Estimating Power Spectral Density of Unmanned Aerial Vehicle Rotor Noise Using Multisensory Information

Benjamin Yen, Yusuke Hioka, Brian Mace
Acoustics Research Centre, Department of Mechanical Engineering, University of Auckland
Auckland, New Zealand
Email: benjamin.yen@ieee.org, yusuke.hioka@ieee.org, b.mace@auckland.ac.nz

***Abstract*—A method to accurately estimate the power spectral density (PSD) of an unmanned aerial vehicle (UAV) is proposed, in anticipation of being used for a UAV-mounted audio recording system that clearly captures target sound while suppressing rotor noise. The method utilises UAV rotor characteristics as well as microphone recorded signals to combat practical limitations seen in a previous study. The proposed method was evaluated on a simulation platform modelled after the UAV used in the previous study. Results showed that the proposed method was able to estimate the rotor noise PSD to within 1.3-3.3 dB log spectral distortion (LSD) regardless of the presence of surrounding sound sources.**

***Index Terms*—Microphone array, unmanned aerial vehicle, source enhancement, power spectral density, rotor noise**

I. Introduction

Unmanned aerial vehicles (UAVs) have recently gained huge popularity for a wide range of applications. Among these, it has shown to be exceptionally popular in filming [1], [2], due to its small size and manoeuvrability. However, audio recording using UAVs have shown to be challenging due to the high noise levels radiated from the UAV rotors. This adds an extra layer of challenge on top of removing potentially already present ambient noise sources, such as wind, traffic, or even unwanted speech.

Algorithms to extract clear signals from their noisy mixtures have been studied and developed over many decades. One of the major frameworks is the use of microphone arrays [3]. Several studies have attempted for localising sound sources [4]–[8], or separating sound sources [9] using UAV-mounted audio recording devices, however, very few focus on improving audio recording quality. A recent study presented by the authors in [10] tackled speech enhancement utilising the well-known framework *beamforming with post Wiener filter* [3]. The key to success is to accurately estimate the power spectral density (PSD) of each sound source for calculating the Wiener filter. While results were promising, the improvement was still marginal, as the algorithm only used audio signals from microphones which are already contaminated by a significant amount of rotor noise. Furthermore, under practical scenarios, each microphone will receive a weighted mixture of signals from all present sound sources in the environment. Therefore, minimising such errors introduced by practical complications, as a mean of accurately estimating the UAV rotor noise PSDs is necessary.

Fortunately, such noise is not random in nature. Furthermore, aerodynamic and aeroacoustic studies have shown that there is a clear pattern closely linked to the rotor's behaviour, such as rotor speed [11], [12]. These can be measured using appropriate sensors without being affected by acoustic signals. Therefore, the rotor noise PSD can potentially be estimated more accurately if the PSD is estimated from these non-acoustical parameters of the rotors.

Recently, machine learning (ML) based methods for source enhancement have gained considerable popularity. For example, many studies utilised deep or recurrent neural networks (DNN/RNN), or deep RNNs (DRNN) for monaural speech separation [13], [14], multichannel speech enhancement [15], [16], and speech recognition [17], [18], due to its potential to model complex relationship between input features extracted from acoustic signals observed by microphones and the target sound source with high accuracy. Some ML studies estimate the PSD of sound sources using the output of beamformers [19], [20], and has found to deliver promising performance.

In this study, a method that utilises multisensory information collected from the UAV to estimate the PSD of the UAV rotor noise is proposed, taking both the rotor's state and microphone signals into account. To model the mapping function from these rotor state information to the rotor noise PSD, a regression tree [21] is used. It is a commonly used non-parametric regression based ML technique that has the advantage to estimate relatively complex patterns without being overly computationally expensive. Since UAV noise has a coherent, distinct pattern, such a method may be sufficient for this purpose, especially when it is to be incorporated in real-time signal processing for actual UAV recordings.

II. UAV System and Observation Model

A. Observation Model

Fig. 1 shows an overview of the UAV system used in this study. The main objective for UAV based recording systems is to extract clear target source signal from its microphone array recordings, while suppressing surrounding interfering *spatially coherent* noise arriving from L sources in different angles (including noise generated by U ($\leq L$) UAV rotors) as well as

ISBN 978-90-827970-1-5 © EURASIP 2018

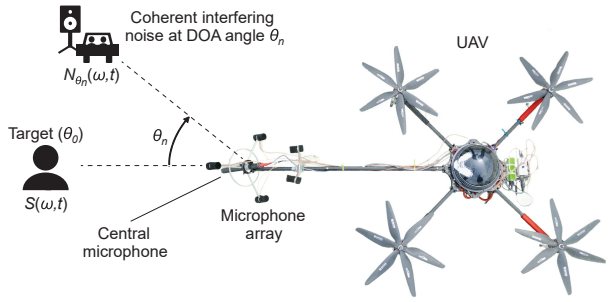

Fig. 1. Audio recording UAV overview (top view).

ambient *spatially incoherent* noise [10]. Assuming M-sensor microphone array observes signals, the short-time Fourier transform (STFT) of the array's input signals is expressed in a vector form given by

$$\begin{aligned}\mathbf{x}(\omega,t) &:= \left[X_1(\omega,t), \quad \cdots, \quad X_M(\omega,t)\right]^T \\ &= \mathbf{a}_{\theta_0}(\omega)S(\omega,t) + \sum_{u=1}^{U}\mathbf{a}_{\theta_u}(\omega)N_{\theta_u}(\omega,t) \\ &+ \sum_{n=U+1}^{L}\mathbf{a}_{\theta_n}(\omega)N_{\theta_n}(\omega,t) + \mathbf{v}(\omega,t), \end{aligned} \tag{1}$$

where ^{T} denotes the transpose, and $X_m(\omega,t)$ is the STFT of the input signal of the m-th microphone. Similarly, $S(\omega,t)$, $N_{\theta_u}(\omega,t)$ and $N_{\theta_n}(\omega,t)$ are the STFT of the target source located at angle θ_0, the u-th rotor noise source located at angle θ_u, and the n-th spatially coherent interfering noise source located at angle θ_n. ω and t denote the angular frequency and frame index, respectively. $\mathbf{a}_\theta(\omega,t)$ is a vector of transfer functions (TF) between the source located at angle θ and microphone m, i.e. $\mathbf{a}_\theta(\omega) = [A_{1,\theta}(\omega),\cdots,A_{M,\theta}(\omega)]^T$, and $\mathbf{v}_\theta(\omega,t) = [V_1(\omega,t),\cdots,V_M(\omega,t)]^T$ is a vector of incoherent noises, where $V_m(\omega,t)$ is the incoherent noise observed by the m-th microphone. The target source, all coherent noise sources, and the incoherent noise are assumed to be mutually uncorrelated.

Since the UAV is most likely to be used in outdoor environments, sound propagation would be similar to a free field where reflections are negligible. Therefore $A_{m,\theta}(\omega)$ is modelled by plane waves given by $A_{m,\theta}(\omega) = e^{j\omega\tau_{\theta,m}}$, where $\tau_{\theta,m}$ is the time delay of the sound arrival from angle θ at the m-th microphone with respect to a reference point on the coordinate system. The problem assumes that the angles of the target source and all noise sources are given *a priori*.

B. Source enhancement using beamforming with postfiltering

Following previous works, the method proposed in this study also adopts the beamforming with postfiltering framework. This section briefly explains the framework. Fixed beamformers are applied to the input signal in order to emphasise the sources in each direction, providing

$$Y_l(\omega,t) = \mathbf{w}_{\theta_l}^H(\omega)\mathbf{x}(\omega,t), \tag{2}$$

where ^{H} denotes the Hermitian transpose. The minimum variance distortionless response (MVDR) beamformer [22] is commonly used for the beamformer design, the filter weights vector $\mathbf{w}_{\theta_l}(\omega)$ at angle θ_l of which is given by

$$\mathbf{w}_{\theta_l}(\omega) = \frac{R^{-1}\mathbf{a}_{\theta_l}(\omega)}{\mathbf{a}_{\theta_l}{}^H(\omega)R^{-1}\mathbf{a}_{\theta_l}(\omega)}. \tag{3}$$

Here, R is the normalised noise covariance matrix modelled using the plane wave equation [3], and θ_l denotes the angle at which the directivity of the l-th beamformer points. The beamformer's output $Y_l(\omega,t)$ is then calculated. Despite noticeable improvements in enhancing the target source signal, $Y_l(\omega,t)$ still contains a significant amount of rotor noise, and other undesired source signals present. Thus, Wiener postfilter is applied to the beamformer's output. To design the Wiener filter, PSD of the target sound source and other noise have to be estimated. From (1) and (2), the PSD of the beamformer's output can be approximated as

$$\begin{aligned}\phi_{Y_l}(\omega,t) \cong \phi_{Y_l,S}(\omega,t) &+ \sum_{u=1}^{U}\phi_{Y_l,N_{\theta_u}}(\omega,t) \\ &+ \sum_{n=U+1}^{L}\phi_{Y_l,N_{\theta_n}}(\omega,t) + \phi_{Y_l,\bar{V}}(\omega,t), \end{aligned} \tag{4}$$

where $\phi_{Y_l,S}(\omega,t)$, $\phi_{Y_l,N_{\theta_u}}(\omega,t)$, $\phi_{Y_l,N_{\theta_n}}(\omega,t)$ and $\phi_{Y_l,\bar{V}}(\omega,t)$ are the PSD of the target source, u-th UAV rotor noise, n-th non-UAV rotor coherent interfering noise, and incoherent noise respectively. The Wiener filter coefficients are then estimated as

$$H = \frac{\hat{\phi}_{Y_l,S}}{\hat{\phi}_{Y_l,S} + \sum_{u=1}^{U}\hat{\phi}_{Y_l,N_{\theta_u}} + \sum_{n=U+1}^{L}\hat{\phi}_{Y_l,N_{\theta_u}} + \hat{\phi}_{Y_l,V}}, \tag{5}$$

where the $\hat{\cdot}$ operator denotes an estimate. Note that ω and t are omitted for brevity in (5). Finally, the Wiener filter output $Z(\omega,t)$ is obtained as

$$Z(\omega,t) = H(\omega,t)Y_0(\omega,t). \tag{6}$$

III. Proposed Method

Although the ultimate goal is to utilise the estimated rotor noise PSD to filter $Y_0(\omega,t)$ to obtain a high-quality target source signal, this study focuses on estimating the rotor noise PSD alone. Naturally, rotor noise is the dominating sound for most UAV audio recordings. However, it is also the most predictable, and is directly related to its UAV characteristics, allowing non-acoustic approaches to estimate its PSD without disturbances from other sound sources, thereby improving estimation accuracy and robustness. Inherently, this should also make PSD estimation of the remaining sound sources easier. The proposed method is designed to estimate $\phi_{Y_l,N_{\theta_u}}(\omega,t)$ taking advantage of these non-acoustic information. Due to the complexity of the PSDs spectral pattern relationship, an approach based on ML is adopted. This section discusses the general framework, feature preparation and the mapping function used.

ISBN 978-90-827970-1-5 © EURASIP 2018

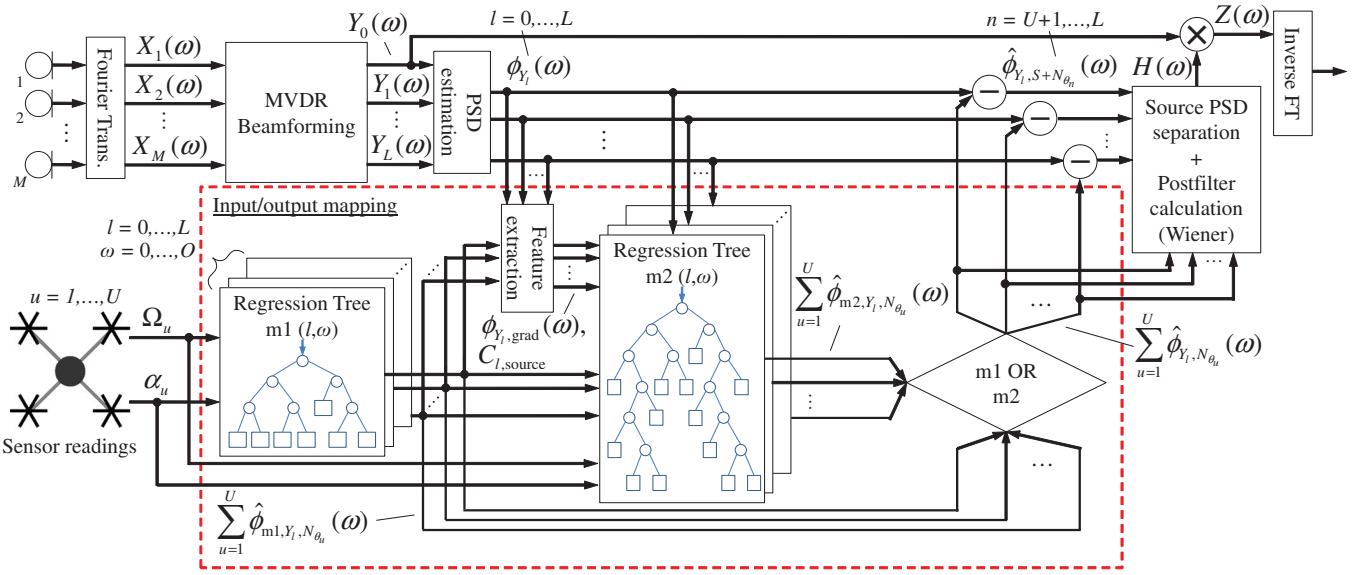

Fig. 2. Framework of the proposed method.

A. General framework and input features

Fig. 2 shows a block diagram of the proposed method. Note that t is omitted for brevity in the figure. The method retains the general beamforming to postfiltering framework, with the section highlighted by the red dashed-box showing the input/output mapping function utilised, which is the core of the rotor noise PSD estimator. Two configurations of ML-based mapping functions (m1 and m2) are investigated. The motivation behind each configuration and their input feature usage are presented below.

Configuration m1: The idea behind config. m1 is to use rotor characteristics to facilitate estimation of the rotor noise PSD, captured via sensor data. Thus config. m1 uses

- Rotor speed and acceleration ($\Omega_u(t)$, $\alpha_u(t)$).

While m1 performed well (see Section IV-B for details), upon closer observation of the rotor noise, it became clear that rotor speed/acceleration cannot fully capture the characteristics of the PSD spectra. Excluding hardware imperfections, additional factors include, but are not limited to practical limitations such as aliasing, or potentially uncaptured aeroacoustic effects. Therefore, an extension of config. m1, config. m2, attempts to capture these effects, in anticipation of further refining PSD estimation accuracy. Ideally, such information is modelled analytically. However, direct beamformer output PSD data has shown to fit well for the purpose of this study.

Configuration m2: In addition to $\Omega_u(t)$ and $\alpha_u(t)$, config. m2 also utilises

- Current/past frames of beamformer output PSD and PSD gradient ($\phi_{Y_l}(\omega,t)$, $\phi_{Y_l,\mathrm{grad}}(\omega,t)$).
- Beamformer expected output PSD ($\phi_{\mathrm{m1},Y_l,N_{\theta_u}}(\omega,t)$): $\phi_{Y_l}(\omega,t)$ contains rotor noise, as well as target and interfering noise signals, and thus will not be practical without a reference rotor noise only PSD. The output of config. m1 fulfils this requirement. Consequently, config. m1 cascades onto config. m2.
- Cue for non-UAV rotor noise activity ($C_{l,\mathrm{source}}(t)$): A logical input that utilises $\phi_{\mathrm{m1},Y_l,N_{\theta_u}}(\omega,t)$ to detect non-UAV rotor noise related activity (e.g. speech or wind). This is estimated as follows

$$C_{l,\mathrm{source}} = \begin{cases} 1 & c_l > c_{\mathrm{th}} \\ 0 & c_l < c_{\mathrm{th}} \end{cases} \tag{7}$$

$$c_l = \int_0^{2\pi} \max\left(\frac{\phi_{Y_l}(\omega) - \hat{\phi}_{\mathrm{m1},Y_l,N_{\theta_u}}(\omega) - \phi_{\mathrm{th}}}{\left|\phi_{Y_l}(\omega) - \hat{\phi}_{\mathrm{m1},Y_l,N_{\theta_u}}(\omega) - \phi_{\mathrm{th}}\right|}, 0 \right) d\omega, \tag{8}$$

where c_{th} is the overall activity threshold, and $\phi_{c_{\mathrm{th}}}$ is the frequency band activity threshold in dB, both heuristically tuned. Note that t is omitted for brevity in (7) and (8).

Note that for both the method in [10] and config. m1 only utilised a single form of input feature, while config. m2 incorporated both forms. In summary, the method from [10] is "acoustic single-sensory", m1 is "non-acoustic single-sensory", and m2 is "multisensory (acoustic/non-acoustic)".

B. Input/output mapping

A model is required to map the input features to the rotor noise PSD $\phi_{Y_l,N_{\theta_u}}(\omega,t)$. This is essentially a regression problem. Whilst rotor noise PSDs have a clear, distinct pattern, due to the multitude of tonal bands, each with their distinct behaviour, the overall response possesses non-linear characteristics, making it difficult to formulate a global solution. Non-parametric regression is a simple yet effective solution to this problem. However, techniques of this category can become computationally expensive. Regression trees [21] work well under these conditions, and have shown to fit well with the problem of this study, and is therefore used for initial proof of

ISBN 978-90-827970-1-5 © EURASIP 2018

TABLE I
SIMULATION PARAMETERS.

Sampling rate	48 kHz
STFT length (overlap shift)	2048 (1024)
Forgetting factor	0.3
# of beamformers	1
$\phi_{c_{\text{th}}}$ threshold (dB)	15
c_{th} threshold	2

TABLE II
SIMULATION DATA SPECIFICATIONS (* DENOTES m2 ONLY).

	Training	Testing
UAV speed range (revolutions per minute)	3000-4000	3000-3300
# of target source patterns	12 (6 male, 6 female)*	4 (2 male, 2 female)
# of interfering noise patterns	8 (4 traffic, 4 music)*	4 (2 traffic, 2 music)
# of int. noise angles	N/A	12
Input SRNR (dB)	$-10, 0, 10$*	$-10, 0, 10$
Input SINR (dB)	$-10, 5, 20$*	$-10, 0, 10, 20$
# of datasets	20	544
# of observations for each dataset	8994 (192 sec)	276 (6 sec)
Total # of observations	179880 (8994×20)	150144 (276×544)

concept. The regression trees are prepared for each independent O frequency bins, and for each beamformer output. Thus, a total of $L \times O$ regression trees are used, each satisfying a minimum mean squared error (MMSE). The regression trees are first grown, followed by pruning via validation using a separate test dataset to avoid overfitting. Finally, an optimal set of trees are selected for testing.

IV. SIMULATION

A. Simulation setup and test parameters

A simulation platform that mimics the UAV characteristics from the previous study [10] was developed to evaluate the performance of the proposed method. This includes the microphone array used (see Fig. 1), which consists of a circular front array of three equidistant microphones with a central microphone, spanning a radius of 0.16 meters, and with the array plane facing the target source. It also consists of a back array of two microphones, 0.2 meters away from the front array, that is 0.14 meters apart from each other. To provide controllable and repeatable testing conditions, a UAV rotor noise synthesiser using an additive variable-frequency wave approach (loosely similar to [23]) was developed to simulate the rotor noise as a function of rotor speed. This noise was then mixed with a corpus of target source patterns before multiplying with the TF.

Some simplifications to the UAV configuration were made. Firstly, the PSD estimation of the proposed method focused only on the beamformer output for the target source $Y_0(\omega, t)$, with the source limited to speech. Thus, only the front array was used. Furthermore, all sound sources were assumed as omnidirectional point sources in an ideal free field. Therefore, they were multiplied with components in $\mathbf{a}_\theta(\omega)$ described in

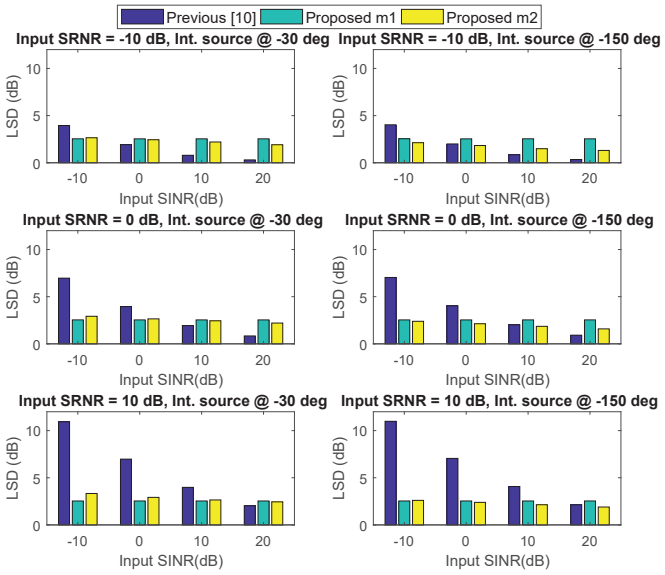

Fig. 3. PSD estimation accuracy for different input SRNR and SINR.

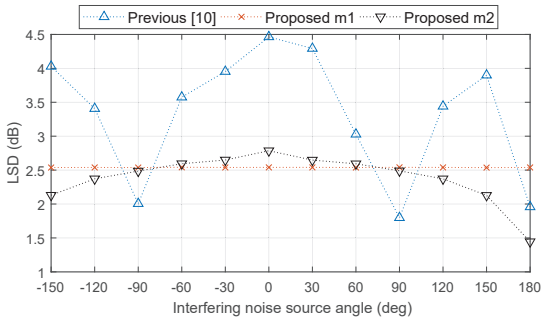

Fig. 4. PSD estimation accuracy for different interfering noise angle (at input SRNR/SINR of 0 dB).

Section II-A. Lastly, only one rotor was used, with a speed range replicating UAV under hover or slow travel.

Simulation parameter specifications are summarised in Table I. Two metrics are introduced to quantify the sound source input conditions: i) signal-to-rotor-noise-ratio (SRNR) and ii) signal-to-interfering-noise-ratio (SINR) [10]. SRNR quantifies the power ratio of the target source to the rotor noise, and SINR quantifies this ratio to the coherent interfering noise sources. These are measured at the central microphone in the front array (see Fig. 1). To evaluate the performance error of the proposed method against the previous method [10], the log spectral distortion (LSD) [24] of the estimated rotor noise PSD will be compared against the true rotor noise PSD (i.e. no presence of target or interfering noises). Table II summarises the parameters used to train and evaluate the ML mapping functions.

B. Results and Discussion

Fig. 3 summarises the results for different input SRNRs and SINRs. As expected, the previous method was greatly affected by surrounding sound sources; the higher the SRNR and lower the SINR, the higher the LSD. This is due to the higher overall power of the surrounding sound sources relative

ISBN 978-90-827970-1-5 © EURASIP 2018

to the rotor noise, leading to larger estimation error of the rotor noise PSD. This effect, on the other hand, was minimal for the proposed method, and is especially apparent with config. $\mathrm{m1}$. This is expected since it only takes rotor speed/acceleration as inputs, both immune to sound. However, overall config. $\mathrm{m2}$ had lower estimation errors than $\mathrm{m1}$, showing that the additional input features in $\mathrm{m2}$ played an important role in refining the estimation.

Fig. 4 shows the results for different interfering noise source angles. Similar to results in Fig. 3, interfering noise angles had practically no effect towards config. $\mathrm{m1}$, as opposed to the previous method, where a high degree of variance was seen. While performance was relatively similar, config. $\mathrm{m2}$ was able to deliver a lower overall LSD for many occasions significantly over $\mathrm{m1}$. This is perhaps due to taking beamformer output PSDs as an input, where its sharp directivity effectively rejected disturbances away from the target source angle. Although the previous method gave very low LSDs at angles of ± 90 and 180 degrees, this is expected as the sound sources were well separated from each other, which is ideal for the method's spatially dependent characteristic. However, it should be noted that such scenario is only possible if the interfering noise sources remained stationary at the desired angle.

Overall, the proposed methods reduced the fluctuation of performance seen in the previous method by using the parameters of the rotor's motion, meaning the method has successfully distinguished the rotor noise and other sound sources.

V. Conclusion

A UAV-characteristics-driven method was developed for accurate estimation of the UAV rotor noise PSD. The method utilised multisensory information from the UAV to estimate the rotor noise PSD for each frequency bin via regression tree. The performance of the method was evaluated via simulation and has shown fair robustness over influences from target source and other interfering noises. However, it should be noted that given its simplified problem setup, this study was merely intended for evaluating the method's conceptual feasibility, and thus development and optimisation for practical use remains part of future study. This includes experimental testing, optimisation of input features and using more elaborate input/output mapping techniques.

References

[1] B. Canis, "Unmanned aircraft systems (UAS): Commercial outlook for a new industry," *Congressional Research Service, Washington*, 2015.

[2] R. Verrier, "Drones are providing film and TV viewers a new perspective on the action," *Los Angales Times*, June 2017.

[3] M. Brandstein, D. Ward, A. Lacroix, and A. Venetsanopoulos, Eds., *Microphone Arrays*, Digital Signal Processing. Springer Berlin Heidelberg, Berlin, Heidelberg, 2001, DOI: 10.1007/978-3-662-04619-7.

[4] 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 *Acoustics 2012*, 2012.

[5] K. Furukawa, K. Okutani, K. Nagira, T. Otsuka, K. Itoyama, K. Nakadai, and H.G. Okuno, "Noise correlation matrix estimation for improving sound source localization by multirotor UAV," in *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*. 2013, pp. 3943–3948, IEEE.

[6] T. Ishiki and M. Kumon, "A microphone array configuration for an auditory quadrotor helicopter system," in *Safety, Security, and Rescue Robotics (SSRR), 2014 IEEE International Symposium on*. 2014, pp. 1–6, IEEE.

[7] T. Ishiki and M. Kumon, "Design model of microphone arrays for multirotor helicopters," in *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*. 2015, pp. 6143–6148, IEEE.

[8] K. Washizaki, M. Wakabayashi, and M. Kumon, "Position estimation of sound source on ground by multirotor helicopter with microphone array," in *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on*. 2016, pp. 1980–1985, IEEE.

[9] 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 *Intelligent Robots and Systems (IROS), 2016 IEEE/RSJ International Conference on*. 2016, pp. 1299–1304, IEEE.

[10] Y. Hioka, M. Kingan, G. Schmid, and K.A. Stol, "Speech enhancement using a microphone array mounted on an unmanned aerial vehicle," in *Acoustic Signal Enhancement (IWAENC), 2016 IEEE International Workshop on*. 2016, pp. 1–5, IEEE.

[11] J.E. Ffowcs Williams and D. L. Hawkings, "Theory relating to the noise of rotating machinery," *Journal of Sound and Vibration*, vol. 10, no. 1, pp. 10–21, 1969.

[12] G. Sinibaldi and L. Marino, "Experimental analysis on the noise of propellers for small UAV," *Applied Acoustics*, vol. 74, no. 1, pp. 79–88, Jan. 2013.

[13] P-S. Huang, M. Kim, M. Hasegawa-Johnson, and P. Smaragdis, "Joint Optimization of Masks and Deep Recurrent Neural Networks for Monaural Source Separation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 12, pp. 2136–2147, Dec. 2015.

[14] X-L. Zhang and D. Wang, "A Deep Ensemble Learning Method for Monaural Speech Separation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 5, pp. 967–977, May 2016.

[15] S. Sivasankaran, A. A. Nugraha, E. Vincent, J. A. Morales-Cordovilla, S. Dalmia, I. Illina, and A. Liutkus, "Robust asr using neural network based speech enhancement and feature simulation," in *2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, Dec 2015, pp. 482–489.

[16] S. Araki, T. Hayashi, M. Delcroix, M. Fujimoto, K. Takeda, and T. Nakatani, "Exploring multi-channel features for denoising-autoencoder-based speech enhancement," in *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*. 2015, pp. 116–120, IEEE.

[17] Y. Tu, J. Du, Y. Xu, L. Dai, and C-H. Lee, "Deep neural network based speech separation for robust speech recognition," in *Signal Processing (ICSP), 2014 12th International Conference on*. 2014, pp. 532–536, IEEE.

[18] G. Hinton, L. Deng, D. Yu, G. Dahl, A-R. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T.N. Sainath, and B. Kingsbury, "Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, Nov. 2012.

[19] K. Niwa, Y. Koizumi, T. Kawase, K. Kobayashi, and Y. Hioka, "Pinpoint extraction of distant sound source based on DNN mapping from multiple beamforming outputs to prior SNR," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2016, pp. 435–439.

[20] T. Kawase, K. Niwa, K. Kobayashi, and Y. Hioka, "Application of neural network to source PSD estimation for wiener filter based array sound source enhancement," in *Acoustic Signal Enhancement (IWAENC), 2016 IEEE International Workshop on*. 2016, pp. 1–5, IEEE.

[21] L. Breiman, J. Friedman, C.J. Stone, and R.A. Olshen, *Classification and regression trees*, CRC press, 1984.

[22] 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.

[23] K. Tsuge, K. Kanamaru, T. Kido, and N. Masuda, "A study of noise in vehicle passenger compartment during acceleration," in *SAE Surface Vehicle Noise and Vibration Conference*. may 1985, SAE International.

[24] L.R. Rabiner and B-H. Juang, *Fundamentals of speech recognition*, PTR Prentice Hall, 1993.

ISBN 978-90-827970-1-5 © EURASIP 2018