Acoustic detection of unmanned aerial vehicles using biologically inspired vision processing

Jian Fang, Anthony Finn, Ron Wyber, et al.

ARTICLES YOU MAY BE INTERESTED IN

Acoustic detection of unmanned aerial vehicles using biologically inspired vision processing[a)]

Jian Fang,[1] Anthony Finn,[1,b)] Ron Wyber,[2] and Russell S. A. Brinkworth[3]

[1]*Science, Technology, Engineering and Mathematics, University of South Australia, Adelaide, South Australia 5095, Australia*

[2]*Midspar Systems, Farrer Place, Oyster Bay, New South Wales 2225, Australia*

[3]*College of Science and Engineering, Flinders University, Clovelly Park, South Australia 5042, Australia*

ABSTRACT:

Robust detection of acoustically quiet, slow-moving, small unmanned aerial vehicles is challenging. A biologically inspired vision approach applied to the acoustic detection of unmanned aerial vehicles is proposed and demonstrated. The early vision system of insects significantly enhances signal-to-noise ratios in complex, cluttered, and low-light (noisy) scenes. Traditional time-frequency analysis allows acoustic signals to be visualized as images using spectrograms and correlograms. The signals of interest in these representations of acoustic signals, such as linearly related harmonics or broadband correlation peaks, essentially offer equivalence to meaningful image patterns immersed in noise. By applying a model of the photoreceptor stage of the hoverfly vision system, it is shown that the acoustic patterns can be enhanced and noise greatly suppressed. Compared with traditional narrowband and broadband techniques, the bio-inspired processing can extend the maximum detectable distance of the small and medium-sized unmanned aerial vehicles by between 30% and 50%, while simultaneously increasing the accuracy of flight parameter and trajectory estimations. © 2022 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

https://doi.org/10.1121/10.0009350

I. INTRODUCTION

Passive distributed acoustic sensor arrays have been used for detecting and tracking moving aircraft,[1–8] ground vehicles,[9–15] and unmanned aerial vehicles (UAV).[16–25] Several array configurations have been explored, including the small aperture circular array,[4] the L-shaped planar array,[7] a tetrahedral array,[6] and the widely distributed small arrays.[10–12] By analysing the acoustic signals received by the individual microphones, a variety of applications can be realized, including object classification, target tracking, and simultaneous localization and mapping (SLAM).[26] With the fast development of UAV platforms, these systems have become a very useful tool for a wide variety of applications,[27] including structure inspection,[28] surveillance,[29] 3D mapping,[30] and acoustic tomography.[31] Nevertheless, an unauthorized UAV may pose a threat to an airport, individuals, or military bases. Therefore, long range detection and precise location of the UAV becomes important for safety and security purposes.

The acoustic signature emitted by UAVs makes their passive detection and tracking possible. Depending on the spectral components of the acoustic signal, two traditional processing techniques for flight parameter estimation are readily available. For propeller-driven aircraft and helicopters that emit strong harmonic tones a narrowband processing technique[20,32] may be used to estimate the flight parameters. The approach is based on identification of the instantaneous frequency of the motion-induced, Doppler shifted acoustic signature of the aircraft. A broadband processing technique,[20] which measures the temporal variation of the time delays between multiple microphone pairs, may also be used. Compared with the narrowband technique, the broadband approach is more flexible because it can handle UAVs that do not emit strong narrowband tones or fixed harmonic frequencies. Accurate estimation of flight parameters is best achieved through application of both methods when a UAV overflies the array. Indeed, using such approaches several researchers have reported detection ranges for aircraft and UAVs in excess of 2 km.[2,17,21,33] When a UAV is far away from the microphones, however, the signal is weak compared to noise and both broad and narrowband approaches struggle to achieve reliable results. This raises a challenge for UAV detection, localization, and tracking, as observation of the acoustic signal at long range is usually highly desirable.

Similar signal conditions exist in the natural world. For instance, the spread of luminance in naturally lit scenes typically covers a very large dynamic range, and the details in dark regions are immersed in noise.[34] In this regard, insect

---

visual systems, such as that of the hoverfly, have proven to be a powerful information capture system.[35] Similar to the visual scenes, acoustic signals can be converted into "images" using spectrograms and correlograms. These traditional techniques thus permit us to visualize the (one-dimensional) acoustic signal as (two-dimensional) images based on a corresponding time-frequency analysis. With the narrowband method, detected acoustic signals are usually visualized as spectrograms and the frequency of harmonics extracted. For the broadband technique correlograms are used, from which the time delays can be obtained. In this sense, the acoustic signal of interest, in the form of harmonics or correlation peaks, can be presented as patterns in two-dimensional arrays or matrices and analysed using vision processing techniques.

The particular model used in this study is described in more detail below in Sec. III and the references therein. It derives from a fully elaborated biologically inspired vision (BIV) model that, in addition to enabling scene agnostic, sub-pixel motion detection via electro-optic and infrared sensor modalities, can apply signal conditioning to sensor output prior to downstream processing. The model is based on multiple layers of non-linear dynamic adaptive components measured or suspected from responses of neurons in pathways of the hoverfly brain. Uniquely, however, we have transitioned the techniques from biological finding and theoretical simulation to embedded hardware implementation running in real time.[36] The approach enhances and suppresses elements of related and unrelated signal and noise, providing crisp sub-pixel/low-amplitude signal detection and classification for these difficult target sets.[37]

This vision-inspired method differs substantially from techniques that draw upon the biology of auditory systems.[38,39] The technique also differs from the well-known biologically inspired convolution neural network (CNN) method.[40,41] A CNN is an adaptive algorithm that is based on interconnected nodes (neurons) arranged in a layered structure to resemble the human brain. The CNN learns from pre-supplied data and can thus be trained to classify the data by breaking down the input layers into layers of abstraction. It is trained using many examples, by the way in which the individual neurons are connected and by the strength (weight) of those connections. The weights are automatically adjusted during training according to specified learning rules until the algorithm performs appropriately. The last layer of the CNN indicates whether the target class (in our case a UAV signature) was present within the data.

This means that while the hidden layers of a CNN may extract and enhance features the core task is typically classification, and any decision making is based on patterns in the training data. This contrasts with the proposed algorithm in this paper, which is focused on enhancement only: classification of the data is still required post facto. This means the two approaches are not mutually exclusive. In fact, there is reason to believe a CNN trained on outputs from the BIV model could be smaller and more accurate than one trained on raw data (due to enhancement of the signals relative to the noise).

This paper is organized into six sections including the present one. Section II briefly describes the traditional methods of UAV acoustic signal visualization, including the narrowband and broadband processing. Section III describes the principles of the bio-vision technique. In Sec. IV field trial used to gather the experimental data is explained. The results of the comparisons between the traditional and bio-inspired methods are given in Sec. V. Finally, in Sec. VI this work is summarized and possible future investigations suggested.

II. UAV ACOUSTIC SIGNAL VISUALIZATION

A. Notations and assumptions

Passive acoustic localization[9,33] is based on the following assumptions. First, the atmosphere is assumed to be an iso-speed sound propagation medium, with the speed of sound denoted as c. Second, the UAV is assumed to travel in a straight line at constant speed V and altitude h for the duration of the inter-observation period, as shown in Fig. 1. The position of the UAV at time τ is $\mathbf{u}(\tau) = [x(\tau), y(\tau), z(\tau)]^T$, which can be expressed in Cartesian coordinates as[9]

$$x(\tau) = d_c \cos \alpha_c + (\tau - \tau_c)V \sin \alpha_c,$$
$$y(\tau) = d_c \sin \alpha_c - (\tau - \tau_c)V \cos \alpha_c,$$
$$z(\tau) = h, \tag{1}$$

where τ_c, d_c, and α_c are the time, the ground range and the bearing angle at the closest point of approach (CPA) to the origin $[0, 0, 0]^T$, respectively, h is the UAV altitude, and $R_c = \sqrt{d_c^2 + h^2}$ is the slant range of CPA. In this case, the UAV trajectory can be explicitly described by five flight parameters $\{V, \tau_c, h, d_c, \alpha_c\}$. Therefore, the UAV position can also be expressed as $\mathbf{u}(\tau; \mathbf{k})$, where $\mathbf{k} = [V, \tau_c, h, d_c, \alpha_c]^T$ is the flight parameter vector. From Eq. (1), the UAV velocity is $\dot{\mathbf{u}} = [V \sin \alpha_c, -V \cos \alpha_c, 0]^T$. Note that these assumptions are only required to be satisfied during the inter-observation period, which is typically short compared to flight dynamics.

The UAV position $\mathbf{u}(\tau)$ can be described in spherical coordinates. Equivalently, $\mathbf{u}(\tau) = r(\tau) \cdot \mathbf{q}(\tau)$, where $r(\tau)$ is the slant range at time τ, and $\mathbf{q}(\tau)$ is the direction-of-arrival (DOA) unit vector defined as

$$\mathbf{q}(\tau) = [\cos \theta(\tau) \cos \phi(\tau), \cos \theta(\tau) \sin \phi(\tau), \sin \theta(\tau)]^T, \tag{2}$$

in which $\theta(\tau)$ and $\phi(\tau)$ are the bearing and elevation angles, respectively. The non-linear relationship between the bearing and elevation angles and the source location is given by

$$\theta(\tau) = \tan^{-1}\left(\frac{y(\tau)}{x(\tau)}\right),$$
$$\phi(\tau) = \tan^{-1}\left(\frac{z(\tau)}{\sqrt{x^2(\tau) + y^2(\tau)}}\right). \tag{3}$$

The acoustic array here comprises a general form, that has a wide aperture with N_{RS} small aperture arrays, each

J. Acoust. Soc. Am. **151** (2), February 2022
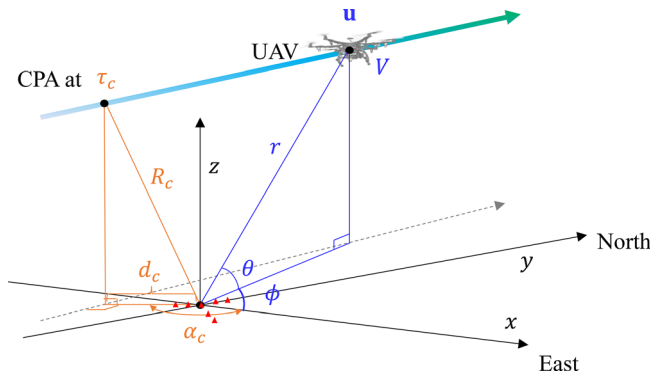
Fang *et al.* 969

FIG. 1. (Color online) Flight model and the geometrical configuration of acoustic array. The UAV flies past the array in a straight line at constant speed and altitude. Red triangles: passive acoustic array; grey dashed line with arrow: trajectory projection on the *x*-*y* plane.

with N_{CH} microphones (each microphone representing a different channel). The j-th channel of the i-th small acoustic array is denoted as $S_{i,j}, 1 \leq i \leq N_{RS}, 1 \leq j \leq N_{CH}$ located at position $\mathbf{s}_{i,j} = [x_{i,j}, y_{i,j}, z_{i,j}]^T$. The central position of the i-th small microphone array is denoted as $\mathbf{s}_i = [x_i, y_i, z_i]^T$, which can be obtained through $\mathbf{s}_i = (1/N_{CH})\sum_{j=1}^{N_{CH}} \mathbf{s}_{i,j}$. In this paper, one small microphone array is located at the origin point of the coordinate system (see Fig. 6 later).

B. Narrowband technique

The narrowband technique suits acoustic signals that contain strong narrowband tones. The acoustic signals are visualized as spectrograms using time-frequency analysis. Figure 2 shows the power spectral density (in dB) of a UAV observed by a single ground microphone. As well as the harmonic tones emitted by the UAV, the wind noise and acoustical Lloyd's mirror effects are also visible. The wind noise exists throughout the observation period, its spectral width varying in accordance with different gust intensities. Although the maximum frequency of wind noise contamination can reach 500 Hz and simply high-pass filtering the signal could eliminate wind noise, considerable harmonic information could also be discarded, which is not desirable.
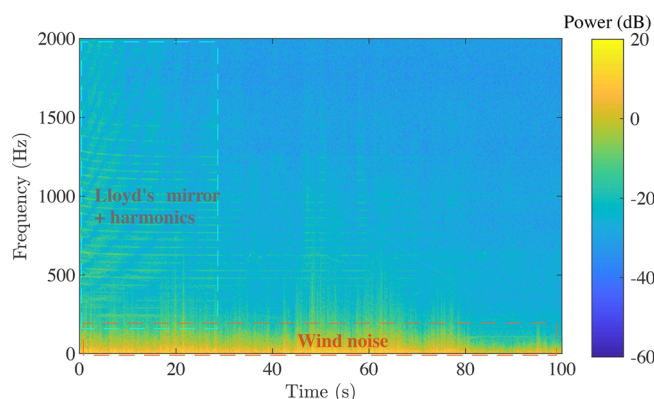
The Lloyd's mirror effect, generated by reflection of acoustic waves by the ground, is clearly visible when the UAV is near the array. This introduces a slow-change variation in the spectrogram, making some harmonic components hard to detect. For instance, in Fig. 2 the amplitude of the harmonic signal around 600 Hz at time 10 s is almost as low as the noise floor, which results in a breakpoint in the harmonics of this order. It is noted that the Lloyd's mirror effect can be utilized in estimating some flight parameters.[42,43] However, the Lloyd's mirror effect is distinct within only a short period when the airborne source flies over the acoustic sensor and thus not suitable for long-distance detection.

To mitigate the wind noise and Lloyd's mirror effect, the cepstrum filtering technique is used.[44] The (power) cepstrum of a signal $x(t)$ is defined as

$$\tilde{x}(q) = \mathcal{F}^{-1}\left\{ \log |X(f)|^2 \right\}, \tag{4}$$

where $\mathcal{F}(\cdot)$ is the Fourier transform and $X(f) = \mathcal{F}\{x(t)\}$ is the spectrum of signal $x(t)$. q is the quefrency variable of the cepstrum. After performing the cepstrum transform on the spectrogram $X(f, t)$, the outputs form a new image called the cepstrogram $\tilde{x}(q, t)$.

Figure 3 shows the cepstrogram for the same UAV considered in Fig. 2. Unlike the spectrogram, in the cepstrogram the power of wind noise is concentrated within a fixed quefrency range. The Lloyd's mirror is distributed within $0 < q < 0.01$ s and $0 \leq t < 20$ s, corresponding to the UAV being within a range of about 300 m from the array. The harmonic components are distributed near $q \approx 0.02$ s, which is clearly distant from both the Lloyd mirror and the wind noise. Therefore, by applying a bandpass filter that retains the harmonic components on the cepstrogram we can effectively distill the essential harmonics. Figure 4(a), which contains only the harmonics, shows the result of cepstrum filtering. Figure 4(b) shows the background spectrogram and contains only the Lloyd mirror and wind noise. Using this approach, we can eliminate the influence of the unwanted signals.

FIG. 2. (Color online) Power spectral density vs time (spectrogram) for a UAV transit observed on a ground microphone 2 m above ground.
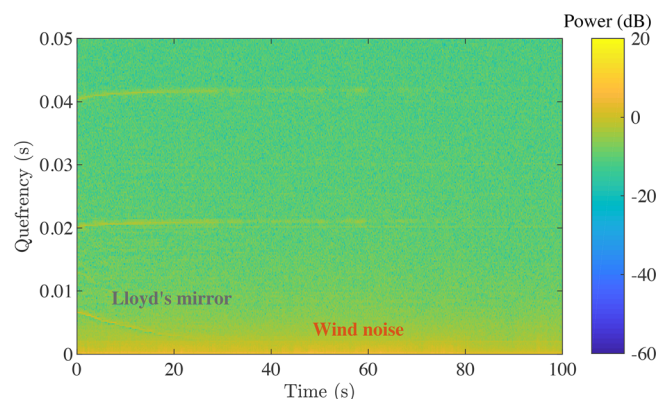
FIG. 3. (Color online) The cepstrogram for the same UAV considered in Fig. 2.

970   J. Acoust. Soc. Am. **151** (2), February 2022
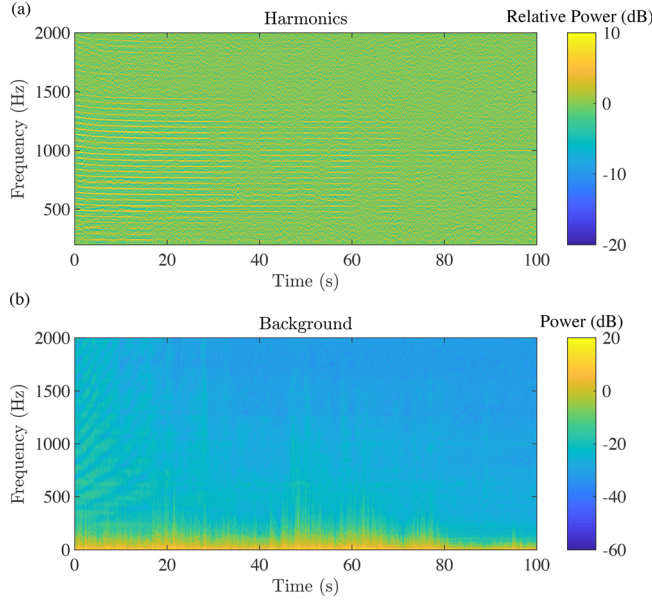
Fang *et al.*

(a)

(b)

FIG. 4. (Color online) Result of cepstrum filtering applied on the spectrogram in Fig. 2. (a) Harmonic spectrogram which contains acoustic harmonics only; (b) Background spectrogram, which contains the wind noise and Lloyd mirror.

After cepstrum filtering, the harmonic spectrogram is then processed for pitch detection. A number of traditional methods are available for high-accuracy pitch detection. For example, in time domain the zero-crossing rate (ZCR),[45] robust algorithm for pitch tracking (RAPT)[46] and YIN estimators[47] may be used, while in the frequency domain there are component frequency ratios,[48] cepstrum analysis,[44] optimum comb filters,[49] and harmonic product spectrum (HPS).[50] In this paper, since the acoustic signal has been visualized into spectrogram, the HPS method for pitch estimation is adopted.

The tonal frequency emitted by the UAV at time τ is received by the i, j-th acoustic sensor at time t. This can be expressed as[23]

$$
\begin{aligned}
f_{i,j}(t;\mathbf{k}) &= f_{i,j}(\tau + \Delta\tau_{i,j}(\tau;\mathbf{k});\mathbf{k}) \\
&= f_0 \left\{ 1 + \frac{[\mathbf{p}(\tau;\mathbf{k}) - \mathbf{s}_{i,j}]^T \dot{\mathbf{p}}}{||\mathbf{p}(\tau;\mathbf{k}) - \mathbf{s}_{i,j}||} \right\}^{-1},
\end{aligned}
\tag{5}
$$

where $\Delta\tau_{i,j}(\tau;\mathbf{k})$ is the travel time from the UAV to the i,j-th microphone, i.e.,

$$
\Delta\tau_{i,j}(\tau;\mathbf{k}) = ||\mathbf{u}(\tau;\mathbf{k}) - \mathbf{s}_{i,j}||/c.
\tag{6}
$$

Using the relation $t = \tau + \Delta\tau_{i,j}(\tau;\mathbf{k})$ and substituting it into Eq. (6), we have

$$
f_{i,j}(t;\mathbf{k}) = f_0 \left[\frac{c^2}{c^2 - v^2}\right] \cdot \left\{ 1 - \frac{v^2(t - \tau_c) + vy'_{i,j}}{\sqrt{(c^2 - v^2)\left[(d_c + x'_{i,j})^2 + h^2\right] + c^2\left[v(t - \tau_c) + y'_{i,j}\right]^2}} \right\},
\tag{7}
$$

where $\mathbf{s}'_{i,j} = [x'_{i,j}, y'_{i,j}, z'_{i,j}]^T = \mathbf{R}(\alpha_c)\mathbf{s}_{i,j}$. $\mathbf{R}(\alpha)$ is the rotation matrix written as

$$
\mathbf{R}(\alpha) = pr \begin{bmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}.
\tag{8}
$$

For the acoustic sensor located at the origin, Eq. (7) reduces to the same form as in Ref. 32.

It is worth noting that not all UAVs exhibit strong acoustic harmonics. Consequently, the narrowband processing is not suitable for all UAV signatures.

C. Broadband technique

The other traditional method for detecting and locating acoustic signatures is based on broadband processing. Unlike the spectrograms in the narrowband technique, the broadband processing uses correlograms to visualise the acoustic signals. Suppose the acoustic signal emitted by the UAV at time τ arrives at two microphones (the m-th and n-th channels) of the i-th array at time t, the time delay between these channels can be estimated from the correlogram $C_{i,mn}(\beta,t)$, which is obtained through the generalized cross correlation and phase transform (GCC-PHAT) method[51,52]

$$
C_{i,mn}(\beta,t) = \mathcal{F}^{-1}\left\{ W(f)\frac{X_{i,m}(f,t)X_{i,n}^*(f,t)}{|X_{i,m}(f,t)X_{i,n}^*(f,t)|} \right\},
\tag{9}
$$

where β is the lag. $X_{i,m}(f,t)$ and $X_{i,n}(f,t)$ are the complex-valued spectrograms of the m-th and n-th channels of the i-th array, respectively. $W(f)$ is the spectral windowing function. The time delay between these two channels can be obtained by searching the peak in the correlogram, i.e.,

$$
\delta_{i,mn}(t) = \arg\max_{\beta} C_{i,mn}(\beta,t).
\tag{10}
$$

When the flight trajectory is fully described by Eq. (1), the time delay $\delta_{i,mn}(t)$ is also referred as $\delta_{i,mn}(t;\mathbf{k})$. According to Eq. (6), it can be rewritten as

J. Acoust. Soc. Am. **151** (2), February 2022

Fang *et al.* 971

$$\delta_{i,mn}(t;\mathbf{k}) = \Delta\tau_{i,m}(\tau;\mathbf{k}) - \Delta\tau_{i,n}(\tau;\mathbf{k})$$
$$= \frac{||\mathbf{u}(\tau;\mathbf{k}) - \mathbf{s}_{i,m}|| - ||\mathbf{u}(\tau;\mathbf{k}) - \mathbf{s}_{i,n}||}{c}, \quad (11)$$

where $\Delta\tau_{i,m}(\tau;\mathbf{k})$ and $\Delta\tau_{i,n}(\tau;\mathbf{k})$ are the travel time from the UAV to the microphone channels $S_{i,m}$ and $S_{i,n}$, respectively. Since the channels were placed near to each other within an array, the relationship between the sound emitted at time τ and received at time t can be treated as a plane wave across the small array. Thus, considering $t = \tau + \Delta\tau_i(\tau;\mathbf{k})$ and solving Eq. (6) along with Eq. (1), the relationship between τ and t for the i-th acoustic array centered at $\mathbf{s}_i$ can be expressed as

$$\tau = \tau_c - \frac{y_i'}{v} + \frac{c^2\left[v(t-\tau_c)+y_i'\right]}{v(c^2-v^2)}$$
$$- \frac{\sqrt{(c^2-v^2)\left[(d_c+x_i')^2+h^2\right]+c^2\left[v(t-\tau_c)+y_i'\right]^2}}{c^2-v^2}.$$
$$(12)$$

When the centre of the small microphone array is located at the origin point, i.e., $\mathbf{s}_i = \mathbf{s}_i' = [0,0,0]^T$, Eq. (12) reduces to the same form as Ref. 9.

Besides the correlogram, the acoustic signal in broadband processing can also be visualized into the global coherence field (GCF).[53] A GCF shows the plausibility for an acoustic source. For the i-th acoustic array, it can be calculated as

$$G_i(\mathbf{u},t) = \frac{1}{M}\sum_{k=1}^{M} C_{i,k}(\beta_i(\mathbf{u}),t), \quad (13)$$

where $M = \binom{2}{N_{CH}}$ is the number of microphone pairs. $\beta_i(\mathbf{u})$ is the expected at time delay of the i-th small acoustic array when the acoustic source is at position $\mathbf{u}$. The GCF can also be presented as a 2-dimensional image in terms of the bearing and elevation angles, based on the DOA unit vector $\mathbf{q}$ defined in Eq. (2).

III. BIO-VISION PROCESSING

A complete biologically inspired vision (BIV) model is a multi-stage non-linear system with adaptive feedback both within and between stages. This model has previously been used only on electromagnetic data but has shown great promise in estimating optic flow[54] and detecting targets in clutter[34] in both visual[36] and infrared[37] portions of the spectrum. Even using the first stage of the model in isolation has yielded improved clarity in poor lighting conditions[55,56] and better target detection.[57,58]

The photoreceptor cell (PRC), which is responsible for dynamic range reduction of the input signal, provides the first stage of the biological visual system. Photoreceptors dynamically adjust the dark and bright regions of input images through temporal pixel-wise operations.[59]

Figure 5 shows the elaborated mathematical model of the bio-vision photoreceptor.[34] It includes four stages: (1) the adaptive filtering; (2) the low-pass divisive feedback, which is also called as the DeVries-Rose stage; (3) the exponential low-pass divisive feedback, which is known as the Weber stage; and (4) the non-linear Naka-Rushton transform. This four-stage model is functionally equivalent to the processing conducted in a primate cone.[60] The detailed implementation is described as follows.

A. Adaptive filter

This stage includes low-pass filtering with a dynamic cut-off frequency, the value of which depends on the adaptation state (long-term average value of the element). This is followed by a variable gain control that acts to reduce the differences over a wide range of acoustic intensities by using a larger gain when the signal is low compared to when it is high. In visual images the power of the signal component is approximately inversely proportional to its spatial frequency, whereas the noise component is essentially white, i.e., constant over all frequencies. There is therefore a frequency above which the signal-to-noise ratio (SNR) falls below an acceptable level and a low-pass filter should be employed. The threshold at which the SNR drops varies in accordance with the intensity (brightness) of the input signal because the majority of the noise is in the sensor itself and therefore independent of the scene being observed. Similarly, in acoustic applications the elements that relate to low amplitude acoustic signals must be more heavily filtered than those that receive high amplitude inputs since their
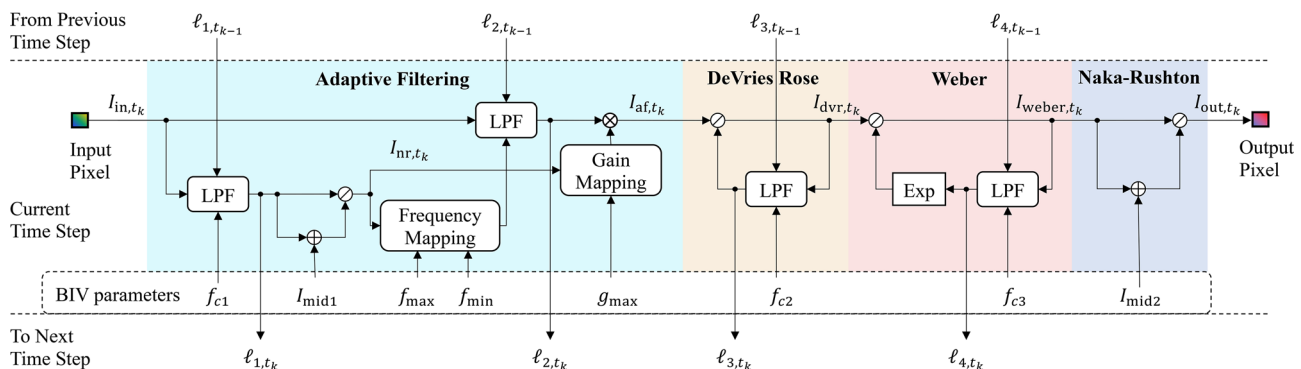

FIG. 5. (Color online) Mathematical model of the photoreceptor cells of a hoverfly's early vision system.

SNR will be low and can be increased by the reduction of high frequency signal components.

To begin, the adaptation state is calculated as

$$\ell_{1,t_k} = f_{\text{LPF}}(I_{\text{in}}, f_{c_1}, \ell_{1,t_{k-1}})$$
$$= \frac{2\pi f_{c1}}{f_r} I_{\text{in},t_k} + \left(1 - \frac{2\pi f_{c1}}{f_r}\right)\ell_{1,t_{k-1}}, \qquad (14)$$

where f_{c1} is the corner frequency of this stage and f_r is the frame rate, which is the reciprocal of the time step interval. Then the low-passed filtered result ℓ_{1,t_k} passes through a tone-mapping (bit depth normalisation) process, which is realized via a Naka-Rushton transform[61] in order to estimate the adaptation state, I_{nr,t_k}, expressed as

$$I_{\text{nr},t_k} = \frac{\ell_{1,t_k}}{\ell_{1,t_k} + I_{\text{mid1}}}, \qquad (15)$$

where I_{mid1} is the mid-point value chosen from the empirical data set. By using the adaptation state it is possible to independently classify each element of the incoming data by its average intensity, and hence estimate the filtering required to improve the SNR and the gain required to amplify low intensity sections if the input signal. The adaptive LPF is realized through

$$\ell_{2,t_k} = f_{\text{LPF}}(I_{\text{in},t_k}, f_{\text{fm},t_k}, \ell_{2,t_{k-1}}), \qquad (16)$$

where f_{fm,t_k} is the adaptive corner frequency calculated as

$$f_{\text{fm},t_k} = (f_{\text{max}} - f_{\text{min}})I_{\text{nr},t_k} + f_{\text{min}}. \qquad (17)$$

In Eq. (17), f_{max} and f_{min} are the bio-vision parameters corresponding to the maximal and minimal adaptation rates of the temporal LPF, respectively, and are set depending on the level of noise in the sensor relative to the intensity of the recorded signal.

Last, an operation to compress the dynamic range is applied, using a non-linear adaptive gain which is given by

$$g_{\text{af},t_k} = (g_{\text{max}} - 1)(1 - I_{\text{nr},t_k}) + 1. \qquad (18)$$

The gain factor g_{af,t_k} compresses the dynamic range by amplifying the lower values more than the higher ones.

The output signal of this stage is expressed as

$$I_{\text{af},t_k} = g_{\text{af},t_k}\ell_{2,t_k}. \qquad (19)$$

The initialization of this stage is derived from the steady-state response, i.e., $\ell_{1,t_0} = \ell_{2,t_0} = I_{\text{in},t_0}$, $I_{\text{nr},t_0} = I_{\text{in},t_0}/(I_{\text{in},t_0} + I_{\text{mid1}}$, $I_{\text{af},t_0} = I_{\text{af},t_0}[(g_{\text{max}} - 1)I_{\text{mid1}}/(I_{\text{in},t_0} + I_{\text{mid1}}) + 1]$, where I_{in,t_0} is the initial input to the bio-vision model.

B. Low-pass divisive feedback (DeVries-Rose)

The second stage incorporates rapid, short-term adaption of the input intensities, allowing the element to respond, but quickly adapt, to adjust for any changes. As shown in Fig. 5, it has a feedback loop with a LPF such that

$$\ell_{3,t_k} = f_{\text{LPF}}(I_{\text{dvr},t_{k-1}}, f_{c2}, \ell_{3,t_{k-1}}), \qquad (20)$$

where $I_{\text{dvr},t_{k-1}}$ is the output of the previous time step, and f_{c2} is the corner frequency of this stage. Function f_{LPF} is defined in Eq. (14). The filtered signal from the previous time step serves as the denominator in the divisive feedback loop (i.e., divided by the input of this stage), which can be written as

$$I_{\text{dvr},t_k} = I_{\text{af},t_k}/\ell_{3,t_k}. \qquad (21)$$

The steady state-behaviour follows a square root, thus the initialization of this stage at time step t_0 is $I_{\text{dvr},t_0} = \sqrt{I_{\text{af},t_0}}$, which is also called the DeVries-Rose law. Due to the low-pass filter, this stage produces overshoots and undershoots at incremental and decremental steps, respectively. The result of this stage is an output that responds to a change in the input with no time delay but will decay over time, asymptotically approaching the square-root of the input value, if the input remains unchanged. This preserves the presence and temporal coherence of changes while simultaneously reducing the required bandwidth of the signal. The operation is similar to a leaky high-pass filter with the addition of compressive non-linearity. Weighting with current and previous iterations, this process supports temporal coherency when processing input sequences.

C. Exponential low-pass divisive feedback (Weber)

The third stage is the Weber model, which contains an exponential operation in the feedback loop. It is parametrically similar to the previous DeVries-Rose stage, but provides long-term, slow adaptation due to the lower corner frequency in the filter, but includes the exponential to alter the rate of change of the system to a disturbance. As with the previous stage the temporal filtering is not on the main signal path, only in the feedback. This allows the model to adapt to slow changes in intensity while maintaining temporal coherency and resistance to high-frequency changes in the overall scene. The low-pass feedback loop in this stage is

$$\ell_{4,t_k} = f_{\text{LPF}}(I_{\text{weber},t_{k-1}}, f_{c3}, \ell_{4,t_{k-1}}), \qquad (22)$$

where $I_{\text{weber},t_{k-1}}$ is the output at previous time step and f_{c3} is the corner frequency of this stage. The filtered signal is first manipulated by an exponential operation and then divided by the input signal of the current time step. It is written as

$$I_{\text{weber},t_k} = I_{\text{dvr},t_k}/\alpha e^{\ell_{4,t_k}}, \qquad (23)$$

where α is the exponential sensitivity of the system.

The initialization of this stage comes from the steady state function that $I_{\text{weber},t_0} = I_{\text{dvr},t_0}/\exp(I_{\text{weber},t_0})$ with the solution $I_{\text{weber},t_0} = W(I_{\text{dvr},t_0})$, where $W(\cdot)$ is the Lambert W function. When $\ln(x) \ll x$, the solution can be approximated as $I_{\text{weber},t_0} = \ln(I_{\text{dvr},t_0})$. This logarithmic relationship is referred to as the Weber law, which is the name of this stage. The exponential scaling enables this stage to perform

significant non-linear rescaling of the signal, which can drastically reduce the intensity of the largest elements.

D. Non-linearity (Naka-Rushton)

The final stage of the photoreceptor is the static non-linearity Naka-Rushton transform process expressed as

$$I_{\mathrm{out},t_k} = \frac{I_{\mathrm{weber},t_k}}{I_{\mathrm{weber},t_k} + I_{\mathrm{mid2}}}, \tag{24}$$

where I_{mid2} is an empirically selected positive offset. As a result, the response becomes increasingly non-linear as the intensity values rise.

IV. FIELD TRIAL EXPERIMENT

A. Field trial equipment

Field trials were conducted at a site known as Evett's Field at the Woomera Test Range, South Australia. The terrain is flat and open, has sandy/rocky ground, no grass and sparse vegetation. Aside from an equipment hut 300 m west of the microphone array there are no substantial scattering objects obstructing line of sight propagation. As shown in Fig. 6(a), Evett's Field has two runways (Runway #1 and Runway #2), each with a length of around 2 km. The acoustic array was located to the south eastern end of the two

runways. An array of 49 microphones was deployed in a fractal pattern of 7 groups of 7 smaller arrays (Fig. 6). Each small array comprised a microphone at its centre (height 2 m) and two sets of three microphones (height 0.15 m) at radii of 1 m and 5 m, each separated in angle by 120°. The 7 smaller arrays were themselves arranged in a similar pattern of equilateral triangles, the inner triangle having 50 m sides, the outer 100 m sides. The position of each microphone was located using real time kinematic carrier phase differential global positioning system, which has a 1σ accuracy of ± 0.03 m.

The sound fields at each array were measured using ECM800 10 mV/Pa condenser microphones sampled at 44.1 kHz using an 8-channel, 24-bit Data Acquisition (DAQ) recorder with 107 dB spurious free dynamic range. Accurate time stamping of the data were obtained from a GPS-derived one pulse per second (1PPS), sampled using channel one of the DAQ.

B. Flight scenarios

Different classes of UAVs were used during the trials, including a DJI Matrice 600 (15 kg, 1.7 m diameter, hexa-rotor), a Skywalker X-8 (3.5 kg, 2.1 m wingspan), and a DJI Mavic Air (0.5 kg, 0.2 m, quad-rotor). The flight scenarios reported in this paper are shown in Table I. The Matrice 600 was equipped with an acoustic payload which continuously
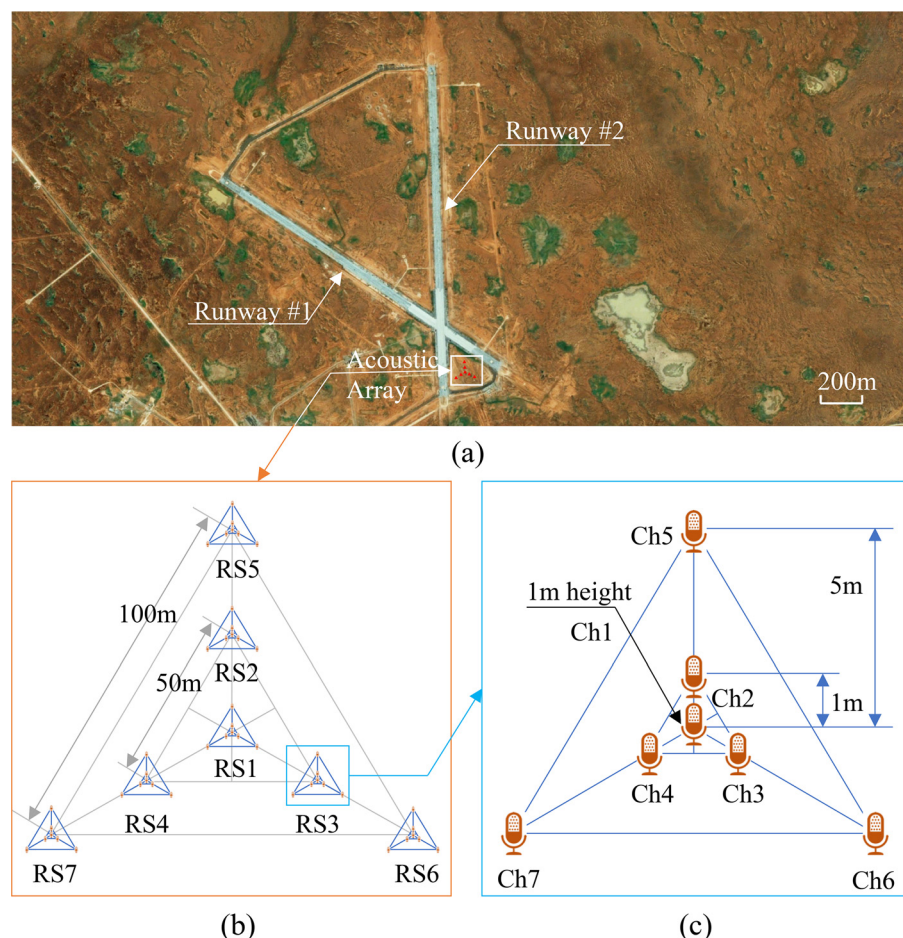

FIG. 6. (Color online) (a) Field trial deployment. (b) The distribution of large acoustic array and (c) small acoustic array (north is up for all images).

TABLE I. Flight scenarios. (N: North, S: South.)

Label	UAV	Payload	Height(m)	Profile
FS1	Matrice 600	Acoustic	100	Runway#2 S →N
FS2	Skywalker X-8	N/A	210	Runway#2 S →N
FS3	Mavic Air	N/A	100	Runway#2 S →N

generated a fixed frequency set of strong narrowband tones superimposed onto a broadband random component, with time-invariant narrowband energy extending from 50 Hz to at least 5 kHz in 50 Hz steps. This payload simulated a UAV propelled by a petrol-driven engine with constant tonal output regardless of flight dynamics, i.e., an idealised such signature. The Skywalker X-8 (petrol-driven) and Mavic Air (electrically powered) were flown without acoustic payloads. The flight scenarios are described throughout the paper as FS1 (Matrice 600), FS2 (Skywalker X-8), and FS3 (Mavic Air), all of which flew along Runway #2 in a northerly direction. All the UAVs listed in Table I were mounted with an iMet XQ UAV sensor, which recorded the UAV's GPS location and local meteorological factors (temperature, pressure, relative humidity) at a rate of 1 Hz. The GPS data from the iMet sensor were used as ground truth for the UAV flight trajectories.

V. RESULTS AND DISCUSSION

Once the data acquisition was completed, the signals were processed using both the narrowband and broadband techniques. Note that as the narrowband processing is only suitable for UAVs with strong harmonic signals, the technique was only applied to FS1. For the other two flight scenarios, there are not obvious fixed harmonic tones. This was because, despite FS2 having a petrol engine, the changing demands on the engine during flight resulted in an acoustic signature with high variability in the dominant frequencies. Consequently, FS2 and FS3 were not suitable for narrowband processing. The broadband technique, however, was applied to all three flight scenarios. The processing details and results, including the improvements obtained by applying the proposed bio-vision technique, are described below.

A. Narrowband technique

The data from each microphone was pre-filtered through a low-pass finite impulse response (FIR) anti-aliasing filter (AAF) with a passband cut-off of 5 kHz, passband ripple of 0.1 dB and stop band attenuation of 100 dB, prior to downsampling by a factor of 5. The spectrograms were obtained through time-frequency analysis, with a digital Fourier transform block size of 4096 samples, 75% overlap between two consecutive data blocks, Hann windowing, and 2 times zero-padding. When combined with the bio-processing model the BIV parameters were set as $f_{c1} = 1$ Hz, $I_{mid1} = 0.02$, $f_{max} = 2$ Hz, $f_{min} = 0.5$ Hz, $g_{max} = 40$, $f_{c2} = 1$ Hz, $f_{c3} = 1$ Hz, $I_{mid2} = 0.02$. These parameters were selected using an empirical process and were not optimised against any quantifiable criteria. Figure 7(a) shows the normalized (with respect to maximum power spectral density) spectrogram of FS1 (Matrice 600 with acoustic payload) obtained from the first channel of the microphone array RS1, while Fig. 7(b) shows the same spectrogram after bio-processing. With the help of adaptive filtering and nonlinear transforms, the bio-processing has enhanced the related acoustic harmonics and suppressed the unrelated noise. Two particular regions, which correspond to ranges of <300 m and around 1000 m from the array and are marked as Z1 and Z2 in Fig. 7, are expanded to more clearly show the improvement of the bio-processing approach. Z1 represents the high-frequency region when the UAV was near the acoustic array. In this case, the narrowband harmonics were low in power, but quickly varied due to the high harmonic order. Z2 denotes the low-frequency region when the UAV is far away from the array. Figures 7(c) and 7(e) show enlargements of the Z1 and Z2 regions of
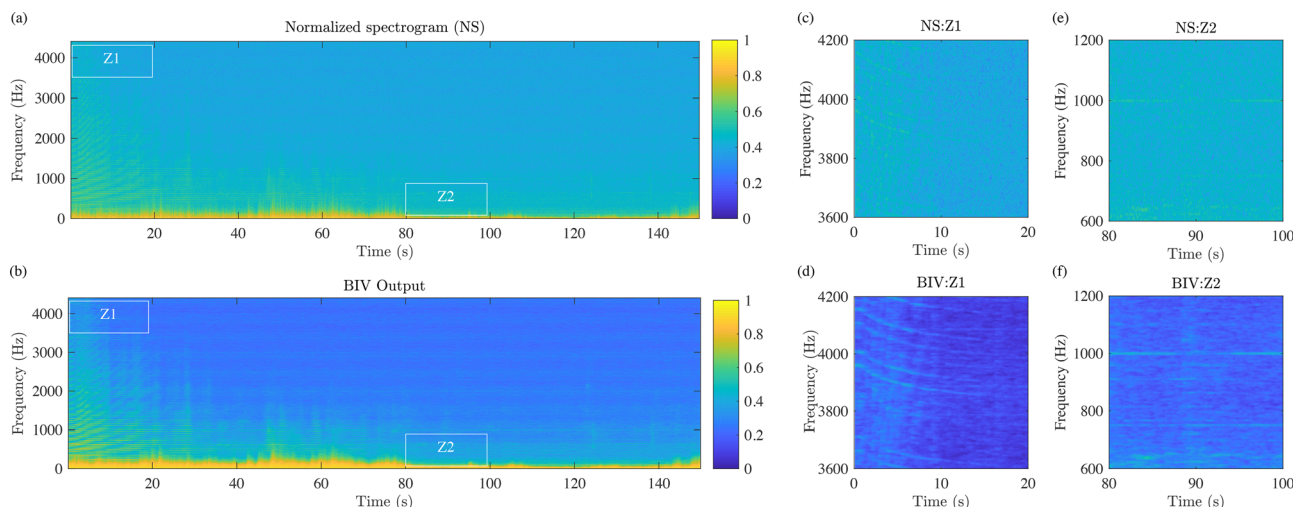

FIG. 7. (Color online) Narrowband spectrograms with (a) and without (b) biologically inspired vision (BIV) processing. Images on the right show enlarged regions for Z1 and Z2. BIV processing led to improved contrast between the signal harmonics and the background.

J. Acoust. Soc. Am. **151** (2), February 2022

Fang *et al.*   975

the normalized spectrogram, respectively. Some harmonics are barely visible as their amplitudes are insufficient to be clearly identified from the background. Figures 7(d) and 7(f) are the same two regions after BIV processing. The former illustrates a more distinct, clearer set of harmonics up to $t = 20$ s and the latter demonstrates an obvious harmonic signal around $t = 90$ s.

Both the original spectrogram and the one processed using BIV were then passing through the same cepstrum filter, which removed the spectral signal with quefrency range $|q| < 0.01$ s to eliminate the influence of wind noise and Lloyd's mirror. As described in Sec. II, the Doppler frequency (pitch) received by each microphone was obtained by searching the peaks in the HPS algorithm. For each small microphone array, the pitch data were complemented by the first four channels (Ch1–Ch4), since they were located very close to each other. Figures 8(a) and 8(b) depict the estimated pitch frequency for the acoustic signal of FS1 obtained from RS1 without and with bio-vision processing, respectively. To quantitatively evaluate the improvement of our proposed method, we adopted the peak signal-to-noise ratio (PSNR) ξ, which is defined as

$$\xi = 20 \log_{10}\left[\frac{\max(I)}{\text{RMSE}(n_I)}\right] \quad \text{(dB)}, \quad (25)$$

where $\max(I)$ is the maximum of the intensity and RMSE (n_I) is the root mean square error of the intensity noise. The data points in Fig. 8 were used for flight parameter estimation by the non-linear least squares (NLS) regression given by

$$\hat{\mathbf{k}} = \arg\max_{\mathbf{k}} \sum_{m=1}^{N_s} \sum_{k=1}^{N_t} w_m(t_k)\left[\hat{f}_m(t_k) - f_m(t_k; \mathbf{k})\right]^2, \quad (26)$$

where $\hat{\mathbf{k}} = [\hat{V}, \hat{\tau}_c, \hat{h}, \hat{d}_c, \hat{\alpha}_c]^T$ are the estimates of $\mathbf{k}$, $w_m(t_k)$ is the weighting function related to the PSNR ξ of the m-th microphone of time step t_k. The NLS fitting results are marked as the red solid lines, which stop at the low tracking confidence regions with severe data deviations and PSNRs lower than $\xi = 6$ dB. The pitch estimated by traditional methods has the highest PSNR of 20.6 dB and the maximum visible period of 70.9 s, as shown in Fig. 8. As a contrast, the PSNR with BIV processing can be as high as 30.1 dB with a maximum visible period of 96.2 s, resulting in a 10 dB improvement on PSNR, and a 35.7% improvement on the maximum visible period.

The flight parameters estimated from the NLS regression are depicted in Table II. The error values are 1σ for the traditional and bio-vision methods and derived from the iMet GPS sensor performance envelope for iMet (ground truth) data. The estimates of flight parameters appear slightly biased, mainly because of the wind noise. Compared with the iMet sensor data, however, the estimation of both traditional and bio-vision processing are acceptable.

The tracks of the UAV trajectory estimated by the acoustic array are superimposed onto the satellite photograph (Fig. 9). The red triangles are the locations of the small microphone arrays (RS1–RS7). In Fig. 9(a), the gray circles are the trajectory measured by the iMet XQ sensor GPS records, while the blue circles represent the UAV measured acoustically as it travels about 1134.5 m along the Runway #2, corresponding to a maximum slant range of 1147.5 m. As a comparison, the green circles in Fig. 9(b) show the measured UAV trajectory up to 1509.1 m, corresponding to a slant range of 1519.2 m. This indicates that for flight scenario 1, the maximum detection range was improved by approximately 33%.

B. Broadband technique

For the broadband technique, the acoustic data were visualized as correlograms. For each small microphone array, 7 channels (Ch1–Ch7) formed $N_p = 21$ sensor pairs. The correlograms were implemented through the GCC-PHAT algorithm in the frequency domain, with a FFT window size of 8192 points, 2 times zero-padding, 50% overlapping, and Kaiser windowing. Note that there was no downsampling in this stage. The time step Δt of the correlograms was 0.093 s, resulting in a frame rate f_r of 10.77 Hz. When
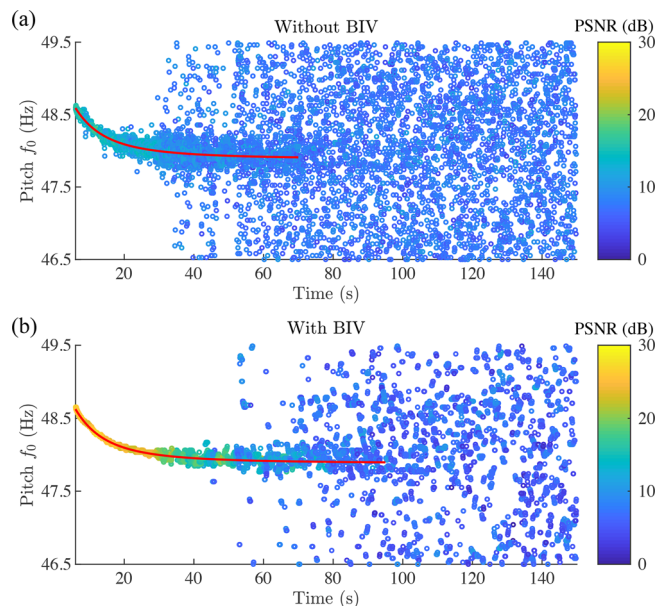

FIG. 8. (Color online) Fundamental frequency estimation for the acoustic signal (a) without BIV processing and (b) with BIV processing for the centre microphone array (RS1) of flight scenario 1. Red solid line: NLS fit. BIV processing resulted in a higher PSNR overall, and the ability to track the signal for longer.

TABLE II. Flight parameter estimation with narrowband technique.

Method	$\hat{V}$(m/s)	$\hat{\tau}_c$(s)	$\hat{d}_c$(m)	$\hat{h}$(m)	$\hat{\alpha}_c$(deg.)
iMet data	15.0	−4.2	136.3	101.2	182.0
Traditional	15.0	−4.6	142.9	97.9	182.2
Bio-vision	15.0	−4.2	144.1	99.5	182.1

976   J. Acoust. Soc. Am. **151** (2), February 2022
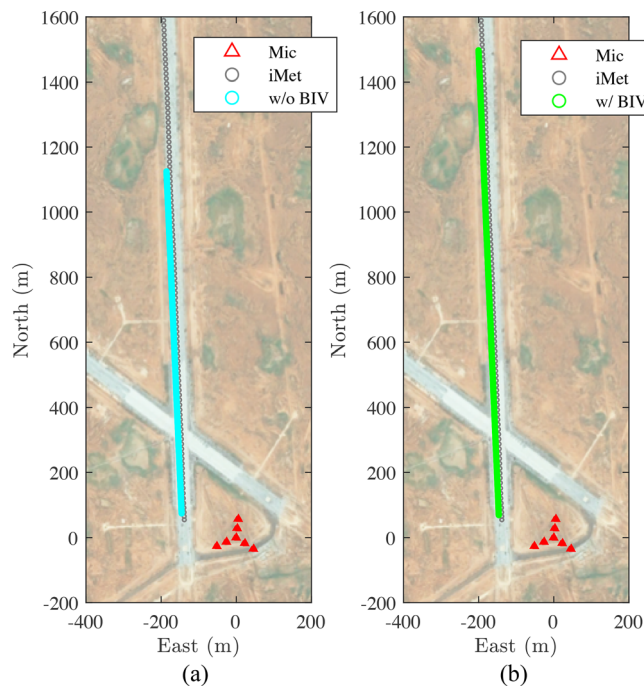
Fang *et al.*

FIG. 9. (Color online) Estimated flight trajectory by narrowband technique of FS1 (a) without BIV processing and (b) with BIV processing. While the accuracy of the flight profile was almost equally accurate in both methods, during the period that the drone was tracked, BIV processing resulted in an increase in the tracking duration and hence range.

FIG. 11. (Color online) Estimated time delays of 21 sensor pairs (a) without BIV processing, and (b) with BIV processing colour-coded according to peak signal-to-noise ratio (PSNR). Not only did the bio-vision processing improve the PSNR of the correlograms, it also made the tracking of the peak more coherent, even at low PSNRs.

processing with the bio-vision, the input images were the correlograms, and the BIV parameters were set as f_{c1} = 1.5 Hz, I_{mid1} = 0.7, f_{max} = 2.5 Hz, f_{min} = 0.1 Hz, g_{max} = 10, f_{c2} = 1.5 Hz, f_{c3} = 1.5 Hz, I_{mid2} = 0.7. Once again, these parameters were not optimised against any objective criteria. Figures 10(a) and 10(c) show the correlograms of the first microphone pair (Ch1-Ch2) and the last microphone pair (Ch6-Ch7), respectively. The former represents the minimum distance between two microphones, the latter the maximum. The two corresponding correlograms
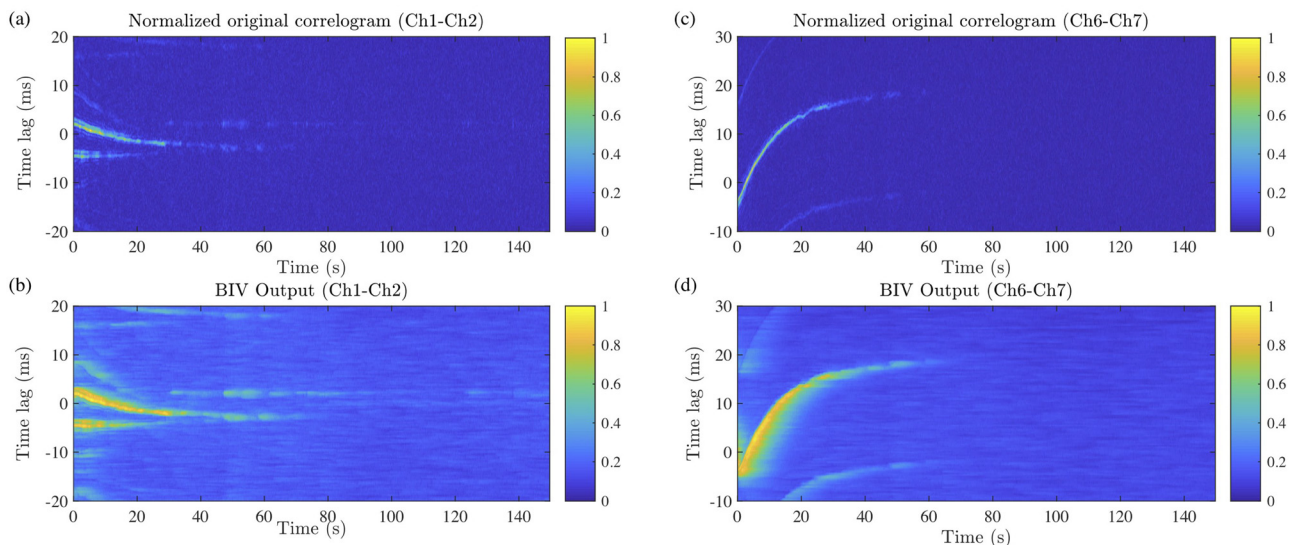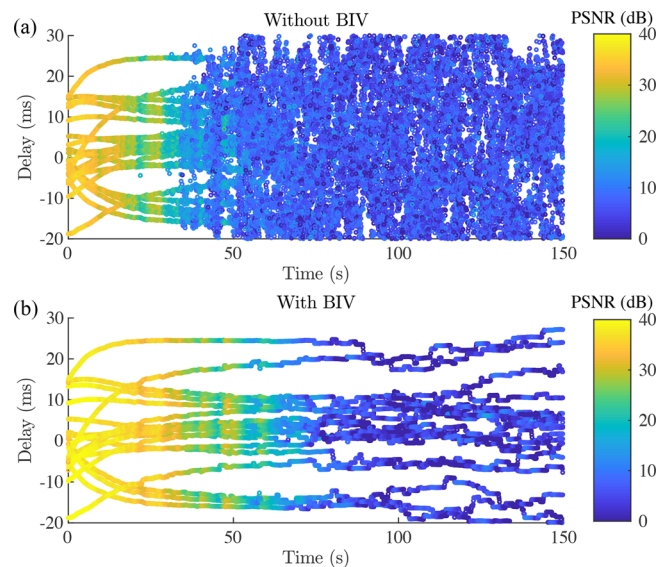
processed using bio-vision are shown in Fig. 10(b) and 10(d). It is worth noting that the correlation peak has a higher amplitude and is temporally prolonged due to the amplification and filtering present in the first stage of the bio-inspired processing chain. There is also a "shadow" after the fast varying correlation peaks [as in Fig. 10(d)]. This is mainly due to the low-pass filter in the divisive feedbacks stages of the bio-vision model. However, the influence of this effect is negligible, since the delays were estimated through the peak searching of the correlograms and if anything increased the local contrast of said peaks.

Figures 11(a) and 11(b) demonstrate the estimated time delays $\delta\tau_{mn}$ of the 21 sensor pairs without and with bio-vision processing, respectively. The PNSRs were also

FIG. 10. (Color online) Correlograms for RS1 and FS1 (a) without and (b) with BIV processing.

J. Acoust. Soc. Am. **151** (2), February 2022

Fang *et al.*   977

calculated via Eq. (25). With the traditional broadband method, the maximal PNSR is about 30 dB, while the bio-vision achieved a PSNR improvement of around 10 dB. Figure 11(a) illustrates that without bio-processing, the correlogram peaks oscillate violently after 54.7 s, while with bio-processing, they remain stable up to 85.2 s. This result indicates an improvement in the tracking of 56%, which is even higher than the improvement for the narrowband processing.

The GCF from the correlograms were also computed according to Eq. (13). Figure 12 shows the GCF of FS1 with and without bio-vision at time 6 s and 80 s, respectively. The red crosses show the estimated bearing and elevation angles, and the white lines the trajectory traces obtained from the iMet data. This demonstrates that the BIV processing resulted in higher contrast and more accurate peaks in the GCF making the estimation of bearing and elevation more accurate at longer ranges. When the UAV was near the acoustic array, i.e., $T = 6$ s, the crosses for the traditional (a) and bio-processed (b) lie on the iMet trace, indicating both provide an accurate estimate of θ and ϕ. Note that the highlighted area of Fig. 12(b) is larger than that in (a), which was mainly due to the prolonged low-pass filtering effect in

the BIV (as in Fig. 10). When the UAV was far away from the acoustic array, i.e., $T = 80$ s, traditional GCF lost track and provided an incorrect estimate, as in Fig. 12(c). However, with bio-vision processing the UAV was still visible in Fig. 12(c), and the estimate matches well with the iMet trace.

Similar broadband processing was conducted for all three flight scenarios (FS1–FS3). The flight parameters were estimated through the NLS regression given by

$$\hat{\mathbf{k}} = \arg \min_{\mathbf{k}} \sum_{m=1}^{7} \sum_{k=1}^{N_t} w_m(t_k) ||\hat{\mathbf{d}}_m(t_k) - \mathbf{d}_m(t_k; \mathbf{k})||^2, \quad (27)$$

where $\hat{\mathbf{d}}_m(t_k) = [\hat{\delta}_{12}^{(m)}(t_k), \hat{\delta}_{13}^{(m)}(t_k), \ldots, \hat{\delta}_{67}^{(m)}(t_k)]^T$ is the estimated delay vector for the m-th small microphone array at time step t_k. $\mathbf{d}_m(t_k; \mathbf{k}) = [\delta_{12}^{(m)}(t_k; \mathbf{k}), \delta_{13}^{(m)}(t_k, \mathbf{k}), \ldots, \delta_{67}^{(m)}(t_k, \mathbf{k})]^T$. $w_m(t_k)$ is the weighting function related to the corresponding PSNR.

The flight parameters obtained from the NLS regression are shown in Table III. For all flight scenarios, an accurate estimate of flight parameters could be obtained using both traditional and bio-vision methods, although in a few cases
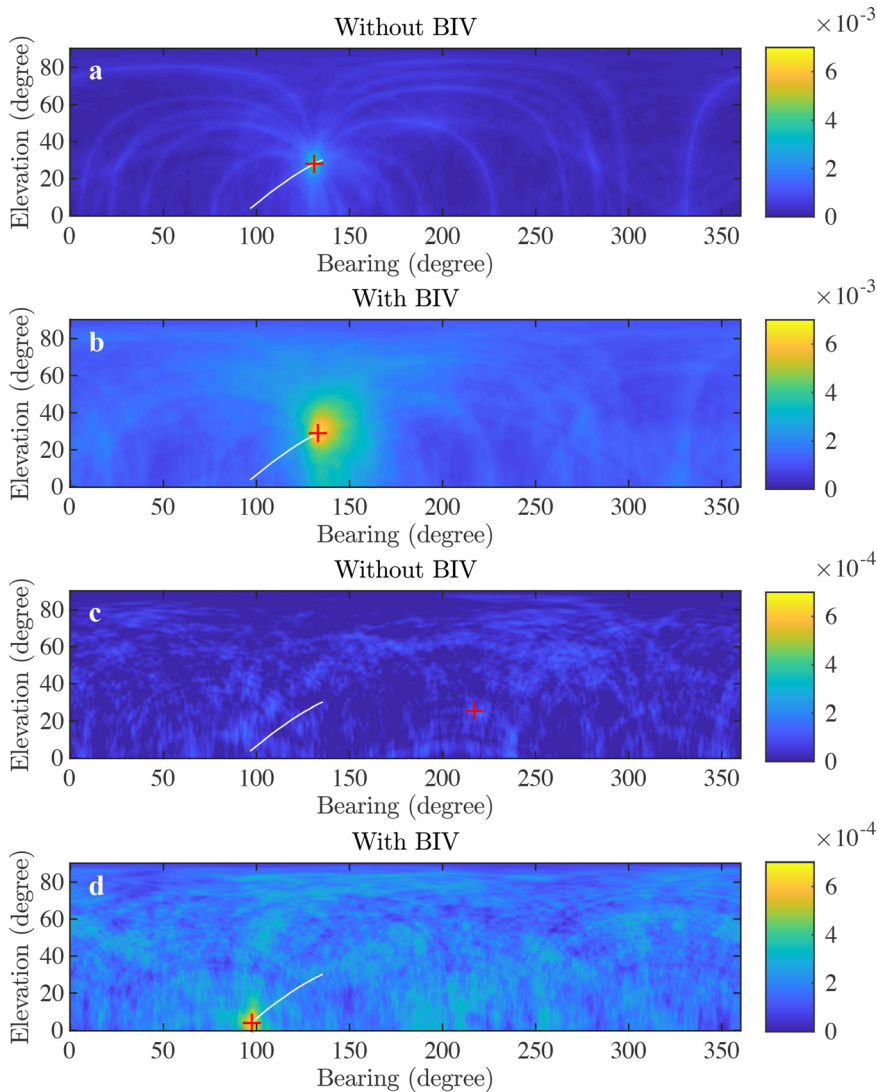

FIG. 12. (Color online) Global coherence field (GCF) of Matrice 600 with acoustic payload (FS1) (a) without BIV at $t = 6$ s, (b) with BIV at $t = 6$ s, (c) without BIV at $t = 80$ s, and (d) with BIV at $t = 80$ s. The red cross in each image represents the global maximum and the white line the historical trajectory of this maximum over time.

TABLE III. Flight parameter estimation with broadband technique.

FS#	Method	$\hat{V}$(m/s)	$\hat{\tau}_c$(s)	$\hat{d}_c$(m)	$\hat{h}$(m)	$\hat{\alpha}_c$(deg.)
3*FS1	iMet data	15.0	−4.1	136.3	101.2	182.0
	Traditional	15.1	−4.3	131.8	97.8	181.8
	Bio-vision	15.2	−4.4	134.9	99.9	181.5
3*FS2	iMet data	20.4	−7.5	147.8	210.1	179.9
	Traditional	20.9	−7.1	148.5	211.9	179.9
	Bio-vision	21.0	−7.3	155.0	214.8	180.3
3*FS3	iMet data	15.3	−5.7	130.2	98.3	181.2
	Traditional	15.0	−5.3	122.1	100.3	181.3
	Bio-vision	15.0	−5.8	123.6	102.0	181.5

TABLE IV. The acoustic detection range for different flight scenarios.

FS#	UAV type	R_{max} w/o BIV	R_{max} w/ BIV	Improvement
FS1	Matrice 600	915.7 m	1360.5 m	49%
FS2	Skywalker X-8	901.1 m	1204.7 m	34%
FS3	Mavic Air	258.7 m	336.7 m	30%

the accuracy of the bio-vision method is worse than that of the traditional approach. However, these differences in accuracy were minor and are traded against large detection range extensions. The reason for the discrepancies between the BIV and traditional results is likely due to use of low-pass filters by the former, which induces a small phase delay within the data.

As in the narrowband technique, we plot the track of UAV trajectory by broadband technique with the satellite photograph as in Fig. 13. With the traditional method, the maximum detectable slant range R_{max} of the Matrice 600 (FS1), Skywalker X-8 (FS2), and Mavic Air (FS3) were 915.7 m, 901.1 m and 258.7 m, respectively. After bio-processing, the maximum detectable range of these FSs were 1360.5 m, 1204.7 m, and 336.7 m, indicating distance improvements of 48.6%, 33.7%, and 30.2%, respectively (see Table IV). It is worth noting that the Mavic Air has a much shorter detection range because of its smaller size, lower signal intensity, and higher spectral signature compared with the other medium size UAVs.

Although all figures shown in this paper relate to flights beginning with the UAV starting close to the microphones (i.e., the SNR declines from that point on, and at towards the end of the run any angular change is small), overall the analysis is drawn from both the outward and the return legs of the flights, i.e., it includes trajectories where the SNR starts low and increases as the target approaches the observer. This was to eliminate any potential influence of "track extrapolation."

In contrast to many traditional imaging systems that operate with a single global or regional gain and attempt to capture the world as faithfully as possible, the BIV operates at multiple local time scales, uses pixel-wise integration and manipulation, and employs self-adapting non-linear feed-back between its stages. This enables it to process all parts of the data in parallel, whilst simultaneously allowing scene-independent adaptations between its components as there is no concept of spatial (and thus spectral) structure in the initial processing stages. Data elements considered dynamic are accentuated, whilst static ones are condensed. This allows the huge dynamic range of the real world to be compressed into manageable bandwidth for optimal information transmission and downstream processing across a diverse range of environments. Consequently, although extraneous coherent noise sources (e.g., interferers such as petrol generators) will be accentuated relative to any background noise, consistent noise sources (such as a constant generator) will be suppressed relative to variable sources (such as a moving UAV), in the same way moving objects are enhanced relative to stationary ones within the visual system. Consequently, unless the temporal-spectral properties of extraneous signals completely overlay those of the UAV targets, in which case they are indistinguishable, the BIV processing will make it easier to discriminate interferers from the signals of interest. A detailed examination of
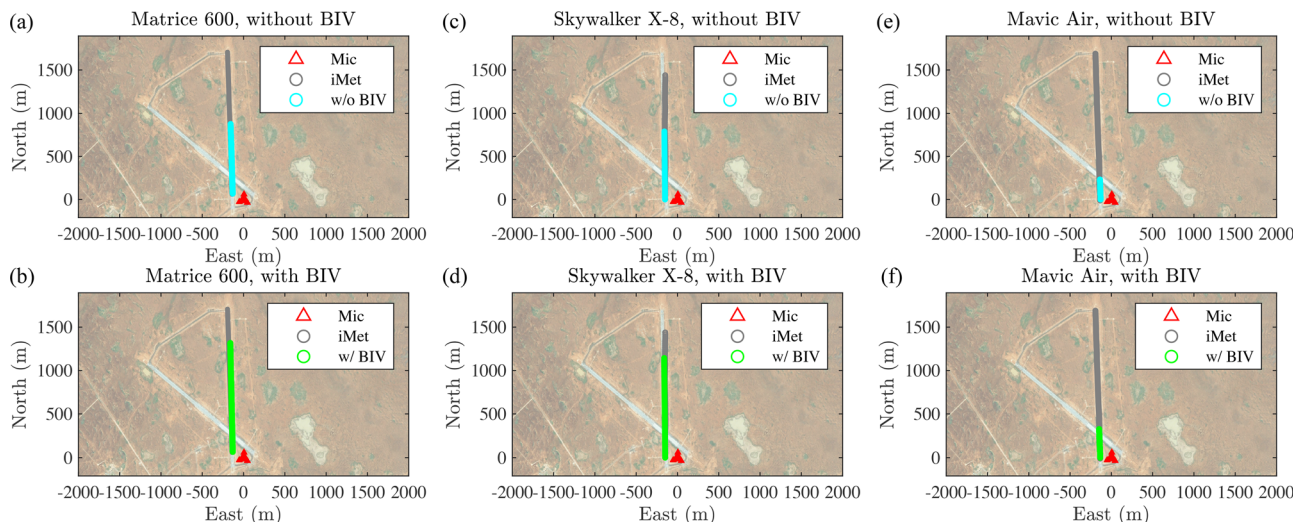

FIG. 13. (Color online) Estimated flight trajectory derived using broadband technique for FS1, FS2, and FS3, respectively.

the topics of the influence of interference and track extrapolation is beyond the scope of this paper and will be published elsewhere.

VI. CONCLUSION

This paper presents the use of a bio-inspired signal processing technique for detecting the acoustic signature of UAVs. Two standard time-frequency processing methods, based on narrowband and broadband techniques, were considered. Such approaches are commonly used by other researchers in this field, and the ranges reported (prior to the addition of any BIV signal conditioning) are similar to other publicly reported findings for such experiments.[16–18,23–25] The photoreceptor model of the insect vision system was applied in conjunction with both these traditional methods. Field trials using three different types of UAV (fixed and rotary wing), and various flight scenarios show that for narrowband processing the bio-vision technique improved the maximum detection range by a factor of 33%, while for broadband processing the bio-inspired method achieved range extension of between 30% and 49%, depending on the UAV model/type and flight scenario.

Recently BIV processing has been shown to greatly increase the detection range of UAVs in both visual[62] and infrared[37] data. However, this is the first time such a finding has been translated to acoustic detection.

Compared with the traditional methods, the bio-vision method also achieves comparable accuracy in flight parameter estimation, indicating that the proposed method is accurate and reliable. Since BIV is a pre-processing (signal conditioning) technique it augments, not replaces, existing detection and tracking methods. This means that BIV can be integrated with other more complex UAV detection algorithms. Furthermore, since it is causal and made up only of relatively simple mathematical operations, BIV is also suitable for real-time applications. Optimisation of BIV parameters against a defined goal would likely lead to a further increase in performance, as has been observed in a different context.[63] However, such improvements are beyond the scope of this paper. Future work includes verification using more UAVs and flight scenarios, fusion of the narrowband and broadband techniques, including further components in the BIV processing pathway,[64] and application of the BIV to the real and imaginary components of the analytic signals which would allow more accurate determination of the mainlobe.

ACKNOWLEDGMENTS

[1]F. M. Dommermuth, "A simple procedure for tracking fast maneuvering aircraft using spatially distributed acoustic sensors," J. Acoust. Soc. Am. **82**(4), 1418–1424 (1987).

[2]B. G. Ferguson, "Variability in the passive ranging of acoustic sources in air using a wavefront curvature technique," J. Acoust. Soc. Am. **108**(4), 1535–1544 (2000).

[3]H. Chen and J. Zhao, "On locating low altitude moving targets using a planar acoustic sensor array," Appl. Acoust. **64**(11), 1087–1101 (2003).

[4]I. Hafizovic, C. C. Nilsen, and M. Kjolerbakken, "Acoustic tracking of aircraft using a circular microphone array sensor," in *Proceedings of the IEEE International Symposium on Phased Array Systems and Technology* (ARRAY) (2010), pp. 1025–1032.

[5]K. W. Lo, "Flight parameter estimation using time delay and intersensor multipath delay measurements from a small aperture acoustic array," J. Acoust. Soc. Am. **134**(1), 17–28 (2013).

[6]M. Genesca, U. P. Svensson, and G. Taraldsen, "Estimation of aircraft angular coordinates using a directional-microphone array–An experimental study," J. Acoust. Soc. Am. **137**(4), 1914–1922 (2015).

[7]K. W. Lo and B. G. Ferguson, "Flight path estimation using frequency measurements from a wide aperture acoustic array," IEEE Trans. Aerosp. Electron. Syst. **37**(2), 685–694 (2001).

[8]B. G. Ferguson and K. W. Lo, "Turboprop and rotary-wing aircraft flight parameter estimation using both narrow-band and broadband passive acoustic signal processing methods," J. Acoust. Soc. Am. **108**(4), 1763–1771 (2000).

[9]K. W. Lo and B. G. Ferguson, "Broadband passive acoustic technique for target motion parameter estimation," IEEE Trans. Aerosp. Electron. Syst. **36**(1), 163–175 (2000).

[10]L. M. Kaplan, P. Molnar, and Q. Le, "Bearings-only target localization for an acoustical unattended ground sensor network," Proc. SPIE **4393**, 40–51 (2001).

[11]R. J. Kozick and B. M. Sadler, "Source localization with distributed sensor arrays and partial spatial coherence," IEEE Trans. Signal Process. **52**(3), 601–616 (2004).

[12]L. M. Kaplan and Q. Le, "On exploiting propagation delays for passive target localization using bearings-only measurements," J. Franklin Inst. **342**(2), 193–211 (2005).

[13]M. R. Azimi-Sadjadi, N. Roseveare, and A. Pezeshki, "Wideband DOA estimation algorithms for multiple moving sources using unattended acoustic sensors," IEEE Trans. Aerosp. Electron. Syst. **44**(4), 1585–1599 (2008).

[14]D. Lindgren, G. Hendeby, and F. Gustafsson, "Distributed localization using acoustic Doppler," Sign. Process. **107**, 43–53 (2015).

[15]J. Huang, X. Zhang, Q. Zhou, E. Song, and B. Li, "A practical fundamental frequency extraction algorithm for motion parameter estimation of moving targets," IEEE Trans. Instrum. Meas. **63**(2), 267–276 (2014).

[16]C. Dumitrescu, M. Minea, I. M. Costea, I. Cosmin Chiva, and I. Semenescu, "Development of an acoustic system for UAV detection," Sensors **20**(17), 4870 (2020).

[17]T. Blanchard, J.-H. Thomas, and K. Raoof, "Acoustic localization and tracking of a multi-rotor unmanned aerial vehicle using an array with few microphones," J. Acoust. Soc. Am. **148**(3), 1456–1467 (2020).

[18]A. Sedunov, H. Salloum, A. Sutin, N. Sedunov, and S. Tsyuryupa, "UAV passive acoustic detection," *2018 IEEE International Symposium on Technologies for Homeland Security (HST)* (2018), pp. 1–6.

[19]T. Pham and N. Srour, "TTCP AG-6: Acoustic detection and tracking of UAVs," *Proceedings of SPIE 5417, Unattended/Unmanned Ground, Ocean, and Air Sensor Technologies and Applications VI* (1 September 2004).

[20]K. W. Lo and B. G. Ferguson, "Tactical unmanned aerial vehicle localization using ground-based acoustic sensors," in *Proceedings of Intelligent Sensors, Sensor Networks and Information Processing Conference, ISSNIP* (2004), pp. 475–480.

[21]A. Finn and S. Franklin, "Acoustic sense and avoid for UAV's," in *Proceedings of Intelligent Sensors, Sensor Networks and Information Processing Conference, ISSNIP*, Adelaide, Australia (6–9 December, 2011).

[22]V. E. Ostashev, D. K. Wilson, A. Finn, and E. Barlas, "Theory for spectral broadening of narrowband signals in the atmosphere and experiment with an acoustic source onboard an unmanned aerial vehicle," J. Acoust. Soc. Am. **145**(6), 3703–3714 (2019).

[23]K. J. Rogers and A. Finn, "Accurate group velocity estimation for unmanned aerial vehicle-based acoustic atmospheric tomography," J. Acoust. Soc. Am. **141**(2), 1269–1281 (2017).

980   J. Acoust. Soc. Am. **151** (2), February 2022

Fang *et al.*

[24]B. Yang, E. T. Matson, A. H. Smith, J. E. Dietz, and J. C. Gallagher, "UAV detection system with multiple acoustic nodes using machine learning models," in *2019 Third IEEE International Conference on Robotic Computing (IRC)* (2019), pp. 493–498.

[25]P. Casabianca and Y. Zhang, "Acoustic-based UAV detection using late fusion of deep neural networks," Drones **5**(3), 54 (2021).

[26]C. Evers and P. A. Naylor, "Acoustic SLAM," IEEE/ACM Trans. Audio Speech Lang. Process. **26**(9), 1484–1498 (2018).

[27]H. Shakhatreh, A. H. Sawalmeh, A. Al-Fuqaha, Z. Dou, E. Almaita, I. Khalil, N. S. Othman, A. Khreishah, and M. Guizani, "Unmanned aerial vehicles (UAVs): A survey on civil applications and key research challenges," IEEE Access **7**, 48572–48634 (2019).

[28]R. R. Murphy, E. Steimle, M. Hall, M. Lindemuth, D. Trejo, S. Hurlebaus, Z. Medina-Cetina, and D. Slocum, "Robot-assisted bridge inspection," J. Intell. Robot Syst. **64**(1), 77–95 (2011).

[29]S. Minaeian, J. Liu, and Y. Son, "Vision-based target detection and localization via a team of cooperative UAV and UGVs," IEEE Trans. Syst. Man. Cybern.: Syst. **46**(7), 1005–1016 (2016).

[30]F. Nex and F. Remondino, "UAV for 3D mapping applications: A review," Appl. Geomat. **6**(1), 1–15 (2014).

[31]K. Rogers and A. Finn, "Three-dimensional UAV-based atmospheric tomography," J. Atmos. Ocean. Technol. **30**(2), 336–344 (2013).

[32]K. W. Lo, "Flight parameter estimation using instantaneous frequency and time delay measurements from a three-element planar acoustic array," J. Acoust. Soc. Am. **139**(5), 2386–2398 (2016).

[33]B. G. Ferguson, "A ground-based narrow-band passive acoustic technique for estimating the altitude and speed of a propeller-driven aircraft," J. Acoust. Soc. Am. **92**(3), 1403–1407 (1992).

[34]S. D. Wiederman, R. S. A. Brinkworth, and D. C. O'Carroll, "Performance of a bio-inspired model for the robust detection of moving targets in high dynamic range natural scenes," J. Comput. Theor. Nanosci. **7**, 911–920 (2010).

[35]R. S. A. Brinkworth, E.-L. Mah, and D. C. O'Carroll, "Bioinspired pixel-wise adaptive imaging," Proc. SPIE **6414**, 641416 (2007).

[36]P. S. M. Skelton, A. Finn, and R. S. A. Brinkworth, "Consistent estimation of rotational optical flow in real environments using a biologically-inspired vision algorithm on embedded hardware," Image Vision Comput. **92**, 103814 (2019).

[37]M. Uzair, R. S. A. Brinkworth, and A. Finn, "Detecting small size and minimal thermal signature targets in infrared imagery using biologically inspired vision," Sensors **21**(5), 1812 (2021).

[38]R. N. Miles and R. R. Hoy, "The development of a biologically-inspired directional microphone for hearing aids," Audiol. Neurotol. **11**(2), 86–94 (2006).

[39]A. Reid, J. Windmill, and D. Uttamchandani, "Bio-inspired sound localization sensor with high directional sensitivity," Procedia Eng. **120**, 289–293 (2015).

[40]Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proc. IEEE **86**(11), 2278–2324 (1998).

[41]R. Yamashita, M. Nishio, Do, R. K. Gian, and K. Togashi, "Convolutional neural networks: An overview and application in radiology," Insights Imag. **9**(4), 611–629 (2018).

[42]K. W. Lo, S. W. Perry, and B. G. Ferguson, "Aircraft flight parameter estimation using acoustical Lloyd's mirror effect," IEEE Trans. Aerosp. Electron. Syst. **38**(1), 137–151 (2002).

[43]K. W. Lo, B. G. Ferguson, Y. Gao, and A. Maguer, "Aircraft flight parameter estimation using acoustic multipath delays," IEEE Trans. Aerosp. Electron. Syst. **39**(1), 259–268 (2003).

[44]A. M. Noll, "Cepstrum pitch determination," J. Acoust. Soc. Am. **41**(2), 293–309 (1967).

[45]C. Roads, *The Computer Music Tutorial* (MIT Press, Cambridge, 1996).

[46]D. Talkin, "A robust algorithm for pitch tracking (RAPT)," in *Speech Coding and Synthesis*, edited by W. B. Kleijn and K. K. Paliwal (Elsevier, Amsterdam, 1995). pp. 495–518.

[47]A. de Cheveigné and H. Kawahara, "YIN, a fundamental frequency estimator for speech and music," J. Acoust. Soc. Am. **111**(4), 1917–1930 (2002).

[48]M. Piszczalski and B. A. Galler, "Predicting musical pitch from components frequency ratios," J. Acoust. Soc. Am. **66**(3), 710–720 (1979).

[49]J. A. Moorer, "On the transcription of musical sound by computer," Comput. Music J. **1**(4), 32–38 (1977).

[50]M. R. Schroeder, "Period histogram and product spectrum: New methods for fundamental-frequency measurement," J. Acoust. Soc. Am. **43**(4), 829–834 (1968).

[51]C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," IEEE Trans. Acoust. Speech Sign. Process. **24**(4), 320–327 (1976).

[52]H.-G. Kang, M. Graczyk, and J. Skoglund, "On pre-filtering strategies for the GCC-PHAT algorithm," in *Proceedings of the International Workshop on Acoustic Signal Enhancement* (2016), pp. 1–5.

[53]M. Omologo, P. Svaizer, and R. De Mori, "Acoustic transduction," in *Spoken Dialogue with Computer* (Academic Press, New York, 1998). pp. 1–46.

[54]R. S. A. Brinkworth and D. C. O'Carroll, "Robust models for optic flow coding in natural scenes inspired by insect biology," PloS Comput. Biol. **5**(11), e1000555 (2009).

[55]K. Haltis, M. Sorell, and R. Brinkworth, "A biologically inspired smart camera for use in surveillance applications," in *Crime Prevention Technologies and Applications for Advancing Criminal Investigation*, edited by C.-T. Li and A. T. S. Ho (IGI Global, Hershey, PA, 2012), Chap. 13.

[56]D. Griffiths, "Biologically inspired high dynamic range imaging for use in machine vision," Ph.D. thesis, School of Engineering, University of South Australia, Australia (2017).

[57]R. S. A. Brinkworth, E. L. Mah, J. P. Gray, and D. C. O'Carroll, "Photoreceptor processing improves salience facilitating small target detection in cluttered scenes," J. Vision **8**(11), 8 (2008).

[58]Z. M. Bagheri, S. D. Wiederman, B. S. Cazzolato, S. Grainger, and D. C. O'Carroll, "Performance of an insect-inspired target tracker in natural conditions," Bioinspir. Biomim. **12**(2), 025006 (2017).

[59]J. H. van Hateren and H. P. Snippe, "Information theoretical evaluation of parametric models of gain control in blowfly photoreceptor cells," Vision Res. **41**(14), 1851–1865 (2001).

[60]J. H. van Hateren and H. P. Snippe, "Phototransduction in primate cones and blowfly photoreceptors: Different mechanisms, different algorithms, similar response," J. Comp. Physiol. A **192**(2), 187–197 (2006).

[61]K. I. Naka and W. A. H. Rushton, "S-potentials from colour units in the retina of fish (cyprinidae)," J. Physiol. **185**(3), 536–555 (1966).

[62]M. Uzair, R. S. Brinkworth, and A. Finn, "Bio-inspired video enhancement for small moving target detection," IEEE Trans. Image Process. **30**, 1232–1244 (2021).

[63]P. S. M. Skelton, A. Finn, and R. S. A. Brinkworth, "Improving an optical flow estimator inspired by insect biology using adaptive genetic algorithms," in *2020 IEEE Congress on Evolutionary Computation (CEC)* (2020), pp. 1–10.

[64]A. Melville-Smith, A. Finn, and R. S. A. Brinkworth, "Enhanced micro target detection through local motion feedback in biologically inspired algorithms," *2019 Digital Image Computing: Techniques and Applications (DICTA)* (2019), pp. 1–8.