DOA Estimation of Audio Sources in Reverberant Environments

ICASSP, 20–25 March 2016

Jesper Rindom Jensen[*,1], Jesper Kjær Nielsen[2,3], Richard Heusdens[1,4], and Mads Græsbøll Christensen[1]

[*]jrj@create.aau.dk

[1]Audio Analysis Lab, AD:MT Aalborg University Denmark

[2]Dept. Electronic Systems Aalborg University Denmark

[3]Bang & Olufsen A/S Denmark

[4]Dept. Microelectronics TU Delft Netherlands

Partly funded by the Danish Council for Independent Research, grant ID: DFF – 1337-00084, InnovationsFonden, and the Villum Foundation.

AALBORG UNIVERSITY
DENMARK

Agenda

Introduction

Signal Model

Proposed DOA Estimators
- Unstructured amplitudes
- Structured amplitudes

Experimental Results

Conclusions

Introduction

- DOA of audio/speech useful for, e.g., surveillance and beamforming.
- Reverberation have a detrimental impact on estimation.
- Most existing DOA estimator do not (explicitly) account for reverberation.
- Performance with reverberation is therefore limited.
- Some methods (e.g., SRP-PHAT) are relatively robust against reverb without accounting for it directly.
- We propose reverb robust DOA estimators based on simple reverb model.
- **Model:** direct-path + early reflections + noise.

Signal Model

An acoustic source is sampled using a microphone array:

$$y_k(n) = (s' * g_k)(n) + v_k'(n) = s_k(n) + v_k'(n), \quad (1)$$

where

$s'(n)$: clean source signal

$g_k(n)$: room impulse response from source to mic k

$v_k'(n)$: additive noise (interferers, sensor noise, etc.)

Remarks:

- Focus on reverb robust DOA estimation.
- Noise, $v_k'(n)$ assumed white Gaussian.

Vector Model

With K microphones recording N samples each, we get

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1^T & \mathbf{y}_2^T & \cdots & \mathbf{y}_K^T \end{bmatrix}^T = \mathbf{s} + \mathbf{v}'. \quad (2)$$

where

$$\mathbf{y}_k = \begin{bmatrix} y_k(0) & \cdots & y_k(N-1) \end{bmatrix}^T$$

$\mathbf{s}$ & $\mathbf{v}'$: desired signal and noise vectors (defined as $\mathbf{y}$)

Further model specifications:

- desired signal assumed quasi-periodic,
- a ULA structure is assumed.

Periodic Signal Model

Clean desired signal modeled as

$$s'(n) = \sum_{l=-L}^{L} \alpha_l e^{jl\omega_0 n}, \tag{3}$$

with

α_l: complex amplitude of l'th harmonic,
ω_0: fundamental frequency [rad/sample],
L: model order, i.e., number of harmonics.

Important observation:

Widely used broadband model is a special case of (3), i.e., for

$$\omega_0 = 2\pi/N \qquad \wedge \qquad L = \lfloor N/2 \rfloor. \tag{4}$$

Array Model

We assume the source of interest to be in the far-field.

For a ULA, TDOA of source r between mic 1 and k is then

$$\tau_{r,k} = k\frac{d\sin\theta_r}{c} = k\eta_r, \tag{5}$$

with

d: spacing between two adjacent mics,

θ_r: DOA of source r,

c: sound propagation speed.

Complete Signal Model

Observation modeled as multiple early reflections in noise:

$$\mathbf{y} = \sum_{r=1}^{R} \mathbf{H}(\eta_r)\alpha_r + \mathbf{v}, \quad (6)$$

where

$$\begin{aligned}
R&: \text{ number of early reflections} \\
\mathbf{H}(\eta_r) &= [\mathbf{Z}^T \; (\mathbf{Z}\mathbf{D}_2(\eta_r))^T \; \cdots \; (\mathbf{Z}\mathbf{D}_K(\eta_r))^T]^T \\
\mathbf{Z} &= \begin{bmatrix} \mathbf{z}_1 & \cdots & \mathbf{z}_L & \mathbf{z}_1^* & \cdots & \mathbf{z}_L^* \end{bmatrix} \\
\mathbf{z}_l &= \begin{bmatrix} 1 & e^{jl\omega_0} & \cdots & e^{j(N-1)l\omega_0} \end{bmatrix}^T \\
\mathbf{D}_k(\eta_r) &= \operatorname{diag}\left(\begin{bmatrix} \mathbf{d}_k^T(\eta_r) & \mathbf{d}_k^H(\eta_r) \end{bmatrix}\right) \\
\mathbf{d}_k(\eta_r) &= \begin{bmatrix} e^{-j\omega_0 k\eta_r} & \cdots & e^{-jL\omega_0 k\eta_r} \end{bmatrix}^T
\end{aligned}$$

Estimation problem: find η_1 from observations!

Reverb Robust DOA Estimation

- We propose two methods for DOA estimation with reverb.
- Idea is to estimate DOAs of both direct-path and early reflections.
- Bias of direct-path estimate reduced in this way.
- Both methods are based on nonlinear least squares:
  1. a method where amplitudes of direct-path and reflections are assumed independent.
  2. a method where the relation between the amplitudes is modeled.
- Estimation of multiple DOAs facilitated by an iterative approach.

Nonlinear Least Squares
Unstructured amplitudes

With unstructured amplitudes, the NLS estimator is

$$\{\widehat{\boldsymbol{\eta}}, \widehat{\overline{\boldsymbol{\alpha}}}\} = \arg \min_{\{\boldsymbol{\eta},\overline{\boldsymbol{\alpha}}\}} \|\mathbf{y} - \overline{\mathbf{H}}(\boldsymbol{\eta})\overline{\boldsymbol{\alpha}}\|_2^2, \tag{7}$$

with

$$\begin{aligned} \boldsymbol{\eta} &= [\eta_1 \ \cdots \ \eta_R]^T \\ \overline{\mathbf{H}}(\boldsymbol{\eta}) &= \begin{bmatrix}\mathbf{H}(\eta_1) & \cdots & \mathbf{H}(\eta_R)\end{bmatrix} \\ \overline{\boldsymbol{\alpha}} &= \begin{bmatrix}\boldsymbol{\alpha}_1^T & \cdots & \boldsymbol{\alpha}_R^T\end{bmatrix}^T \end{aligned}$$

Solving for $\overline{\boldsymbol{\alpha}}$ gives

$$\widehat{\boldsymbol{\eta}} = \arg\min_{\boldsymbol{\eta}} \Big\| \big(\underbrace{\mathbf{I} - \overline{\mathbf{H}}(\boldsymbol{\eta})(\overline{\mathbf{H}}(\boldsymbol{\eta})^H \overline{\mathbf{H}}(\boldsymbol{\eta}))^{-1} \overline{\mathbf{H}}(\boldsymbol{\eta})^H}_{\mathbf{P}^{\perp}_{\overline{\mathbf{H}}(\boldsymbol{\eta})}} \big) \mathbf{y} \Big\|_2^2. \tag{8}$$

Iterative Procedure
Unstructured amplitudes

Consider a modified observed signal model:

$$\mathbf{y}_r = \mathbf{y} - \sum_{q=1, q \neq r}^{R} \mathbf{H}(\widehat{\eta}_q)\widehat{\alpha}_q, \tag{9}$$

This suggests:

$$\widehat{\alpha}_r = (\mathbf{H}^H(\eta_r)\mathbf{H}(\eta_r))^{-1}\mathbf{H}(\eta_r)^H\mathbf{y}_r, \tag{10}$$

$$\widehat{\eta}_r = \arg\min_{\eta_r} \|\mathbf{P}^{\perp}_{\mathbf{H}(\eta_r)}\mathbf{y}_r\|_2^2. \tag{11}$$

This enables iterative DOA estimation [Li&Stoica,1996], termed RNLS.

Algorithm
Unstructured amplitudes

Step (1): Assume $R = 1$. Estimate η_1 and α_1 from $\mathbf{y}_1 = \mathbf{y}$ as described before.

Step (2): Assume $R = 2$. Estimate η_2 and α_2 from $\mathbf{y}_2$ using parameter estimates from Step (1). Re-estimate η_1 and α_1 from $\mathbf{y}_1$. Iterate until “practical convergence”.

Step (3): Assume $R = 3$. Estimate η_3 and α_3 from $\mathbf{y}_3$ using parameters from Step (2). Re-estimate η_1 and α_1 from $\mathbf{y}_1$. Re-estimate η_2 and α_2 from $\mathbf{y}_2$. Iterate until “practical convergence”.

Remaining steps: Continue similarly to the previous steps until R is equal to the number of early reflections.

Nonlinear Least Squares
Structured amplitudes

An alternative model with amplitude relations can be formulated

$$\mathbf{y} = \sum_{r=1}^{R} \gamma_r \mathbf{H}(\eta_r)\mathbf{T}_r\boldsymbol{\alpha} + \mathbf{v}, \tag{12}$$

where

γ_r: attenuation of reflection r ($\gamma_1 = 1$)

η_r: delay of reflection r ($\eta_1 = 0$)

$\boldsymbol{\alpha}$: direct-path harmonic amplitudes

$$\mathbf{T}_r = \text{diag}\left(\begin{bmatrix}\mathbf{t}_r^T & \mathbf{t}_r^H\end{bmatrix}\right)$$

$$\mathbf{t}_r = \begin{bmatrix}e^{j\omega_0\xi_r} & \cdots & e^{jL\omega_0\xi_r}\end{bmatrix}^T$$

Iterative Procedure

Structured amplitudes

Again, consider a modified observed signal model:

$$\mathbf{y}_r = \mathbf{y} - \sum_{q=1, q\neq r}^{R} \widehat{\gamma}_q \mathbf{H}(\widehat{\eta}_q)\widehat{\boldsymbol{\alpha}} \tag{13}$$

With this, LS amplitudes and attenuations estimates are

$$\widehat{\boldsymbol{\alpha}} = [\mathbf{H}^H(\eta_1)\mathbf{H}(\eta_1)]^{-1}\mathbf{H}^H(\eta_1)\mathbf{y}_1 \quad (r=1) \tag{14}$$

$$\widehat{\gamma}_r = \frac{\text{Re}\{\widehat{\boldsymbol{\alpha}}^H \mathbf{T}_r^H \mathbf{H}^H(\eta_r)\mathbf{y}_r\}}{\widehat{\boldsymbol{\alpha}}^H \mathbf{T}_r^H \mathbf{H}^H(\eta_r)\mathbf{H}(\eta_r)\mathbf{T}_r\widehat{\boldsymbol{\alpha}}} \quad (r = 2, \ldots, R). \tag{15}$$

Iterative Procedure

Structured amplitudes

DOA of direct-path is then estimated by ($r = 1$)

$$\widehat{\eta}_1 = \arg\min_{\eta_1} \|\mathbf{P}^{\perp}_{\mathbf{H}(\eta_1)}\mathbf{y}_1\|_2^2. \tag{16}$$

Early reflection DOAs and delays estimated jointly ($r = 2, \ldots, R$)

$$\{\widehat{\eta}_r, \widehat{\xi}_r\} = \arg\min_{\eta_r, \xi_r} \|\mathbf{y}_r - \widehat{\gamma}_r \mathbf{H}(\eta_r)\mathbf{T}_r\widehat{\alpha}\|_2^2. \tag{17}$$

This method is termed RNLS-S.

Remarks

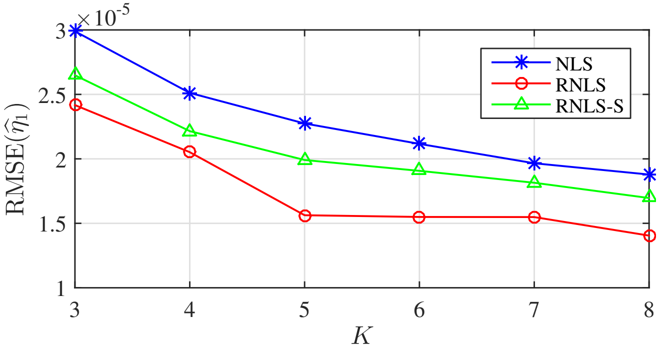
- Implemented using iterative procedure as for RNLS.
- More complex (2d estimation for reflections), but more realistic.

Experimental Results
Synthetic data

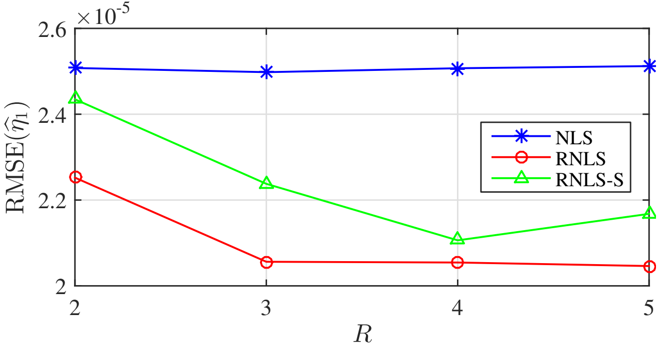
- Evaluated the method on synthetic data.
- Setup:
  - $f_0 = 255.2$ Hz, $f_s = 8$ kHz
  - $L = 6$ (unit amplitude + random phase)
  - f_0 assumed known
  - signal synthesized spatially using RIR generator
  - $d = 0.05$ cm, SNR$=$ 40 dB, $N = 200$
  - source DOA varied $(-80^\circ, -75^\circ, \ldots, 80^\circ)$
  - source-array distance: 2.5 m.
- Average results depicted to the right.

Experimental Results

Synthetic data

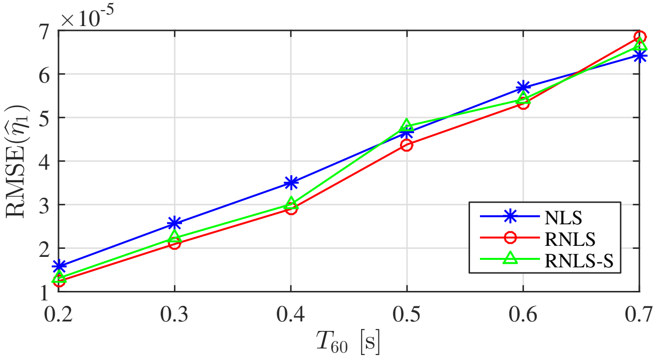


Experimental Results
Synthetic data



Experimental Results

Synthetic data



Experimental Results
Real data

- Also evaluated on a real and moving speech source.
- Four seconds of female speech used (synthesized spatially using RIR generator).
- Pitch and model order estimated using an NLS estimator [Christensen,2009].
- Setup: $R = 4$, $T_{60} = 0.3$ s, $K = 4$.

NLS	RNLS	RNLS-S	SRP-PHAT
$3.8 \cdot 10^{-5}$	$3.6 \cdot 10^{-5}$	$3.6 \cdot 10^{-5}$	$5.4 \cdot 10^{-5}$

Conclusions

- Considered DOA estimation of audio/speech with reverb.
- Proposed NLS estimator based on model of reverbed signal (direct-path + early refl. + noise).
- DOA of direct-path and early reflections estimated iteratively.
- Contributions from reflections subtracted when localizing direct-path $\rightarrow$ reduces estimation bias.
- Experiments confirm proposed method can outperform state of the art in terms of RMSE on both synthetic and real data.
- May be used in room geometry estimation, as DOAs of early reflections are also estimated.