Echo-aware signal processing for audio scene analysis

Diego Di Carlo

November 29, 2020

PhD supervisors: Antoine Deleforge
Nancy Bertin

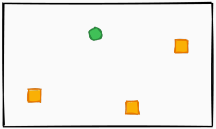
Jury members: Laurent Girin (reviewer - president)
Simon Doclo (reviewer)
Fabio Antonacci (examiner)
Renaud Seguier (examiner)

Université de Rennes 1, IRISA/INRIA, Panama research group

Acoustic Echo Estimation

State of the Art

Key ingredient – *Cross relation identity*



$$\tilde{x}_1 = \tilde{h}_1 * \tilde{s}$$

$$\tilde{h}_2 * \tilde{x}_1 = \tilde{h}_2 * \tilde{h}_1 * \tilde{s} = \tilde{h}_1 * \tilde{h}_2 * \tilde{s} = \tilde{h}_1 * \tilde{x}_2$$

Ideas:

1. Sampled version of $\tilde{x}_1, \tilde{x}_2$ are available: x_1, x_2
2. Echo TOAs $\propto$ sampling frequency
3. Find echoes $\rightarrow$ **find sparse vectors** h_1, h_2 of length L
4. Modeled as **Lasso**-like problem

$= \texttt{Toeplitz}(x_i)h_j \in \mathcal{O}(L^2)$

$$\hat{h}_1, \hat{h}_2 \in \underset{h_1, h_2 \in \mathbf{R}^n}{\arg\min} \; \|x_1 * h_2 - x_2 * h_1\|_2^2 + \lambda \mathcal{P}(h_1, h_2) \quad \text{s.t.} \quad \mathcal{C}(h_1, h_2)$$

$\mathcal{P}(h_1, h_2) \longrightarrow$ sparse promoting regularizer

$\mathcal{C}(h_1, h_2) \longrightarrow$ constraints e.g. nonnegativity anchor

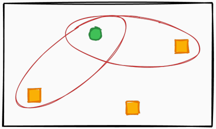
✓ [Tong et al., 1994] ✓ [Lin et al., 2008] ✓ [Aissa-El-Bey and Abed-Meraim, 2008]
✓ [Kowalczyk et al., 2013] ✓ [Crocco and Del Bue, 2016]

State of the Art

Key ingredient – *Cross relation identity*

$$\tilde{x}_1 = \tilde{h}_1 * \tilde{s}$$

$$\tilde{h}_2 * \tilde{x}_1 = \tilde{h}_2 * \tilde{h}_1 * \tilde{s} = \tilde{h}_1 * \tilde{h}_2 * \tilde{s} = \tilde{h}_1 * \tilde{x}_2$$



Ideas:

1. Sampled version of $\tilde{x}_1, \tilde{x}_2$ are available: x_1, x_2
2. Echo TOAs $\propto$ sampling frequency
3. Find echoes $\rightarrow$ **find sparse vectors** h_1, h_2 of length L
4. Modeled as **Lasso**-like problem

$$= \texttt{Toeplitz}(x_i)h_j \in \mathcal{O}(L^2)$$

$$\hat{h}_1, \hat{h}_2 \in \underset{h_1,h_2 \in \mathbf{R}^n}{\arg\min} \; \|x_1 * h_2 - x_2 * h_1\|_2^2 + \lambda \mathcal{P}(h_1, h_2) \quad \text{s.t.} \quad \mathcal{C}(h_1, h_2)$$

$\mathcal{P}(h_1, h_2) \longrightarrow$ sparse promoting regularizer $\quad \mathcal{C}(h_1, h_2) \longrightarrow$ constraints e.g. nonnegativity anchor

✓ [Tong et al., 1994] ✓ [Lin et al., 2008] ✓ [Aissa-El-Bey and Abed-Meraim, 2008]
✓ [Kowalczyk et al., 2013] ✓ [Crocco and Del Bue, 2016]

Proposed approach: analytical & off-grid

Observation 1: the cross relation remains true in the frequency domain

$$\mathcal{F}x_1 \cdot \mathcal{F}h_2({}^{n}/_{F_s}) = \mathcal{F}x_2 \cdot \mathcal{F}h_1({}^{n}/_{F_s}) \qquad n = 0 \dots N-1$$

Proposed approach: analytical & off-grid

Observation 1: the cross relation remains true in the frequency domain

$$\mathcal{F}x_1 \cdot \mathcal{F}h_2(n/F_s) = \mathcal{F}x_2 \cdot \mathcal{F}h_1(n/F_s) \qquad n = 0 \dots N-1$$

Observation 2: $\mathcal{F}\delta_{\text{echo}}$ **is known in closed-form**

Proposed approach: analytical & off-grid

Observation 1: the cross relation remains true in the frequency domain

$$\mathcal{F}x_1 \cdot \mathcal{F}h_2({}^{n}/{}_{F_s}) = \mathcal{F}x_2 \cdot \mathcal{F}h_1({}^{n}/{}_{F_s}) \qquad n = 0 \ldots N-1$$

Observation 2: $\mathcal{F}\delta_{\text{echo}}$ **is known in closed-form**

Observation 3: $\mathcal{F}\mathrm{x_i}$ **can be (well) approximated by DFT**

$$\mathbf{X}_i = \mathtt{DFT}(x_i) \simeq \mathcal{F}x_i(nF_s) \qquad n = 0 \ldots N-1$$

Proposed approach: analytical & off-grid

Observation 1: the cross relation remains true in the frequency domain

$$\mathcal{F}x_1 \cdot \mathcal{F}h_2(n/F_s) = \mathcal{F}x_2 \cdot \mathcal{F}h_1(n/F_s) \qquad n = 0 \dots N-1$$

Observation 2: $\mathcal{F}\delta_{\text{echo}}$ **is known in closed-form**

Observation 3: $\mathcal{F}\mathrm{x_i}$ **can be (well) approximated by DFT**

$$\mathbf{X}_i = \mathtt{DFT}(x_i) \simeq \mathcal{F}x_i(nF_s) \qquad n = 0 \dots N-1$$

Idea: Recover echoes by matching a finite number of frequencies

$$\underset{h_1,h_2 \in \text{measure space}}{\arg\min} \ \frac{1}{2}\|\mathbf{X}_1 \cdot \mathcal{F}h_2(f) - \mathbf{X}_2 \cdot \mathcal{F}h_1(f)\|_2^2 + \lambda\|h_1 + h_2\|_{\mathrm{TV}} \quad \text{s.t.} \begin{cases} h_1(\{0\}) = 1 \\ h_l \geq 0 \end{cases}$$

Proposed approach: analytical & off-grid

Observation 1: the cross relation remains true in the frequency domain

$$\mathcal{F}x_1 \cdot \mathcal{F}h_2(n/F_s) = \mathcal{F}x_2 \cdot \mathcal{F}h_1(n/F_s) \qquad n = 0 \dots N-1$$

Observation 2: $\mathcal{F}\delta_{\text{echo}}$ **is known in closed-form**

Observation 3: $\mathcal{F}\mathrm{x_i}$ **can be (well) approximated by DFT**

$$\mathbf{X}_i = \mathtt{DFT}(x_i) \simeq \mathcal{F}x_i(nF_s) \qquad n = 0 \dots N-1$$

Idea: Recover echoes by matching a finite number of frequencies

$$\underset{h_1, h_2 \in \substack{\text{measure} \\ \text{space}}}{\arg\min} \ \tfrac{1}{2} \|\mathbf{X}_1 \cdot \mathcal{F}h_2(f) - \mathbf{X}_2 \cdot \mathcal{F}h_1(f)\|_2^2 + \lambda \|h_1 + h_2\|_{\text{TV}} \quad \text{s.t.} \begin{cases} h_1(\{0\}) = 1 \\ h_l \geq 0 \end{cases}$$

$\sim$ **Lasso** problem, but $\mathcal{F}h_2(f)$ is a continuous function.

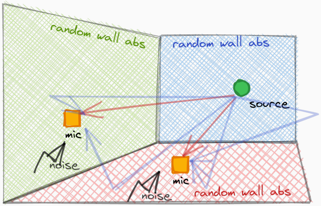
Instance of a **BLasso** problem [Bredies and Carioni, 2020]

Solved with Sliding Frank-Wolfe algorithm [Denoyelle et al., 2019]

Experimental results

Methods

- BSN — SIMO BCE[Lin et al., 2007]
- IL1C: iteratively-weighted ℓ_1 constraint SIMO BCE [Crocco and Del Bue, 2015]
- Blaster: Proposed off-grid approach

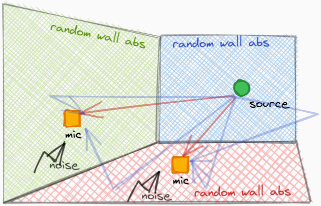


Baseline method are xvalidated on other dataset

Experimental results

Methods

- BSN — SIMO BCE[Lin et al., 2007]
- IL1C: iteratively-weighted ℓ_1 constraint SIMO BCE [Crocco and Del Bue, 2015]
- `Blaster`: Proposed off-grid approach



Baseline method are xvalidated on other dataset

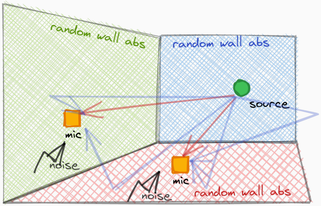
Dataset

- $\mathcal{D}^{\text{SNR}}$: $SNR \in [0, 20]$ dB, $\text{RT}_{60} = 400$ ms
- $\mathcal{D}^{\text{RT60}}$: $\text{RT}_{60} = [100, 1000]$ ms, $SNR = 20$ dB

Experimental results

Methods

- BSN — SIMO BCE[Lin et al., 2007]
- IL1C: iteratively-weighted ℓ_1 constraint SIMO BCE [Crocco and Del Bue, 2015]
- `Blaster`: Proposed off-grid approach



Baseline method are xvalidated on other dataset

Dataset

- $\mathcal{D}^{\text{SNR}}$: $SNR \in [0, 20]$ dB, $\text{RT}_{60} = 400$ ms
- $\mathcal{D}^{\text{RT60}}$: $\text{RT}_{60} = [100, 1000]$ ms, $SNR = 20$ dB

Metrics

Performance per # of echoes

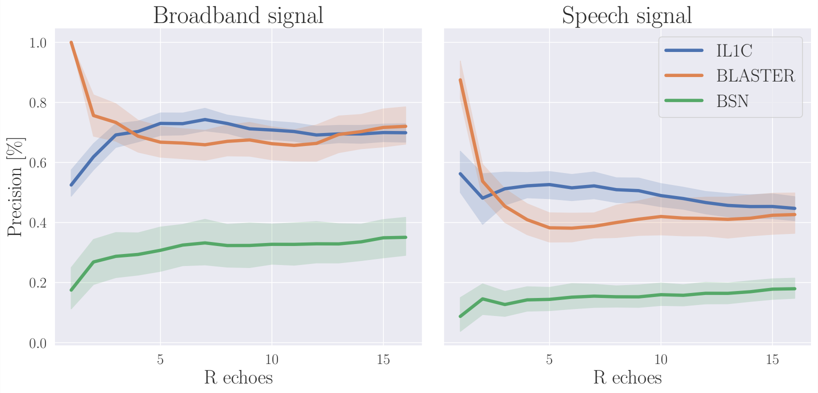


Figure 1: $\mathrm{RT}_{60} = 400$ ms and SNR $= 20$ dB.

✗ Sensitive to # echoes

✗ Sensitive source signal

✓ Good for 2 echoes

Performance per # of echoes

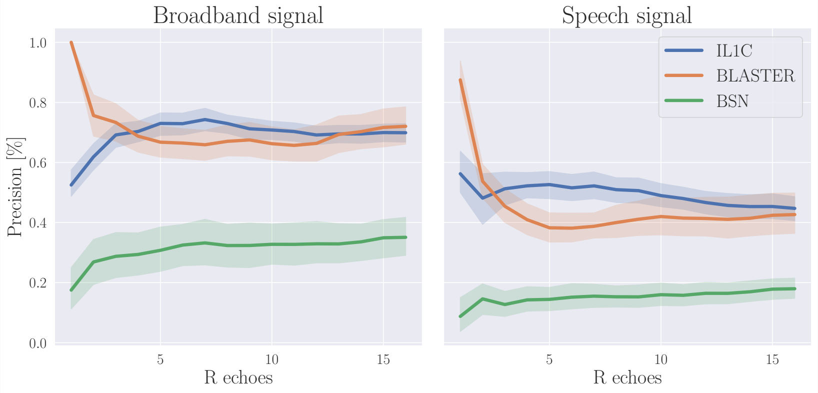


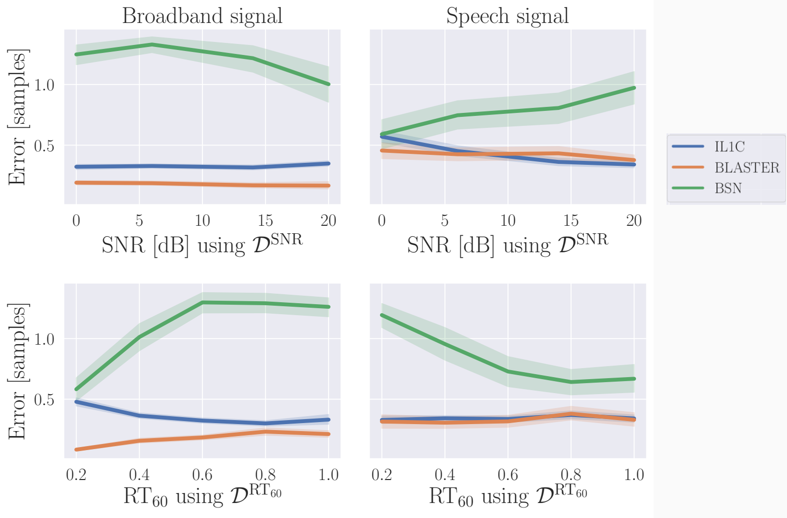
Figure 1: $\mathrm{RT}_{60} = 400$ ms and SNR $= 20$ dB.

✗ Sensitive to # echoes

✗ Sensitive source signal

✓ Good for 2 echoes [Scheibler et al., 2018,

Error per Dataset/Signal while recovering 7 echoes



✓Lower RMSE ✓ Robustness to SNR and RT_{60} ✗ Source signal dependent

References i

Aissa-El-Bey, A. and Abed-Meraim, K. (2008).
Blind simo channel identification using a sparsity criterion.
In *2008 IEEE 9th Workshop on Signal Processing Advances in Wireless Communications*, pages 271–275. IEEE.

Bredies, K. and Carioni, M. (2020).
Sparsity of solutions for variational inverse problems with finite-dimensional data.
Calculus of Variations and Partial Differential Equations, 59(1):14.

Crocco, M. and Del Bue, A. (2015).
Room impulse response estimation by iterative weighted l 1-norm.
In *2015 23rd European Signal Processing Conference (EUSIPCO)*, pages 1895–1899. IEEE.

Crocco, M. and Del Bue, A. (2016).
Estimation of tdoa for room reflections by iterative weighted l 1 constraint.
In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3201–3205. IEEE.

Denoyelle, Q., Duval, V., Peyré, G., and Soubies, E. (2019).
The sliding frank–wolfe algorithm and its application to super-resolution microscopy.
Inverse Problems, 36(1):014001.

Di Carlo, D., Deleforge, A., and Bertin, N. (2019).
Mirage: 2d source localization using microphone pair augmentation with echoes.
In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 775–779. IEEE.

Kowalczyk, K., Habets, E. A., Kellermann, W., and Naylor, P. A. (2013).
Blind system identification using sparse learning for tdoa estimation of room reflections.
IEEE Signal Processing Letters, 20(7):653–656.

Lin, Y., Chen, J., Kim, Y., and Lee, D. D. (2007).
Blind sparse-nonnegative (bsn) channel identification for acoustic time-difference-of-arrival estimation.
In *2007 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 106–109. IEEE.

Lin, Y., Chen, J., Kim, Y., and Lee, D. D. (2008).
Blind channel identification for speech dereverberation using l1-norm sparse learning.
In *Advances in Neural Information Processing Systems*, pages 921–928.

Scheibler, R., Di Carlo, D., Deleforge, A., and Dokmanić, I. (2018).
Separake: Source separation with a little help from echoes.
In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6897–6901. IEEE.

Tong, L., Xu, G., and Kailath, T. (1994).
Blind identification and equalization based on second-order statistics: A time domain approach.
IEEE Transactions on information Theory, 40(2):340–349.