Sound Source Localization and Reconstruction Using a Wearable Microphone Array and Inertial Sensors

Fusion 2020 Virtual Conference
Advances in Motion Estimation I Inertial Sensors

Clas Veibäck, Martin A. Skoglund, Fredrik Gustafsson and Gustaf Hendeby

Automatic Control
Department of Electrical Engineering
Linköping University

Introduction

LINKÖPING
UNIVERSITY

Introduction

- Prototype of a wearable microphone array

Introduction

- Prototype of a wearable microphone array
- Evaluate the LinDoA method for Direction of Arrival and Source Separation

Introduction

- Prototype of a wearable microphone array
- Evaluate the LinDoA method for Direction of Arrival and Source Separation
- Track sound sources in global frame using IMU

Introduction

- Prototype of a wearable microphone array
- Evaluate the LinDoA method for Direction of Arrival and Source Separation
- Track sound sources in global frame using IMU
- Develop LinDoA method to manage multiple sources

Introduction

- Prototype of a wearable microphone array
- Evaluate the LinDoA method for Direction of Arrival and Source Separation
- Track sound sources in global frame using IMU
- Develop LinDoA method to manage multiple sources
- Simulations and experiments

Introduction

- Prototype of a wearable microphone array
- Evaluate the LinDoA method for Direction of Arrival and Source Separation
- Track sound sources in global frame using IMU
- Develop LinDoA method to manage multiple sources
- Simulations and experiments
- Hearing aid applications

Advantages of LinDoA

- Taylor series expansion of signals

Advantages of LinDoA

- Taylor series expansion of signals
- Slow sampling for DoA

Advantages of LinDoA

- Taylor series expansion of signals
- Slow sampling for DoA
- Wideband and narrowband

Advantages of LinDoA

- Taylor series expansion of signals
- Slow sampling for DoA
- Wideband and narrowband
- Small arrays

Advantages of LinDoA

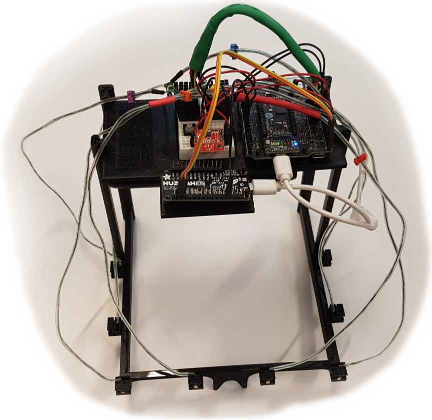
- Taylor series expansion of signals
- Slow sampling for DoA
- Wideband and narrowband
- Small arrays
- Near and far field

Advantages of LinDoA

- Taylor series expansion of signals
- Slow sampling for DoA
- Wideband and narrowband
- Small arrays
- Near and far field
- Parallelisation

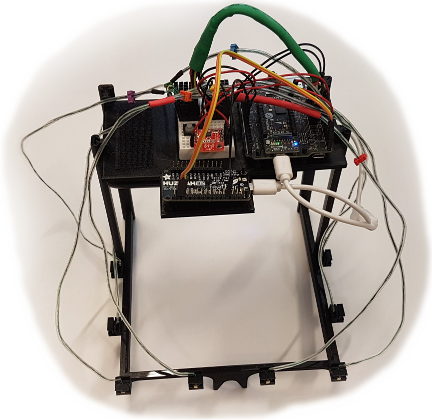
Array Frame Prototype

- 8 microphones



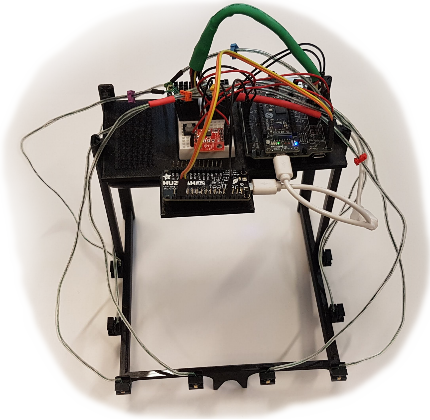
Array Frame Prototype

- 8 microphones
- IMU



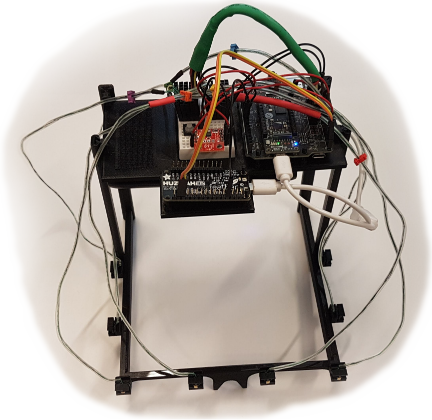
Array Frame Prototype

- 8 microphones
- IMU
- Embedded DSP



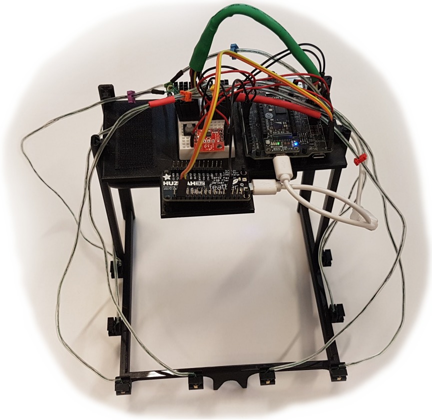
Array Frame Prototype

- 8 microphones
- IMU
- Embedded DSP
- Mobile: Battery & WiFi



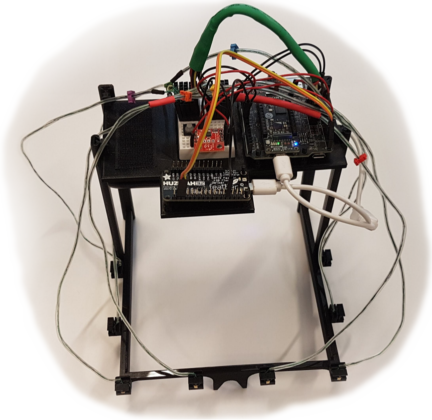
Array Frame Prototype

- 8 microphones
- IMU
- Embedded DSP
- Mobile: Battery & WiFi
- 3D-printed frame



Array Frame Prototype

- 8 microphones
- IMU
- Embedded DSP
- Mobile: Battery & WiFi
- 3D-printed frame
- Open source and design



Signal Models

Signal Models - Single Source

The signal model used in the single source case is

$$
\begin{aligned}
&y^n(t) = s(t + \tau_n) + e^n(t), \quad n = 1, \dots, N, \\
&e^n(t) \sim \mathcal{N}(0, \sigma_s^2).
\end{aligned}
$$

Signal Models - Taylor Series Expansion

An Lth order Taylor series expansion of the signal gives

$$s(t+\tau) \approx \sum_{l=0}^{L} \bar{\tau}_l \, s^{(l)}(t)$$

where $\bar{\tau}_l = \dfrac{\tau^l}{l!}$

Signal Models - Taylor Series Expansion

An Lth order Taylor series expansion of the signal gives

$$s(t+\tau) \approx \sum_{l=0}^{L} \bar{\tau}_l \, s^{(l)}(t) = \mathbf{h}^T(\tau)\mathbf{x}(t),$$

where $\bar{\tau}_l = \dfrac{\tau^l}{l!}$, the vector of signal derivatives is

$$\mathbf{x}(t) = \begin{bmatrix} s(t) & s^{(1)}(t) & \dots & s^{(L)}(t) \end{bmatrix}^T,$$

and the vector of delays is

$$\mathbf{h}(\tau) = \begin{bmatrix} 1 & \tau & \bar{\tau}_2 & \dots & \bar{\tau}_L \end{bmatrix}^T.$$

Signal Models - Approximation

The signal model

$$\begin{aligned} y^n(t) &= s(t+\tau_n) + e^n(t), \qquad n = 1, \dots, N, \\ e^n(t) &\sim \mathcal{N}(0, \sigma_s^2). \end{aligned}$$

Signal Models - Approximation

The signal model is then approximated as

$$y^n(t) = \mathbf{h}(\tau_n)\mathbf{x}(t) + e^n(t), \qquad n = 1, \ldots, N,$$
$$e^n(t) \sim \mathcal{N}(0, \sigma_r^2).$$

Signal Models - Approximation

The signal model is then approximated and discretized as

$$y_k^n = \mathbf{h}(\tau_n)\mathbf{x}_k + e_k^n, \qquad n = 1, \dots, N,$$
$$e_k^n \sim \mathcal{N}(0, \sigma_r^2), \qquad k = 1, \dots, K.$$

Signal Models - Approximation

The signal model is then approximated and discretized as

$$y_k^n = \mathbf{h}(\tau_n)\mathbf{x}_k + e_k^n, \qquad n = 1, \ldots, N,$$
$$e_k^n \sim \mathcal{N}(0, \sigma_r^2), \qquad k = 1, \ldots, K.$$

In vector form the model reduces to

$$\mathbf{y}_k = \mathbf{H}(\boldsymbol{\tau})\mathbf{x}_k + \mathbf{e}_k,$$
$$\mathbf{e}_k \sim \mathcal{N}(\mathbf{0}, \sigma_r^2 \mathbf{I}_N), \qquad k = 1, \ldots, K.$$

where $\mathbf{y}_k \triangleq \begin{bmatrix} y_k^1 & \ldots & y_k^N \end{bmatrix}^T$ and $\boldsymbol{\tau} \triangleq \begin{bmatrix} \tau_1 & \ldots & \tau_N \end{bmatrix}^T$.

Estimation

LINKÖPING
UNIVERSITY

Estimation - LinDoA

Least-squares solution gives, for $k = 1, \dots, K$,

$$\hat{\mathbf{x}}_k(\boldsymbol{\tau}) = (\mathbf{H}^T(\boldsymbol{\tau})\mathbf{H}(\boldsymbol{\tau}))^{-1}\mathbf{H}^T(\boldsymbol{\tau})\mathbf{y}(t),$$
$$\text{cov}(\hat{\mathbf{x}}_k(\boldsymbol{\tau})) = (\mathbf{H}^T(\boldsymbol{\tau})\mathbf{H}(\boldsymbol{\tau}))^{-1}\sigma_r^2.$$

Estimation - LinDoA

Least-squares solution gives, for $k = 1, \ldots, K$,

$$\hat{\mathbf{x}}_k(\boldsymbol{\tau}) = (\mathbf{H}^T(\boldsymbol{\tau})\mathbf{H}(\boldsymbol{\tau}))^{-1}\mathbf{H}^T(\boldsymbol{\tau})\mathbf{y}(t),$$
$$\operatorname{cov}(\hat{\mathbf{x}}_k(\boldsymbol{\tau})) = (\mathbf{H}^T(\boldsymbol{\tau})\mathbf{H}(\boldsymbol{\tau}))^{-1}\sigma_r^2.$$

This method is denoted LinDoA and estimates signal derivatives from samples in space.

Estimation - LinDoA

Least-squares solution gives, for $k = 1, \ldots, K$,

$$\hat{\mathbf{x}}_k(\boldsymbol{\tau}) = (\mathbf{H}^T(\boldsymbol{\tau})\mathbf{H}(\boldsymbol{\tau}))^{-1}\mathbf{H}^T(\boldsymbol{\tau})\mathbf{y}(t),$$
$$\mathrm{cov}(\hat{\mathbf{x}}_k(\boldsymbol{\tau})) = (\mathbf{H}^T(\boldsymbol{\tau})\mathbf{H}(\boldsymbol{\tau}))^{-1}\sigma_r^2.$$

This method is denoted LinDoA and estimates signal derivatives from samples in space.

Each estimate is independent from other samples in time.

Estimation - Constrained LinDoA

While efficient, the independence in time allows for inconsistencies.

Estimation - Constrained LinDoA

While efficient, the independence in time allows for inconsistencies.

To enforce consistency, constraints in time are added on the form

$$\underline{\mathbf{I}}\mathbf{x}_{k+1} = \underline{\mathbf{F}}\mathbf{x}_k,$$

where $\underline{\mathbf{I}}$ is the identity matrix with the last row removed.

Estimation - Constrained LinDoA

While efficient, the independence in time allows for inconsistencies.

To enforce consistency, constraints in time are added on the form

$$\underline{\mathbf{I}}\mathbf{x}_{k+1} = \underline{\mathbf{F}}\mathbf{x}_k,$$

where $\underline{\mathbf{I}}$ is the identity matrix with the last row removed. $\underline{\mathbf{F}}$ can, e.g., be

$$\underline{\mathbf{F}} = \begin{bmatrix} 1 & T & \bar{T}_2 & \dots & \bar{T}_L \\ 0 & 1 & T & \dots & \bar{T}_{L-1} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & T \end{bmatrix}.$$

Estimation - Constrained LinDoA

This results in the following Constrained Least-Squares problem

$$(\hat{\mathbf{x}}_1(\boldsymbol{\tau}), \ldots, \hat{\mathbf{x}}_K(\boldsymbol{\tau})) = \underset{\mathbf{x}_1, \ldots, \mathbf{x}_K}{\arg\min} \quad \sum_{k=1}^{K} \|\mathbf{y}_k - \mathbf{H}(\boldsymbol{\tau})\mathbf{x}_k\|^2,$$

$$\text{s.t.} \quad \underline{\mathbf{I}}\mathbf{x}_{k+1} = \underline{\mathbf{F}}\mathbf{x}_k, \quad k = 1, \ldots, K-1.$$

Estimation - Time-Delay Estimation

The time delays can be estimated as

$$\hat{\boldsymbol{\tau}} = \arg\min_{\boldsymbol{\tau}} \sum_{k=1}^{K} \|\mathbf{y}_k - \mathbf{H}(\boldsymbol{\tau})\hat{\mathbf{x}}_k(\boldsymbol{\tau})\|^2,$$

where $\hat{\mathbf{x}}_k(\boldsymbol{\tau})$ is computed using a variant of LinDoA.

Estimation - Time-Delay Estimation

The time delays can be estimated as

$$\hat{\boldsymbol{\tau}} = \arg\min_{\boldsymbol{\tau}} \sum_{k=1}^{K} \|\mathbf{y}_k - \mathbf{H}(\boldsymbol{\tau})\hat{\mathbf{x}}_k(\boldsymbol{\tau})\|^2,$$

where $\hat{\mathbf{x}}_k(\boldsymbol{\tau})$ is computed using a variant of LinDoA.

This can be solved using, e.g., numerical search.

Estimation - Multiple Sources

By superposition, M sources are incorporated as,

$$y^n(t) = \sum_{m=1}^{M} s_m(t + \tau_{nm}) + e^n(t), \quad n = 1, \ldots, N.$$

Estimation - Multiple Sources

By superposition, M sources are incorporated as,

$$y^n(t) = \sum_{m=1}^{M} s_m(t + \tau_{nm}) + e^n(t), \quad n = 1, \ldots, N.$$

A model can still be obtained on the form $\mathbf{y}_k = \mathbf{H}(\boldsymbol{\tau})\mathbf{x}_k + \mathbf{e}_k$.

Estimation - Multiple Sources

By superposition, M sources are incorporated as,

$$y^n(t) = \sum_{m=1}^{M} s_m(t + \tau_{nm}) + e^n(t), \quad n = 1, \dots, N.$$

A model can still be obtained on the form $\mathbf{y}_k = \mathbf{H}(\boldsymbol{\tau})\mathbf{x}_k + \mathbf{e}_k$.

By design, $\mathbf{H}(\boldsymbol{\tau})$ is now rank-deficient, resulting in an unobservable model using LinDoA.

Estimation - Multiple Sources

By superposition, M sources are incorporated as,

$$y^n(t) = \sum_{m=1}^{M} s_m(t + \tau_{nm}) + e^n(t), \quad n = 1, \ldots, N.$$

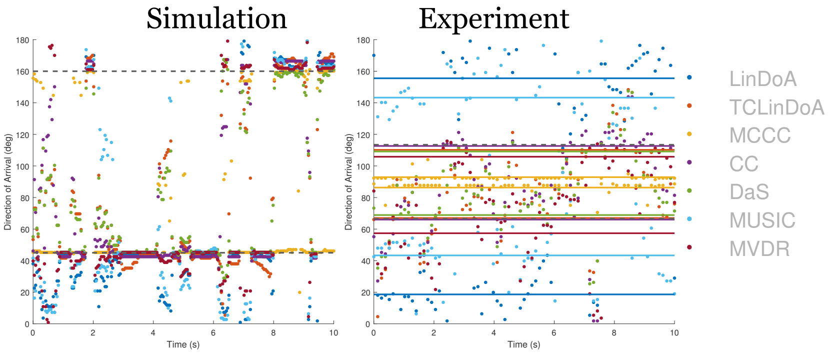
A model can still be obtained on the form $\mathbf{y}_k = \mathbf{H}(\boldsymbol{\tau})\mathbf{x}_k + \mathbf{e}_k$.

By design, $\mathbf{H}(\boldsymbol{\tau})$ is now rank-deficient, resulting in an unobservable model using LinDoA.

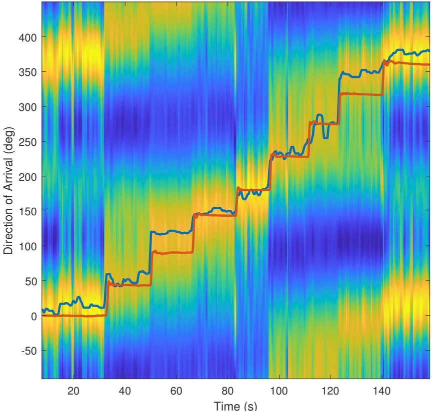
However, the signals are still observable using Constrained LinDoA.

Results

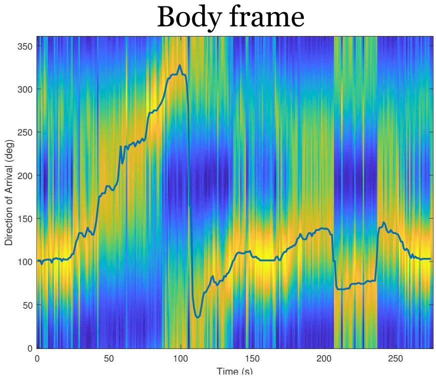
Two Sources Direction of Arrival



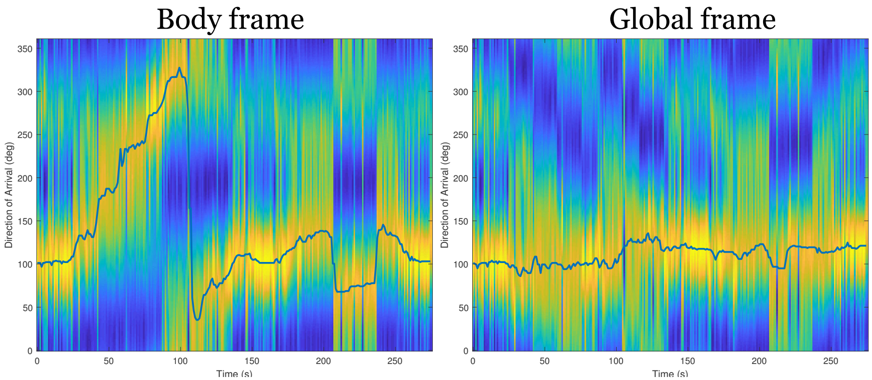
Single Source Constrained LinDoA compared with IMU



Single Source DoA with IMU Integration



Single Source DoA with IMU Integration



Source Separation

Delay and Sum

	Estimated Left	Estimated Right
True Left	0.7863	0.3114
True Right	0.5290	0.8802

Source Separation

Delay and Sum

	Estimated Left	Estimated Right
True Left	0.7863	0.3114
True Right	0.5290	0.8802

Constrained LinDoA

	Estimated left	Estimated right
True left	0.8942	0.0975
True right	0.1061	0.9534

Conclusion

LINKÖPING UNIVERSITY

Conclusions

- A head-worn microphone array prototype was developed.

Conclusions

- A head-worn microphone array prototype was developed.
- 3D-printing and modular design allows for rapid development.

Conclusions

- A head-worn microphone array prototype was developed.
- 3D-printing and modular design allows for rapid development.
- Integrating the array with an IMU allows for global tracking.

Conclusions

- A head-worn microphone array prototype was developed.
- 3D-printing and modular design allows for rapid development.
- Integrating the array with an IMU allows for global tracking.
- Variants of LinDoA were developed to handle multiple sources.

Conclusions

- A head-worn microphone array prototype was developed.
- 3D-printing and modular design allows for rapid development.
- Integrating the array with an IMU allows for global tracking.
- Variants of LinDoA were developed to handle multiple sources.
- LinDoA capable of tracking a single source in reverberant environment.

Conclusions

- A head-worn microphone array prototype was developed.
- 3D-printing and modular design allows for rapid development.
- Integrating the array with an IMU allows for global tracking.
- Variants of LinDoA were developed to handle multiple sources.
- LinDoA capable of tracking a single source in reverberant environment.
- Sound source separation of two sources performs well.

Possible Directions for Future Work

- Theoretical analysis of method

Possible Directions for Future Work

- Theoretical analysis of method
- Dereverberation and calibration using a HRTF

Possible Directions for Future Work

- Theoretical analysis of method
- Dereverberation and calibration using a HRTF
- Additional sensors (e.g. camera for face tracking)

Possible Directions for Future Work

- Theoretical analysis of method
- Dereverberation and calibration using a HRTF
- Additional sensors (e.g. camera for face tracking)
- Evaluating form factors, number of microphones and dimensions

Possible Directions for Future Work

- Theoretical analysis of method
- Dereverberation and calibration using a HRTF
- Additional sensors (e.g. camera for face tracking)
- Evaluating form factors, number of microphones and dimensions
- Multi-target tracking framework

Thank you for listening!

`gitlab.liu.se/veiback-public/lindoa`
`gitlab.liu.se/veiback-public/array-frame`

www.liu.se

LINKÖPING UNIVERSITY