Separation of Bird Calls and DOA estimation using a 4-Microphone Array

Paul M. Baggenstoss
Fraunhofer FKIE, Wachtberg, Germany
Email: paul.baggenstoss@fkie.fraunhofer.de

Karl-Heinz Frommolt
Museum für Naturkunde, Berlin
Email: Karl-Heinz.Frommolt@mfn.berlin

Olaf Jahn
Museum für Naturkunde, Berlin
Email: Olaf.Jahn@mfn.berlin

Frank Kurth
Fraunhofer FKIE, Wachtberg, Germany
Email: frank.kurth@fkie.fraunhofer.de

***Abstract*—This paper describes an approach to separate time-overlapping bird vocalizations in 4-channel recordings and overcome the limitations of classical beamforming. As a first step, the short-time-Fourier transform (STFT) of the input time series is computed. The beam response is calculated for each time-frequency (TF) bin by conventional beamforming with amplitude-based sidelobe suppression and then converted into a red-green-blue (RGB) color vector, producing the directional spectrogram (D-SPEC) in which color represents direction of arrival (DOA). The scalar color value (hue) is then clustered into bird individuals using a probabilistic approach (Gaussian mixture model). In a final step, spatial processing is used to promote grouping of nearby TF bins into the same cluster. Results are tested using annotated field recordings in a challenging scenario.**

I. Background and Motivation

Automated survey methods are popular to study a wide spectrum of research questions in ecology, ethology, and conservation science, and are rapidly gaining importance for documenting and understanding the impacts of ongoing environmental change on biodiversity [1], [2]. Among the most widely used methods is passive acoustic monitoring, which allows for low-impact continuous observation of sound-producing animal species such as insects, anurans, birds, and mammals [3] and facilitates the automated identification of their sound signals, even in complex acoustic environments [4]. A specialized application of passive acoustic monitoring is the localization of terrestrial wildlife, which has been used for different purposes in studies on animal behavior and ecology [5]. Here we describe an approach to separate simultaneous bird vocalizations into directional spectrograms, ideally containing only the signals of single individuals, and to estimate their direction of arrival (DOA). The sound source separation method is based on short (10-20 sec.) 4-channel audio recordings that are gathered by small cable-synchronized microphone arrays connected to an autonomous recording unit (ARU). The birds are assumed stationary during the recording.

Our aim is to show that the method produces robust DOA estimates even with compact microphone arrays and that it has the potential to support species' abundance estimation in realistic scenarios. We briefly explain the experimental setup and provide the algorithms needed to derive DOA estimates and directional spectrograms (D-SPEC). Furthermore, we validate our DOA bearings using time-difference-of-arrival (TDOA) calculations. Finally, we discuss some advantages of the approach and present ideas for further improvements.

This work was sponsored by the German Federal Ministry of Education and Research (Grants 16LC1903F and 16LC1903J).

II. Experimental Approach and Array Specification

The field study was conducted in a floodplain area within the Lower Oder Valley National Park (Brandenburg, Germany). We used four autonomous recording units (ARUs), each consisting of a TASCAM DR-680 portable multi-track recorder and four Beyerdynamic condenser microphones. At a height of 1.5 m, the microphones were mounted under a 40×40 cm PVC plate and arranged in a 27×27 cm square (cf. Fig. 1. in [6]). Each microphone had a cardioid response pattern with the zero (null) directed at the array center (Fig. 1). The microphones were directed at 45, 135, 225, and 315 degrees using a precision compass. The ARUs were placed with distances between units of 52.2 to 118.3 m. We ensured that the altitude above the ground and geographic orientation were exactly the same for all ARUs. The recordings were made in the night from April 24-25, 2015, at 48 kHz sample rate and 16 bit data depth, but down-sampled to 24 kHz for processing.

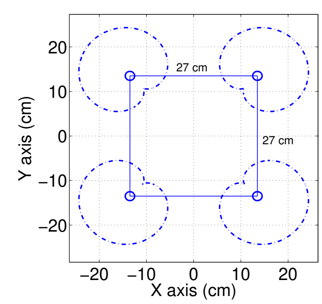

Fig. 1. The microphone array used in each ARU.

III. Directional Spectrogram (D-SPEC)

A. Beamforming

The resolution afforded by classical beamforming, or even adaptive beamforming using such a small array would be unable to separate most overlapping bird vocalizations. However, to efficiently communicate in noisy environments territorial birds tend to avoid interference using spatial, temporal, or acoustic partitioning of vocalizations [7], [8]. This means that simultaneous overlap in time and frequency occurs only rarely. To take advantage of this, data is first processed by short-time Fourier transform (STFT) and then beamforming is done independently on each time-frequency (TF) bin, in which it is assumed that just one source is present. For a signal originating at a distant point at a given direction, frequency-domain beamforming [9] calculates the inner product between the array signal at one TF bin (which in this case is a 4-dimensional complex vector) with the idealized output including cardioid microphone pattern. We used a directional resolution of 3 degrees, resulting in 120 steering directions. For a length N FFT, there are $n = N/2 + 1$ frequency bins, producing a complex beamformer output of size $n \times T \times 120$, where T is the number of time steps.

To get an idea of the difficulty of beamforming with only 4 microphones, refer to Figure 2. An intensity plot of the STFT amplitude is shown for a segment exhibiting a loud chirp which was known to originate from an angle of 330 degrees (Inset A). Seven test points along the trajectory of the chirp were manually selected, approximately evenly spaced in frequency, and the beam response was calculated at each test point. In Inset B, the seven beam responses are calculated without consideration of the microphone cardioid response. The responses show many sidelobes at various angles, with no correspondence except at the true location, where they all reach a peak. In Inset C, the microphone response is included and the correct location of 330 degrees becomes more apparent, demonstrating the importance of including the cardioid pattern in the steering vector. Finally, in Inset D, the beam responses are processed by the sidelobe suppression explained below.

B. Sidelobe Suppression

Let $b_{k,i}$, $1 \leq k \leq n$, $1 \leq i \leq m$ be the beamformer amplitude at frequency bin k and direction i, where there are $m = 120$ directions. Let $\tilde{b}_{k,i} = \frac{b^p_{k,i}}{\sum_{j=1}^{m} b^p_{k,j}}$ and $a_k = \left(\sum_{i=1}^{m} \tilde{b}^p_{k,i}\right)^{1/p}$. Quantity $\tilde{b}_{k,i}$ is the sidelobe-suppressed beam pattern, whereas a_k is the amplitude estimate for frequency k (computed using p-norm). A power of $p = 16$ was used. This type of sidelobe suppression relies on the assumption that just one source is present in each TF bin. All of these quantities are calculated for each of the T time steps, but the time index is not shown for simplicity of notation.

C. Color Determination

To convert the length-m sidelobe-suppressed beam amplitude response at each time-frequency bin into a color, we multiply it by a $3\times m$ RGB (red-green-blue) color map matrix, in which the colors vary cyclically in the 360-degree range[1]. The resulting D-SPEC is an array of size $T \times n \times 3$ (Fig. 3).

[1]This type of color map is commonly known as HSV because the hue varies linearly and cyclically.

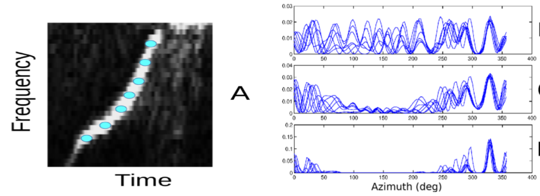

Fig. 2. (A) Intensity plot of the STFT of a loud chirp originating from an angle of 330 deg and beam-patterns computed at seven test points, (B) without and (C) with cardioid pattern, and (D) with sidelobe suppression.

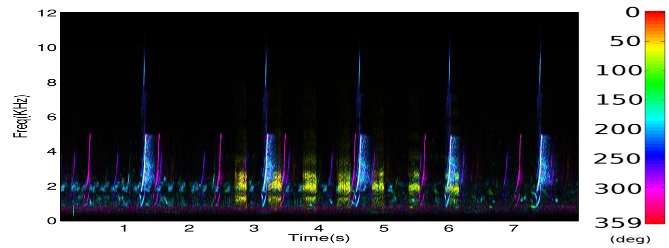

Fig. 3. Left: D-SPEC example, Right: Color key (degrees azimuth). Compare with original 4-channel spectrograms in Figure 7.

IV. Clustering

To the human eye, vocalizations from at least 4 birds are visible in Figure 3. Automatic separation requires clustering.

Before clustering, information is gathered from the D-SPEC including $C(t, k, l)$ which is the $T \times n \times 3$ D-SPEC itself, $\{a_{t,k}\}$ which are the spectrogram amplitudes (time index t is now included), and $\{z_{t,k}\}$ which is the matrix of hue values. The hue is obtained by converting the 3-dimensional RGB value at a TF bin to HSV. The hue (first element) is a value between 0 and 1 that represents the range 0 to 360 degrees. Both $\{a_{t,k}\}$ and $\{z_{t,k}\}$ are $T\times n$ matrices. This data is reduced significantly by keeping only bins where $\{a_{t,k}\}$ is above an empirically-defined threshold. The TF bin locations of the threshold crossings are denoted by t_i and k_i, where i ranges from 1 to N_s, which is the number of threshold crossings. The hue values are $\{z_i\} \triangleq \{z_{t_i,k_i}\}$, and the amplitudes are $a_i \triangleq a_{t_i,k_i}$. A histogram of hue values is shown in Figure 4 for a typical field recording of 10 seconds.

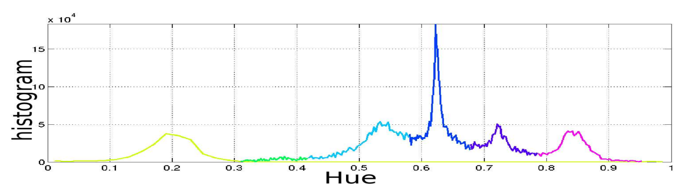

Fig. 4. Example histogram of hue values. Lines are drawn in a color corresponding to the average hue in nearest cluster.

A. Initialization

An initial clustering of the hue values $\{z_i\}$ is accomplished with K-means [10] using a high initial number of clusters (for example $M = 9$). The cluster count M is reduced in a later step where similar clusters are merged.

Let $\mathbf{W} = \{W_{i,j}\}$ be the $N_s \times M$ cluster probability matrix where $W_{i,j} = P(j|i)$ is the probability that sample i belongs to cluster j. Initially, these values are equal to one or zero according to the cluster membership determined by the initial K-means clustering. Recall that index i ranges over the N_s STFT bins that have exceeded a threshold and can point to any time or frequency location.

To improve upon the K-means clustering, a statistical approach is used to represent the distribution of hues in each cluster by a C-component Gaussian mixture model (GMM). The GMM probability density for the hue value z of cluster j is equal to $g_j(z) = \sum_{c=1}^{C} \alpha_{j,c}\ (2\pi\sigma_{j,c}^2)^{-1/2} \exp\left\{-\frac{[z-\mu_{j,c}]_1^2}{2\sigma_{j,c}^2}\right\}$, where $\alpha_{j,c},\ \mu_{j,c},\ \sigma_{j,c}^2$ are the weight, mean, and variance parameters for cluster j, component c, and where $[\ \ ]_1$ is the modulo-1 operator that results in a signed hue value in the range $[-.5,\ .5]$. Signed modulo-1 arithmetic is needed for comparing hue values. For example, $[0.95 - 0.05]_1 = -0.1$. We used $C = 2$ components. Parameter $\sigma_{j,c}^2$ takes an initial value of 0.1, while $\mu_{j,c}$ can be set to randomly-determined values, for example by choosing C randomly-selected hue values from cluster j. Weights $\alpha_{j,c}$ are initialized to $1/C$. The complete collection of clustering parameters are given by $\Theta = \left[\{W_{i,j}\}, \{\alpha_{j,c},\},\ \{\mu_{j,c},\},\ \{\sigma_{j,c}^2\}\right]$, with $i \in [1, N_s]$, $j \in [1, M]$, $c \in [1, C]$.

B. GMM Estimation

To complete the statistical clustering, it is necessary to estimate the GMM using an iterative expectation-maximization (E-M) algorithm [11]. The GMM estimation iteration begins by calculating the hue log-likelihood matrix $L_{i,j,c}$, which is the $N_s \times M \times C$ matrix

$$L_{i,j,c} = -\frac{1}{2}\log(2\pi\sigma_{j,c}^2) - \frac{[z_i - \mu_{j,c}]_1^2}{2\sigma_{j,c}^2}. \qquad (1)$$

From $\{L_{i,j,c}\}$, the component membership probabilities are estimated for each sample i as $\hat{P}(c|i,j) = \frac{\alpha_{j,c}\exp(L_{i,j,c})}{g_j(z_i)}$, where $g_j(z_i) = \sum_{c=1}^{C} \alpha_{j,c}\exp(L_{i,j,c})$. In the next step, the cluster membership probabilities are estimated as

$$\hat{P}(j|i) = \frac{W_{i,j}\ g_j(z_i)}{\sum_{j'=1}^{M} W_{i,j'}\ g_{j'}(z_i)}. \qquad (2)$$

This is the E-step of the E-M algorithm. Next, in the M-step, the GMM component probabilities are re-estimated:

$$\alpha_{j,c} = P(c|j) = \frac{\sum_{i=1}^{N_s} a_i^q\ \hat{P}(j|i)\ \hat{P}(c|i,j)}{\sum_{c'=1}^{C}\sum_{i=1}^{N_s} a_i^q\ \hat{P}(j|i)\ \hat{P}(c'|i,j)}, \qquad (3)$$

where q is an exponent parameter that determines how much amplitude affects the estimation. We used a value of $q = 0.5$. Next, the cluster weights are estimated

$$P(j) = \frac{\sum_{i=1}^{N_s} a_i^q\ \hat{P}(j|i)}{\sum_{j'=1}^{M}\sum_{i=1}^{N_s} a_i^q\ \hat{P}(j'|i)}, \qquad (4)$$

and, the GMM parameters are updated using

$$\delta\mu_{j,c} = \frac{\sum_{i=1}^{N_s} a_i^q\ \hat{P}(j|i)\ \hat{P}(c|i,j)[z_i - \mu_{j,c}]_1}{\sum_{i=1}^{N_s} a_i^q\ \hat{P}(j|i)\ \hat{P}(c|i,j)}, \qquad (5)$$

which is the estimated change in $\mu_{j,c}$. The updated values of $\mu_{j,c}$ are $\mu_{j,c} := \mu_{j,c} + \delta\mu_{j,c},\ 1 \le j \le M,\ 1 \le c \le C$. Finally, the variances are updated:

$$\sigma_{j,c}^2 = \frac{\sum_{i=1}^{N_s} a_i^q\ \hat{P}(j|i)\ \hat{P}(c|i,j)[z_i - \mu_{j,c}]_1^2}{\sum_{i=1}^{N_s} a_i^q\ \hat{P}(j|i)\ \hat{P}(c|i,j)} + \sigma_0^2, \qquad (6)$$

where σ_0^2 is added to prevent variance tending to zero. We used $\sigma_0^2 := 0.00015$. The sample-wise cluster-probabilities are set equal to the estimates:

$$W_{i,j} = \hat{P}(j|i),\ 1 \le i \le N_s,\ 1 \le j \le M. \qquad (7)$$

The algorithm then loops back to compute hue log-likelihoods (eq. 1). Typically, about 100 GMM updates are executed. Every 10 iterations or so, the algorithm seeks to reduce the number of clusters by cluster-merging (Section IV-C).

C. Cluster Merging

In cluster merging, a measure of similarity between clusters is computed: $S_{j,k} = \sum_{c=1}^{C} \alpha_{j,c} g_k(\mu_{j,c})$, where $S_{j,k}$ estimates the probability that data from cluster j could be a member of cluster k by treating the GMM component means $\mu_{j,c}$ as potential data samples from cluster j occurring with probability $P(c|j)$. Before applying to a threshold, we normalize it and take the log: $\tilde{S}_{j,k} = \log\left(\frac{S_{j,k}}{\max_{j'} S_{j',k}}\right)$. If $\tilde{S}_{j,k}$ is above a threshold (we used -0.15), then the weakest of the two clusters, i.e. with lower weight $P(j)$, is eliminated so that its members will likely belong to the stronger cluster on the next iteration. All cluster pairs are tested.

D. Spatial Processing

So far, cluster membership probability is based only on the hue of the sample (2). Now, we include information about the likely cluster membership of neighboring TF bins. Let $\mathbb{N}_i$ be the set of neighbors of sample i (not including sample i itself). A sample i' is a neighbor if the squared distance $d^2(i,i') = (t_i - t_{i'})^2 + (k_i - k_{i'})^2$ is less than the square of the neighborhood radius. A radius of about 6 to 10 is used, which is measured in pixels (i.e. TF bins), with time and frequency dimensions treated the same. A neighborhood weighting function is defined as $\eta(i,i') = \exp(-d^2(i,i'))$. In spatial iterations, the hue-based sample-wise cluster probabilities (2) are replaced by spatial sample-wise cluster probabilities

$$P_s(j|i) = \left(\frac{\sum_{i' \in \mathbb{N}_i} \eta(i,i') P^{\gamma}(j|i')}{\sum_{j'=1}^{M}\sum_{i' \in \mathbb{N}_i} \eta(i,i') P^{\gamma}(j'|i')}\right) \cdot P(j|i), \qquad (8)$$

where parameter γ determines the amount of spatial influence. We used $\gamma = 30$. In a spatial-processing iteration, $W_{i,j}$ is updated to equal $P_s(j|i)$, after it has been normalized so that it sums to 1 over j. This repeats about 5 times. The GMM hue model is not changed.

Figure 5 provides an example of spatial processing starting with (left) a cropped section of the original D-SPEC (Fig. 3) and continuing with (center) an artificial state map created from matrix $\mathbf{W}$ in which the cluster identity of each TF bin ($\arg\max$ over cluster) is shown in a different color. In the area between 3.2 and 3.4 seconds (left), the reverberation of the call from one Spotted Crake (blue) overlaps with the call of a Mallard (yellow/green). Pixels in this area are assigned to three or four different clusters. On the right, the artificial state map of the same area is shown after spatial processing, resulting in more consistent spatial distribution of cluster identities.

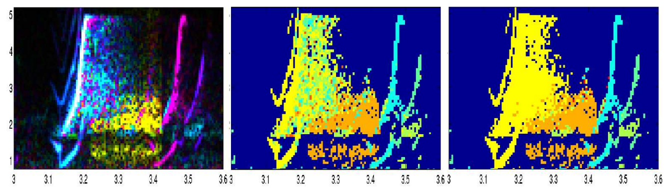

Fig. 5. Example of spatial processing. Left: original D-SPEC. Center: cluster-identity map prior to spatial processing. Right: after spatial processing.

V. Example of Source Separation

Cluster-specific spectrograms are created by multiplying D-SPEC RGB values by the corresponding column of the cluster probability matrix $\mathbf{W}$ after clustering is finished. This results in spectrograms that are ideally associated with one individual (Fig. 6). Information about the clusters in Figure 6 is tabulated in Table I. The angle shown for each cluster is the mean cluster azimuth, calculated by taking the weighted mean of the hue values in the cluster, then converting to angle.

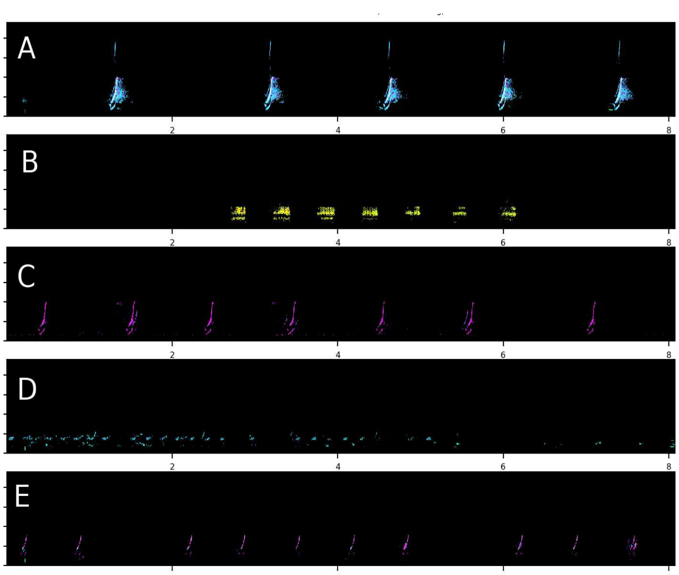

Fig. 6. Example of source separation for the 8 second recording at one ARU, also seen in Figures 3 and 7. Additional details are given in Table I.

Cluster	Species	Angle	Weight
A	Spotted Crake (*Porzana porzana*)	221.3	56.0%
B	Mallard (*Anas platyrhynchos*)	79.5	18.0%
C	Spotted Crake (*Porzana porzana*)	292.1	12.0%
D	European Tree Frog (*Hyla arborea*)	188.0	7.0%
E	Spotted Crake (*Porzana porzana*)	256.4	4.0%

TABLE I
Annotation and cluster details for Figure 6.

A. Creation of Cluster Time Series

It is a well-known approach to re-create time series from the STFT by "overlap-add". Since the cluster probability estimates $\{W_{i,j}\}$ correspond to a STFT time-frequency bin that has exceeded the initial threshold, it is possible to multiply the STFT by the entries in matrix $\mathbf{W}$, a kind of individual-specific mask, prior to re-creation of the time series. This will produce cluster-specific time series. Naturally, all STFT bins that do not exceed the threshold will not be used in the re-synthesis.

VI. Validation

To validate the source separation, we compared the D-SPEC clusters (Fig. 6) with manually annotated spectrograms of the original 4-channel recordings (Fig. 7). The correspondence between the individuals in the figures, in the form "annotation from Fig. 7" = "cluster from Fig. 6" is as follows: `P1=A`, `mall=B`, `P3=C`, `P2=E`. Parts of individual `P4` may be seen in cluster `D` and the European Tree Frogs, which dominate cluster D were not annotated. Apart from that, the other clusters were correctly assigned to individuals. To validate the angular measurements based on D-SPEC clusters, we separately estimated the direction of the dominant individual (Fig. 6, cluster A) using three different ARUs and the built-in TDOA-function of Avisoft SASLab Pro software, and compared them with the D-SPEC angles. Only the two microphone channels with highest amplitudes were used for DOA estimation, as it was clear that the signals arrived from that side of the array. We calculated the DOA from TDOA estimates using basic geometry and assuming a plane wave front. Measurements were made separately in six consecutive 10-second intervals and are summarized in Table II, which shows the mean and standard deviation of the angles reported by the two methods (TDOA and D-SPEC) in the six intervals. Note that the TDOA-based measurement is very stable, showing no variations over the six intervals. Angles reported by D-SPEC are based on weighted average of all the TF pixels assigned to a cluster and can include reverberation and sometimes other individuals at the same direction. Despite this, there is very good agreement.

ARU	TDOA		D-SPEC		Diff
	Mean (deg)	Stdev	Mean (deg)	Stdev	
I	224.9	0	226.8	2.4	1.9
II	117.1	0	119.5	5.6	2.4
III	149.6	0	153.8	4.6	4.2

TABLE II
Statistical summary of TDOA-based and D-SPEC angular measurements for the dominant individual at three ARUs.

VII. Discussion

In bioacoustic research, microphone arrays have been widely used for tasks related to localization and tracking of animals [5]. Our approach has shown that even an array consisting of only four microphones can be used to reliably separate the voices of individual birds. Although more research is needed for the practical application of this method, its potential for biodiversity studies is obvious. With our method

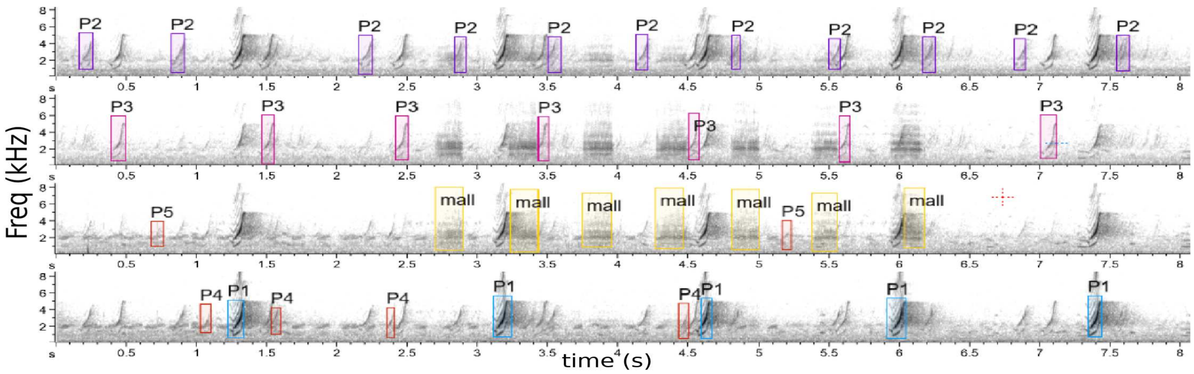

Fig. 7. Annotation of original 4-channel field recording. Each spectrogram is made from one microphone. Direction can be guessed by comparing the amplitudes across microphones.

we were able to separate at least 5 animals based on calling direction. The method also allows the reliable assignment of individual calls to a sequence, even if these are superimposed by other sounds.

Although our microphone arrangement was not optimal for beamforming-based direction determination, the DOA could be estimated with high accuracy. The 4-microphone array in our study was originally designed for classical localization using TDOA estimates [6], and therefore had a larger inter-microphone spacing (Fig. 1). Spotted Crake calls have a peak frequency of 2.5 kHz, meaning that DOA estimation by beamforming alone would be extremely difficult, due to strong sidelobes (Fig. 2B). Sidelobes could be avoided by matching the microphone spacing to the wavelength of the signal, but this would result in a very wide beam and lower localization sharpness [12]. Despite wide beams and high sidelobes, our approach of TF bin clustering using sidelobe suppression (Fig. 2D), which exploits the assumption of only one source present at a bin, allows DOA estimates to be made at individual bins with fine resolution. By averaging over the large number of DOA estimates (i.e. color) in a cluster, and applying spatial smoothing (Fig. 5), a robust DOA estimate is achieved. Our microphone array design also allows good independent angular validation by means of acoustic triangulation. In addition, the directional spectrograms (Figs. 3 and 6A-E) may lead to tools for signal annotation and automated species identification.

VIII. Conclusions

We anticipate that the sound source separation approach presented here, opens new opportunities for applications in the field of automated biodiversity monitoring, specifically for abundance estimation of a wide spectrum of sound-producing animals. Additional field work could reveal the most appropriate array construction, regarding the number, spacing, and geometrical arrangement of the microphones. A crucial prerequisite for widespread and permanent use under field conditions would be, however, the availability of weatherproof cardioid microphones, making sound-wave-deflecting constructions for rain protection superfluous. Weatherproof microphones would also facilitate 3D DOA estimation with arrays of 4 or more microphones mounted in tetragonal or similar arrangements [13], [14].

References

[1] W. Turner, "Sensing biodiversity," *Science*, vol. 346, no. 6207, pp. 301–302, 2014.

[2] A. Bush, R. Sollmann, A. Wilting, K. Bohmann, B. Cole, H. Balzter, C. Martius, A. Zlinszky, S. Calvignac-Spencer, C. Cobbold, *et al.*, "Connecting earth observation to high-throughput biodiversity data. nat ecol evol 1: 0176," *Nat Ecol Evol*, 2017.

[3] R. Gibb, E. Browning, P. Glover-Kapfer, and K. E. Jones, "Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring," *Methods in Ecology and Evolution*, vol. 10, no. 2, pp. 169–185, 2019.

[4] N. Priyadarshani, S. Marsland, and I. Castro, "Automated birdsong recognition in complex acoustic environments: a review," *Journal of Avian Biology*, vol. 49, no. 5, pp. jav–01447, 2018.

[5] T. A. Rhinehart, L. M. Chronister, T. Devlin, and J. Kitzes, "Acoustic localization of terrestrial wildlife: Current practices and future opportunities," *Ecology and Evolution*, vol. 10, pp. 6794–6818, 2020.

[6] K.-H. Frommolt and K.-H. Tauchert, "Applying bioacoustic methods for long-term monitoring of a nocturnal wetland bird," *Ecological Informatics*, vol. 21, pp. 4–12, 2014.

[7] J. W. Popp, R. W. Ficken, and J. A. Reinartz, "Short-term temporal avoidance of interspecific acoustic interference among forest birds," *The Auk*, vol. 102, no. 4, pp. 744–748, 1985.

[8] H. Brumm, "Signalling through acoustic windows: nightingales avoid interspecific competition by short-term adjustment of song timing," *Journal of Comparative Physiology A*, vol. 192, no. 12, pp. 1279–1285, 2006.

[9] G. DeMuth, "Frequency domain beamforming techniques," in *ICASSP '77. IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 713–715, 1977.

[10] S. Banerjee, A. Choudhary, and S. Pal, "Empirical evaluation of k-means, bisecting k-means, fuzzy c-means and genetic k-means clustering algorithms," in *2015 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE)*, pp. 168–172, 2015.

[11] D. M. Titterington, A. F. M. Smith, and U. E. Makov, *Statistical Analysis Of Finite Mixture Distributions*. John Wiley & Sons, 1985.

[12] H. Wang, C. Chen, A. Ali, S. Asgari, R. Hudson, K. Yao, D. Estrin, and C. Taylor, "Acoustic sensor networks for woodpecker localization," in *Advanced Signal Processing Algorithms, Architectures, and Implementations XV*, vol. 5910, p. 591009, International Society for Optics and Photonics, 2005.

[13] Z. Harlow, T. Collier, V. Burkholder, and C. E. Taylor, "Acoustic 3d localization of a tropical songbird," in *IEEE China Summit and International Conference on Signal and Information Processing (ChinaSIP)*, pp. 220–224, 2013.

[14] R. W. Hedley, Y. Huang, and K. Yao, "Direction-of-arrival estimation of animal vocalizations for monitoring animal behavior and improving estimates of abundance," *Avian Conservation and Ecology*, vol. 12, no. 6, 2017.