Relative Transfer Matrix for Drone Audition Applications: Source Enhancement

Wageesha N. Manamperi, Thushara D. Abhayapala
The Australian National University, Canberra, Australia
E-mail: {*wageesha.manamperi, thushara.abhayapala*}@*anu.edu.au*

***Abstract*—The Relative Transfer Matrix (ReTM), recently introduced as a generalization of the relative transfer function for multiple sources and multiple microphones, shows promising performance when applied to speech denoising and speaker separation in a noisy reverberant room. This work utilizes the ReTM to propose a novel framework for drone noise suppression. Difficulties in noise cancellation for drone audition applications arise due to its self-generated noise that causes an adverse noisy environment of low Signal-to-Drone Noise Ratio (SDNR) levels. In this paper, we divide the drone on-board microphones into two multichannel groups to approximately estimate the drone noise from one group to the other with known drone noise ReTM for denoising. We demonstrate the ReTM spatial mapping ability in both indoor and outdoor experiments using hovering drone on-board microphone recordings with low average magnitude spectrum error. Finally, we validate the method in a real-life environment for source signal enhancement over different S$\overline{\text{D}}$NR conditions and offer both improved speech intelligibility and signal-to-distortion ratio.**

I. Introduction

Making drones hear is a fundamental requirement of drone audition functions, such as sound source localization [1]–[5], source separation [6], source tracking [7], [8], and signal enhancement [9]–[13]. However, (i) its self-generated noise, and (ii) on-board microphones being closer to the drone noise sources compared to the desired sound source on the ground, cause a highly adverse noisy environment with an extremely low signal-to-drone noise ratio (S$\overline{\text{D}}$NR) (defined as the power ratio between the source signal and the drone noise) level [14]. As a result, drone noise reduction is in itself an active problem in acoustic signal processing that enables drone audition applications, e.g., search and rescue missions.

Many early drone noise reduction approaches are based on the beamforming [9], [15]–[19]. Other popular algorithms are spatial filtering-based methods [20], [21], blind source separation [20], [22], [23], spherical sector harmonics-based methods [24], as well as supervised approaches including template-based approaches [25], correlation matrix-based methods [26], and reference-based methods [27], [28]. Recently favored are machine learning-based techniques, the well-known form of supervised approaches, for drone noise suppression [29]–[32]. For example, a denoising autoencoder based on fully convolutional neural networks to cancel drone noise was discussed in [32]. While these methods suppress the drone noise in low S$\overline{\text{D}}$NR levels, they suffer from difficulties in obtaining an accurate estimate of the drone noise signals from on-board multichannel recordings due to the contribution of the noise source signals. This paper utilizes the recently proposed Relative Transfer Matrix (ReTM) [33], which is defined as the generalization of the relative transfer function for multiple simultaneous sound sources and multiple microphones, to propose a signal-independent solution that solely depends on the drone noise sources position in a static acoustic environment for drone noise cancellation.

ReTM is a new spatial feature containing relevant cues of the multiple sound sources, and independent of the emitted source signals, is defined by allocating the receivers into two multi-microphone groups [33]. Related works [34]–[36] highlight the promising performances of the ReTM methods when applied to speech enhancement, and speaker separation in a multi-source noisy reverberant environment.

In this paper, we present a novel algorithm for drone noise reduction at very low S$\overline{\text{D}}$NR levels using the ReTM. The derivation of this method is similar to [34] in a room scenario that exploits covariance matrices of the multiple noise sources-only signals by dividing the receivers into two multi-microphone groups based on the ReTM presented in [33]. Compared to [34], the proposed algorithm is (i) applied for drone noise estimation, (ii) analyzed the estimation accuracy in terms of the average magnitude spectrum error using two drone acoustic datasets, i.e., DREGON [37] and AVQ dataset [11] in indoor and outdoor environments, respectively, and (iii) presented the speech denoising performance at extreme S$\overline{\text{D}}$NR conditions (< -10 dB). We verify the proposed method's applicability for source enhancement via a real-life experiment conducted outdoors, and show increased speech intelligibility and signal-to-distortion ratio (SDR) over differing S$\overline{\text{D}}$NR level.

II. Problem Formulation

Let there be Q microphones mounted on a drone capturing the sound produced by a target speech source. The signal received by the q^{th} microphone due to the source in time domain is

$$m_q(n) = d_q(n) * s(n) + \sum_{\ell=1}^{\mathcal{L}} h_{q,\ell}(n) * v_\ell(n), \tag{1}$$

where $d_q(n)$ is the drone-related impulse response from the sound source to the q^{th} microphone, $s(n)$ is the source signal, $v_\ell(n)$ is the drone noise due to the ℓ^{th} motor, $h_{q,\ell}(n)$ is the

Thanks to Australian Research Council (ARC) Discovery Project Grant DP200100693 for funding.

impulse response from the ℓ^{th} motor to the q^{th} microphone and $\mathcal{L}$ is the number of motors on the drone and '$*$' is the convolution operator.

In the short-time Fourier transform (STFT) domain, (1) can be written as

$$M_q(f,t) = d_q(f)S(f,t) + \sum_{\ell=1}^{\mathcal{L}} H_{q,\ell}(f)V_\ell(f,t), \quad (2)$$

where $M_q(f,t)$, $S(f,t)$, and $V_\ell(f,t)$ are the STFTs of $m_q(n)$, $s(n)$ and $v_\ell(n)$, respectively, $d_q(f)$ is the drone-related transfer function [38] for the q^{th} microphone, $H_{q,\ell}(f)$ is the noise transfer function and t and f are the frame and frequency indices, respectively with $t \in \{1,\ldots,T\}$ and $f \in \{1,\ldots,F\}$ where T and F are the number of time frames and frequency bins, respectively. We simplify (2) further and can be expressed in matrix form as

$$\mathbf{M}(f,t) = \mathbf{d}(f)S(f,t) + \mathbf{H}(f)\mathbf{V}(f,t), \quad (3)$$

where $\mathbf{M}(f,t) = [M_1(f,t),\ldots,M_Q(f,t)]^T$ and $\{\cdot\}^T$ is the matrix transpose, and $\mathbf{d}(f)$, is the $[Q \times 1]$ vector of drone-related transfer functions, $\mathbf{H}(f)$ is the Q by $\mathcal{L}$ drone noise sources transfer function matrix, and $\mathbf{V}(f,t)$ are the $\mathcal{L}$ drone noise source signals that similarly defined.

The problem discussed in this paper is to suppress the drone noise by estimating the relative transfer matrix (ReTM) of the $\mathcal{L}$ drone noise sources from the received microphone signals, which we explain next.

III. Drone Noise Estimation and Reduction with Relative Transfer Matrix

This section uses the ReTM to propose a drone noise reduction algorithm. We first briefly review the ReTM presented in [33], which divides the multiple microphones into two groups to relate the signal received at the first microphone group to the second. We then blindly estimate the drone noise sources ReTM using covariance matrices. Finally, we utilize the drone noise ReTM to map the drone noise from one group to the other for denoising.

A. Drone Noise Sources Relative Transfer Matrix

Consider two microphone groups, $\{A\}$ and $\{B\}$ with Q_A and Q_B microphones, respectively ($Q = Q_A + Q_B$). Let $\mathbf{M}_A(f,t) = [M_1^{(A)}(f,t),\cdots,M_{Q_A}^{(A)}(f,t)]^T$, and $\mathbf{M}_B(f,t) = [M_1^{(B)}(f,t),\cdots,M_{Q_B}^{(B)}(f,t)]^T$ denote the vector of microphone signals for each group. The received signals at each microphone group in matrix form can be written as

$$\mathbf{M}_A(f,t) = \mathbf{d}_A(f)S(f,t) + \mathbf{H}_A(f)\mathbf{V}(f,t), \quad (4)$$

$$\mathbf{M}_B(f,t) = \mathbf{d}_B(f)S(f,t) + \mathbf{H}_B(f)\mathbf{V}(f,t). \quad (5)$$

The drone noise sources ReTM, $\boldsymbol{\mathcal{R}}_{AB}(f)$, can be defined as in [33]

$$\boldsymbol{\mathcal{R}}_{AB}(f) \triangleq \mathbf{H}_A(f)\mathbf{H}_B(f)^\dagger, \quad (6)$$

where $(\cdot)^\dagger$ is Moore-Penrose inverse, assuming the validity, i.e., $Q_B \geq \mathcal{L}$.

Consider the received signals by microphone groups when the speech is inactive ($S(f,t) = 0$), then the drone noise-only signals can be obtained from the received signals. Following the ReTM mapping as in [33], we can relate the drone noise signals received by microphone group $\{A\}$ with ReTM as

$$\mathbf{M}_A^{(v)}(f,t) = \boldsymbol{\mathcal{R}}_{AB}(f)\mathbf{M}_B^{(v)}(f,t). \quad (7)$$

Note that $\boldsymbol{\mathcal{R}}_{AB}(f)$ is defined by the spatial properties of the noise sources such that it is independent of the sound emitted by the drone motors and propellers. In drone audition applications, the $\boldsymbol{\mathcal{R}}_{AB}(f)$ is constant for a drone hovering scenario.

We note that it is typically straightforward to estimate the ReTM in (6) by assuming that segments of recording containing the drone noise-only signal when the drone flying in the hovering manoeuvre above the ground prior to the speech source being active, which we present in next subsection.

B. Blind Estimation of the ReTM from Received Signals

Here, we process the received signals in (4), and (5) when the speech source is inactive to estimate the drone noise sources ReTM that will be the building block of our proposed drone noise reduction algorithm.

Formatting the ReTM estimation using covariance matrices of

$$\begin{aligned}\boldsymbol{\mathcal{P}}_{AA}^{(v)}(f) &\triangleq E\{\mathbf{M}_A^{(v)}(f,t)\mathbf{M}_A^{(v)^*}(f,t)\},\\ \boldsymbol{\mathcal{P}}_{BA}^{(v)}(f) &\triangleq E\{\mathbf{M}_B^{(v)}(f,t)\mathbf{M}_A^{(v)^*}(f,t)\},\end{aligned} \quad (8)$$

where $E\{\cdot\}$ denotes the expectation which can be obtained by averaging across the time frames, and then with further simplifications as in [33], we approximate

$$\boldsymbol{\mathcal{R}}_{AB}(f) \approx \boldsymbol{\mathcal{P}}_{AA}^{(v)}(f)\left(\boldsymbol{\mathcal{P}}_{BA}^{(v)}(f)\right)^\dagger. \quad (9)$$

We observe from (9) that the drone noise sources ReTM ($\boldsymbol{\mathcal{R}}_{AB}(f)$) only relies on the estimation of the covariance matrices of the drone noise-only signals (no active speech), and does not require the number of noise sources in the mixture. The common approach is to obtain the noise-only segments using a Voice Activity Detection (VAD) algorithm, however, due to drones being loud these algorithms are not recommended. In practice, the drone noise-only segments can be isolated by initiating to activate the speech source after the drone broadcasts a specific signal, such as a tone. Although blind estimation of ReTM is used in this paper, the proposed method can also be implemented by pre-trained or semi-blind estimation.

The next section proposes how to suppress the drone noise to enhance the target speech captured by a hovering drone.

C. Drone Noise Suppression using Drone Noise Sources ReTM

Assuming that the drone noise sources ReTM, $\boldsymbol{\mathcal{R}}_{AB}(f)$ is accurately estimated, thus, we can approximately calculate the drone noise signals at microphone group $\{A\}$ by following the ReTM mapping in (7), we approximate that

$$\boldsymbol{\mathcal{R}}_{AB}(f)\mathbf{M}_B^{(v)}(f,t) \approx \mathbf{H}_A(f)\mathbf{V}(f,t). \quad (10)$$

We can subtract (10) from the received signals/noisy speech at the microphone group $\{A\}$ to remove the drone noise from the group $\{A\}$ microphone signals as

$$\mathbf{M}_A(f,t) - \boldsymbol{\mathcal{R}}_{AB}(f)\mathbf{M}_B(f,t) \approx \mathbf{d}_A(f)S(f,t), \tag{11}$$

thus, enhancing the target speech. However, in practice, the approximation of (10) is not accurate. To make the exposition concise, we omit the dependency of time (t) and frequency (f) in the rest of this section.

However, we accomplish a 'distorted' version of the target speech signal S by expanding the left side of (11), such that

$$\begin{aligned}\hat{\mathbf{S}} &= \mathbf{M}_A - \boldsymbol{\mathcal{R}}_{AB}\mathbf{M}_B,\\ &= \mathbf{d}_A S + \mathbf{H}_A\mathbf{V} - \boldsymbol{\mathcal{R}}_{AB}(\mathbf{d}_B S + \mathbf{H}_B\mathbf{V}),\\ &= \mathbf{d}_A S + \mathbf{H}_A\mathbf{V} - \boldsymbol{\mathcal{R}}_{AB}\mathbf{d}_B S - \underbrace{\boldsymbol{\mathcal{R}}_{AB}\mathbf{H}_B}_{\mathbf{H}_A}\mathbf{V},\\ &= \mathbf{d}_A S + \mathbf{H}_A\mathbf{V} - \boldsymbol{\mathcal{R}}_{AB}\mathbf{d}_B S - \mathbf{H}_A\mathbf{V},\\ &= \mathbf{d}_A S - \boldsymbol{\mathcal{R}}_{AB}\mathbf{d}_B S,\\ &= \underbrace{[\mathbf{d}_A - \boldsymbol{\mathcal{R}}_{AB}\mathbf{d}_B]}_{\text{distortion}}S,\end{aligned} \tag{12}$$

where $\hat{\mathbf{S}}$ is a $Q_A \times 1$ vector consists Q_A copies of estimated target speech signal S. We obtain a complete suppression of the drone noise from the received signals in (12), along with an accurate estimate of drone noise source ReTM, $\boldsymbol{\mathcal{R}}_{AB}$. However, we note that this is a 'distorted' version of the target speech signal S in terms of both the drone noise sources transfer matrix and transfer function of the speech source, but in future, the ReTM may be exploited using multiple copies of denoised speech to do further enhancement.

IV. Experiments

In this section, we illustrate the proposed method's performance using real-life multichannel recordings of a hovering drone from two publicly available drone acoustic datasets; (i) DREGON [37], and (ii) AVQ [11]. The experimental procedures and results are described. First, the DREGON dataset was used to measure the drone noise estimation performance in an indoor environment. Second, the AVQ dataset was examined to analyze the algorithm performance in an outdoor environment.

The proposed method's performance is analyzed using a qualitative metric in terms of the average magnitude spectrum error of the first microphone in group $\{A\}$ as in [33]

$$\epsilon_1(f) = \operatorname*{mean}_t 10\log_{10}\frac{|\hat{M}_1^{(A)}(f,t) - M_1^{(A)}(f,t)|^2}{|M_1^{(A)}(f,t)|^2},$$

where $\hat{M}_1^{(A)}(f,t)$ denotes the estimated drone noise signal at the first microphone of the group $\{A\}$ calculated using (7) with the drone noise sources ReTM, which we will refer to as *'remote'* signals from here on. We note that microphones are randomly grouped as group $\{A\}$ and $\{B\}$, such that no microphone picks twice [33]. But in the future, we discuss the microphone group selection to provide a complete analysis of the performance variation.

A. Indoor Environments

1) Experimental Recordings: We used *hovering* audio signals ('DREGON_hovering_nosource_room2.wav.') of the '*In Flight Noise-Only Recordings*' from the DREGON dataset in [37] conducted in a $10 \times 10 \times 2.5$ m rectangular room ($T_{60} = 150$ ms). These recordings were made with a flying drone without any speech sources. The DREGON dataset has a constellation of 8 microphones in a cubic-shaped structure placed below the motor-propeller plane of the drone. We assigned 4 microphones on the bottom (channels $\{0, 2, 4, 6\}$) to group $\{A\}$, and 4 microphones on the top (channels $\{1, 3, 5, 7\}$) to group $\{B\}$. The audio recordings were processed directly in the short-time Fourier domain for a 2^{18} window size, 44.1 kHz sampling, and 10 second duration that was long enough to satisfy the multiplicative transfer function [39].

2) Results: We discuss the magnitude spectrum error of the drone noise estimation using the ReTM in Fig. 1 and reconstruction of the drone noise signals at microphone group $\{A\}$ in Fig. 2. Fig. 1 shows the magnitude spectrum of the measured signal $M_1^{(A)}$, and remote signal $\hat{M}_1^{(A)}$ for a single time frame as well as their error averaged over all time frames. We observe that there is no difference between $M_1^{(A)}$ and $\hat{M}_1^{(A)}$, which is equivalent to being a perfect mapping. This is further confirmed by a very low error observed throughout the frequency band of $20 - 16$k Hz. The results suggest that the drone noise sources ReTM exactly map the drone noise source's spatial properties and accurately estimate the drone noise signals at the microphone channel 0.

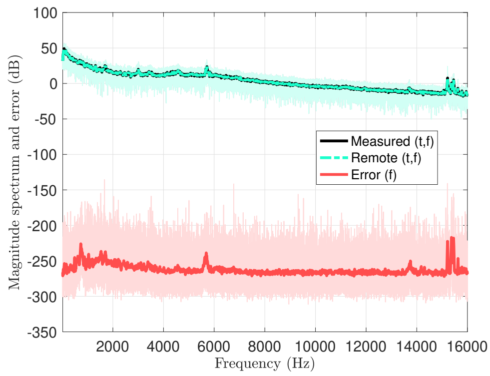

Fig. 1. Magnitude spectrum and error averaged over time for in flight noise-only recordings from the DREGON dataset [37].

We present both time domain and spectrogram plots of the measured and remote signals in Fig. 2. Strong similarity is once again shown for hovering recordings confirming the accuracy of the estimated drone noise sources ReTM. We share a link to the audio files on GitHub[1].

B. Outdoor Environments

1) Experimental Recordings: We utilized an AVQ dataset [11] recorded outdoors from a fixed drone position on a

[1] https://github.com/wnilmini/ReTM_Source_Enhancement/tree/main/Drone%20noise%20Estimation

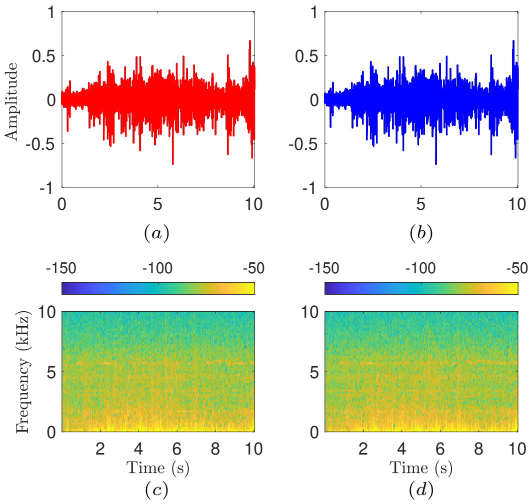

Fig. 2. Measured and remote drone noise signals at channel 0 at microphone group $\{A\}$: (a,c) measured signals, (b,d) remote signals, (a,b) time domain plots, and (c,d) spectrogram plots.

tripod at the height of 1.8 m with two static speech sources (scenario: 'subset 1'). We selected microphone recordings from the *'sequence 2'*, drone noise-only signals with a constant speed of 100% for ReTM mapping. The AVQ dataset has 8 microphones in a circular array located 15 cm above the body of the drone. We assigned 4 channels $\{4, 8, 1, 6\}$ to group $\{A\}$, and 4 channels $\{3, 5, 7, 2\}$ to group $\{B\}$. The 30 second long audio recordings were processed directly in the short-time Fourier domain for a 2^{20} window size, 44.1 kHz sampling rate.

2) Results: Fig. 3 displays the magnitude spectrum error of the drone noise estimation using the ReTM for the outdoor environment. A similar result is also obtained for the AVQ dataset as the DREGON dataset despite being outdoors, supporting that the drone noise sources ReTM is modeled correctly. As intended, both time domain and spectrogram plots of the remote signals in Fig. 4 are also seen to be an accurate estimate of the measured signals.

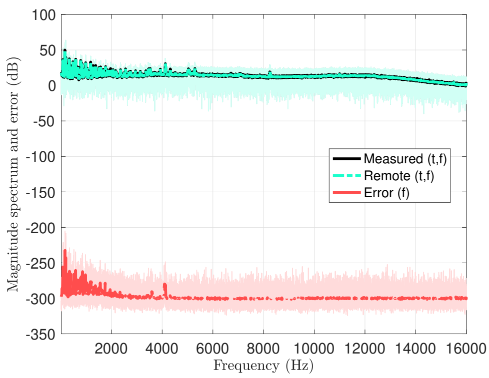

Fig. 3. Magnitude spectrum and error averaged over time for in flight noise-only recordings from the AVQ dataset [11].

These results suggest that the drone noise at microphone group $\{A\}$ can be accurately estimated using the drone noise at microphone group $\{B\}$ with the drone noise sources ReTM in both the indoor and outdoor environments. This method can be easily generalized to different types of drones using the drone noise-only recordings obtained when the drone is stationary e.g., hovering manoeuvre. In the next section, ReTM-based drone noise reduction algorithm will perform for a target speech enhancement.

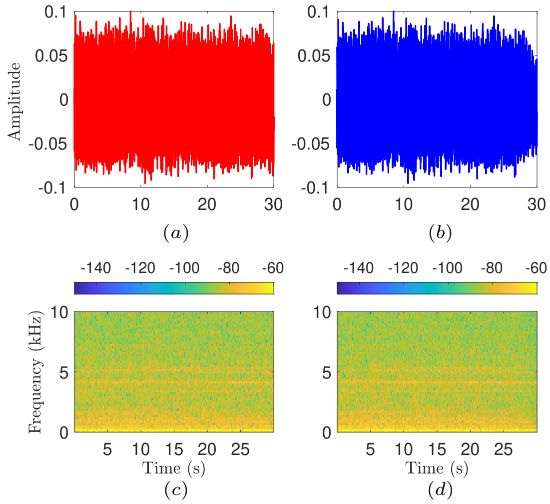

Fig. 4. Measured and remote drone noise signals at channel 4 at microphone group $\{A\}$: (a,c) measured signals, (b,d) remote signals, (a,b) time domain plots, and (c,d) spectrogram plots.

V. Application in Source Signal Enhancement

This section briefly examines the proposed drone noise reduction algorithm in an outdoor environment in the presence of a speech source at very low $\overline{\text{SDNR}}$ conditions (< -10 dB). Our intention is to analyze the speech enhancement performance over different $\overline{\text{SDNR}}$ levels in terms of signal distortion and speech intelligibility. We evaluate the quality of the enhanced speech using objective measures, (i) Signal-to-Distortion Ratio (SDR), using the BSS-Eval toolbox [40] and (ii) the Short-Time Objective Intelligibility score (STOI) [41]. We want to iterate that we do not consider comparing the proposed algorithm with baseline methods in this paper as our main focus here is to quantify the signal distortion introduced by the ReTM, in such a case, we expect that the proposed method can be performed lower STOI value than the state-of-the-art methods as in [10].

We used the AVQ dataset in [11] for the evaluation. The multichannel time-domain recordings are produced by adding the drone noise signals recorded at the microphone array separately (discussed in Section IV-B) with speech-only recordings from the '*sequence 4*' of the same subset. Then, the speech-only recordings are scaled to simulate different $\overline{\text{SDNR}}$ levels from -30 dB to -15 dB with an increment of 5 dB. We also used the same parameter settings and microphone group selection as in Section IV-B.

Fig. 5 shows the time-domain and spectrogram plots of the recorded and enhanced signals at the channel 4 (remote channel) of the microphone group $\{A\}$ at $\bar{\text{SDNR}}$ level of -20 dB. Immediately we observe that the proposed method significantly reduced the strong drone noise in the mixture signal (Fig. 5b & d). Additionally, a clear preservation of the speech spectrum is observed at the lower frequencies in Fig. 5d. We share a link to the enhanced signals[2].

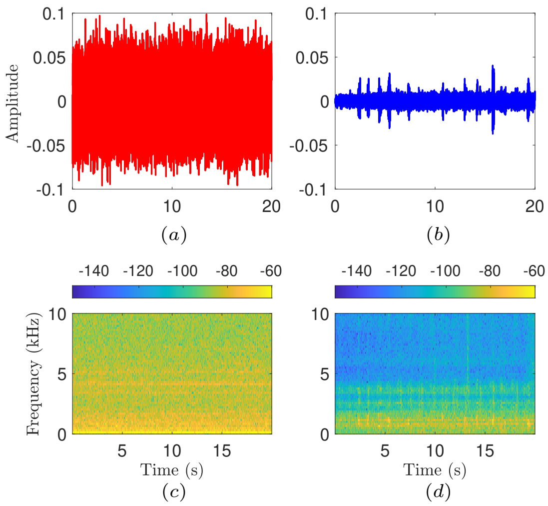

Fig. 5. Recorded and enhanced signals at channel 4 at microphone group $\{A\}$ at $\bar{\text{SDNR}}$ level of -20 dB: (a,c) recorded signals, (b,d) enhanced signals, (a,b) time domain plots, and (c,d) spectrogram plots.

TABLE I
SIGNAL ENHANCEMENT PERFORMANCE EVALUATION USING IMPROVEMENTS IN SDR AND INPUT/OUTPUT STOI FOR VARIOUS $\bar{\text{SDNR}}$ LEVELS

$\bar{\text{SDNR}}$ level	SDR (dB)	STOI (%)	
		Input	Output
-15 dB	2.04	20.55	47.86
-20 dB	-5.19	19.44	43.08
-25 dB	-6.88	17.14	40.15
-30 dB	-8.73	15.75	36.43

We will now examine the speech enhancement performance with the proposed drone noise reduction algorithm over -30 dB and -15 dB $\bar{\text{SDNR}}$ range. Table I depicts the improvements in SDR and input/output STOI scores as a function of $\bar{\text{SDNR}}$ levels. We observe that both the SDR and STOI performance have degraded with worsening $\bar{\text{SDNR}}$ condition. However, the STOI improvement is nearly double compared to the input scores over all $\bar{\text{SDNR}}$ levels. Again, SDR is slightly better with higher $\bar{\text{SDNR}}$ levels. Here we reiterate that this is due to distortion introduced with the drone noise sources ReTM in (12). In contrast, from the informal listening, we find that the enhanced speech signals are able to greatly increase the understandability of the speech content compared to the recorded signals. The results confirm that the proposed method better enhance the source signal in extreme noisy environments although both STOI and SDR are gradually degrading with decreased $\bar{\text{SDNR}}$ level.

[2]https://github.com/wnilmini/ReTM_Source_Enhancement/tree/main/Source%20Enhancement

VI. Conclusion

This paper proposed a novel algorithm to reduce the drone noise for drone audition applications using ReTM. We separate the multiple microphones mounted on a drone into two multi-microphone groups to estimate the drone noise from one group to the second. Extensive experimental studies using both indoor and outdoor real-life datasets confirmed the accuracy of the proposed method. This was shown by the very lower magnitude spectrum error obtained for the full frequency band. STOI and SDR values were shown to be improved for speech enhancement at very low $\bar{\text{SDNR}}$ levels in outdoor environments. In the future, the ReTM may be exploited using multiple copies of enhanced speech together filtering approaches to minimize the spectral distortion further.

References

[1] M. Wakabayashi, H. G. Okuno, and M. Kumon, "Drone audition listening from the sky estimates multiple sound source positions by integrating sound source localization and data association," *Adv. Robot.*, vol. 34, no. 11, pp. 744–755, Jun. 2020.

[2] T. Ohata, K. Nakamura, T. Mizumoto, T. Taiki, and K. Nakadai, "Improvement in outdoor sound source detection using a quadrotor-embedded microphone array," in *Proc. IEEE/RSJ Int. Conf. on Intell. Robots and Syst.*, Sep. 2014, pp. 1902–1907.

[3] W. Manamperi, T. D. Abhayapala, J. Zhang, and P. Samarasinghe, "Drone audition: Sound source localization using on-board microphones," *IEEE/ACM Trans. on Audio, Speech, and Lang. Process.*, vol. 30, pp. 508–519, Jan. 2022.

[4] K. Hoshiba, K. Washizaki, M. Wakabayashi, *et al.*, "Design of uav-embedded microphone array system for sound source localization in outdoor environments," *Sensors*, vol. 17, no. 11, p. 2535, 2017.

[5] B. Yen and Y. Hioka, "Noise power spectral density scaled snr response estimation with restricted range search for sound source localisation using unmanned aerial vehicles," *EURASIP J. on Audio, Speech, and Music Process.*, no. 1, pp. 1–26, Dec. 2020.

[6] T. Morito, O. Sugiyama, R. Kojima, and K. Nakadai, "Partially shared deep neural network in sound source separation and identification using a uav-embedded microphone array," in *Proc. IEEE/RSJ Int. Conf. on Intell. Robots and Syst.*, Oct. 2016, pp. 1299–1304.

[7] T. Yamada, K. Itoyama, K. Nishida, and K. Nakadai, "Sound source tracking by drones with microphone arrays," in *Proc. IEEE/SICE Int. Symp. on Syst. Integr.*, Jan. 2020, pp. 796–801.

[8] L. Wang, R. Sanchez-Matilla, and A. Cavallaro, "Tracking a moving sound source from a multi-rotor drone," in *Proc. IEEE/RSJ Int. Conf. on Intell. Robots and Syst.*, Oct. 2018, pp. 2511–2516.

[9] Y. Hioka, M. Kingan, G. Schmid, and K. A. Stol, "Speech enhancement using a microphone array mounted on an unmanned aerial vehicle," in *Proc. IEEE Int. Workshop on Acoust. Signal Enhancement*, Sep. 2016, pp. 1–5.

[10] W. N. Manamperi, T. D. Abhayapala, P. N. Samarasinghe, and J. A. Zhang, "Drone audition: Audio signal enhancement from drone embedded microphones using multichannel wiener filtering and gaussian-mixture based post-filtering," *Appl. Acoust.*, vol. 216, p. 109 818, Jan. 2024.

[11] L. Wang, R. Sanchez-Matilla, and A. Cavallaro, "Audio-visual sensing from a quadcopter: Dataset and baselines for source localization and sound enhancement," in *Proc. IEEE/RSJ Int. Conf. on Intell. Robots and Syst.*, Nov. 2019, pp. 5320–5325.

[12] W. N. Manamperi, T. D. Abhayapala, J. Zhang, and P. N. Samarasinghe, "Drone audition: Improved GMM wiener filtering approach for signal enhancement," in *Proc. INTER-NOISE and NOISE-CON Congr. and Conf.*, Aug. 2023, pp. 3584–3594.

[13] B. Yen, Y. Hioka, G. Schmid, and B. Mace, "Multi-sensory sound source enhancement for unmanned aerial vehicle recordings," *Appl. Acoust.*, vol. 189, no. 108590, Feb. 2022.

[14] W. Manamperi, T. D. Abhayapala, J. Zhang, and P. N. Samarasinghe, "Estimating drone motor related acoustic transfer function: A preliminary investigation," in *Proc. Asia-Pacific Signal and Inf. Process. Assoc. Annu. Summit and Conf.*, Dec. 2020, pp. 156–160.

[15] T. Ishiki and M. Kumon, "A microphone array configuration for an auditory quadrotor helicopter system," in *Proc. IEEE Int. Symp. on Safety, Security, and Rescue Robot.*, Oct. 2014, pp. 1–6.

[16] T. Ishiki and M. Kumon, "Design model of microphone arrays for multirotor helicopters," in *Proc. IEEE/RSJ Int. Conf. on Intell. Robots and Syst.*, Sep. 2015, pp. 6143–6148.

[17] D. Salvati, C. Drioli, G. Ferrin, and G. L. Foresti, "Beamforming-based acoustic source localization and enhancement for multirotor UAVs," in *Proc. Eur. Signal Process. Conf.*, Sep. 2018, pp. 987–991.

[18] P. Misra, A. A. Kumar, P. Mohapatra, and P. Balamuralidhar, "Aerial drones with location-sensitive ears," *IEEE Commun. Mag.*, vol. 56, no. 7, pp. 154–160, Jul. 2018.

[19] Y. Hioka, M. Kingan, G. Schmid, R. McKay, and K. A. Stol, "Design of an unmanned aerial vehicle mounted system for quiet audio recording," *Appl. Acoust.*, vol. 155, no. 7, pp. 423–427, Dec. 2019.

[20] L. Wang and A. Cavallaro, "Microphone-array ego-noise reduction algorithms for auditory micro aerial vehicles," *IEEE Sensors J.*, vol. 17, no. 8, pp. 2447–2455, Feb. 2017.

[21] L. Wang and A. Cavallaro, "Acoustic sensing from a multirotor drone," *IEEE Sensors J.*, vol. 18, no. 11, pp. 4570–4582, Apr. 2018, ISSN: 1530-437X.

[22] L. Wang and A. Cavallaro, "Ear in the sky: Ego-noise reduction for auditory micro aerial vehicles," in *Proc. IEEE Int. Conf. on Adv. Video and Signal Based Surveillance*, Aug. 2016, pp. 152–158.

[23] L. Wang and A. Cavallaro, "A blind source separation framework for ego-noise reduction on multi-rotor drones," *IEEE/ACM Trans. on Audio, Speech, and Lang. Process.*, vol. 28, pp. 2523–2537, Aug. 2020.

[24] H. Bi, F. Ma, T. D. Abhayapala, and P. N. Samarasinghe, "Spherical sector harmonics based directional drone noise reduction," in *Proc. IEEE Int. Workshop on Acoust. Signal Enhancement*, Sep. 2022, pp. 1–5.

[25] P. Marmaroli, X. Falourd, and H. Lissek, "A UAV motor denoising technique to improve localization of surrounding noisy aircrafts: Proof of concept for anti-collision systems," in *Proc. Acoust.*, Apr. 2012, pp. 1–6.

[26] K. Furukawa, K. Okutani, K. Nagira, *et al.*, "Noise correlation matrix estimation for improving sound source localization by multirotor UAV," in *Proc. IEEE/RSJ Int. Conf. on Intell. Robots and Syst.*, Nov. 2013, pp. 3943–3948.

[27] S. Yoon, S. Park, Y. Eom, and S. Yoo, "Advanced sound capturing method with adaptive noise reduction system for broadcasting multicopters," in *Proc. IEEE Int. Conf. Consumer Electron.*, Jan. 2015, pp. 26–29.

[28] S. Yoon, S. Park, and S. Yoo, "Two-stage adaptive noise reduction system for broadcasting multicopters," in *Proc. IEEE Int. Conf. Consumer Electron.*, Jan. 2016, pp. 219–222.

[29] T. Morito, O. Sugiyama, R. Kojima, and K. Nakadai, "Partially shared deep neural network in sound source separation and identification using a uav-embedded microphone array," in *Proc. IEEE/RSJ Int. Conf. on Intell. Robots and Syst.*, Oct. 2016, pp. 1299–1304.

[30] Z. W. Tan, A. H. Nguyen, and A. W. Khong, "An efficient dilated convolutional neural network for UAV noise reduction at low input SNR," in *Proc. Asia-Pacific Signal and Inf. Process. Assoc. Annu. Summit and Conf.*, Nov. 2019, pp. 1885–1892.

[31] L. Wang and A. Cavallaro, "Deep learning assisted time-frequency processing for speech enhancement on drones," *IEEE Trans. on Emerging Topics in Computational Intelligence*, vol. 5, no. 6, pp. 871–881, Aug. 2020.

[32] C. Chun, K. M. Jeon, T. Kim, and W. Choi, "Drone noise reduction using deep convolutional autoencoder for UAV acoustic sensor networks," in *Proc. IEEE Int. Conf. on Mobile Ad Hoc and Sensor Syst. Workshops*, Nov. 2019, pp. 168–169.

[33] T. D. Abhayapala, L. Birnie, M. Kumar, D. Grixti-Cheng, and P. N. Samarasinghe, "Generalizing the relative transfer function to a matrix for multiple sources and multichannel microphones," in *Proc. Eur. Signal Process. Conf.*, Sep. 2023, pp. 336–340.

[34] M. Kumar and et al., "Speech denoising in multi-noise source environments using multiple microphone devices via relative transfer matrix," in *Eur. Signal Process. Conf. (in press)*, Sep. 2024, pp. 336–340.

[35] W. N. Manamperi and T. D. Abhayapala, "Successive speaker relative transfer function estimation through relative transfer matrix in noisy reverberant environments," in *Asia-Pacific Signal and Inf. Process. Assoc. Annu. Summit and Conf. (in press)*, Dec. 2024.

[36] W. N. Manamperi and T. D. Abhayapala, "Blind separation of multiple speakers in noisy reverberant environments using relative transfer matrix," in *IEEE Int. Conf. on Acoust., Speech and Signal Process. (unpublished)*, 2024.

[37] M. Strauss, P. Mordel, V. Miguet, and A. Deleforge, "DREGON: Dataset and methods for uav-embedded sound source localization," in *Proc. IEEE/RSJ Int. Conf. on Intell. Robots and Syst.*, Oct. 2018, pp. 1–8.

[38] W. N. Manamperi, T. D. Abhayapala, L. I. Birnie, J. A. Zhang, and P. N. Samarasinghe, "Drone audition: On measurements and modeling of drone-related transfer functions," *IEEE/ACM Trans. on Audio, Speech, and Lang. Process. (unpublished)*,

[39] Y. Avargel and I. Cohen, "On multiplicative transfer function approximation in the short-time Fourier transform domain," *IEEE Signal Process. Lett.*, vol. 14, no. 5, pp. 337–340, Apr. 2007.

[40] E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation," *IEEE Trans. on Audio, Speech, and Lang. Process.*, vol. 14, no. 4, pp. 1462–1469, Jun. 2006.

[41] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An algorithm for intelligibility prediction of time–frequency weighted noisy speech," *IEEE/ACM Trans. on Audio, Speech, and Lang. Process.*, vol. 19, no. 7, pp. 2125–2136, Feb. 2011.