Rotor noise-aware noise covariance matrix estimation for UAV audition

Benjamin Yen[1], Yameizhen Li[2], Yusuke Hioka[2]

[1] Nakadai Lab, Tokyo Institute of Technology, Japan
[2] Acoustics Research Centre, University of Auckland, New Zealand

Introduction

- Audio recording using unmanned aerial vehicles (UAVs) have shown to be challenging due to the high noise levels radiated from the UAV rotors.
- The study in [1] showed that a highly accurate noise covariance matrix (NCM) for rotor noise is vital in clear signal extraction performance.
- We propose a NCM estimation method where its *amplitude* is estimated through a multi-sensory rotor noise power spectral density (PSD) estimator (based on [2]), and its *phase* estimated by exploiting the microphone array design utilised in this study.
- We demonstrate that the estimated NCM can be incorporated into several source enhancement algorithms to reduce the effects of rotor noise.

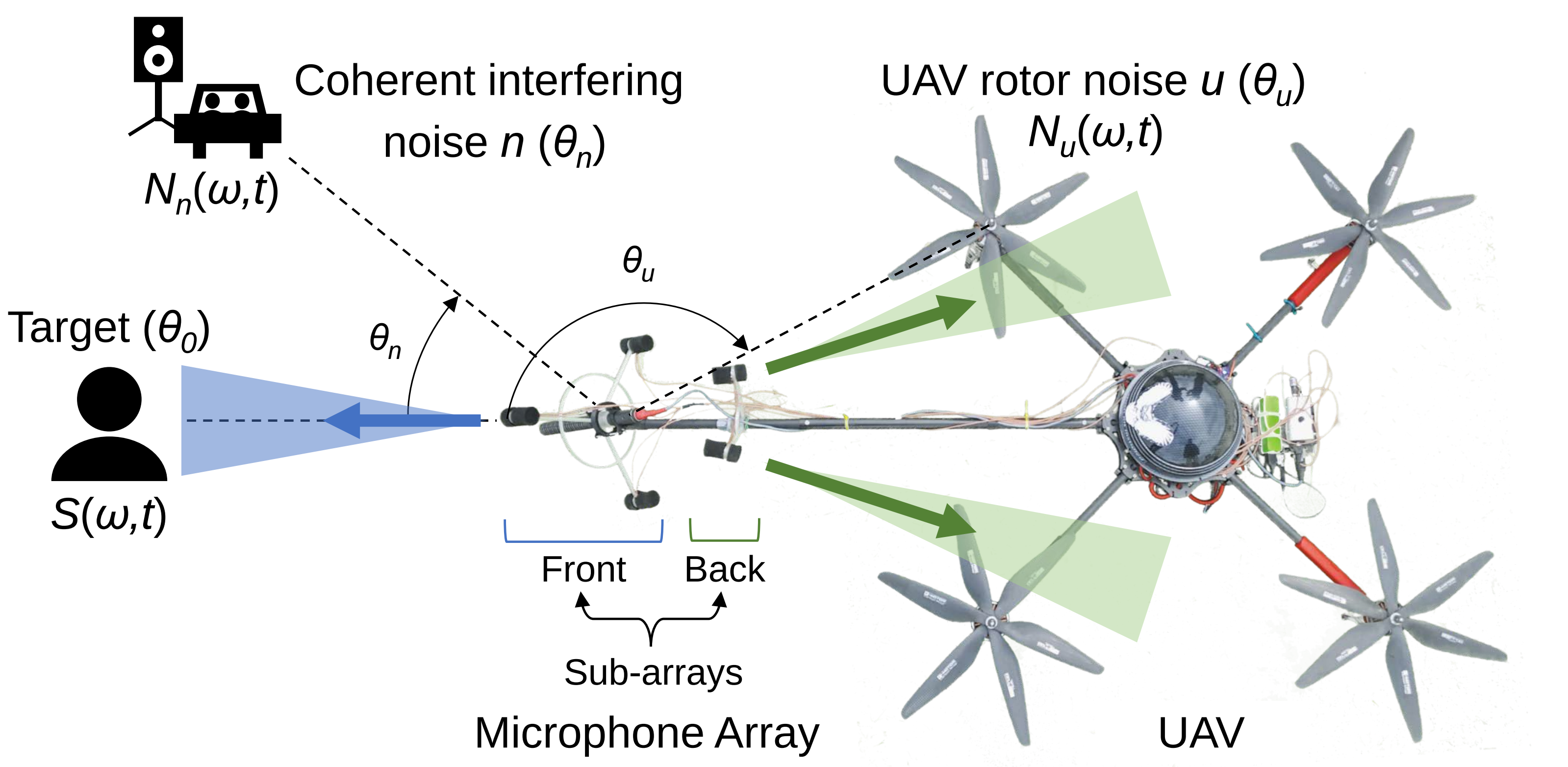

Fig. 1. Audio recording UAV overview (top view).

Problem Setup

The problem assumes an *M*-channel recording $\mathbf{x}(\omega,t)$. This can be broken down as:

- $\mathbf{x}_S(\omega,t)$: microphone signal for target source,
- $\mathbf{x}_{\mathrm{rot}}(\omega,t)=\sum_u \mathbf{x}_u(\omega,t)$: mic. signal for *U spatially coherent* rotor noise sources.
- $\mathbf{x}_{\mathrm{int}}(\omega,t)=\sum_n \mathbf{x}_n(\omega,t)$: mic. signal for *K – U spatially coherent* interfering noise sources

$$\mathbf{x}(\omega,t)=\mathbf{x}_S(\omega,t)+\mathbf{x}_{\mathrm{rot}}(\omega,t)+\mathbf{x}_{\mathrm{int}}(\omega,t) \text{ where } \mathbf{x}(\omega,t)=\left[X_1(\omega,t),\ \dots,\ X_M(\omega,t)\right]^T. \tag{1}$$

$$\mathbf{x}(\omega,t)=\left[\mathbf{x}_{\mathbf{f}}(\omega,t),\mathbf{x}_{\mathbf{b}}(\omega,t)\right]. \tag{2}$$

i.e. the mic array can be broken down into a front and back sub-arrays

$$\mathbf{R}(\omega,t)\cong\mathbf{R}_S(\omega,t)+\mathbf{R}_{\mathrm{rot}}(\omega,t)+\mathbf{R}_{\mathrm{int}}(\omega,t). \tag{3}$$

i.e. the covariance matrices can be approximated like that of the mic array signals

Rotor noise-aware NCM estimation method

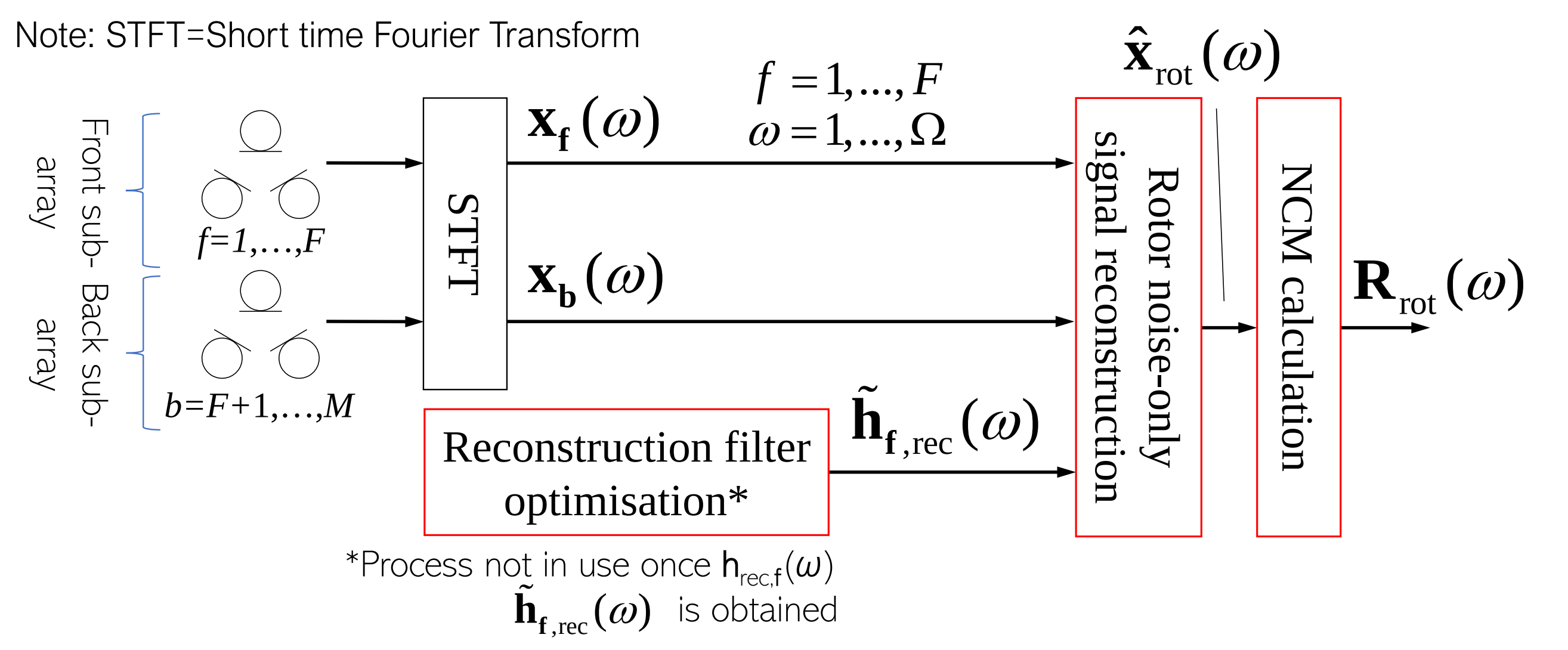

Fig. 2. Framework of the proposed method wo/ rotor noise PSD informed amplitude estimation [2] (**NCM**).

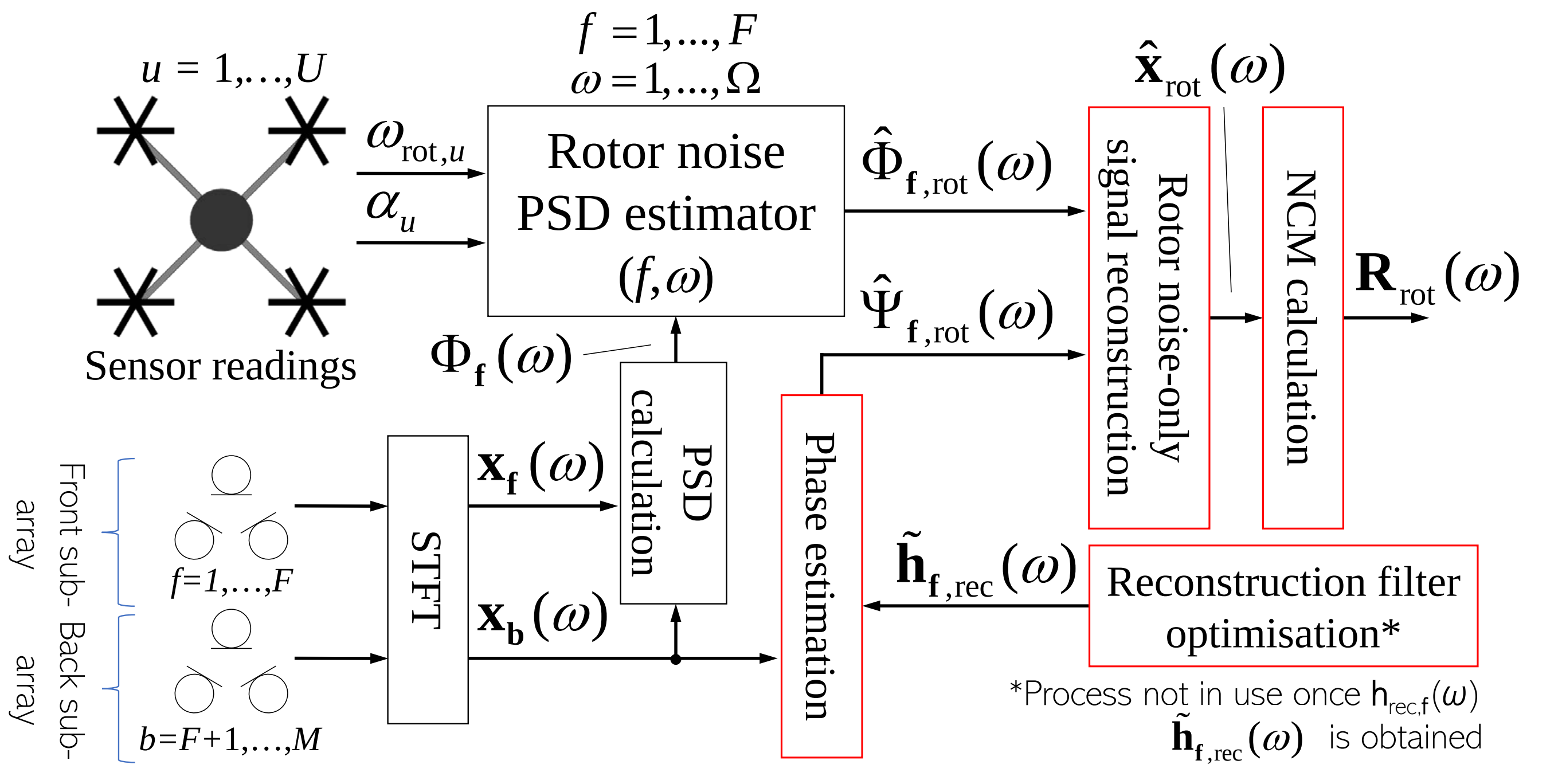

Fig. 3. Framework of the proposed method w/ rotor noise PSD informed amplitude estimation [2] (**iNCM**).

- The NCM estimation method exploits the following aspects of the problem setting and UAV-system:
  1) UAV rotor noise is significantly louder than any other sound sources.
  2) The UAV-system utilise directional microphones.
  3) The relative positions between the UAV rotors and the microphone array are fixed.
- Since the back sub-array which directional microphones faces the UAV rotors (see Fig.1), we assume $\mathbf{x}_{\mathbf{b}}(\omega,t)\approx\mathbf{x}_{\mathbf{b},\mathrm{rot}}(\omega,t)$, leading to a filter $\mathbf{h}^H_{f,\mathrm{rec}}(\omega)$ that reconstruct rotor noise components in the front sub-array $\mathbf{x}_{\mathbf{f}}(\omega,t)$ using $\mathbf{x}_{\mathbf{b}}(\omega,t)$ as follows:

$$\hat{X}_{f,\mathrm{rot}}(\omega,t)=\mathbf{h}^H_{f,\mathrm{rec}}(\omega)\mathbf{x}_{\mathrm{b}}(\omega,t) \quad \text{where} \quad f\in\mathbf{f}, \tag{4}$$

$$\varepsilon(\omega,t)=X_{f,\mathrm{rot}}(\omega,t)-\mathbf{h}^H_{f,\mathrm{rec}}\mathbf{x}_{\mathbf{b},\mathrm{rot}}(\omega,t), \tag{5}$$

$$J(\mathbf{h}^H_{f,\mathrm{rec}}(\omega))=E\left[|\varepsilon(\omega,t)|^2\right], \tag{6}$$

$$\frac{dJ(\mathbf{h}_{f,\mathrm{rec}}(\omega))}{d\mathbf{h}^H_{f,\mathrm{rec}}(\omega)}=0, \quad (7) \quad \xrightarrow{\qquad} \quad \tilde{\mathbf{h}}_{f,\mathrm{rec}}(\omega)=2\mathbf{R}_{\mathbf{b}}^{-1}(\omega)\mathbf{r}_f(\omega), \quad (8)$$

$$\hat{\mathbf{x}}_{\mathrm{rot}}(\omega,t)=\left[\hat{\mathbf{x}}_{\mathbf{f},\mathrm{rot}}(\omega,t)\mathbf{x}_{\mathbf{b},\mathrm{rot}}(\omega,t)\right]^T, \tag{9}$$

$$\boxed{\hat{\mathbf{R}}_{\mathrm{rot}}(\omega,t)=E\left[\hat{\mathbf{x}}_{\mathrm{rot}}(\omega,t)\hat{\mathbf{x}}^H_{\mathrm{rot}}(\omega,t)\right]. \quad (10)}$$

- However, in practice, the back sub-array's signal is by no means a perfect rotor noise reference. Hence, to further improve the NCM's effectiveness, we replace its amplitude component with one estimated from a machine learning-based multi-sensory rotor noise PSD estimator [2] (**iNCM**) as follows:

$$\hat{\mathbf{R}}_{\mathrm{rot}}(\omega,t)=E\left[\left|\hat{\mathbf{x}}_{\mathrm{rot}}(\omega,t)\right|\left|\hat{\mathbf{x}}_{\mathrm{rot}}(\omega,t)\right|^T\odot\left(e^{j\hat{\Psi}_{\mathrm{rot}}(\omega,t)}\left(e^{j\hat{\Psi}_{\mathrm{rot}}(\omega,t)}\right)^H\right)\right], \tag{11}$$

$$\hat{\Phi}_{\mathbf{f},\mathrm{rot}}(\omega,t)=\mathrm{diag}\left(E\left[\left|\hat{\mathbf{x}}_{\mathbf{f},\mathrm{rot}}(\omega,t)\right|\left|\hat{\mathbf{x}}_{\mathbf{f},\mathrm{rot}}(\omega,t)\right|^T\right]\right). \tag{12}$$

Experiments and Results

- Proposed NCM evaluated through outdoor experiments with the UAV system shown in Fig. 4.
- Target speech and impulse response recorded as shown in Fig. 5.
- UAV rotor noise recorded through a flying drone (see Fig. 6).
- Evaluated by the following metrics:
  - i) Signal-to-rotor-noise-ratio improvement (SRNRi)
  - ii) Signal-to-interfering-noise-ratio improvement (SINRi).
  - iii) Short-time intelligibility measure improvement (SRNRi).
  - iv) Perceptual eval. of speech quality improvement (PESQi).

Findings from Table 1:

- The proposed NCM and iNCM improves the MVDR-DWF baseline over most metrics, especially SRNRi and PESQi.
- Performance is especially apparent at target-to-UAV distances ≤ 5m.

Findings from Table 2:

- While the proposed NCM and iNCM delivered lower SRNRi than its SP-MWF baseline, it has significantly improved STOIi and PESQi.
- Results indicate that the proposed method resolves the overfiltering issue from SP-MWF.

Overall: The proposed NCM improves the audio quality of the resultant signal output.

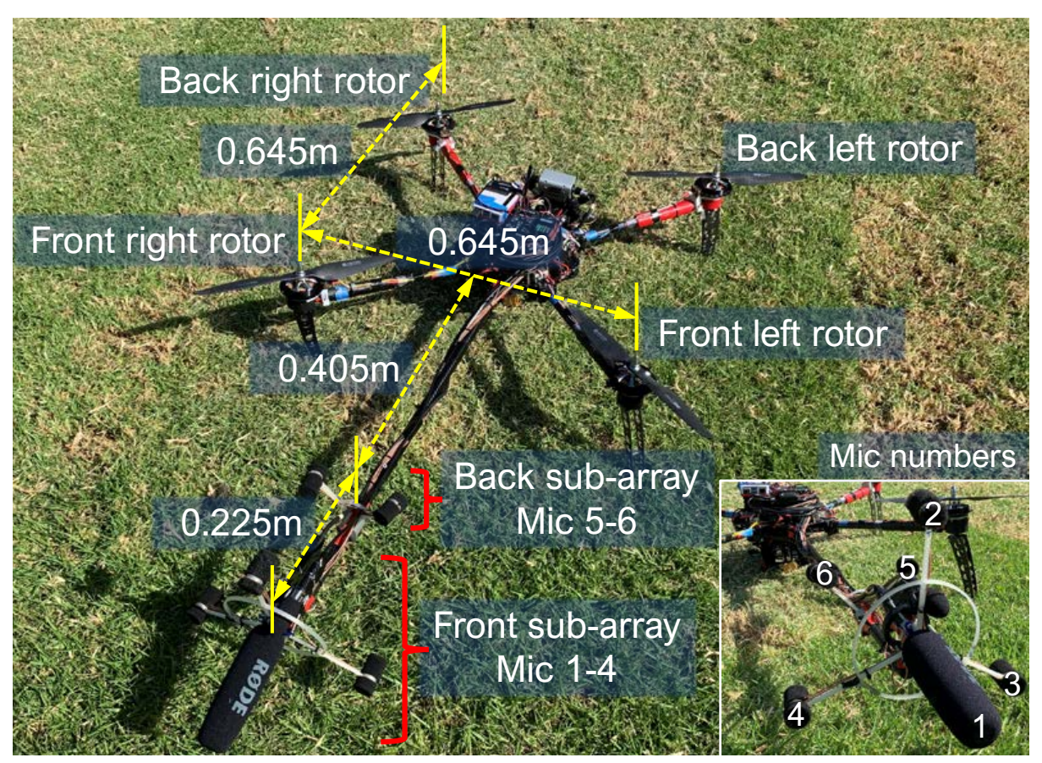

Fig. 4. UAV system setup for experiments.

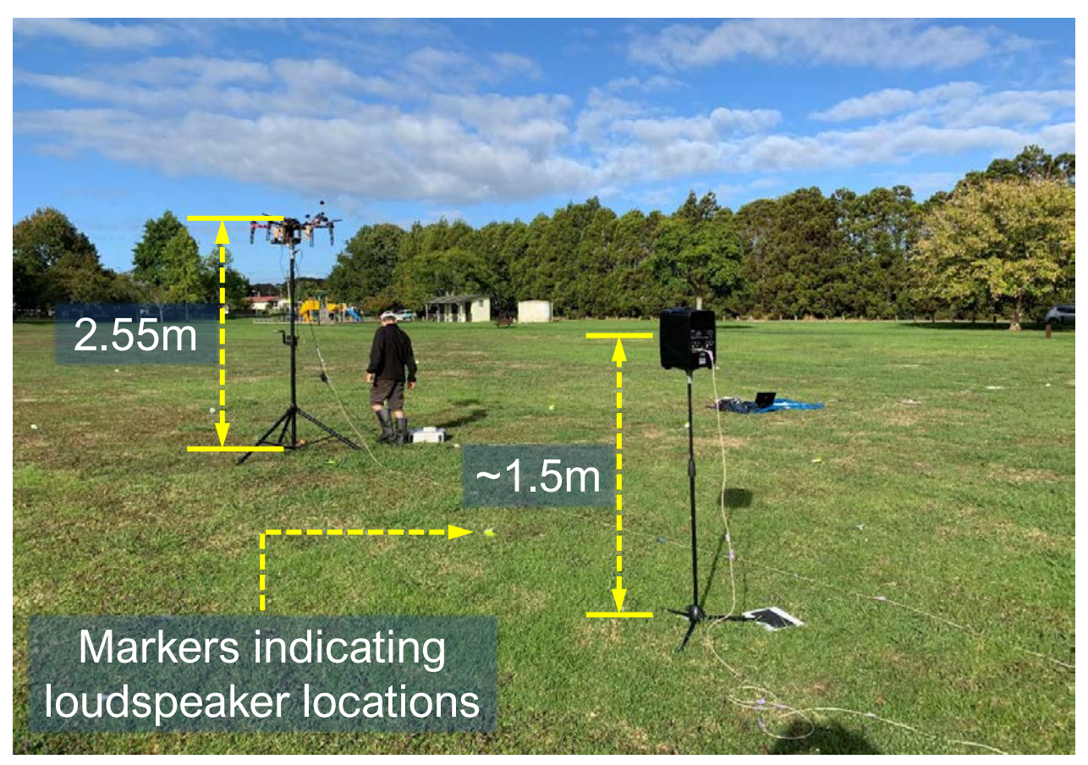

Fig. 5. Experiment setup for recording target speech and impulse response.

Fig. 6. Recording UAV in-flight rotor noise.

Table 1. Results for beamforming with postfiltering-based methods. **Bold**=Outperformed baseline, *Italics*=highest mean performance. (MVDR=minimum variance distortionless response beamformer [3])

Metric	Source enhancement framework	Target source sound level (dBA)			
		80			
		Target source-to-UAV distance (m)			
		3	5	10	15
SRNRi	MVDR [3]	10.08	0.98	11.83	15.03
	MVDR-DWF [2]	20.58	18.20	20.39	16.00
	NCM-DWF (proposed)	22.87	14.25	22.19	17.02
	iNCM-DWF (proposed)	**23.87**	***23.28***	***25.43***	*17.76*
SINRi	MVDR [3]	4.16	3.69	8.74	19.12
	MVDR-DWF [2]	7.73	7.99	9.89	14.12
	NCM-DWF (proposed)	6.50	7.52	9.35	13.76
	iNCM-DWF (proposed)	7.00	8.27	9.83	13.00
STOIi	MVDR [3]	0.102	0.104	*0.070*	*0.022*
	MVDR-DWF [2]	0.097	0.111	0.031	-0.001
	NCM-DWF (proposed)	*0.111*	*0.126*	0.046	0.012
	iNCM-DWF (proposed)	0.103	0.111	0.023	-0.002
PESQi	MVDR [3]	0.012	-0.060	-0.065	0.001
	MVDR-DWF [2]	0.054	-0.017	-0.063	-0.009
	NCM-DWF (proposed)	**0.190**	0.014	-0.041	-0.010
	iNCM-DWF (proposed)	**0.192**	*0.050*	-0.053	-0.005

Table 2. Results for speech distortion weighted multichannel WF(SDW-MWF) based methods. **Bold**=Outperformed baseline, *Italics*=highest mean performance. (SP-MWF=spatial prediction SDW-MWF [4])

Metric	Source enhancement framework	Target source sound level (dBA)			
		80			
		Target source-to-UAV distance (m)			
		3	5	10	15
SRNRi	SP-MWF [4]	15.80	19.49	21.32	21.90
	NCM-SP-MWF (proposed)	2.47	2.37	5.61	10.22
	iNCM-SP-MWF (proposed)	7.47	7.50	12.51	16.96
SINRi	SP-MWF [4]	4.40	6.56	11.29	11.34
	NCM-SP-MWF (proposed)	3.28	3.13	6.43	11.06
	iNCM-SP-MWF (proposed)	4.42	4.36	9.42	*14.02*
STOIi	SP-MWF [4]	-0.189	-0.133	-0.133	-0.105
	NCM-SP-MWF (proposed)	**0.021**	**0.032**	**0.014**	***0.013***
	iNCM-SP-MWF (proposed)	**0.040**	**-0.054**	**0.023**	**0.010**
PESQi	SP-MWF [4]	0.008	-0.047	-0.093	-0.002
	NCM-SP-MWF (proposed)	0.222	0.166	*0.276*	*0.229*
	iNCM-SP-MWF (proposed)	0.268	0.137	0.075	0.226

References

[1] Y. Li, B. Yen and Y. Hioka. "Performance evaluation on multi-channel Wiener filter based speech enhancement for unmanned aerial vehicles recordings" in *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, Vol. 263, No. 3, pp. 3584-3594, Aug, 2021.
[2] B. Yen, Y. Hioka, G. Schmid and B. Mace. "Multi-sensory sound source enhancement for unmanned aerial vehicle recordings", *Applied Acoustics*, 189, 108590, 1-22, Feb, 2022.
[3] H. Cox, R. M. Zeskind, and M. M. Owen, "Robust adaptive beamforming," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35, no. 10, pp. 1365–1376, 1987.
[4] B. Cornelis, M. Moonen, and J. Wouters, "Performance analysis of multichannel Wiener filter based noise reduction in hearing aids under second order statistics estimation errors," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 5, pp. 1368–1381, July 2011.

Conclusions

- We demonstrated that by utilising the back sub-array input signals and support from an accurate estimation of the rotor noises' PSD, the front sub-array's rotor noise-only NCM can be effectively estimated.
- Experimental evaluation with a flying UAV shows that the proposed NCM estimation incorporates well with several MWF-based source enhancement frameworks.
- Future work includes further development of the NCM estimation method to improve SINRi, as well as expand experimental evaluation to include moving sound sources with a wider range of environmental settings.