TAMPERE UNIVERSITY OF TECHNOLOGY

Time-of-arrival estimation for blind beamforming

Pasi Pertilä, pasi.pertila (at) tut.fi
www.cs.tut.fi/~pertila/
Aki Tinakari, aki.tinakari (at) tut.fi
Tampere University of Technology
Tampere, Finland

Presentation outline

1) Traditional beamforming / beam steering
2) Ad-hoc microphone arrays
3) Three ad-hoc array beam steering methods
   – Time-of-Arrival (TOA) based solutions
4) Simulation of TOA accuracy
5) Measurements with an array of smartphones
   – Accuracy of TOA estimation
   – Obtained beamforming quality

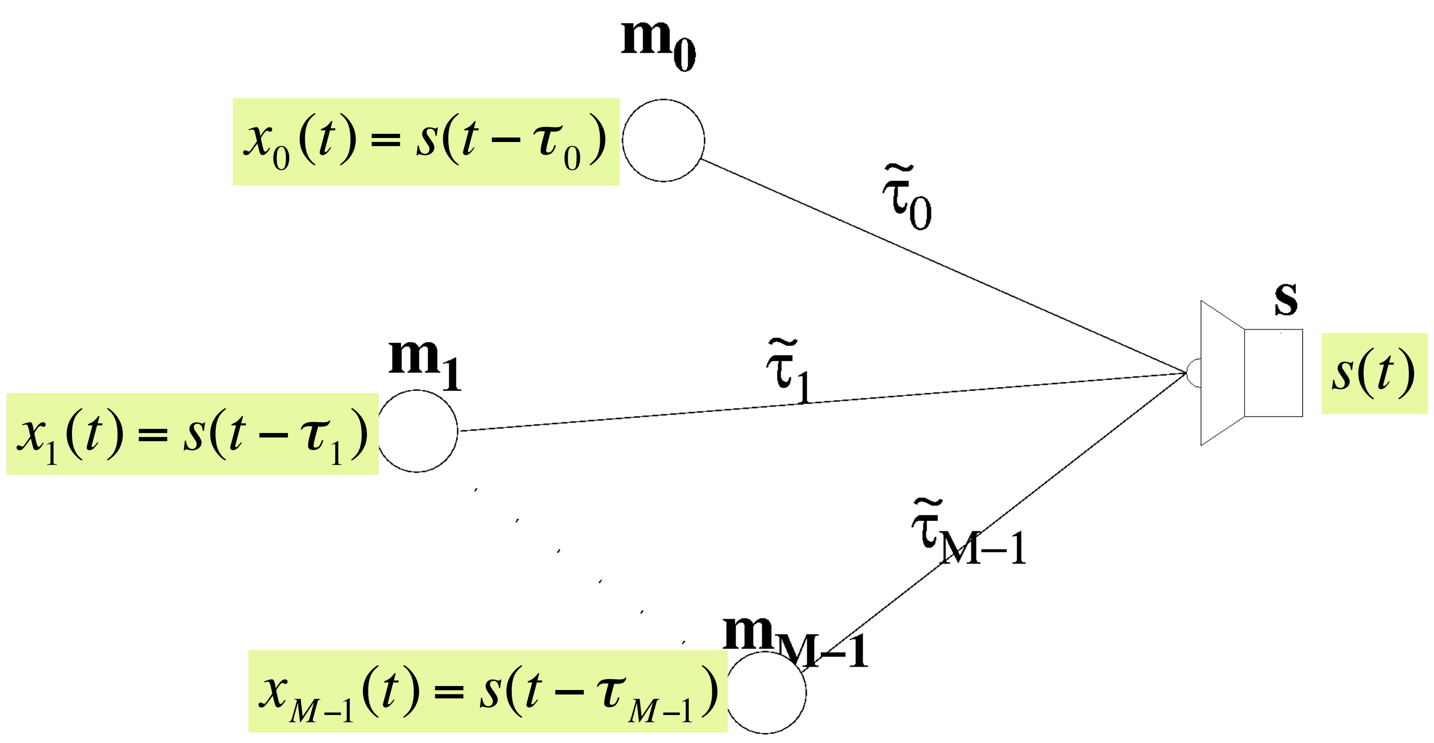
Traditional Beamforming

- Linear combination of microphone signals $X_i(\omega)$, where $i = 1, \ldots, M$
- Requirements for steering the beam:
  1) Array shape is known (mic. position matrix **M**)
  2) Sensors are synchronous (time offset is zero/known)
  3) Direction/position to steer the array is known or can be scanned e.g. based on energy.
- Simple Delay-and-Sum Beamformer **(DSB)**

$$Y(\omega) = \sum_{i=0}^{M-1} \underbrace{\exp(i\omega\tau_i)}_{\text{time-shifting}} X_i(\omega)$$

Signal observation (near field)

- Sound Time-of-Flight (TOF) is $\tilde{\tau}_i = \|\mathbf{m}_i - \mathbf{s}\| c^{-1}$
- Align signals by advancing $x_i(t)$ by $\tilde{\tau}_i$



Ad-Hoc microphone array

- Independent devices equipped with a microphone
- Traditional beamforming requirements unfulfilled
  1. Array geometry is unknown ($\mathbf{M}$ is unknown)
  2. Devices aren't synchronized (unknown time offsets Δ_i)
  3. The space cannot be easily panned to find source direction θ to steer the beam into

Time of Arrival (TOA)

- Signal time-of-arrival (TOA) for and ad-hoc array

$$\tau_i = \underbrace{c^{-1}\left\|\mathbf{s}-\mathbf{m}_i\right\|}_{\text{propagation delay}} + \underbrace{\Delta_i}_{\text{time offset}}$$

- Time-difference-of-Arrival (TDOA) for mics i, j

$$\tau_{i,j} = \tau_i - \tau_j$$

- TDOAs $\tau_{i,j}$ can be measured using e.g. correlation
- Previously considered as source spatial information

A. Brutti and F. Nesta, "Tracking of multidimensional TDOA for multiple sources with distributed microphone pairs," Computer Speech & Language, vol. 27,

- TDOA and TOA vectors are written as

$$\mathbf{y} = [\tau_{1,2}, \tau_{1,3}, \ldots \tau_{M-1,M}]^{\mathrm{T}}, \ \mathbf{y} \in \mathbb{R}^{P\times 1}$$

$$\boldsymbol{\tau} = [\tau_1, \tau_2, \tau_3, \ldots, \tau_M]^{\mathrm{T}}, \ \boldsymbol{\tau} \in \mathbb{R}^{M\times 1}$$

$P=M(M-1)/2$

Time of Arrival (TOA)

- By defining an observation matrix

$$\mathbf{H} = [\ \mathbf{e}_1 - \mathbf{e}_2, \mathbf{e}_1 - \mathbf{e}_3, \ldots, \mathbf{e}_1 - \mathbf{e}_M, \mathbf{e}_2 - \mathbf{e}_3, \ldots,$$
$$\mathbf{e}_2 - \mathbf{e}_M, \ldots, \mathbf{e}_{M-1} - \mathbf{e}_M]^{\mathrm{T}},\ \mathbf{H} \in \mathbb{R}^{P \times M},$$

  – E.g. for three microphones $\mathbf{H} = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix}$

- The linear model between TOA and TDOA is

$$\boxed{\mathbf{y} = \mathbf{H}\boldsymbol{\tau}}$$

- TOA proposed as source spatial representation

Time of Arrival (TOA) – 1st

- **Baseline method (TDOA subset)**:

1. Select a reference microphone (e.g. 1st mic)
2. Use relative delays $\tau_{i,j}$ between the reference $(i = 1)$ and rest $(j = 2, \ldots, M)$ as TOA

$$\boxed{\hat{\boldsymbol{\tau}} = [0, \tau_{1,2}, \tau_{1,3}, \ldots, \tau_{1,M}]^{\mathrm{T}}, \hat{\boldsymbol{\tau}} \in \mathbb{R}^{M}.}$$

– Does not utilize TDOA information between all sensors

Time of Arrival (TOA) – 2nd

- Moore-Penrose inverse solution for TOA

$$\boxed{\hat{\tau}_0 = (\mathbf{H}_0^T \mathbf{H}_0)^{-1} \mathbf{H}_0^T \mathbf{y}.}$$

- $\mathbf{H}_0$ is $\mathbf{H}$ without the first column to account for one missing degree of freedom, i.e. the TOA is relative to 1st sensor (which is set to zero).

+ Utilizes TDOA information between all sensors

Time of Arrival (TOA) – 3rd

- Kalman filtering based TOA estimation

$$\boxed{\begin{aligned}\mathbf{x}_t &= \mathbf{A}\mathbf{x}_{t-1} + \mathbf{q}_t, \\ \mathbf{y}_t &= \mathbf{H}_0\mathbf{x}_t + \mathbf{r}_t,\end{aligned}}$$

(state eq.)

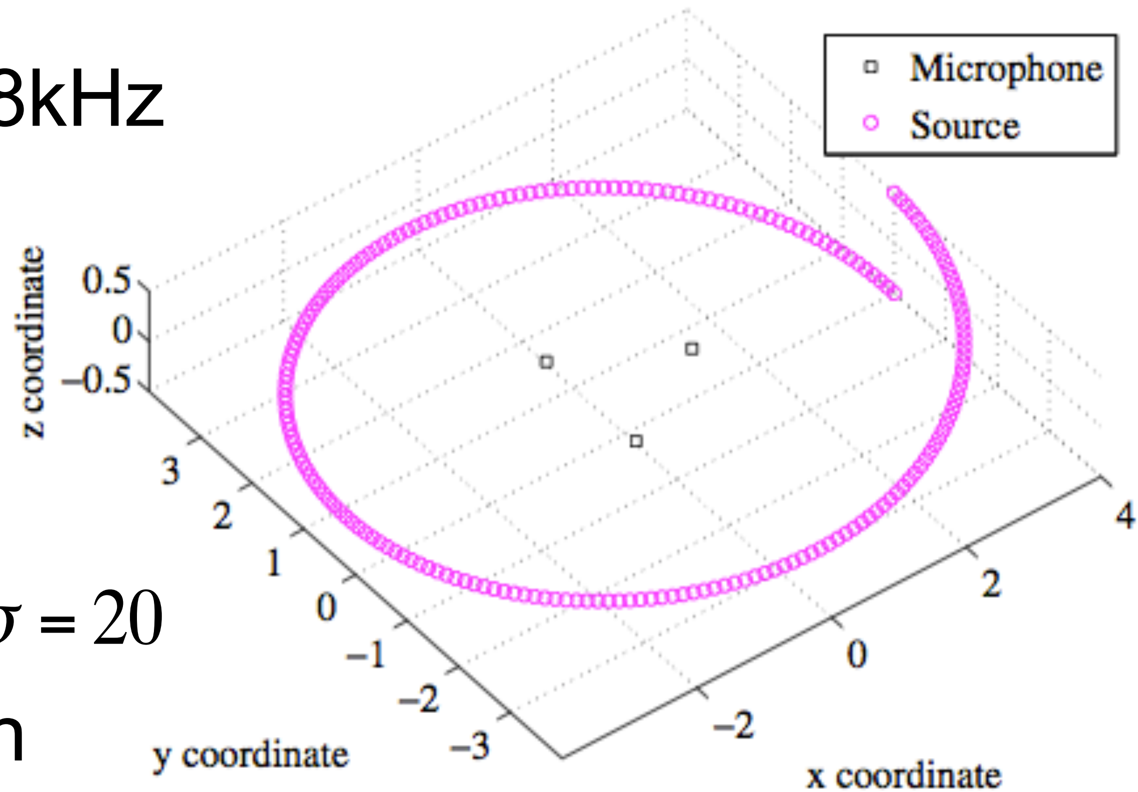
(measurement eq.)

  – x consists of TOA and TOA velocity, $\mathbf{x} = \begin{bmatrix} \tau \\ \dot{\tau} \end{bmatrix}$
  – $\mathbf{A}$ is transition matrix, $\mathbf{q}$, $\mathbf{r}$ are noise
  – Predict $p(x_t|y_{t-1})$ and update $p(x_t|y_t)$ steps.
  – Outlier rejection based on projected measurement likelihood $p(\mathbf{y}_t|\mathbf{x}_t, \text{target present}) = N(\mathbf{y}_t|\mathbf{H}\mathbf{x}_t, \mathbf{R})$

+ Utilizes TDOA information between all pairs
+ Can track speaker during noise contaminated segments.

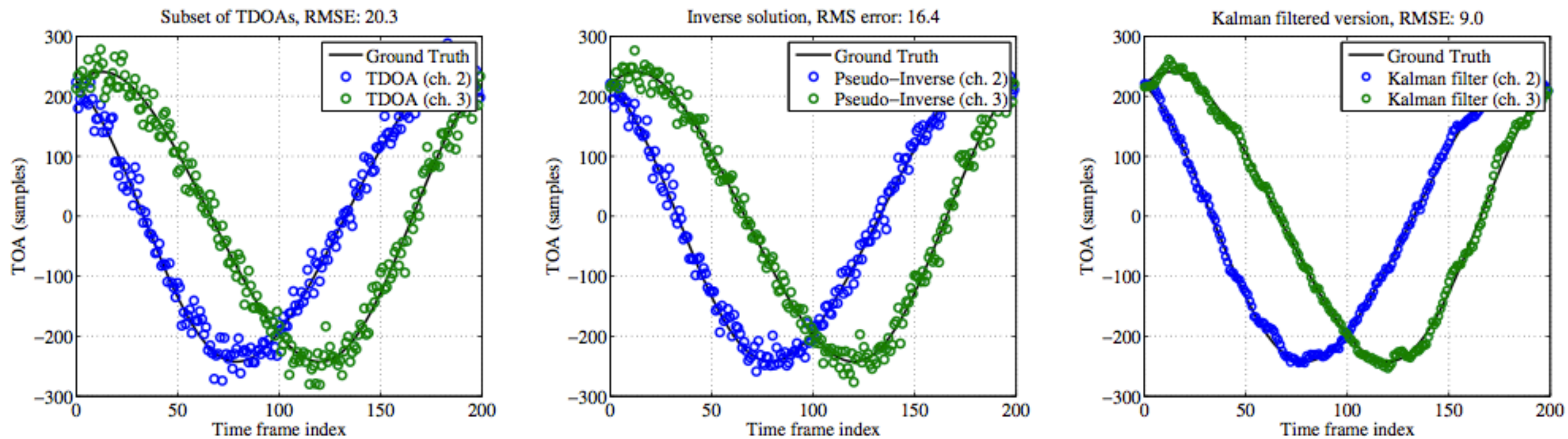
TOA Estimation simulation

- 3 microphones 48kHz
- Source rotates around the array
- Gaussian noise added to TDOA observations τ_{ij}, $\sigma = 20$
- Gaussian noise in offset values Δ_i, $\sigma^2 = 10$

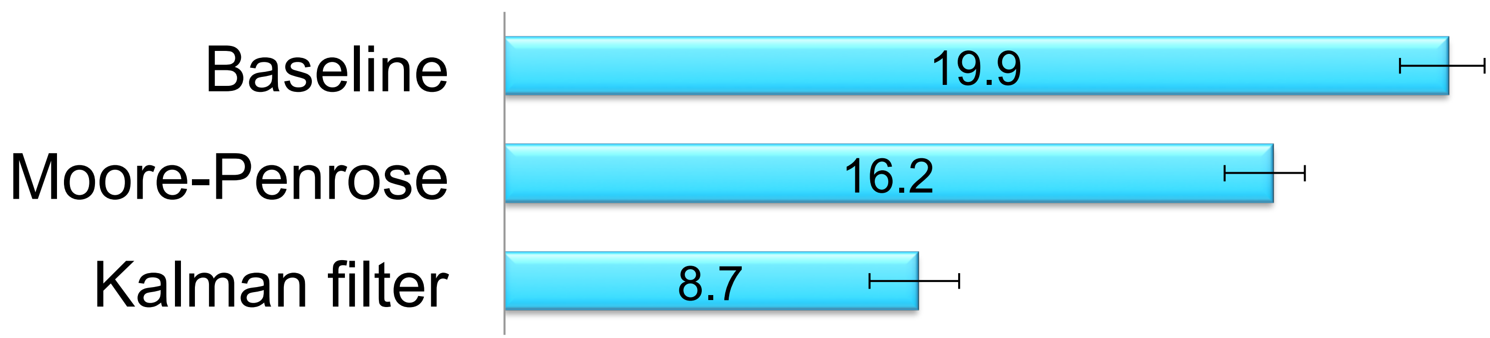


Simulation – TOA accuracy

Baseline (subset of TDOAs) | Moore-Penrose Inverse | Kalman filter



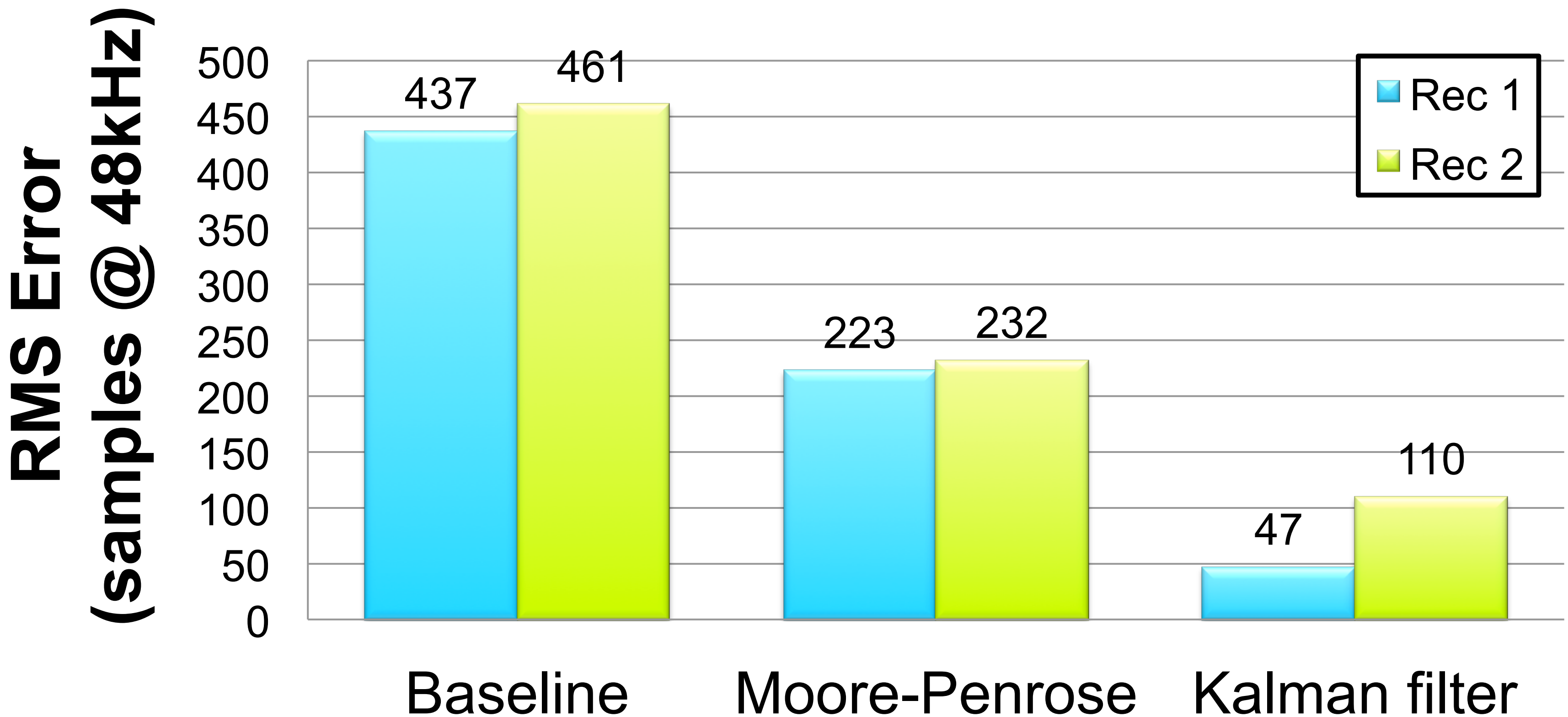
TOA RMS error (samples@48k, 100 trials)



Measurements

- 10 smartphones were used to capture audio
- 9 and 12 second sentences were used
  - Speaker walked around the array
- Reverberation time T60 ~ 370 ms
- Room size: 5.1m × 6.6m
- TDOAs were manually annotated to obtain ground truth TOA.
- Reference signal was captured with headworn microphones.

Performance of TOA estimators in measurements



Obtained beamforming quality

- We used estimated TOAs to steer DSB
- Output *y(t)* quality was evaluated with BSS-metric “Signal-to-Artifacts-Ratio” or SAR*)

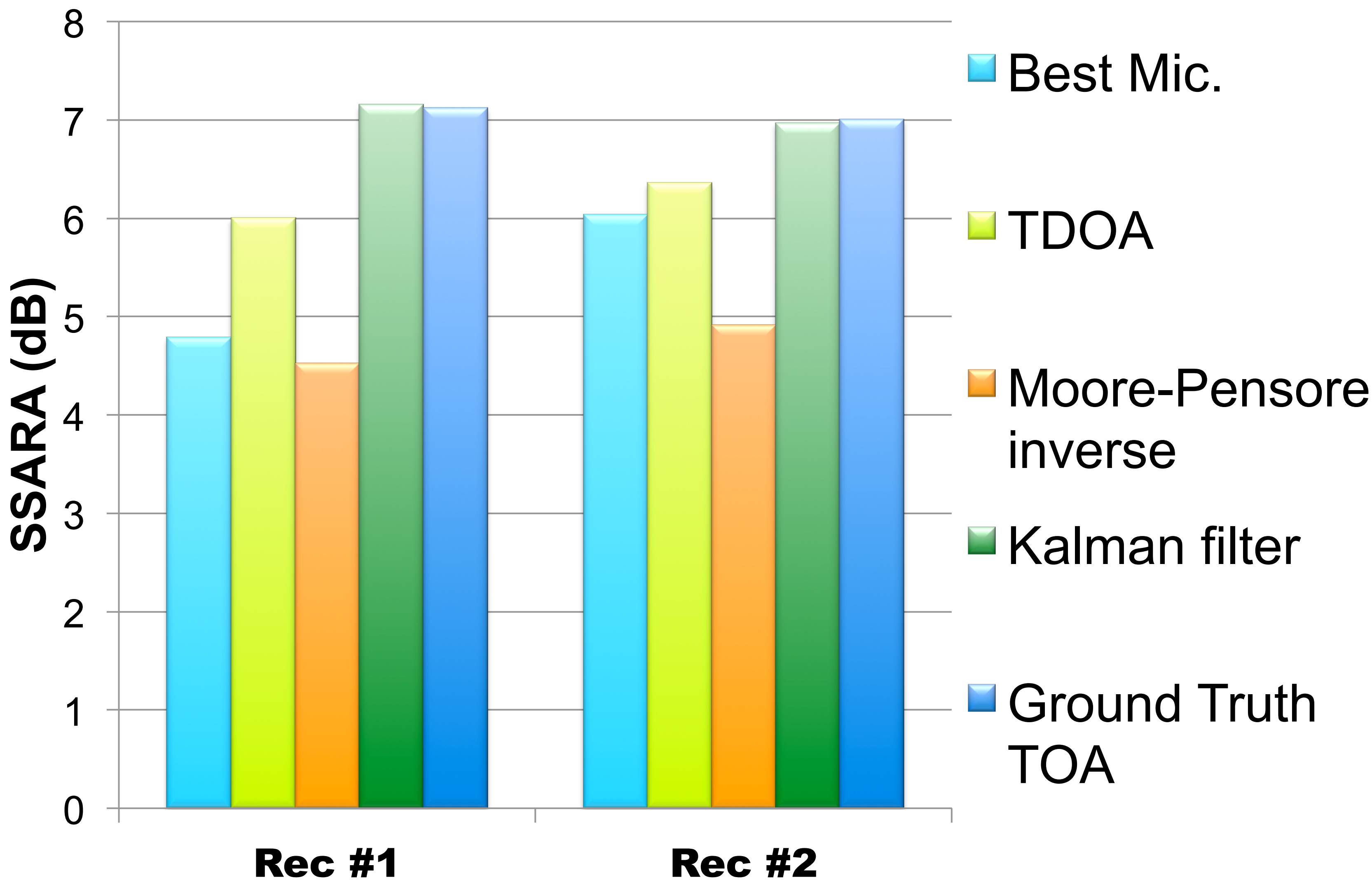
$$\mathrm{SAR} = 20\log_{10}\left(\left\|s_{\mathrm{target}}\right\| / \left\|e_{\mathrm{artifacts}}\right\|\right)$$

$$y(t) = s_{\mathrm{target}}(t) + e_{\mathrm{artifacts}}(t)$$

  – Scored in segments due to speaker movement (gain variation)
  – Only active segments considered (with VAD)
  – Modified metric: Segmental Signal-to-Artifacts Ratio Arithmetic mean (SSARA)

*) http://bass-db.gforge.inria.fr/bss eval/

Objective speech quality



Conclusions

- Proposed TOA as the spatial source information of an ad-hoc microphone array
  - Previous research only considered TDOA
  - Dimension of TOA is $M-1$, for TDOA $M(M-1)/2$
- Three TOA estimation solutions considered
  - TDOA subset (baseline), pseudo-inverse, and Kalman filtering → most accurate
- TOA allows beam-steering towards source
  - w/o mic. positions / synchronization: blindly
  - Kalman filter based TOA provided best objective signal quality for beamforming