aim at reliable direction of arrival (DOA) estimation of early room reflections. We compare a traditional plane wave decomposition (PWD) approach against compressive sensing (CS) with orthogonal matching pursuit (OMP). Since rigid, spherical microphone arrays in range of 4-15 cm radius are nowadays available for practical in-situ room measurements, we concentrate on these designs, rather than randomly sampled or larger aperture microphone arrays. The question arises if CS can be profitably deployed for DOA estimation under these circumstances rather than the PWD approach.

Method

Evaluation is performed by a specific example, cf. [1], using image source model simulations of the early part of room impulse responses. The room has (width$_x$, length$_y$, height$_z$) = (10, 11, 3) m. Frequency independent absorption coefficient $\alpha = 0.15$ is assumed. A point source is located at $\mathbf{x}_\mathrm{S} = (7, 8, 1.8)^\mathrm{T}$m. The microphone array has $N = 14$ or 170 microphones arranged on a spherical Lebedev grid with radius $R_\mathrm{N} = 0.1$ m centered at the position $\mathbf{x}_\mathrm{R} = (2.1, 4.5, 1.8)^\mathrm{T}$m. Plane wave dictionary entries are modeled by a dense $M = 1202$ spherical Lebedev grid. Thus, the acoustic transfer functions written as sensing matrix $\mathbf{A}_{N\times t\times M}$ models M plane wave incidences to N microphone array receivers. All simulations use sampling frequency $f_\mathrm{s} = 16$ kHz and speed of sound $\mathrm{c} = 343$ m/s.

Input Data The direct sound and 31 reflections from up to 4th order mirror image sources arrive at the microphone array before room's mixing time. Thus, the measurement matrix $\mathbf{y}_{N\times t}$ contains room reflection impulse responses from azimuth ϕ_is / colatitude θ_is impinging onto ϕ_n / θ_n of the rigid microphone array. Uncorrelated white Gaussian noise is—if any—added to simulate sensor noise.

PWD DOA estimation via PWD includes matched filtering in the spherical harmonics transform (SHT) domain, yielding a high, frequency independent white noise gain [1]. The PWD matrices $\mathbf{p}^N_{\phi_\mathrm{m}\times\theta_\mathrm{m}\times t}$ are obtained by inverse SHT for all plane wave dictionary entries ϕ_m / θ_m.

CS-OMP DOA estimation by means of OMP uses implementation in Python's scikit-learn package. A sliding window with length $t_w = 0.75$ ms is utilized to solve $\mathbf{y}_{N\cdot t_w\times 1} = \mathbf{A}_{N\cdot t_w\times M} \cdot \mathbf{x}_{M\times 1}$ for $\mathbf{x}_{M\times 1}$ with 3 non-zero coefficients using the OMP. This seems to be arbitrary, but is linked to the temporal coincidence of certain reflections. The choice is highly dependent on the room characteristics. Then, finding a specific OMP solution for all specific t, the sparse matrix $\mathbf{x}_{M\times t}$ with $3\ M$ non-zero coefficients is obtained. Explicitly rearranging for plane wave incidence angles ϕ_m / θ_m yields the DOA matrices $\mathbf{x}^N_{\phi_\mathrm{m}\times\theta_\mathrm{m}\times t}$ using CS-OMP beamforming.

Results

We aim at direct comparison of both, PWD and CS-based DOA matrices using consistent setups. These raw data could be used to compute DOA estimates by e.g. a blob or peak detection. Since the degree of freedom for variation of parameters is huge, we restrict the discussion to the three prototypical scenarios shown right.

Color In the shown figures, color and size of the scatter circles indicate level. In the top left corner of each subplot two reference circles are included, not to be mistaken with estimated reflections: The large yellow circle indicates the size for 0 dB level, the small black circle -10 dB. Circle's radius in between is varied linearly in terms of dB. A small white dot and a green number right above it indicate the reference position of an image source. Numbering is sorted according to traveling time.

Level Normalization The level is normalized to 0 dB for the direct sound and exhibits a visible range from -10 dB to 1 dB. Under the assumption that mirror image sources / reflections behave like point sources, their amplitude decay $\propto 1/r$ over distance r can be compensated for, actually being exact for images sources of 1st order, only.

Conclusion

We compared DOA by matched filter plane wave decomposition and by compressive sensing using orthogonal matching pursuit with a sliding time window and 3 non-zero coefficients per window for early room reflections acquired by spherical microphone array. By consistent comparison of 3D data over time, the implemented compressive sensing approach not necessarily yields outperforming results compared to traditional plane wave decomposition for the chosen scenario. Especially, when aiming at the exact count and robust DOA of early room reflections, compressive sensing might be not the optimal solution. This is not a flaw of the technique itself, but rather due to violation of prerequisites, such as low mutual coherence of the sensing matrix $\mathbf{A}$ and signal sparsity in the problem under discussion. For both, degrees of freedom in changing parameters is rather low.

References

[1] Spors, S.; Rettberg, T. (2018): On the Estimation of Acoustic Reflection Coefficients from In-Situ Measurements using a Spherical Microphone Array. *Proc. of 44th DAGA*, Munich, 1326-1329.

[2] Xenaki, A.; Gerstoft, P.; Mosegard, K. (2014): Compressive beamforming. *J. Acoust. Soc. Am.* **136**(1), 260-271.

[3] Mignot, R.; Daudet, L.; Ollivier, F. (2013): Room Reverberation Reconstruction: Interpolation of the Early Part Using Compressed Sensing. *IEEE Trans. Audio, Speech, Language Process.* **21**(11), 2301-2312.

[4] Verburg, S.A.; Fernandez-Grande, E. (2018): Reconstruction of the sound field in a room using compressive sensing. *J. Acoust. Soc. Am.* **143**(6), 3770-3779.

`https://github.com/spatialaudio/doa-early-room-reflections-pwd-vs-omp`

PWD vs. CS-OMP

All figures depict early room reflections' DOA as azimuth/colatitude over time, with subplots arranged as:

Top: PWD $|\mathbf{p}_{\phi_\mathrm{m}\times\theta_\mathrm{m}\times t}|$ in dB, Bottom: CS-OMP $|\mathbf{x}_{\phi_\mathrm{m}\times\theta_\mathrm{m}\times t}|$ in dB

Left: azimuth ϕ_m over time t, Right: colatitude θ_m over time t

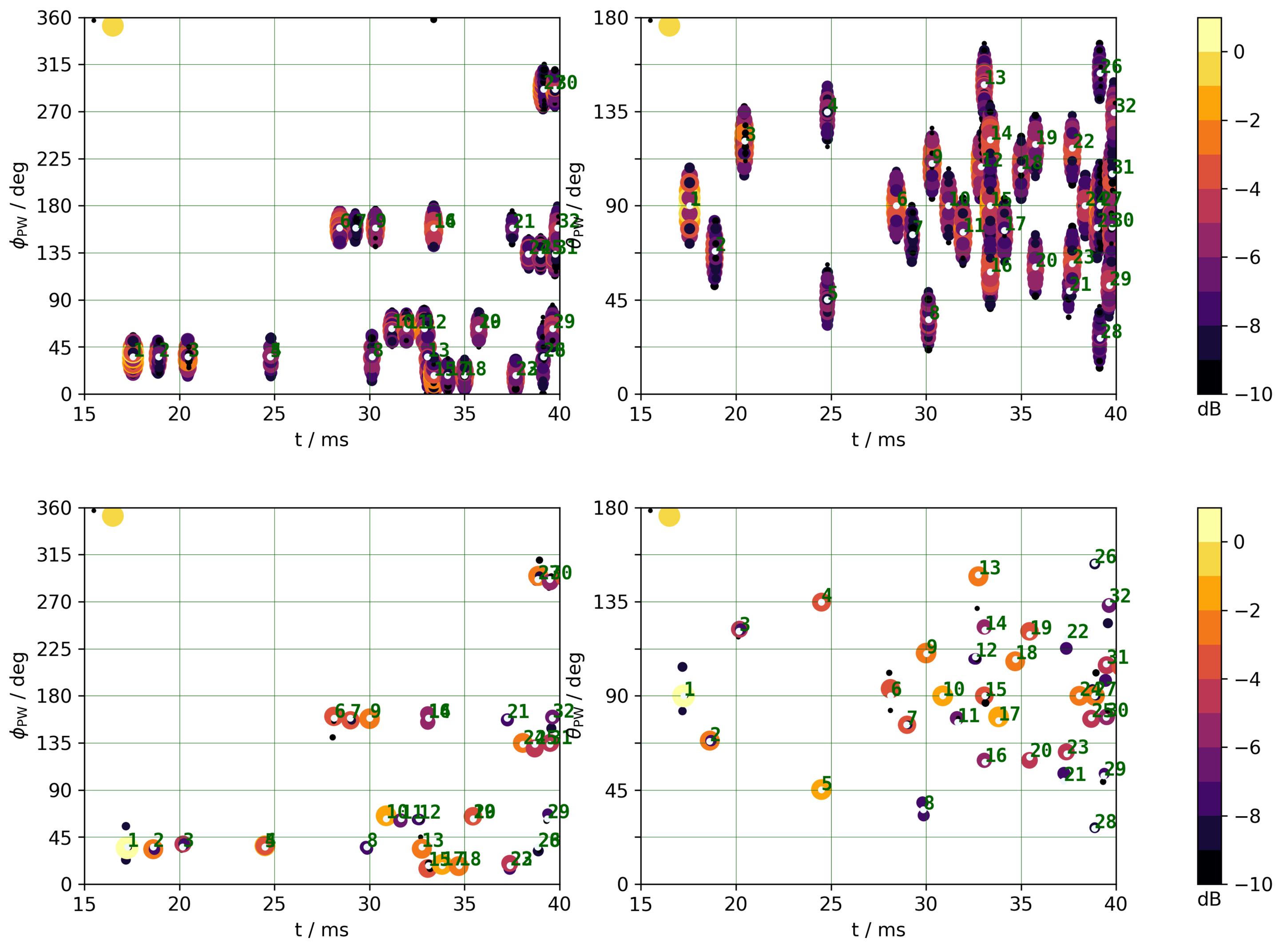

Figure 1: 170 microphones, no additive noise.

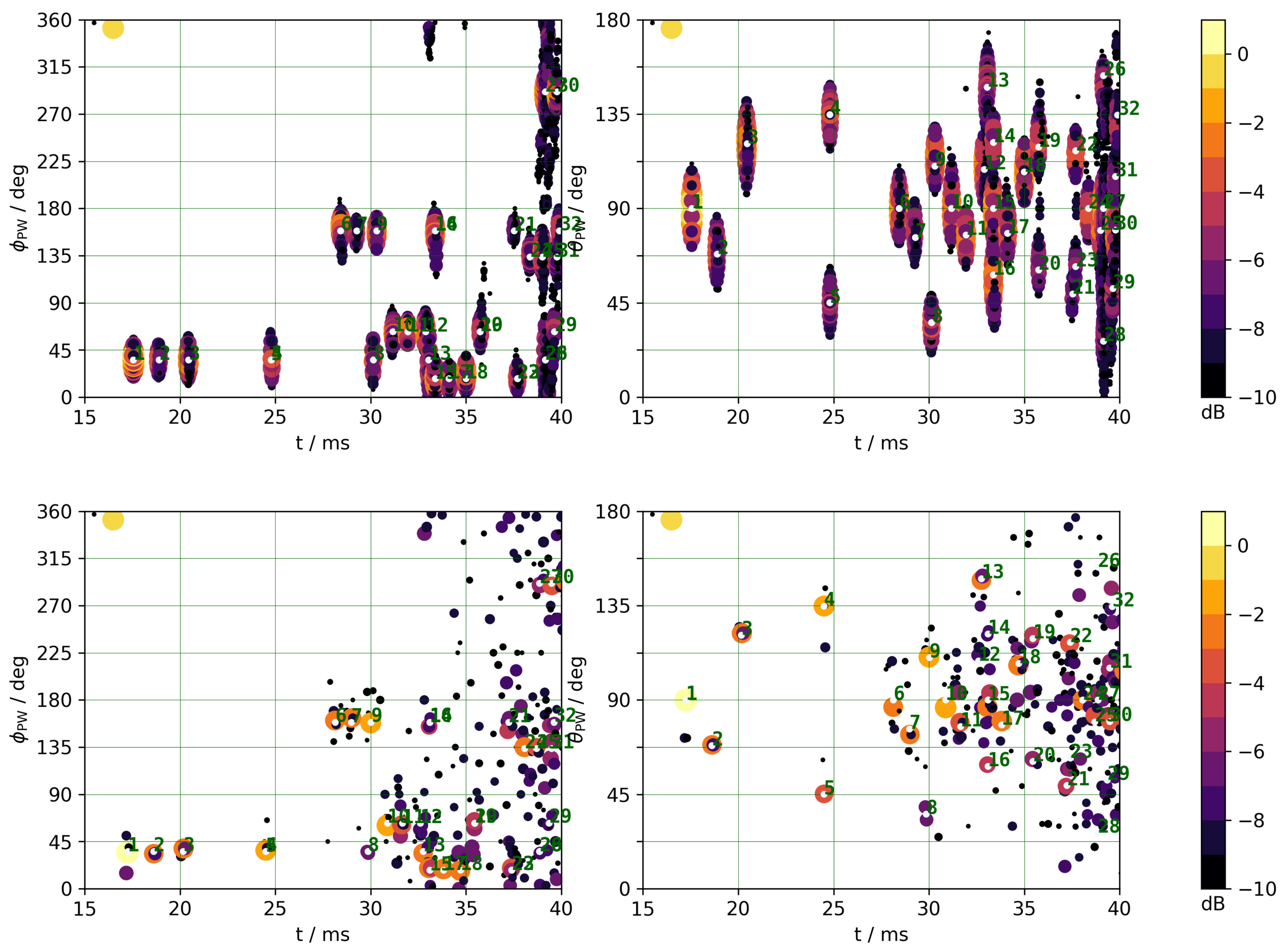

Figure 2: 170 microphones, 10 dB SNR.

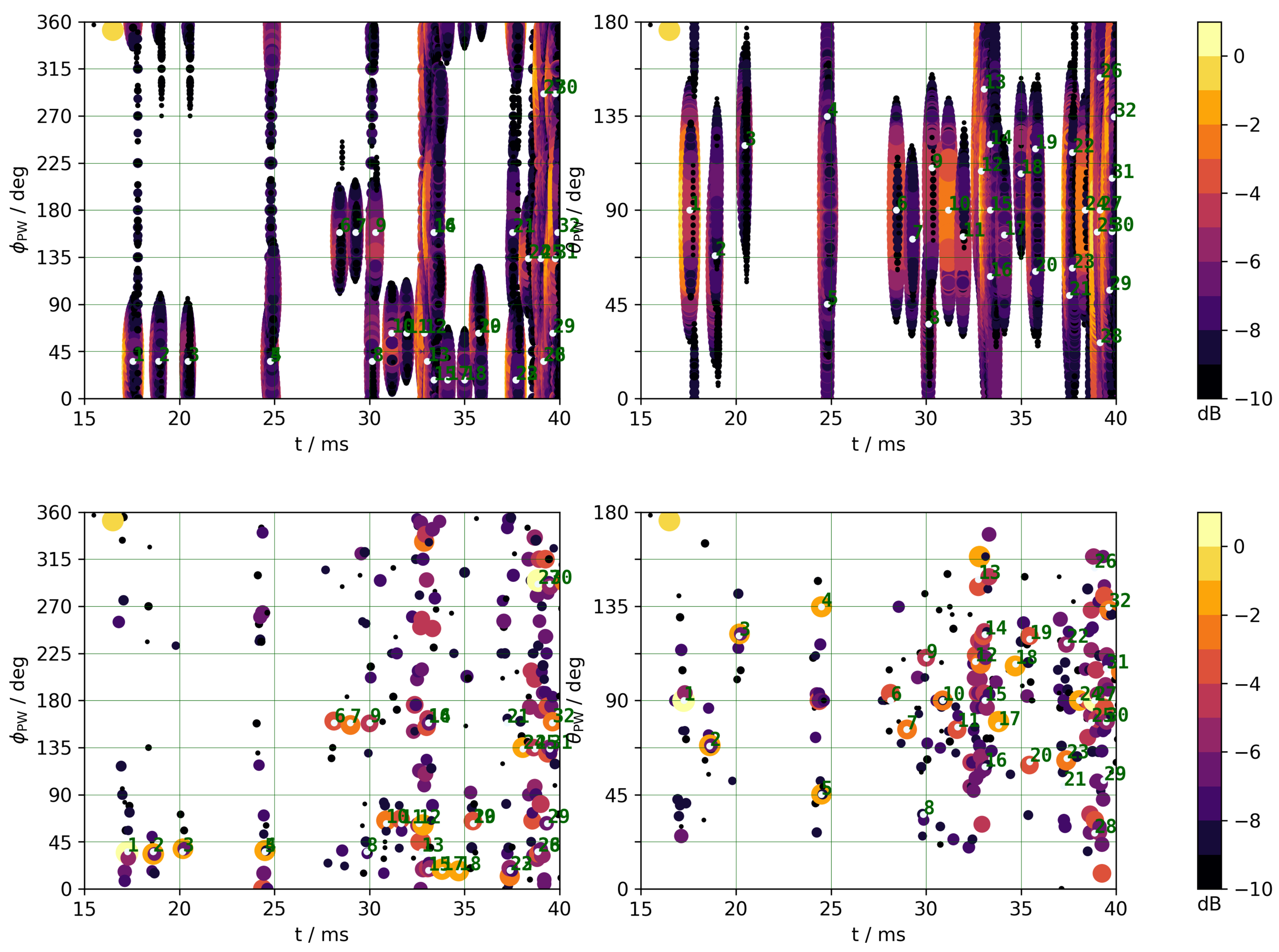

Figure 3: 14 microphones, no additive noise.

23rd International Congress on Acoustics, Aachen, Germany, September 2019