Ear in the sky: ego-noise reduction for auditory micro aerial vehicles

Lin Wang, Andrea Cavallaro
Centre for Intelligent Sensing, Queen Mary University of London
{lin.wang, a.cavallaro}@qmul.ac.uk

Abstract

We investigate the spectral and spatial characteristics of the ego-noise of a multirotor micro aerial vehicle (MAV) using audio signals captured with multiple onboard microphones and derive a noise model that grounds the feasibility of microphone-array techniques for noise reduction. The spectral analysis suggests that the ego-noise consists of narrowband harmonic noise and broadband noise, whose spectra vary dynamically with the motor rotation speed. The spatial analysis suggests that the ego-noise of a P-rotor MAV can be modeled as P directional noises plus one diffuse noise. Moreover, because of the fixed positions of the microphones and motors, we can assume that the acoustic mixing network of the ego-noise is stationary. We validate the proposed noise model and the stationary mixing assumption by applying blind source separation to multi-channel recordings from both a static and a moving MAV and quantify the signal-to-noise ratio improvement. Moreover, we make all the audio recordings publicly available.

1. Introduction

Micro aerial vehicles (MAV) are increasingly used in a wide range of applications, such as search and rescue operations, personal and professional video capturing, and surveillance [1]. Multirotor MAVs can hover above a target area and can be used as a universal sensing platform equipped with a variety of sensors, such as cameras, microphones, laser scanners or ultrasonic radars. While visual sensing has attracted considerable research attention [2, 3], despite its potential impact acoustic sensing using MAVs has been relatively overlooked.

When deploying flying MAVs in search and rescue operations, microphones would be important to locate sound-emitting targets (*e.g.* a person in distress) especially at night, in low visibility or in the presence of visual obstacles and occlusions (*e.g.* a victim under debris) [4, 5, 6]. Flying MAVs can also be deployed for multimedia broadcasting of an event by transmitting the audio and video streams to remote locations [7]. The main barrier for effective MAV-based acoustic sensing is the strong ego-noise [5, 8], which masks the target sound(s) and degrades the overall sound recording quality significantly.

The ego-noise is generated mainly by the motors and propellers, which are closer to the microphones than the emitter of the target sound, thus leading to extremely low signal-to-noise ratios. Moreover, because the rotation speed of each motor dynamically changes during a flight, the ego-noise is nonstationary. Also, the microphones move together with the MAV, thus leading to a dynamic acoustic mixing network. Finally, natural wind increases the noise components captured by the microphones. The above-mentioned noise sources are considerable challenges to existing noise reduction algorithms, which were mainly designed for indoor environments with fixed microphones [9]. Appropriate sound enhancement techniques are therefore necessary for MAV-based acoustic sensing.

In this paper, we investigate the spectral and spatial characteristics of the ego-noise using multiple microphones on the MAV and we propose a noise model that grounds the feasibility of using microphone-array techniques for noise reduction. We model the noise as a sum of multiple directional sounds and a diffuse sound, and we apply blind source separation to multi-microphone recordings made with a static MAV and a moving MAV. Moreover, the experimental results for the moving MAV suggest a stationary mixing network of the ego-noise, which provides valuable insights for developing noise reduction algorithms in dynamic environments.

The paper is organized as follows. Sec. 2 reviews the related work. Sec. 3 discusses our noise component analysis and modeling. Sec. 4 validates the proposed model and presents the noise reduction results using blind source separation. Finally, conclusions are drawn in Sec. 5.

2. Related Work

Currently only a few works have been presented to specifically address the challenging MAV ego-noise reduction problem. These works can be categorized as single-channel or multi-channel approaches. *Single-channel*

978-1-5090-3811-4/16/$31.00

approaches mainly exploit the amplitude of the microphone signal for noise reduction (spectral enhancement). Traditional spectral enhancement approaches typically require the noise to be stationary in order to blindly estimate the noise power spectrum density (PSD). These approaches are not directly applicable to MAV sound recording because the PSD of the ego-noise varies dynamically. Since the MAV ego-noise mainly consists of harmonic components whose fundamental frequency is proportional to the motor rotation speed, a template-based approach was proposed that generates noise spectral templates given the prior knowledge of motor rotation speed [10]. A drawback of this approach is the need of additional sensors to provide information about the motor rotation speed and the MAV behavior. Using amplitude information only, single-channel approaches produce severe signal-of-interest distortion and even fail completely in scenarios with extremely low signal-to-noise ratios (SNR).

Multi-channel approaches mainly exploit the phase information and the correlation among multiple microphones for noise reduction. Delay-and-sum (fixed) beamforming was used for target sound acquisition with a microphone array, which was optimally designed in terms of array size and sensor placement [8, 11]. Fixed beamforming is robust to low SNR and MAV movement. However, it usually requires a large array to get satisfactory noise reduction performance and also requires the knowledge of the target location to steer the beam. A reference-based approach [7] uses reference microphones, which are installed close to the propellers, to pick up motor noises and cancel them with an adaptive filter. The results reported in [7] are promising for a static MAV, but still quite limited for a moving MAV.

The above methods were applied to MAVs and are summarized in Table 1. In addition to these methods, other noise reduction approaches for voice communication or ground robot audition [9] could be used. For instance, non-negative matrix factorization (NMF) [12] was employed for single-channel spectral enhancement for ground robots. NMF estimates the noise bases from pre-recorded training data and then estimates the noise PSD from the noisy recording. Multi-channel adaptive beamforming [13] and blind source separation [14] perform more efficiently than a delay-and-sum beamformer for noise reduction. However, applying these methods to a moving MAV is challenging because the acoustic mixing network changes dynamically. Template-based approaches [15, 16] construct noise correlation matrices as a function of the behavior of the robot and use them to design an adaptive beamformer. This can be seen as a multi-channel extension of the single-channel template [10], which only considers the spectral amplitude. Template-based approaches were applied to source localization with MAVs [17]. However, application to MAV ego-noise reduction has not been reported yet.

Table 1. Single-channel (SC) and multi-channel (MC) noise reduction approaches applied to MAV sound recording.

	Method	Ref.
SC	Template-based spectral enhancement	[10]
MC	Fixed beamforming	[8, 11]
MC	Reference-based noise reduction	[7]

3. Noise component modelling

3.1. Preliminaries

Let an MAV equipped with M microphones capture the sound emitted by a target (*e.g.* a person). The microphone signal, $\boldsymbol{x}(n) = [x_1(n), \cdots, x_M(n)]^{\mathrm{T}}$, contains both the target sound, $\boldsymbol{s}(n) = [s_1(n), \cdots, s_M(n)]^{\mathrm{T}}$, and the ego-noise, $\boldsymbol{v}(n) = [v_1(n), \cdots, v_M(n)]^{\mathrm{T}}$, *i.e.*

$$\boldsymbol{x}(n) = \boldsymbol{s}(n) + \boldsymbol{v}(n). \tag{1}$$

We aim to answer the following questions: (i) What are the spectral characteristics of the constituent components of the ego-noise? (ii) Does the ego-noise show strong correlation among multiple microphones? (iii) Do the spatial characteristics of the ego-noise vary with the movement of the MAV?

To this end, we built a circular microphone array consisting of eight Boya BY-M1 omnidirectional lavelier microphones to be fixed on the top of the MAV (Fig. 1). The diameter of the array is 0.2 m and the distance from the array to the top side of the MAV is 0.15 m. The signals from the eight microphones are sampled simultaneously with a Zoom R24 multi-channel audio recorder, at a sampling rate of 44.1 kHz (downsampled to 8 kHz before processing).

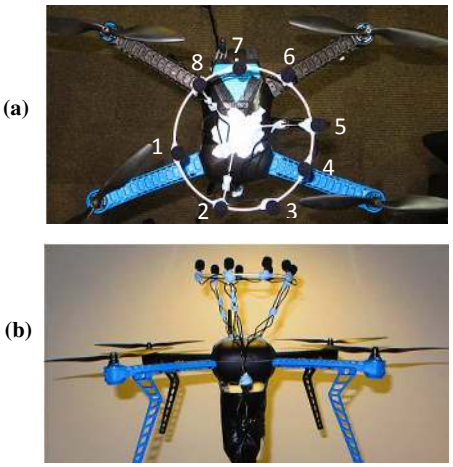


Figure 1. The circular microphone array mounted on the MAV. (a) Top view. (b) Side view.

3.2. Noise components

The ego-noise consists of three main components, namely the mechanical noise generated by the rotation of the motors, the air flow noise generated by the rotation of the propellers cutting the air, and the wind noise blowing directly to the microphones from the propellers.

The propeller noise was analyzed experimentally in order to optimally design the MAV propulsion system [18, 19]. The wind noise may significantly deteriorate the recording quality. The relationship between microphone positioning and noise reduction was examined experimentally in [20], where it was concluded that positions below the MAV receive more wind noise than positions above and beside the MAV. Since the wind from the propellers blows downwards, we positioned the microphones above the MAV to avoid the influence of the self-generated wind (Fig. 1(b)). Thus, we only need to consider the first two noise elements in the experiments: the mechanical noise and the air flow noise.

One favorable factor for ego-noise reduction is that the noise sources are fixed with respect to the positions of the microphones and thus prior knowledge can be used to choose appropriate noise reduction approaches. We thus investigate the correlation information of the microphone signals, which plays crucial role on the noise reduction performance. If the target signal shows high correlation at the microphones while the noise signal shows low correlation, we can employ a simple delay-and-sum beamformer for noise reduction [13]. If both the target and noise signals show high correlation at the microphones, we can employ more advanced algorithms, such as adaptive beamforming or blind source separation, which work more efficiently for noise reduction [13].

The correlation information can be represented with the correlation coefficient between two microphone signals in the time-frequency domain. By applying the short-time Fourier transform (STFT) to two time-domain signals $x_1(n)$ and $x_2(n)$, the time-frequency signals are obtained and represented as $X_1(k,l)$ and $X_2(k,l)$, respectively; where k and l are the frequency and frame indices, respectively. The correlation coefficient $\gamma(k,l)$ is defined as

$$\gamma(k,l) = \left| \frac{\sum_{l'=l-\delta}^{l+\delta} X_1^*(k,l') X_2(k,l')}{\sum_{l'=l-\delta}^{l+\delta} |X_1(k,l') X_2(k,l')|} \right|, \qquad (2)$$

where the superscript * denotes conjugate and $\delta = 3$ indicates the number of consecutive frames that are used for the calculation of the coefficient.

3.3. Noise modelling

We analyze the spectral and spatial characteristics of the ego-noise using real-recorded data. We design two scenarios to investigate the first two types of noise. In the first scenario we record the sound from an MAV without propellers, so that the recording only contains the *mechanical noise* from the rotating motors. In the second scenario we record the sound from an MAV with propellers, so that the recording contains both the *mechanical noise* from the rotating motors and the *airflow noise* from the rotating propellers cutting air. The experiment is conducted in a room of size 6m×5m×3m with a reverberation time of around 200 ms. A 3DR Iris quadcopter [1] is fixed on a tripod at a height of 1.8 m (Fig. 2). The size of the MAV is about 0.55m×0.55m. The motor rotation speed is modified with a remote controller. A loudspeaker is placed 3 m away from the MAV, at a height of 1.3 m.

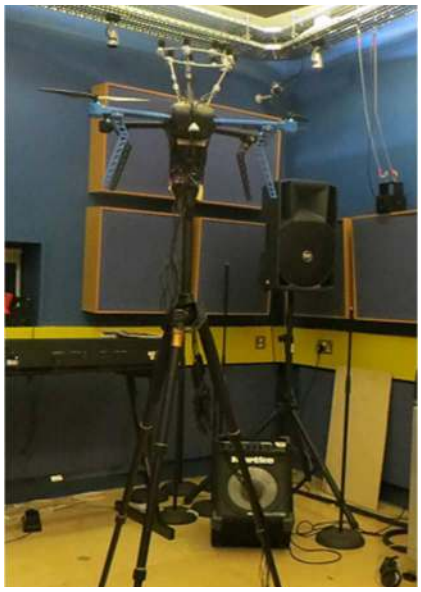
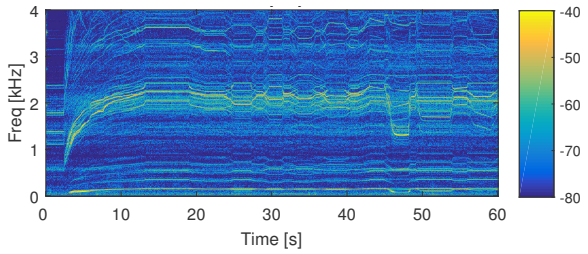


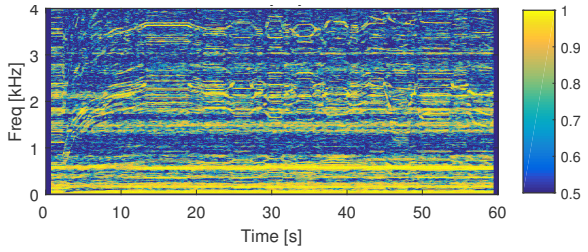
Figure 2. The sound recording setup.

Fig. 3(a) depicts the spectrum of the microphone signal for a one-minute long segment from an *MAV without propellers*. The motor rotation speed is rising monotonically during the first 10 s, remains constant during the following 10 s, and then varies randomly during the last 40 s. The motor noise mainly consists of harmonic components, with energy peaks at multiple isolated frequency bins. The fundamental frequency (pitch) of the harmonic noise varies corresponding to the motor rotation speed: the pitch rises monotonically during the first 10 s, remains stable during the following 10 s, and then varies dynamically during the remaining 40 s. The MAV has four motors, each with a different rotation speed, and the superimposition of the pitches from these motors leads to a complex spectrum structure. Fig. 3(b) depicts the time-frequency correlation between two microphone signals. It can be observed that the two microphones show strong correlation in the time-frequency domain, especially at harmonic frequencies with

[1] https ://store.3drobotics.com/products/iris

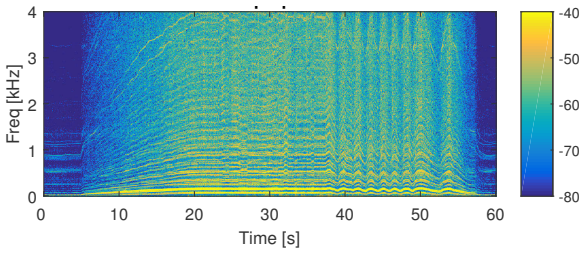


(a)

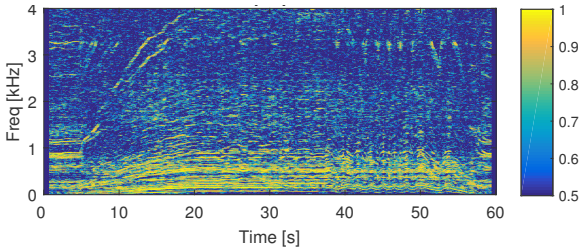


(b)

Figure 3. Sound recording from an MAV without propellers. (a) time-frequency spectrum (dB); (b) time-frequency correlation.



(a)



(b)

Figure 4. Sound recording from an MAV with propellers. (a) time-frequency spectrum (dB); (b) time-frequency correlation.

high energy. We thus assume that the harmonic noise can be modeled as directional point sources located at the engine houses.

Fig. 4(a) depicts the spectrum of the microphone signal for a one-minute long segment from an *MAV with propellers*. The motor rotation speed is rising monotonically during the first 20 s, remains constant during the following 20 s, and then varies randomly during the last 20 s. The noise signal mainly consists of two components: harmonic noise, whose energy peaks at isolated frequency bins, and broadband noise, whose energy spreads throughout the whole frequency band. Similarly to Fig. 3(a), the pitch of the harmonic noise varies based on the motor rotation speed. In addition to this, the harmonic noise is more intense in Fig. 4(a) than in Fig. 3(a) due to the influence of air resistance, which slows the rotation of the motors and propellers, and also introduces broadband noise. The energy of the broadband noise is proportional to the propeller rotation speed: a faster rotation generates a more rapid airflow and thus a stronger noise. Fig. 4(b) depicts the time-frequency correlation between two microphone signals. Strong correlation can be observed only at low frequencies and only at harmonic frequencies with high energy. This is mainly due to the influence of broadband noise. The superimposition of the noise from the four propellers may generate diffuse-like characteristics, *i.e.* with low correlation at high frequencies but high correlation at low frequencies.

Based on the above analysis, we model the ego-noise, $\boldsymbol{v}(n)$, as a sum of multiple directional point-source noises and one directionless diffuse noise:

$$\boxed{\boldsymbol{v}(n) = \sum_{p=1}^{P} \boldsymbol{v}_{d\_p}(n) + \boldsymbol{v}_f(n)} \tag{3}$$

where P is the number of motors, $\boldsymbol{v}_{d\_p}(n)$ is the *directional noise* from the p-th motor and $\boldsymbol{v}_f(n)$ is the *diffuse noise* generated by the P propellers jointly.

This ego-noise model provides us with valuable insights for noise reduction: the directional components could be suppressed with advanced microphone-array techniques such as adaptive beamforming [13] or blind source separation [21]; moreover, because the model consists of $P + 1$ independent noise components, at least $P + 2$ microphones are required to satisfactorily suppress the ego-noise while preserving the target signal. Finally, since the relative locations between the motors and the microphones are fixed, we can presume that the acoustic mixing paths between them remain constant even during MAV movement. This presumption and the validity of the noise model will be verified in the next section.

4. Model validation

4.1. Noise reduction method

Assuming that all the source signals (target and noise) are statistically independent, we use *Blind Source Separation* (BSS) to estimate a demixing system that separates all the sources in the microphone signals. BSS does not need prior knowledge on the source locations and can effectively suppress directional noises that show strong correlation across microphones. However, to estimate the demixing

system satisfactorily BSS typically works in a batch style and requires the acoustic mixing network to be fixed for a certain interval (*e.g.* $>$ 10 s) [21]. The stationary mixing condition is satisfied with physically static sound sources and microphones, *e.g.* an MAV hovering stably in the air and recording a static speaker.

We use a frequency-domain implementation of the BSS algorithm [14], with a filter length $L_w = 1024$ at sampling rate 8 kHz. The noise reduction performance of BSS can be evaluated with the global SNR measure. Suppose the spatial filter corresponding to the target signal is $\boldsymbol{w}(n) = [w_1(n), \cdots, w_M(n)]$, the target signal is estimated as

$$y(n) = \boldsymbol{w}(n) * \boldsymbol{x}(n) = \sum_{p=0}^{L_w-1} \boldsymbol{w}(p)\boldsymbol{x}(n-p)$$
$$= y_s(n) + y_v(n) = \boldsymbol{w}(n) * \boldsymbol{s}(n) + \boldsymbol{w}(n) * \boldsymbol{v}(n), \quad (4)$$

where $*$ denotes the convolutive filtering procedure [14]; y_s and y_v are respectively the target and noise components at the output. The global SNR is calculated in target-signal-active periods $\mathbb{S}$ as [13]

$$\text{SNR} = 10\log_{10}\frac{\sum_{n'\in\mathbb{S}} y_s^2(n')}{\sum_{n'\in\mathbb{S}} y_v^2(n')}. \quad (5)$$

4.2. Discussion

We first evaluate the performance of BSS when the stationary mixing condition is satisfied, *i.e.* the MAV, the microphones and the loudspeaker are physically fixed. The speech signal and noise signal are recorded separately, and the noisy signal is generated by summing the two signals at different input SNRs, which vary from -25 dB to 10 dB, with an interval of 5 dB. We examine the performance of BSS with the number of microphones, M, varying from 2 to 8. For each configuration of M and input SNR, we implement 10 realizations, where in each realization the speech and noise are both randomly chosen from the recording with 12 s duration. The SNR improvement (*i.e.* the difference between input and output SNRs) is calculated by averaging the 10 realizations.

Fig. 5 depicts the SNR improvement achieved by BSS at different input SNRs $\in [-25, 10]$ dB, with an interval of 5 dB, when using different $M \in [2, 8]$ on a *static MAV*. BSS can improve the SNR in all testing scenarios. The amount of SNR improvement varies with M. The SNR improvement is quite limited when $M \leq 3$, but increases significantly when M gets from 3 to 6. Then the amount of increase slows when M gets from 6 to 8. This suggests that the recording contains 6 independent elements: one speech (target signal) element and five noise elements (cf. (3)). Increasing M helps to capture all the elements and thus improves the noise suppression performance effectively.

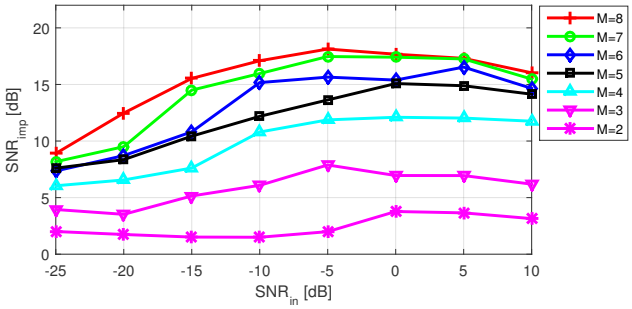


Figure 5. SNR improvement by BSS with a static MAV and signal duration of 12 s when varying the input SNR and the number of microphones, M, of the array.

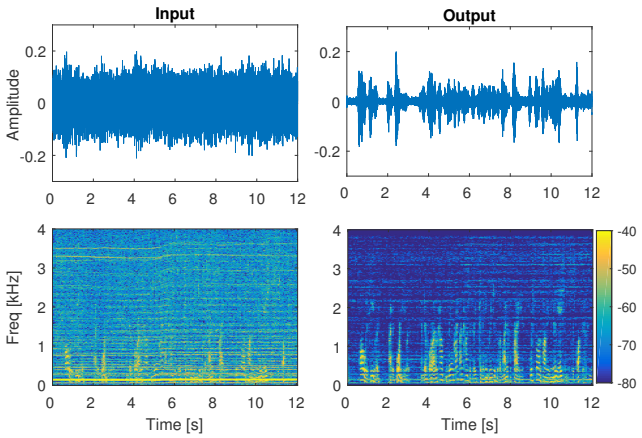


Figure 6. Time-domain waveforms and time-frequency spectra of a microphone signal before and after BSS for a 12 s signal and a static MAV with 8 microphones. The input and output SNRs are -10 dB and 11.5 dB, respectively.

When $M \geq 6$ the SNR improvement still rises slowly since the additional microphones may help suppress uncorrelated noise. However, the improvement when $M \geq 6$ appears less evident. The amount of SNR improvement also varies with the input SNR. The SNR improvement is quite limited when $\text{SNR}_{\text{in}} \leq -20$ dB, but increases significantly when SNR_{in} rises from -20 dB to -5 dB, and then slows or even decreases when the noise becomes less dominant for SNR_{in} from -5 dB to 10 dB.

Fig. 6 shows sample time-domain waveforms and time-frequency spectra for one microphone signal before and after BSS. Before BSS, the speech signal is hardly distinguishable from the noisy background (input SNR: -10 dB). After BSS, the speech signal can be clearly observed in the enhanced output (SNR: 11.5 dB). Moreover, the strong harmonic noises are almost completely removed in the output spectrum.

Fig. 7 compares the performance of BSS, fixed beamforming and adaptive beamforming using the same data used for Fig. 5. A delay-and-sum beamformer is used for fixed beamforming, *assuming the location of the*

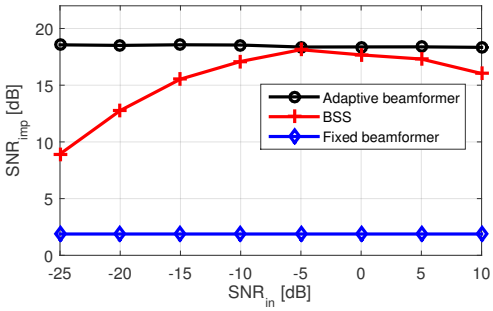


Figure 7. SNR improvement by BSS, fixed beamforming and adaptive beamforming for a static MAV with 8 microphones. Note that adaptive beamforming assumes the noise correlation matrix to be known. *A demo with the audio signals corresponding to this figure is available [22].*

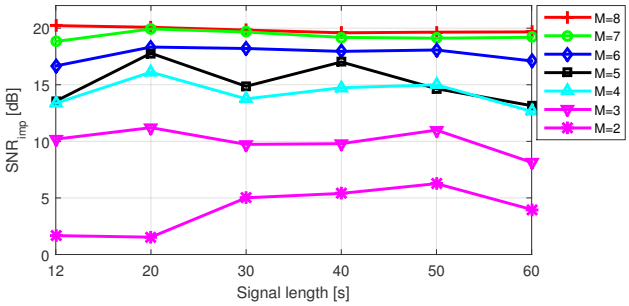


Figure 8. SNR improvement by BSS for a moving MAV. The signal duration is varying from 12 s to 60 s (input SNR: -5 dB) and the number of microphones, M, varies from 2 to 8.

target speaker to be known. The MaxSNR beamformer implemented in [13] is employed for adaptive beamforming, *assuming the noise correlation matrix to be known*. BSS outperforms fixed beamforming, which can only improve the SNR by around 2 dB for all cases. This shows that the directional components of the ego-noise are better suppressed by BSS. The adaptive beamforming results provide a benchmark for this problem as it uses the knowledge of the noise correlation matrix, whose estimation is still an open problem [13]. BSS does not need to know the noise correlation matrix and the obtained noise reduction performance is close to the benchmark at high input SNRs (*e.g.* $\geq$ -10 dB). However, at low SNRs BSS performs worse than the benchmark. We expect to improve the performance of BSS if additional information of the noise correlation matrix is available, *e.g.* estimated by using template-based methods [17].

In summary, the ego-noise contains directional components and thus can be suppressed effectively with BSS. The SNR improvement slows down when the number of microphones is larger than 6 thus suggesting that the ego-noise contains 5 noise components. The above observations are consistent with the model proposed in (3).

Finally, we investigate whether the acoustic transfer paths of the ego-noise remain constant with a *moving MAV*. We move the MAV with the tripod and generate noisy signals by summing speech signals, which are recorded when the MAV is static, and noise signals, which are recorded when the MAV is moving. Since BSS works well only for a stationary mixing network and the acoustic paths between the speaker and microphones are fixed in the simulation, the stationary mixing assumption of the ego-noise can be easily verified based on the BSS performance for the simulated data. We test BSS using different signal durations, which vary from 12 s to 60 s. For each duration, we use different input SNRs, which vary from -25 dB to 10 dB, with an interval of 5 dB, and different M, which varies from 2 to 8. For each configuration of M and input SNR, we implement one realization. Fig. 8 depicts the SNR improvement achieved by BSS at input SNR -5 dB, when using different signal lengths, varying from 12 s to 60 s, and different number of microphones $M \in [2, 8]$. The SNR improvement increases significantly when M is rising from 2 to 6, and the increase slows when M is rising from 6 to 8. This is similar to the observation made for Fig. 5. For each M, the SNR improvement does not degrade evidently when increasing the signal length, thus confirming the stationary mixing assumption of the ego-noise for a moving MAV.

5. Conclusion

In this paper, we modelled the ego-noise of a P-rotor MAV as P directional noises (each coming from a motor) plus one diffuse noise (coming from the propellers) and validated the noise model by applying blind source separation to MAV sound recording. At least $P + 2$ microphones are required to suppress the noise satisfactorily. Due to fixed relative locations between the motors and the microphones, the acoustic mixing network of the ego-noise tends to remain stationary during the movement of the MAV. This finding is very promising for developing efficient noise reduction algorithms in practical applications, where the acoustic mixing network of the target always changes due to MAV movement.

To facilitate comparisons and reproducibility we make all the audio recordings discussed in this paper publicly available [22].

Our future work includes using the proposed noise model to design a new BSS algorithm that works outdoors with moving MAVs and in conditions where natural wind and surrounding noises impose additional challenges, and also real-time implementation on a practical UAV system.

Acknowledgement: This work was supported by the ARTEMIS-JU and the UK Technology Strategy Board (Innovate UK) through the COPCAMS Project, under Grant 332913.

References

[1] D. Floreano and R. J. Wood, "Science, technology and the future of small autonomous drones," *Nature*, vol. 521, pp. 460-466, May 2015.

[2] F. Remondino, L. Barazzetti, F. Nex, M. Scaioni, and D. Sarazzi, "UAV photogrammetry for mapping and 3D modeling - current status and future perspectives," *Int. Archives Photogrammetry, Remote Sensing and Spatial Inform. Sci.*, Zurich, Switzerland, 2011, pp. 25-31.

[3] F. Poiesi and A. Cavallaro, "Distributed vision-based flying cameras to film a moving target," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, Hamburg, Germany, 2015, pp. 2453-2459.

[4] M. Basiri, F. Schill, P. U. Lima, and D. Floreano, "Robust acoustic source localization of emergency signals from micro air vehicles," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, Vilamoura-Algarve, Portugal, 2012, pp. 4737-4742.

[5] K. Okutani, T. Yoshida, K. Nakamura, and K. Nakadai, "Outdoor auditory scene analysis using a moving microphone array embedded in a quadrocopter," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, Vilamoura-Algarve, Portugal, 2012, pp. 3288-3293.

[6] S. Lana, K. Takahashi, and T. Kinoshita, "Consensus-based sound source localization using a swarm of micro-quadrocopters," in *Proc. Robot. Soc. Japan*, Tokyo, Japan, 2015, pp. 1-4.

[7] 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. Consum. Electron.*, Las Vegas, USA, 2015, pp. 26-29.

[8] T. Ishiki and M. Kumon, "A microphone array configuration for an auditory quadrotor helicopter system," in *Proc. IEEE Int. Symp. Safety, Security, Rescue Robot.*, Toyako-cho, Japan, 2014, pp. 1-6.

[9] H. G. Okuno and K. Nakadai, "Robot audition: Its rise and perspectives," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Brisbane, Australia, 2015, pp. 5610-5614.

[10] 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.*, 2012, pp. 1-6.

[11] T. Ishiki and M. Kumon, "Design model of microphone arrays for multirotor helicopters," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, Hamburg, Germany, 2015, pp. 6143-6148.

[12] T. Tezuka, T. Yoshida, and K. Nakadai, "Ego-motion noise suppression for robots based on semi-blind infinite non-negative matrix factorization," in *Proc. IEEE Int. Conf. Robot. Autom.*, Hong Kong, China, 2014, pp. 6293-6298.

[13] L. Wang, T. Gerkmann, and S. Doclo, "Noise power spectral density estimation using MaxNSR blocking matrix," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 9, pp. 1493-1508, Sep. 2015.

[14] L. Wang, "Multi-band multi-centroid clustering based permutation alignment for frequency-domain blind speech separation," *Digit. Signal Process.*, vol. 31, pp. 79-92, 2014.

[15] G. Ince, K. Nakamura, F. Asano, H. Nakajima, and K. Nakadai, "Assessment of general applicability of ego noise estimation," in *Proc. IEEE Int. Conf. Robot. Autom.*, Shanghai, China, 2011, pp. 3517-3522.

[16] G. Ince, K. Nakadai, and K. Nakamura, "Online learning for template-based multi-channel ego noise estimation," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, Vilamoura-Algarve, Portugal, 2012, pp. 3282-3287.

[17] K. Furukawa, K. Okutani, K. Nagira, T. Otsuka, K. Itoyama, K. Nakadai, and H. G. Okun, "Noise correlation matrix estimation for improving sound source localization by multirotor UAV," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, Tokyo, Japan, 2013, pp. 3943-3948.

[18] L. Marino, "Experimental analysis of UAV-propellers noise," in *Proc. AIAA/CEAS Aeroacoustics Conf.*, Stockholm, Sweden, 2010, pp. 1-14.

[19] G. Sinibaldi and L. Marino, "Experimental analysis on the noise of propellers for small UAV," *Appl. Acoust.*, vol. 74, no. 1, pp. 79-88, Jan. 2015.

[20] J. Klapel, *Acoustic Measurements with a Quadcopter: Embedded System Implementations for Recording Audio from Above*, Master Thesis, Norwegian University of Science and Technology, 2014.

[21] S. Makino, T. W. Lee, and H. Sawada, eds. *Blind Speech Separation*, Berlin, Germany: Springer-Verlag, 2007.

[22] http://www.eecs.qmul.ac.uk/∼andrea/ear-in-the-sky.html