Audio-visual sensing from a quadcopter: dataset and baselines for source localization and sound enhancement

Lin Wang, Ricardo Sanchez-Matilla, Andrea Cavallaro

CIS centre for intelligent sensing

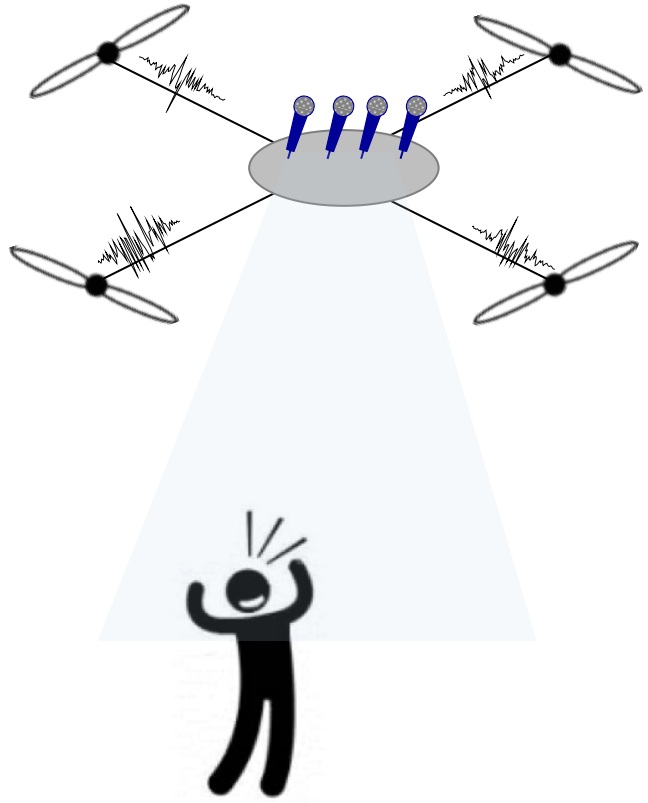
Queen Mary University of London

Outline

- Motivation
- Contributions
- Related work
- The AVQ dataset
- Challenges
- Baseline demos

Introduction

- Sound processing on drones
  - human-robot interaction
  - surveillance
  - multimedia broadcasting
- Acoustic sensing
  - sound source localization
  - sound enhancement
- Main challenges
  - strong ego-noise (SNR < -15 dB)
  - dynamics due to drone changes
  - wind noise
- A new research question
  - audio-visual sensing from drones



Contributions

- **A**udio-**V**isual **Q**uadcopter (**AVQ**) dataset
  - audio-visual dataset from a quadcopter drone
  - first outdoors dataset
  - annotations

- Baseline evaluation
  - sound source localization
  - sound enhancement

Related work

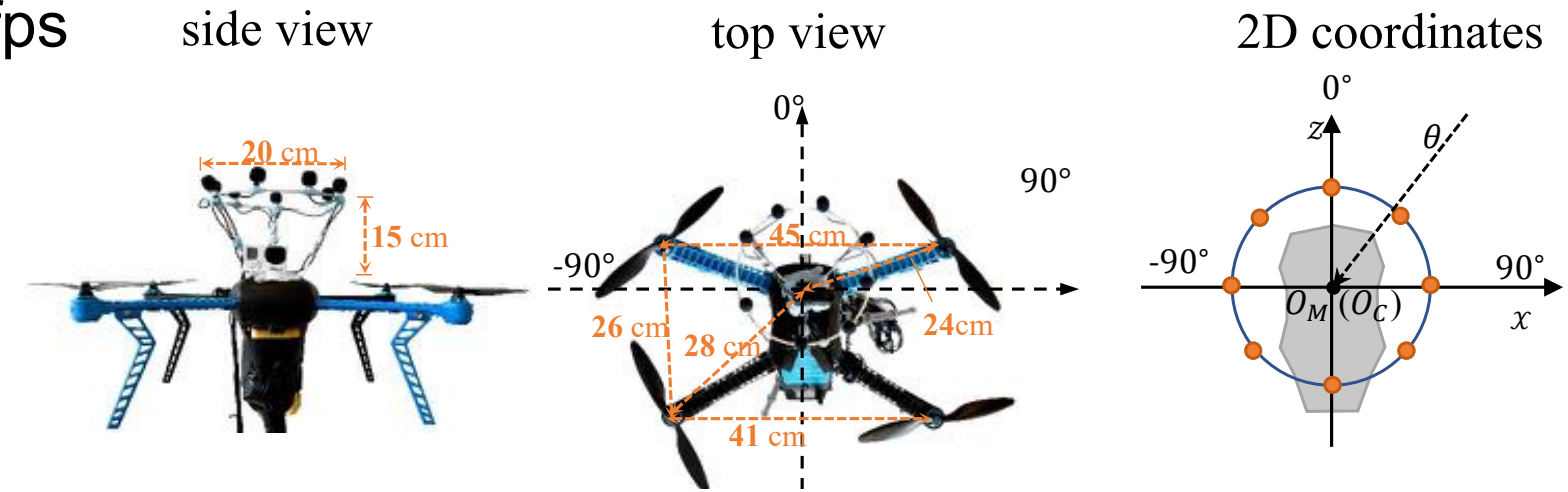
Ref.	Scenario	Drone	Audio	Video
DREGON [1]	Indoors	Mikrokopter drone	8-mic array	-
AIRA-UAS [2]	Indoors	DJI Matrice 100 3DR Solo Parrot Bebop 2	8-mic array	-
AVQ	Outdoors	3DR IRIS	8-mic array	HD @ 30fps

[1] M. Strauss, P. Mordel, V. Miguet, and A. Deleforge, “DREGON: dataset and methods for UAV-embedded sound source localization”, in Proc. IROS, 2018

[2] O. Ruiz-Espitia, J. Martinez-Carranza, and C. Rascon, “AIRA-UAS: an evaluation corpus for audio processing in unmanned aerial system,” in Proc. ICUAS, 2018

Hardware

- 3DR IRIS quadcopter
- Audio:
  - 8-microphone circular array
  - Boya BY-M1 omnidirectional microphones
  - 44.1 KHz
- Video:
  - GoPro camera
  - HD resolution at 30 fps

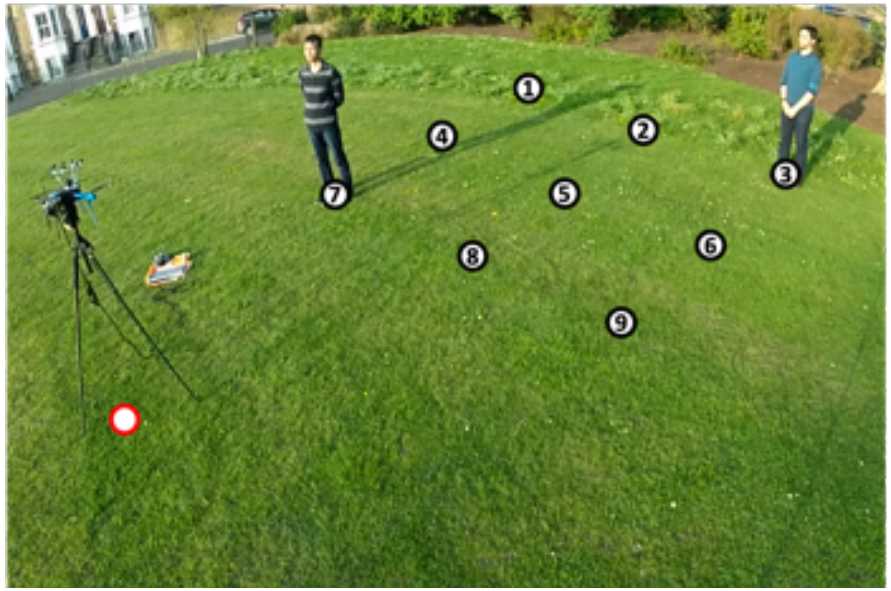


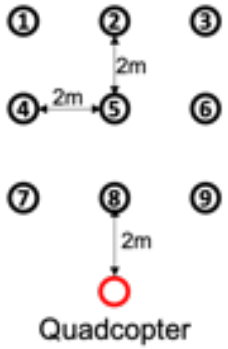
The dataset

- 12 audio-visual sequences
  - 50 minutes in total
- Synchronized and calibrated audio-visual signals
- Annotations
  - speaker location
  - voice activity detection

Property	Options	
Speakers motion	Static	Moving
Drone power	Constant	Dynamic
Recording	Composite mixture	Natural

Static speakers

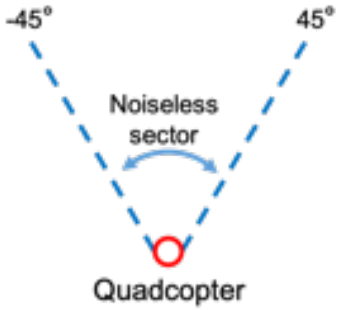




External view

Drone view

Moving speaker



External view

Drone view

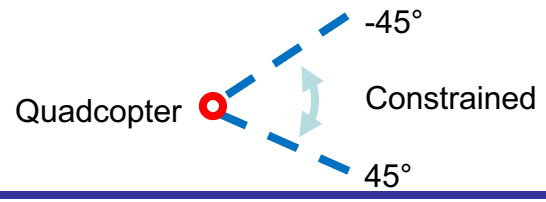
AVQ sequences

	Seq.	Duration [secs]	GT	Type	Drone power	Sound source
Subset 1						

AVQ sequences

Seq.		Duration [secs]	GT	Type	Drone power	Sound source
Subset 1	1	120		Ego-noise only	50%	
	2	120			50%	
	3	40			50%	
	4	797	✓	Speech only	0%	2 sources 9 locations

AVQ sequences

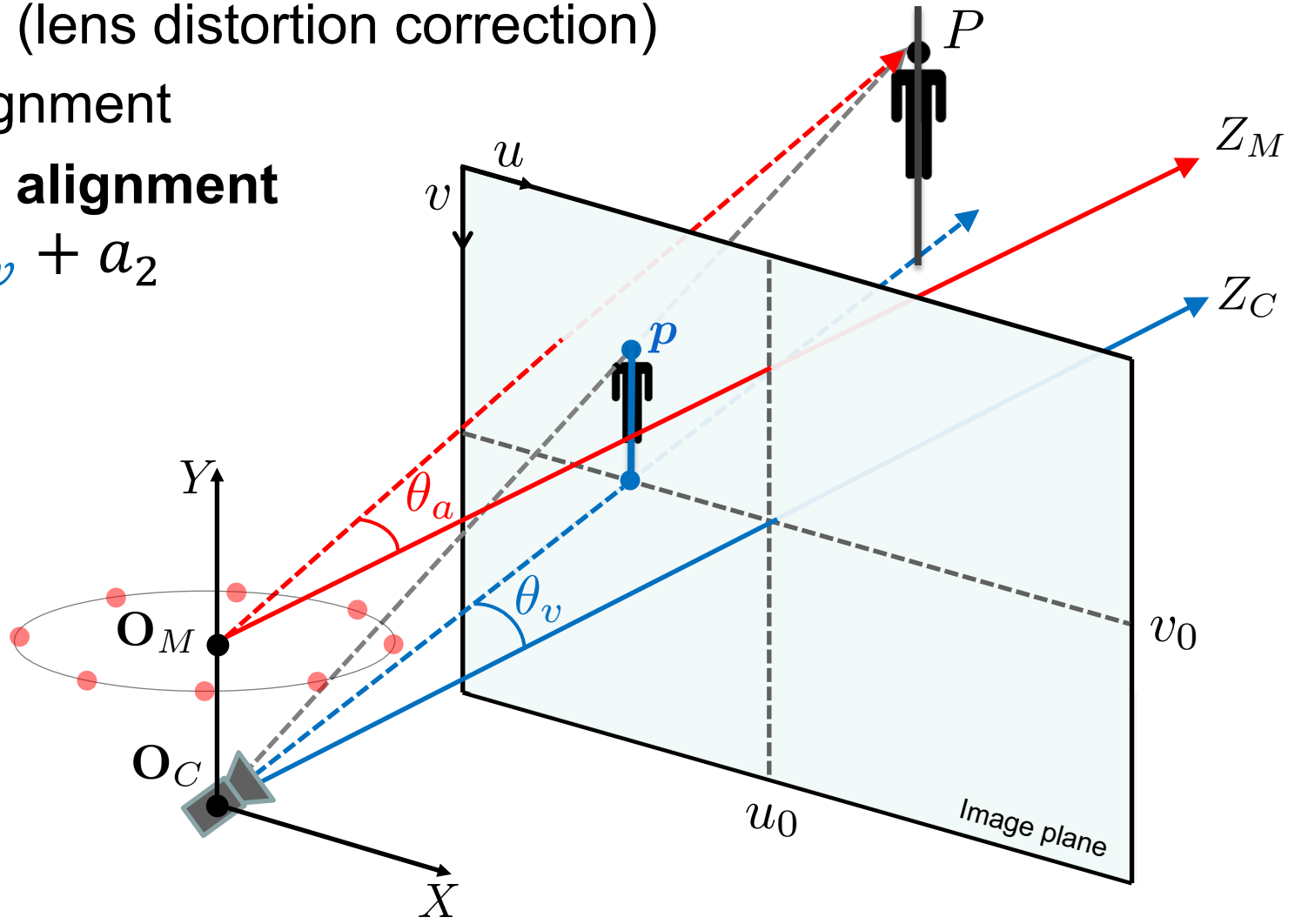


Seq.		Duration [secs]	GT	Type	Drone power	Sound source
Subset 1	1	120		Ego-noise only	50%	
	2	120			50%	
	3	40			50%	
	4	797	✓	Speech only	0%	2 sources 9 locations
Subset 2	1	210		Drone only	100%	
	2	214			50-100%	
	3	215	✓	Speech only	0%	constrained
	4	217	✓		0%	unconstrained
	5	303	✓	Mixture	100%	constrained
	6	271	✓		100%	unconstrained
	7	258	✓		50-100%	constrained
	8	249	✓		50-100%	unconstrained

Annotation of sound source

- Audio-visual calibration
  - Resectioning (lens distortion correction)
  - Temporal alignment
  - **Geometrical alignment**

$$\theta_a = a_1\theta_v + a_2$$

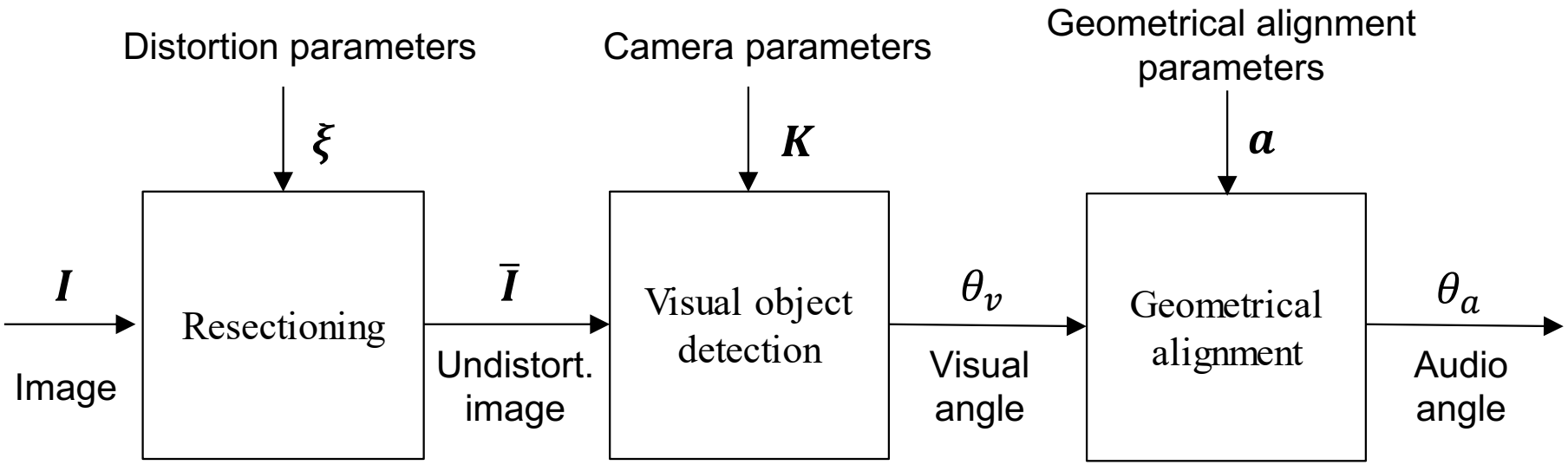


Annotation of sound source

θ_v

$\theta_a = a_1\theta_v + a_2$

θ_a



Application of AVQ

- Baseline performance [3-5]
  - Sound enhancement
  - Source localization
  - Source tracking

[3] R. Sanchez-Matilla, L. Wang, and A. Cavallaro, “Multi-modal localization and enhancement of multiple sound sources from a micro aerial vehicle”, Proc. ACM Multimedia, 2017
[4] L. Wang, R. Sanchez-Matilla, and A. Cavallaro, “Tracking a moving sound source from a multi-rotor drone”, in Proc. IROS, 2018
[5] L. Wang and A. Cavallaro, “Acoustic sensing from a multi-roto drone”, IEEE Sensors, 2018

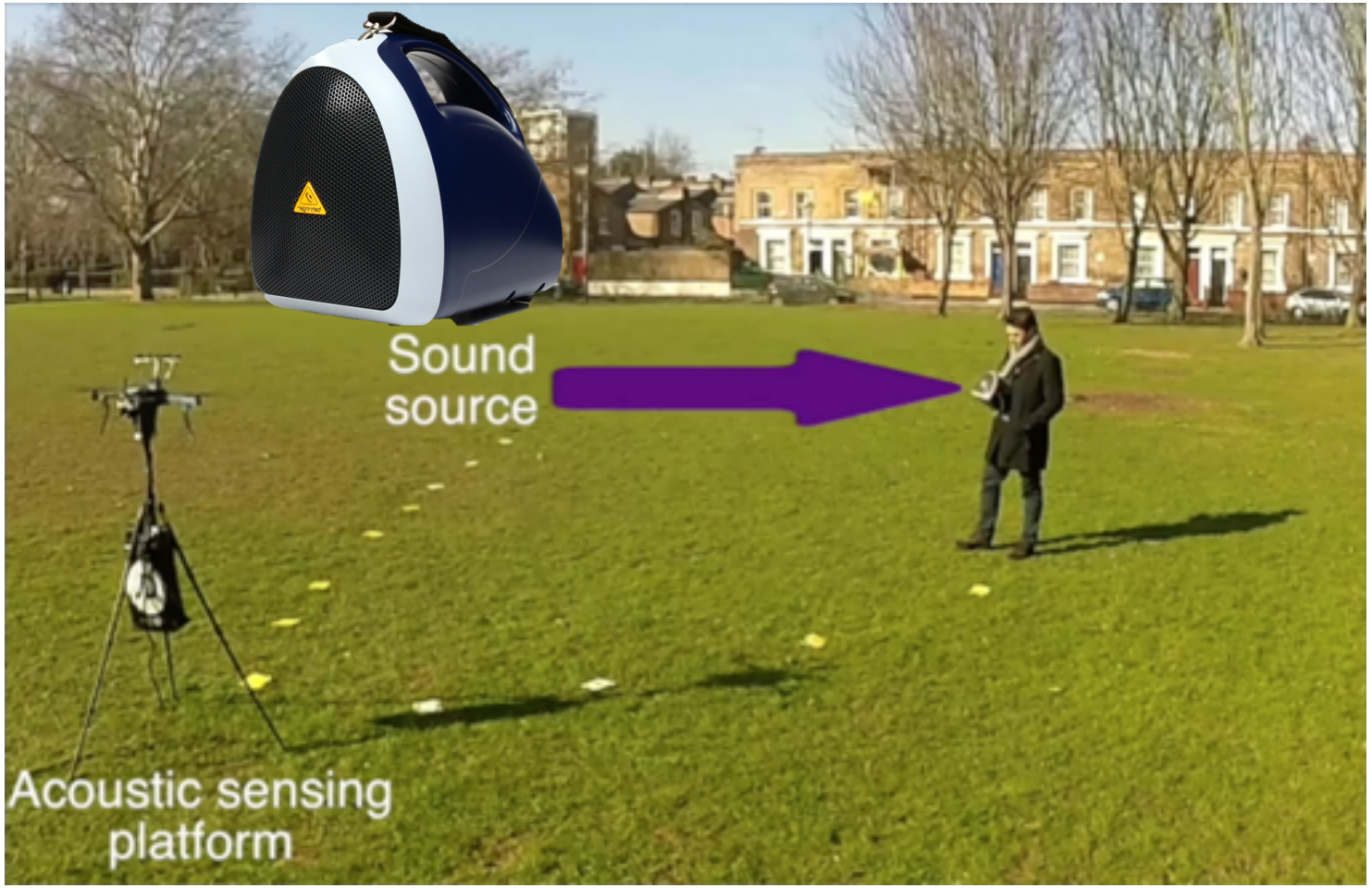
Application of AVQ - Sound enhancement (input)

[3] R. Sanchez-Matilla, L. Wang, and A. Cavallaro, “Multi-modal localization and enhancement of multiple sound sources from a micro aerial vehicle,” Proc. ACM Multimedia, 2017.

Application of AVQ - Sound enhancement (output)

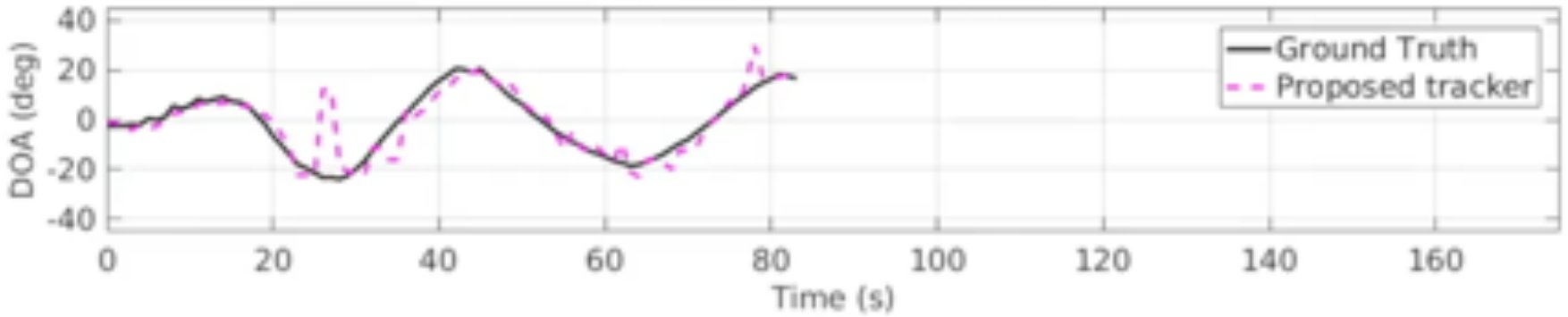
[3] R. Sanchez-Matilla, L. Wang, and A. Cavallaro, “Multi-modal localization and enhancement of multiple sound sources from a micro aerial vehicle,” Proc. ACM Multimedia, 2017.

Application of AVQ - Sound source tracking



[4] L. Wang, R. Sanchez-Matilla, and A. Cavallaro, “Tracking a moving sound source from a multi-rotor drone,” in Proc. IROS, 2018.

Application of AVQ - Sound source tracking



[4] L. Wang, R. Sanchez-Matilla, and A. Cavallaro, “Tracking a moving sound source from a multi-rotor drone,” in Proc. IROS, 2018.

Dataset

http://cis.eecs.qmul.ac.uk/projects/avq/

CIS centre for intelligent sensing

L. Wang, R. Sanchez-Matilla, and A. Cavallaro, “Audio-visual sensing from a quadcopter: dataset and baselines for source localization and sound enhancement”, Proc. IROS, 2019

Queen Mary University of London