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)

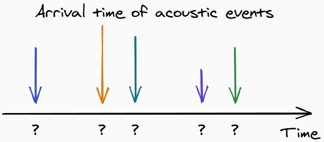
Université de Rennes 1, IRISA/INRIA, Panama research group

Acoustic Echo Estimation

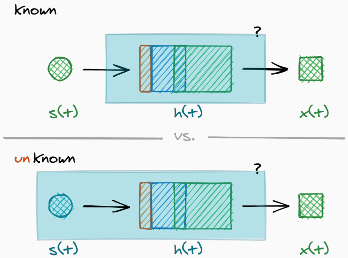
Acoustic Echo Retrieval

Estimating early (strong) reflections for microphones recordings, i.e.,

$$\{\tilde{x}_i\}_i \longrightarrow \{\tau_i^{(r)}, \alpha_i^{(r)}\}_{i,r}$$



Scenarios: the source signal is

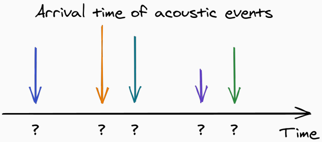


Our case: signal source and passive system of (I microphones)

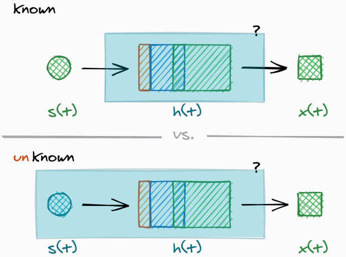
Acoustic Echo Retrieval

Estimating early (strong) reflections for microphones recordings, i.e.,

$$\{\tilde{x}_i\}_i \longrightarrow \{\tau_i^{(r)}, \alpha_i^{(r)}\}_{i,r}$$



Scenarios: the source signal is



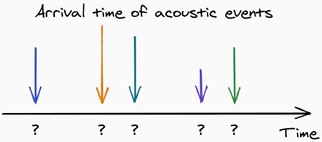
Our case: signal source and passive system of (I microphones)

1

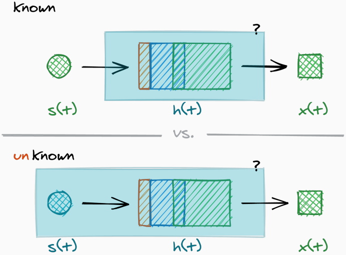
Acoustic Echo Retrieval

Estimating early (strong) reflections for microphones recordings, i.e.,

$$\{\tilde{x}_i\}_i \longrightarrow \{\tau_i^{(r)}, \alpha_i^{(r)}\}_{i,r}$$



Scenarios: the source signal is

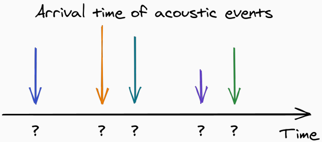


Our case: signal source and passive system of (I microphones)

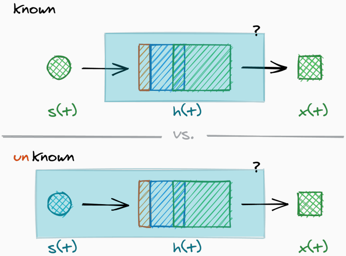
Acoustic Echo Retrieval

Estimating early (strong) reflections for microphones recordings, i.e.,

$$\{\tilde{x}_i\}_i \longrightarrow \{\tau_i^{(r)}, \alpha_i^{(r)}\}_{i,r}$$



Scenarios: the source signal is



Active

- **non-blind** problem
- **intrusive** or specific setups
  (**Application:** sonar, calibration, measurements, etc.)

Passive

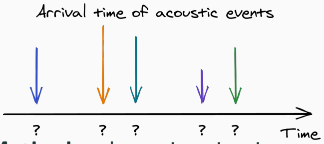
- **blind inverse** problem (harder)
- **passive** and more common setups
  (**Applications:** recording on smart speakers, laptop, etc.)

Our case: signal source and passive system of (I microphones)

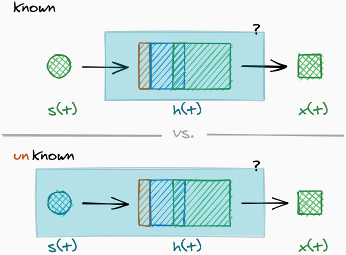
Acoustic Echo Retrieval

Estimating early (strong) reflections for microphones recordings, i.e.,

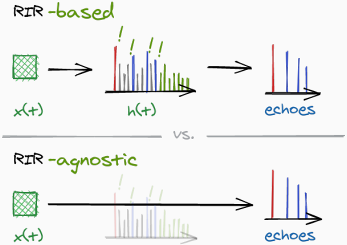
$$\{\tilde{x}_i\}_i \longrightarrow \{\tau_i^{(r)}, \alpha_i^{(r)}\}_{i,r}$$



Scenarios: the source signal is



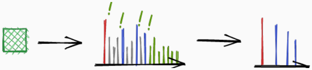
Methods: the estimation is



Our case: signal source and passive system of (I microphones)

Passive Acoustic Echo Retrieval

RIR-based approaches



1. "BCE"[1] problem $\Longrightarrow$ RIRs
2. Peak picking $\Longrightarrow$ Echoes

✓ BCE is well and known studied

✓ it is the state of the art
[Crocco and Del Bue, 2016]

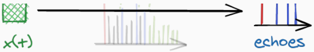
✓ reasonably good for some application

✗ Full RIRs need to be estimated

✗ Peak picking has hyperparameters

✗ Issues due to *on-grid* estimation

RIR-agnostic approaches



1. Direct off-grid estimation of $\left\{\tau_i^{(r)}, \alpha_i^{(r)}\right\}$ e.g., with maximum-likelihood

✓ No full RIRs & no peak picking
- lower complexity
- less hyperparameters

✓ Echo properties are respected
(e.g. Sparsity, Non-negativity)

✗ exploratory
(no standard solver, few works on audio)

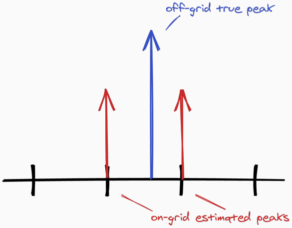
next topic

[1] Blind Channel Estimation

Limitations & bottleneck

Estimation is on-grid

- Sparsity and non-negativity not true "on grid"
- *Body guard* effect [Duval and Peyré, 2017]
  - $\rightarrow$ low recall $\implies$ low accuracy
  - $\rightarrow$ slow convergence
- Pick Picking
  - $\rightarrow$ manually tuned / corrected
  - $\rightarrow$ echo labeling (NP-hard problem)



How about higher sampling rate F_s?

$\rightarrow$ Increase Precision

But, computational bottleneck!

- Bigger vectors and matrices
  - $\longrightarrow$ memory usage
- the higher the sampling frequency
  - $\longrightarrow$ more ill-conditioned

How to solve this?

RIR-agnostic + off-grid

1. Learning-based
2. Analytical

Proposed approach: learning-based & off-grid

Idea: (Deep) Learning-based AER

1. Extend virtually learning-based SSL to AER
2. Estimate first echo estimation (simple but important)
3. Only 2 microphones

Observations:

- This *direct* mapping is difficult, the *inverse* "is not"
  → acoustic simulators: mic/src/room $\longrightarrow$ $\tau_i^{(r)}, \alpha_i^{(r)}, \quad \tilde{h}_i, \quad \tilde{x}_i$
- Acoustic simulator are "simple", versatile and fast
  → many data
- This approach is successfully in *Sound Source Localization*
  → position is related to echoes
  [Nguyen et al., 2018, Perotin et al., 2019] ⚠ Not only DNN

Models

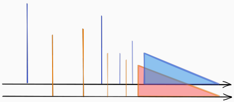
Inputs: Interchannel level and phase difference features[1] from

$$R[f] = \underset{t}{\text{avg.}} \frac{X_2[f,t]}{X_1[f,t]} \approx \underset{t}{\text{avg.}} \frac{H_2[f]S[f,t]}{H_1[f]S[f,t]}$$

$\approx$ the relative transfer function

$\rightarrow$ remove source dependency

Output: Inter and intra arrival delays



4 TOA

$\downarrow$

3 Time Difference of Arrivals (**TDOAs**)[1]

HP: first $\Leftrightarrow$ strongest echo (close-surface)

- Architecture: `MLP`, `CNN` [Chakrabarty and Habets, 2017, Nguyen et al., 2018]
- Loss Function:
  1. RMSE (Multi-label regression) $\rightarrow$ TDOAs
  2. Gaussian log-likelihood $\rightarrow \{\mu_\tau, \sigma^2_\tau\}\ \forall \tau \in$ TDOAs
  3. Student log-likelihood $\rightarrow \{\mu_\tau, \lambda_\tau, \nu_\tau\}\ \forall \tau \in$ TDOAs

  } for data fusion similar to **MDN** [Bishop, 1994]
- Data:
  - Virtually-supervised learning (= data from acoustic simulator)
  - white-noise as source signal + noise

[1] See appendix for detailed computation

Models

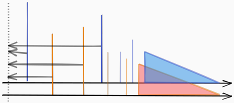
Inputs: Interchannel level and phase difference features[1] from

$$R[f] = \underset{t}{\text{avg.}} \frac{X_2[f,t]}{X_1[f,t]} \approx \underset{t}{\text{avg.}} \frac{H_2[f]S[f,t]}{H_1[f]S[f,t]}$$

$\approx$ the relative transfer function

$\rightarrow$ remove source dependency

Output: Inter and intra arrival delays



4 TOA

$\downarrow$

3 Time Difference of Arrivals (**TDOAs**)[1]

HP: first $\Leftrightarrow$ strongest echo (close-surface)

- Architecture: `MLP`, `CNN` [Chakrabarty and Habets, 2017, Nguyen et al., 2018]
- Loss Function:
  1. RMSE (Multi-label regression) $\rightarrow$ TDOAs
  2. Gaussian log-likelihood $\rightarrow \{\mu_\tau, \sigma^2_\tau\}\ \forall \tau \in$ TDOAs
  3. Student log-likelihood $\rightarrow \{\mu_\tau, \lambda_\tau, \nu_\tau\}\ \forall \tau \in$ TDOAs

  } for data fusion similar to **MDN** [Bishop, 1994]
- Data:
  - Virtually-supervised learning (= data from acoustic simulator)
  - white-noise as source signal + noise

[1] See appendix for detailed computation

Models

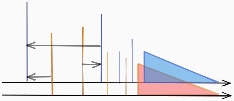
Inputs: Interchannel level and phase difference features[1] from

$$R[f] = \underset{t}{\text{avg.}} \frac{X_2[f,t]}{X_1[f,t]} \approx \underset{t}{\text{avg.}} \frac{H_2[f]S[f,t]}{H_1[f]S[f,t]}$$

$\approx$ the relative transfer function

$\rightarrow$ remove source dependency

Output: Inter and intra arrival delays



4 TOA

$\downarrow$

3 Time Difference of Arrivals (**TDOAs**)[1]

HP: first $\Leftrightarrow$ strongest echo (close-surface)

- Architecture: `MLP`, `CNN` [Chakrabarty and Habets, 2017, Nguyen et al., 2018]
- Loss Function:
  1. RMSE (Multi-label regression) $\rightarrow$ TDOAs
  2. Gaussian log-likelihood $\rightarrow \{\mu_\tau, \sigma^2_\tau\}\ \forall \tau \in$ TDOAs
  3. Student log-likelihood $\rightarrow \{\mu_\tau, \lambda_\tau, \nu_\tau\}\ \forall \tau \in$ TDOAs

  for data fusion similar to **MDN** [Bishop, 1994]
- Data:
  - Virtually-supervised learning (= data from acoustic simulator)
  - white-noise as source signal + noise

[1] See appendix for detailed computation

Models

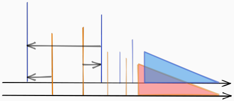
Inputs: Interchannel level and phase difference features[1] from

$$R[f] = \underset{t}{\texttt{avg.}} \frac{X_2[f,t]}{X_1[f,t]} \approx \underset{t}{\texttt{avg.}} \frac{H_2[f]S[f,t]}{H_1[f]S[f,t]}$$

$\approx$ the relative transfer function

$\rightarrow$ remove source dependency

Output: Inter and intra arrival delays

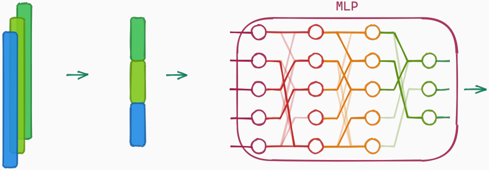


4 TOA

$\downarrow$

3 Time Difference of Arrivals (**TDOAs**)[1]

HP: first $\Leftrightarrow$ strongest echo (close-surface)



- Architecture: `MLP`, `CNN` [Chakrabarty and Habets, 2017, Nguyen et al., 2018]
- Loss Function:
  1. RMSE (Multi-label regression) $\rightarrow$ TDOAs
  2. Gaussian log-likelihood $\rightarrow \{\mu_\tau, \sigma^2_\tau\}\ \forall \tau \in$ TDOAs
  3. Student log-likelihood $\rightarrow \{\mu_\tau, \lambda_\tau, \nu_\tau\}\ \forall \tau \in$ TDOAs

  for data fusion similar to **MDN** [Bishop, 1994]
- Data:
  - Virtually-supervised learning (= data from acoustic simulator)

5

Models

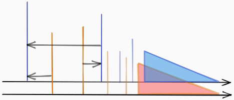
Inputs: Interchannel level and phase difference features[1] from

$$R[f] = \underset{t}{\mathrm{avg.}}\frac{X_2[f,t]}{X_1[f,t]} \approx \underset{t}{\mathrm{avg.}}\frac{H_2[f]S[f,t]}{H_1[f]S[f,t]}$$

$\approx$ the relative transfer function

$\rightarrow$ remove source dependency

Output: Inter and intra arrival delays

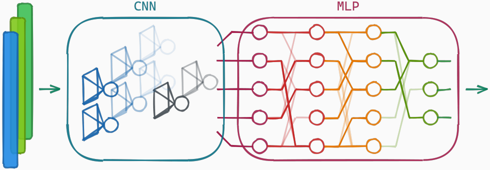


4 TOA

$\downarrow$

3 Time Difference of Arrivals (**TDOAs**)[1]

HP: first $\Leftrightarrow$ strongest echo (close-surface)



- Architecture: `MLP`, `CNN` [Chakrabarty and Habets, 2017, Nguyen et al., 2018]
- Loss Function:
  1. RMSE (Multi-label regression) $\rightarrow$ TDOAs
  2. Gaussian log-likelihood $\rightarrow \{\mu_\tau, \sigma_\tau^2\}\ \forall\tau \in$ TDOAs
  3. Student log-likelihood $\rightarrow \{\mu_\tau, \lambda_\tau, \nu_\tau\}\ \forall\tau \in$ TDOAs

  } for data fusion similar to **MDN** [Bishop, 1994]
- Data:
  - Virtually-supervised learning (= data from acoustic simulator)

Experimental results

Metrics: normalized RMSE (0 = best fit, 1 = random fit)

Proposed Method: MLP, CNN, $\mathtt{CNN}_{\mathcal{N}}$, $\mathtt{CNN}_{\mathcal{T}}$

Baseline: GCC PHAT[1]

Are better than baseline?

(only TDOA on direct path)

→ yes ✓

Echoes' TDOAs?

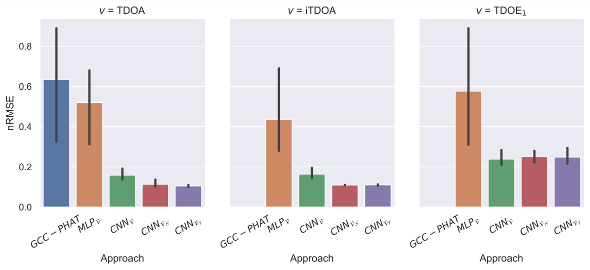
→ yes ✓

→ CNN better than MLP

Are robust to noise?

→ yes ✓

→ $\mathtt{CNN}_{\mathcal{N}}$/$\mathtt{CNN}_{\mathcal{T}}$ better than CNN



⚠Only 2 echoes ⚠work in progress (mor echoes and[6] generalizatior

[1] [Knapp and Carter, 1976]

Experimental results

Metrics: normalized RMSE (0 = best fit, 1 = random fit)

Proposed Method: $\mathtt{MLP}$, $\mathtt{CNN}$, $\mathtt{CNN}_{\mathcal{N}}$, $\mathtt{CNN}_{\mathcal{T}}$

Baseline: $\mathtt{GCC\,PHAT}$[1]

Are better than baseline?

(only TDOA on direct path)

→ yes ✓

Echoes' TDOAs?

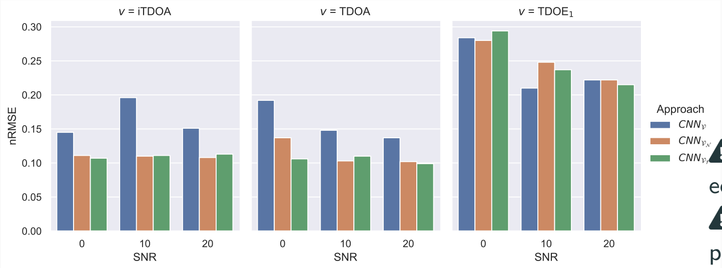
→ yes ✓

→ CNN better than MLP

Are robust to noise?

→ yes ✓

→ $\mathtt{CNN}_{\mathcal{N}}$/$\mathtt{CNN}_{\mathcal{T}}$ better than $\mathtt{CNN}$



⚠Only 2 echoes

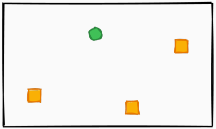
⚠work in progress (mor echoes and[6] generalizatior

[1] [Knapp and Carter, 1976]

State of the Art

(common knowledge)

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:

- Sampled version of $\tilde{x}_1, \tilde{x}_2$ are available (x_1, x_2)

1. Echoes' $\tau_i^{(r)} \propto$ sampling frequency
2. Find echoes $\rightarrow$ **find sparse vectors** h_1, h_2 of length L
3. Modeled as **Lasso**-like problem

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

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

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

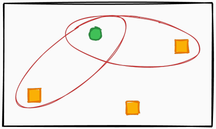
$\mathcal{C}(h_1, h_2) \longrightarrow$ constraints e.g. nonnegativity, anchor

✓ [Tong et al., 1994] ✓ [Lin et al., 2008] ✓ [Aissa-El-Bey and Abed-Meraim, 2008]
✓ [Kowalczyk et al., 2013] ✓ [Crocco and Del Bue, 2015, Crocco and Del Bue, 2016]

State of the Art

(common knowledge)

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:

- Sampled version of $\tilde{x}_1, \tilde{x}_2$ are available (x_1, x_2)

1. Echoes' $\tau_i^{(r)} \propto$ sampling frequency
2. Find echoes $\rightarrow$ **find sparse vectors** h_1, h_2 of length L
3. 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, 2015, 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

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

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**

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

Proposed approach: analytical & off-grid

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

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

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

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

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

Idea: Recover echoes by matching a finite number of frequencies

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

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: $\mathcal{F}\mathrm{x}_\mathrm{i}$ 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

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

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

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

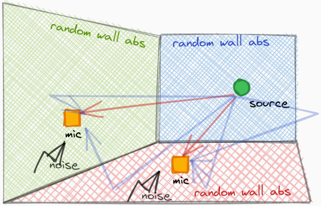
Solved with Sliding Frank-Wolfe algorithm [Denoyelle et al., 2019]

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

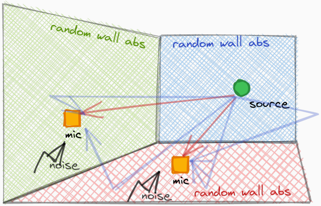
Baseline method are xvalidated on other dataset



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

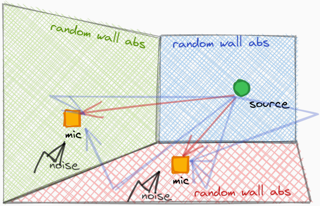
Dataset

- $\mathcal{D}^{\text{SNR}}$: $SNR \in [0, 20]$ 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

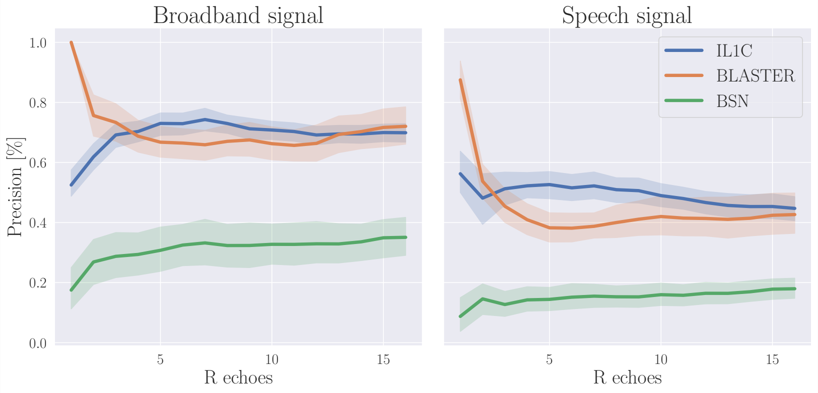


Figure 1: $\mathrm{RT}_{60} = 400$ ms and SNR $= 20$ dB.

✗ Sensitive to # echoes

✗ Sensitive source signal

✓ Good for 2 echoes

Performance per # of echoes

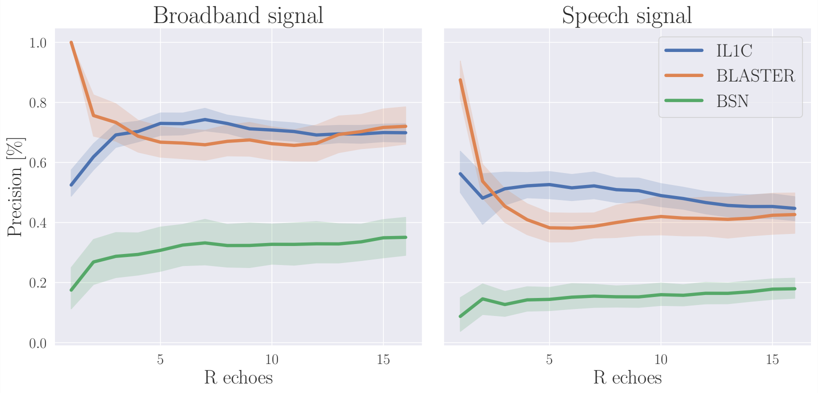


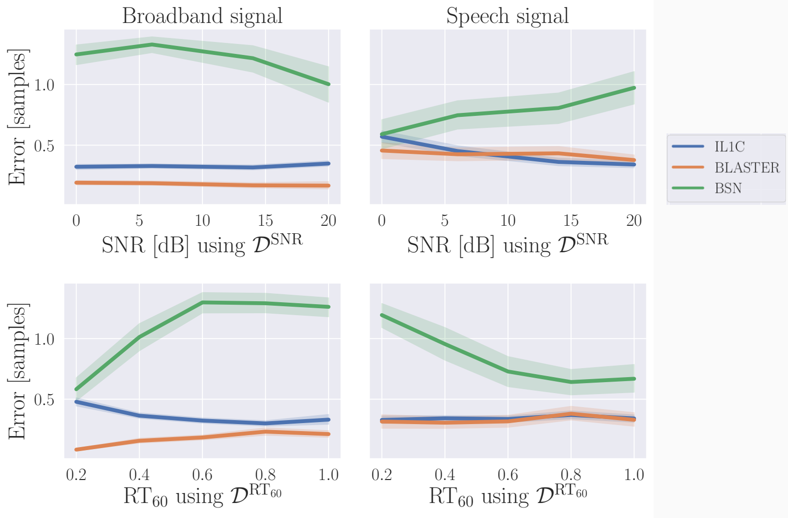
Figure 1: $\mathrm{RT}_{60} = 400$ ms and SNR $= 20$ dB.

✗ 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.

Bishop, C. M. (1994).
Mixture density networks.

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.

Chakrabarty, S. and Habets, E. A. (2017).
Broadband doa estimation using convolutional neural networks trained with noise signals.
In *2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 136–140. IEEE.

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.

Duval, V. and Peyré, G. (2017).
Sparse regularization on thin grids i: the lasso.
Inverse Problems, 33(5):055008.

Knapp, C. and Carter, G. (1976).
The generalized correlation method for estimation of time delay.
IEEE transactions on acoustics, speech, and signal processing, 24(4):320–327.

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.

References v

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.

Nguyen, Q., Girin, L., Bailly, G., Elisei, F., and Nguyen, D.-C. (2018).
Autonomous sensorimotor learning for sound source localization by a humanoid robot.

Perotin, L., Défossez, A., Vincent, E., Serizel, R., and Guérin, A. (2019).
Regression versus classification for neural network based audio source localization.
In *2019 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 343–347. IEEE.

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.