*INTERSPEECH 2020*
October 25–29, 2020, Shanghai, China

# Discriminative Singular Spectrum Analysis for bioacoustic classification

*Bernardo B. Gatto* [1,2], *Eulanda M. dos Santos* [1], *Juan G. Colonna* [1],
*Naoya Sogi* [2], *Lincon S. Souza* [2], *Kazuhiro Fukui* [2]

[1] Federal University of Amazonas
[2] University of Tsukuba

bernardo@cvlab.cs.tsukuba.ac.jp

## Abstract

Classifying bioacoustic signals is a fundamental task for ecological monitoring. However, this task includes several challenges, such as nonuniform signal length, environmental noise, and scarce training data. To tackle these challenges, we present a discriminative mechanism to classify bioacoustic signals, which does not require a large amount of training data and handles nonuniform signal length. The proposed method relies on transforming the input signals into subspaces generated by the singular spectrum analysis (SSA). Then, the difference between the subspaces is used as a discriminative space, providing discriminative features. This formulation allows a segmentation-free approach to represent and classify bioacoustic signals, as well as a highly compact descriptor inherited from the SSA. We validate the proposed method using challenging datasets containing a variety of bioacoustic signals.

**Index Terms**: subspace analysis, Fisher criterion, bioacoustic classification

## 1. Introduction

Environmental monitoring has taken on an increasingly important role by providing means to analyze and evaluate climate changes. Cataloging and counting animals through bioacoustic monitoring offers a large amount of information to understand and solve various problems. For example, recent studies have pointed out that the migratory route of some species of birds has been drastically affected by global warming [1]. Since these animals are susceptible to such changes, it is valuable to study the dynamics of their populations [2].

Ecological monitoring presents many challenges, such as the need for obtaining information from remote access areas and the requirement for specialized types of equipment, which are often expensive. Many authors have presented solutions based on the bioacoustic signal classification to cope with these challenges. Current solutions usually employ a bioacoustic sensor network to capture signals in remote locations. Then, a classification model may be employed to identify and count individuals of each species and send its results, decreasing the data load on the network. Solutions based on bioacoustic signals have minimal impact on the ecosystem and can be implemented with low-cost hardware [3]. Challenges in matching bioacoustic signals include handling different syllable length and alignments, requiring sophisticated feature extraction techniques. Overall, the recordings may also present a high level of redundancy and long segments with no informative data.

Recently, subspace-based methods have been employed to represent bioacoustic signals for classification problems. These methods attempt to group data into clusters called subspaces, where the learning patterns are represented as the linear combination of the basis vectors. These basis vectors can be ranked according to its information contribution, providing a data compression and selection mechanism. Subspace-based methods operate on multiple patterns at once, achieving higher recognition rates than methods that operate on single patterns [4].

Since bioacoustic sensors may collect redundant signals, subspace-based methods are useful for compacting bioacoustic signals, proving a suitable representation [5]. Mutual singular spectrum analysis (MSSA) is a framework designed to handle signals of nonuniform length, achieving competitive accuracy results on supervised learning problems [6]. MSSA employs the eigenvectors obtained by singular spectrum analysis (SSA) to represent the bioacoustic signals. As the eigenvectors span a subspace, the comparison among bioacoustic signals is simplified by the use of canonical angles. This method achieved encouraging results in very challenging databases. Moreover, MSSA requires only one singular value decomposition to represent a bioacoustic signal of any length.

Another advantage of MSSA is its relative robustness to environmental noise with Gaussian characteristics, including an automatic feature extraction mechanism. Since the method operates directly on signals of different sizes, the extraction of syllables becomes unnecessary, which makes the method computationally efficient. As a result, MSSA is a time and memory-efficient method, which is expected in bioacoustic signals classification applications. Also, the linear subspaces express data through the linear combination of the features. Thus, the eigenvectors extracted from a few examples may represent a vast number of signals, allowing subspace-based methods to obtain excellent results even when few learning samples are available.

Despite its capabilities, MSSA has no discriminative mechanism, i.e., the subspaces generated to represent the bioacoustic classes may not be optimal for classification, since the bioacoustic subspaces are computed independently, neglecting the relationship between them. This difficulty may prevent MSSA from achieving even more competitive results [7].

**Contribution:** Motivated by the recent results obtained by MSSA, we develop a discriminative mechanism for bioacoustic subspaces denominated Discriminative Singular Spectrum Analysis (DSSA). DSSA is based on the difference between the subspaces and extracts discriminative features from bioacoustic signals, maintaining the compact representation provided by the subspaces. Thus, DSSA inherits the advantages of MSSA, such as the ability to compare signals of different lengths without segmentation, and robustness to noise.

The proposed DSSA[1] relies on transforming the input signals into subspaces generated by the singular spectrum analysis (SSA). Then, the difference between the subspaces is used as a discriminative space, providing discriminative features. This formulation allows a segmentation-free approach to represent and classify bioacoustic signals, as well as a highly compact descriptor inherited from the SSA. The validation is given using datasets containing a variety of bioacoustic signals.

[1] The code is available upon request.



http://dx.doi.org/10.21437/Interspeech.2020-2134

## 2. Proposed Method

In this work, we use the following notations. Scalars and matrices are denoted by lowercase and uppercase letters, respectively. Calligraphic letters are assigned to subspaces, and Greek letters to eigenvectors and canonical angles. The subspace $\mathcal{S}$ spanned by the eigenvectors $\{\phi_j \in \mathbb{R}^l\}_{j=1}^d$ is $d-$dimensional.

**Problem formulation:** Let us consider a classification problem where a dataset containing supervised signals $\{X_i\}_{i=1}^n$ is available, and each signal is labeled with one of the $c$ available classes. In the proposed method, the supervised signals are expressed by subspaces, and a discriminative space $\mathcal{D}$ is computed to preserve the fundamental information for classification. Now, given an input signal $Y$, its subspace $\mathcal{Q}$ is extracted, providing a compact representation of $Y$. After that, the projection of the subspaces onto $\mathcal{D}$ produces subspaces more adequate for classification, which can be conducted by a $k-$NN classifier.

**SSA for bioacoustic subspace representation:** The traditional SSA consists of 2 stages. The first stage decomposes the signal, and the second stage reconstructs the decomposed series to provide an enhanced signal. We are interested in the decomposition properties presented by SSA. The decomposition produces independent components, which can represent trends, periodic components, or noise, depending on the application [8].

First, SSA transforms an input signal $X = \{x_1, \ldots, x_m\}$ into a matrix structure. This procedure is conducted by selecting a vector of $l$ consecutive sub-signals of $X$ and moving this throughout the input signal. This operation can also be regarded as an embedding and generates the trajectory matrix $H$ with dimensions $l$ by $k$, where $k = m - l + 1$. By embedding $X$ into its time-delayed coordinates results in a sequence of lagged vectors. Finally, this set of lagged vectors is arranged as columns of a trajectory matrix with a Hankel structure, as follows:

$$H = \begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_k \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_l & x_{l+1} & x_{l+2} & \cdots & x_m \end{bmatrix}. \quad (1)$$

During the decomposition stage, a window length $l$ should be set. The parameter $l$ is usually chosen experimentally (unless strong assumptions are made) and depends on the structure of the data. A useful strategy is to set $l$ proportionally to the periodicity of the data, usually, $l \ll k$. The Nyquist rate gives some direction about how to set $l$ appropriately when specific information about the observations is available [9].

By computing the correlations between the entries of $H$, one can obtain a matrix $U$ whose columns form an orthogonal basis of the $l-$dimensional space. The matrix $U$ can be obtained by the eigenvalue decomposition of the $l \times l-$dimensional autocorrelation matrix $A = HH^\top$ as follows: $A = U\Sigma U^\top$. This decomposition presents the related bioacoustic signal $X$ by the matrix $U$ of eigenvectors and corresponding eigenvalues $\Sigma = \mathrm{diag}(\sigma_1, \cdots, \sigma_l)$. One advantage of this new representation is that it presents the most representative components of the signal in an orderly fashion, allowing the selection of the most relevant components in terms of the reconstruction error.

The dimension $p$ ($p \ll l$) is a critical parameter. By selecting an adequate subspace to represent a bioacoustic signal, the trade-off between memory storage and representation can be highly optimized. Similar to PCA, we define $p$ by the cumulative energy of the eigenvectors. For different datasets and applications, it is not trivial to set $p$ uniformly.

**Canonical angles for bioacoustic subspaces:** The canonical angles between two $p-$dimensional subspaces $\mathcal{P}$ and $\mathcal{Q}$ can be computed by the singular values of $W = U^\top V$. The eigenvectors $U$ and $V$ span the bioacoustic subspaces $\mathcal{P}$ and $\mathcal{Q}$, respectively. By using the singular values of $W$, $\{\delta_1, \ldots, \delta_p\}$, the canonical angles can be obtained by the following equation:

$$\{\theta_1, \ldots, \theta_p\} = \{\cos^{-1}(\delta_1), \ldots, \cos^{-1}(\delta_p)\}. \quad (2)$$

Where the first canonical angle $\theta_1$ is the smallest angle between the subspaces spanned by the basis vectors $U$ and $V$. Then, the second angle $\theta_2$ is the smallest angle in the orthogonal direction of $\theta_1$. The remaining angles follow the same concept.

A reasonable method of estimating the similarity between two $p-$dimensional subspaces is by averaging the sum of the canonical angles. This procedure can be achieved as follows:

$$\gamma(\mathcal{P}, \mathcal{Q}) = \frac{1}{p}\sum_{j=1}^{p} \cos^2(\theta_j). \quad (3)$$

The average $\gamma(\cdot,\cdot)$ provides interpretability, since when the bioacoustic subspaces share a large amount of oscillatory components, $\gamma(\cdot,\cdot)$ approaches 1, indicating that these subspaces have very high similarity. Otherwise, $\gamma(\cdot,\cdot)$ approaches zero, suggesting that the subspaces represent distinct bioacoustic classes. Additionally, The canonical angles are computationally efficient, requiring only an SVD computation.

**The difference between bioacoustic subspaces:** The sum subspace $\mathcal{S}$ of two $p-$dimensional subspaces $\mathcal{P}_1$ and $\mathcal{P}_2$ spanned by the basis vectors $U_1$ and $U_2$ can be computed through the sum of the autocorrelation matrices of $U_1$ and $U_2$:

$$G_{(2)} = U_1U_1^\top + U_2U_2^\top. \quad (4)$$

The matrix $G_{(2)}$ contains the sum of the autocorrelation matrices of $U_1$ and $U_2$ and, therefore, its eigenvectors span the sum subspace $S$. The algebraic intuition behind this procedure is that the sum of two subspaces is equal to the span of their union [10]. Therefore, we obtain the basis vectors that span $\mathcal{S}$ by solving the following eigen problem:

$$G_{(2)} = E_{(2)}\Lambda_{(2)}E_{(2)}^\top. \quad (5)$$

In (5), the columns of $E_{(2)} = \{\phi_1, \ldots, \phi_p\}$ are the normalized eigenvectors of $G_{(2)}$, and $\Lambda_{(2)}$ is the diagonal matrix with corresponding eigenvalues $\{\lambda_1, \ldots, \lambda_p\}$ in descending order, where $p = \mathrm{rank}(G_{(2)})$. A practical interpretation of this concept is that the sum subspace $\mathcal{S}$ is composed of the oscillatory components contained in both $\mathcal{P}_1$ and $\mathcal{P}_2$, in addition to their linear combinations. This property allows the bioacoustic subspaces to represent a high amount of data with few basis vectors. We can further decompose $\mathcal{S}$ so the principal subspace $\mathcal{F}$ and the difference subspace $\mathcal{D}$ can be brought in evidence:

$$\mathcal{S} = \mathcal{F} \oplus \mathcal{D}. \quad (6)$$

In Equation 6, the symbol $\oplus$ stands for the decomposition of $\mathcal{S}$ into two subspaces. This decomposition can be accomplished by examining the eigenvalues associated with the eigenvectors spanning the sum subspace [11, 12]. The dimensions of $\mathcal{F}$ and $\mathcal{D}$ can be obtained directly by checking the distribution of the eigenvalues of $G_{(2)}$. Since the sum of two autocorrelation matrices composes $G_{(2)}$, it will present eigenvalues varying from 2.0 to 0.0, where the common structures (principal space $\mathcal{F}$) will be allocated on the first eigenvectors ($1.0 < \lambda_i \leq 2.0$). Then, the difference subspace $\mathcal{D}$ is spanned by the eigenvectors associated with the eigenvalues ranging from 1.0 to 0.0.

**GDS of bioacoustic subspaces:** In a classification problem, $\{\mathcal{P}_i\}_{i=1}^n$ is the set of reference bioacoustic subspaces

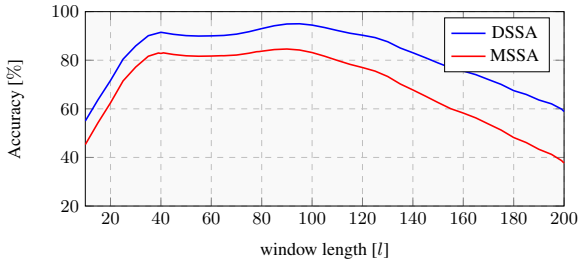



Figure 1: *Accuracy of the methods when $l$ is modified.*

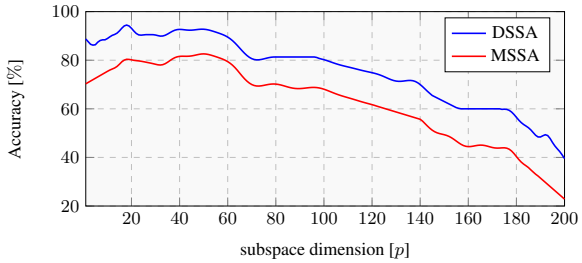



Figure 2: *Accuracy of the methods when $p$ is modified.*

spanned by $\{U_i\}_{i=1}^{n}$. A subspace $\mathcal{D}$ that can act on $\mathcal{P}_i$ can be developed to extract discriminative information. In DSSA, this method is conducted through the removal of the principal subspace that represents the intersection between the different class subspaces. By projecting the subspaces onto $\mathcal{D}$, we expect to obtain useful information for classification. Then, the normalized sum of the matrices $G_{(n)}$ is computed as follows:

$$G_{(n)} = \frac{1}{n}\sum_{i=1}^{n} U_i U_i^{\top}. \tag{7}$$

Since the matrix $G_{(n)}$ has information regarding all the $n$ bioacoustic subspaces, it is interesting to exploit it to extract discriminative elements, by eigen decomposing $G_{(n)}$ as follows:

$$G_{(n)} = E_{(n)}\Lambda_{(n)}E_{(n)}^{\top}. \tag{8}$$

The subset of $E_{(n)}$, $E_{(n)}^{*} = \{\psi_1, \ldots, \psi_d\}$ associated with the smallest eigenvalues $\Lambda_{(n)}$ preserves most of the discriminative information contained in $G_{(n)}$ and can be used to generate $\mathcal{D}$. The optimal subspace dimension $d$ is set experimentally by maximizing the orthogonality degree among the bioacoustic class subspaces projected on $\mathcal{D}$. Once equipped with $\mathcal{D}$ and $\{\mathcal{P}_i\}_{i=1}^{n}$, we can project $\{\mathcal{P}_i\}_{i=1}^{n}$ onto $\mathcal{D}$, obtaining a set of bioacoustic subspaces $\{\dot{\mathcal{P}}_i\}_{i=1}^{n}$, where their oscillatory components are more discriminative than the ones given in $\{\mathcal{P}_i\}_{i=1}^{n}$.

**Fisher score for bioacoustic subspaces:** We adapt the Fisher score [13] to estimate the orthogonality degree between the bioacoustic subspaces. This score is employed to optimize the discriminative subspace $\mathcal{D}$. The Fisher score is broadly employed for model selection and consists of scoring a nested model according to its discriminative importance. More precisely, the Fisher score evaluates a given model regarding the distances between data points in different classes and the distances between data points in the same class. A high Fisher score ensures high between-class and low within-class variability, providing a stable learning model. Since we represent bioacoustic signals by subspaces, we introduce Fisher's formulation in terms of subspaces. The average between-class and within-class variability $F_{(b)}$ and $F_{(w)}$ can be defined as follows:

$$F_{(b)} = \frac{1}{n}\sum_{i=1}^{n} \gamma\big(\hat{\mathcal{P}}_i, \hat{\mathcal{P}}\big), \tag{9}$$

$$F_{(w)} = \frac{1}{r}\sum_{i=1}^{n}\sum_{k=1}^{n_i} \gamma\big(\hat{\mathcal{P}}_i, \mathcal{P}_{ik}\big), \tag{10}$$

where $\hat{\mathcal{P}}_i$ stands for the Karcher mean of the $i-$th class subspace, $\hat{\mathcal{P}}$ is the Karcher mean of the $\hat{\mathcal{P}}_i$ subspaces, $n_i$ is the number of subspaces of the $i-$th class and $r = n \cdot n_i$. Finally, $\gamma(\cdot,\cdot)$ is described in (3). The Karcher mean can be obtained by solving the following optimization problem [14, 15]: $\hat{\mathcal{P}}_i = \arg\min \sum_{k=1}^{n_i} \gamma(\mathcal{P}, \mathcal{P}_k)^2$. Then, $F(\mathcal{M}) = F_{(b)}/F_{(w)}$ is the orthogonality degree for the projection space $\mathcal{M}$. The given score is employed to select the optimal dimension of $\mathcal{D}$.

## 3. Experimental Results

In this section, we first present the employed databases, and then we provide a series of experiments to demonstrate the efficacy of the proposed DSSA. In the first experiment, we evaluate the parameters of DSSA, such as the window length $l$ and the bioacoustic subspace dimension $p$ that result in the best representation. Finally, we visualize the clusters produced by MSSA and DSSA. In the last experiment, we compare the proposed method with existing bioacoustic classification methods.

**Databases employed for the experiments:** The NU-Hive dataset [16] contains 576 files of 10 min duration each, resulting in approximately 96 hours of recordings. The task is to classify whether the bee queen is present or not in the beehive. The samples from two hives have been employed, collected with sampling frequency = 32 kHz by microphones inside the hives.

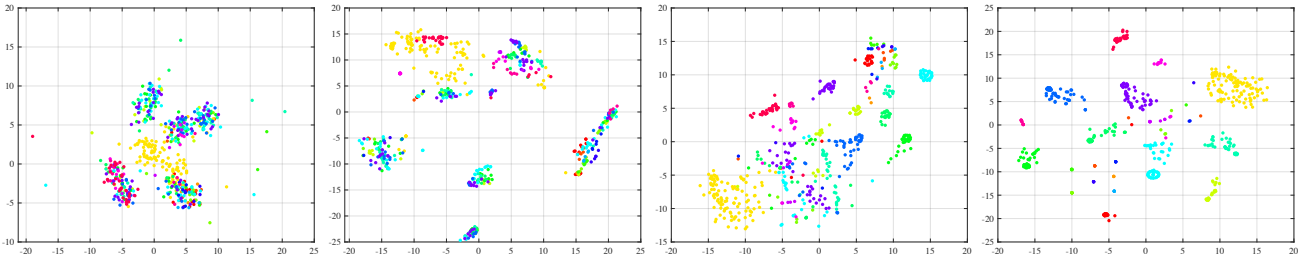
The Mosquito wingbeat dataset [17] contains 626 recordings of 20 different species of mosquitoes. The records reflects the bioacoustic signatures of free-flying and were acquired using mobile phones at different sampling rates (8 kHz $\sim$ 44.1 kHz). The samples were sampled to 44.1 kHz for evaluation.

The Anuran dataset [18] contains 60 recordings of 10 different species of frogs with varying lengths collected under noise conditions, with the number of records per species varying from 3 to 11. It provides a genuine challenge due to the limited number of exemplars caused by cataloging difficulties.

The datasets employed in this experimental section are relatively small; we understand that deep neural methods, such as CNN, cannot be directly employed. An alternative to using deep networks would be to employ methods involving transfer learning, which restricts the range of applications.

**Parameter evaluation:** We employ the NU-Hive dataset to evaluate the window length $l$ of the Hankel matrix, and the number of basis vectors $p$, which maximizes the accuracy of MSSA and DSSA. This analysis is crucial to understand how robust the proposed method is regarding the parameter change compared to MSSA. Also, knowing how $l$ and $p$ affect the model allows the development of new bioacoustic systems in similar datasets.

Figure 1 shows the changes in terms of accuracy in both methods when the window length $l$ varies from 10 to 200. The horizontal axis denotes the window length $l$ used for computing the Hankel matrix. For this experiment, we set $p$ to represent 95% of the variance contained in each bioacoustic subspace. From the results, we can verify that the accuracy of MSSA and DSSA increases, as the window length $l$ increases until it reaches the neighborhood of 40. The value of $l$ that maximizes



(a) *LPC* (b) *MFCC* (c) *MSSA* (d) *DSSA*

Figure 3: *T-SNE embeddings showing distances between the* 20 *classes of the Mosquito wingbeat dataset using* 4 *features.*

the accuracy of both methods is in the range 90 to 95, which produces about 95% and 84% of accuracy on DSSA and MSSA, respectively. This result shows the effect of selecting the proper window length $l$ to represent a subspace. When selecting subspaces with $l > 100$, the accuracy no longer improves, suggesting that a $l > 100$ includes noise data.

Figure 2 shows the effects on the accuracy of the models when the number of eigenvectors $p$ changes from 1 to 200. According to the results, DSSA requires fewer eigenvectors to achieve a higher recognition rate. This note confirms that equipping the model with a discriminative space improves its accuracy. We set $l$ to 95, which was the optimal value obtained in the previous experiment. When $p$ is set to 51 and 18, it produces 96% and 85% of accuracy on DSSA and MSSA, respectively.

**Evaluating the separability of DSSA:** The efficiency of DSSA is verified on the Mosquito wingbeat dataset using the Fisher score. This score approaches 1.0 when the distance between the subspaces of different classes is high, and the distance between the same classes subspaces is low. Differently, it approaches 0.0 otherwise. We also employ two feature descriptors for audio data: MFCC and Linear Prediction Coefficients (LPC) [19]. Figure 3 shows the scatter plots of LPC, MFCC, MSSA, and DSSA produced by the t-SNE. In the plots, each point corresponds to one sample from the Mosquito wingbeat dataset, and the different colors denote the different classes. We employed 20 MFCCs and 12 LPCs to represent the samples since these parameters are commonly used in literature [18]. According to the t-SNE plots, the clusters presented by MFCC are visually more compact, and the different classes are more distinguishable than the ones given by LPC. In contrast, the clusters generated by LPC present a high amount of overlap, which may negatively interfere with the classification accuracy.

For the subspace-based methods, we set $l$ to $l = 200$, and $p$ is set to 9. The dimension $d$ of the discriminative subspace that maximizes the Fisher score is 41. The Fisher score is 0.41 for MSSA and 0.79 for DSSA. These indexes indicate that the discriminative mechanism employed by DSSA for bioacoustic subspaces offers more reliable features for classification than the ones provided by MSSA. According to the results shown by t-SNE, MSSA and DSSA consistently exhibit more compact clusters than those attained by LPC and MFCC. These results suggest that the bioacoustic signals benefit from the feature extraction representation provided by the subspaces. Also, the clusters presented by the DSSA exhibits higher separability among the different classes than the produced by MSSA, suggesting that the discriminative mechanism adopted by DSSA provides more reliable clusters than the ones given by MSSA.

**Evaluating the accuracy of DSSA:** We evaluate DSSA and three baselines under limited training data conditions. This experiment is important since many practical problems can only be solved when the learning model handles scarce training data. The number of training samples is set to 40%, 60%, and 80% on the dataset-I, -II, and -III, respectively. The remaining data were used for testing. In each case, we randomly select the training data and repeat the experiment 10 times. We report the average classification rates attained in each scenario. The CNN configuration corresponds to the reported in [16], where CNN-I and CNN-II stand for CNN operating on MFCC and HHT, respectively. Table 1 shows the classification results of different algorithms by the measurement of accuracy in percentage. The top value is highlighted in bold font and the second-best in italic. The overall performance of DSSA is competitive with the presented by the baselines. In particular, DSSA works well when the training data is limited, confirming the suitability of subspaces for representing and classifying bioacoustic signals under limited resources conditions. The linear combinations of the oscillatory components described by the eigenvectors are able to express distinct patterns even when few data is available.

Table 1: *Classification results of different methods.*

| Datasets | CNN-I | CNN-II | MSSA | DSSA |
|---|---|---|---|---|
| Anuran-I | *61.91* | 57.23 | 55.64 | **66.29** |
| Anuran-II | *68.56* | 65.74 | 57.39 | **69.67** |
| Anuran-III | **76.27** | 73.11 | 62.88 | *75.31* |
| Beehive-I | *87.71* | 85.88 | 74.87 | **89.32** |
| Beehive-II | *92.48* | 90.67 | 81.19 | **93.36** |
| Beehive-III | **97.33** | 94.15 | 85.57 | *96.20* |
| Mosquito-I | 66.33 | *66.67* | 61.93 | **69.15** |
| Mosquito-II | **73.35** | 72.24 | 65.41 | *72.83* |
| Mosquito-III | 75.21 | **76.33** | 69.27 | *75.62* |

## 4. Conclusions and Future Work

We proposed a bioacoustic signal classification method based on subspace representation. We achieved improvements in classification accuracy employing several datasets by introducing a discriminative mechanism using the concept of the difference between subspaces. For future directions, we will investigate diverse types of signals (e.g., speech and seismic signals), which can benefit from the low computational cost and robustness of the proposed method. We will apply DSSA in a lightweight network architecture [20, 21], which may improve its performance.

## 5. Acknowledgements

This work was supported by JSPS KAKENHI Grant Numbers 19K20335 and 19H04129.

# 6. References


[1] D. A. Satterfield, P. P. Marra, T. S. Sillett, and S. Altizer, "Responses of migratory species and their pathogens to supplemental feeding," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 373, no. 1745, p. 20170094, 2018.

[2] L. M. Diele-Viegas, F. P. Werneck, and C. F. D. Rocha, "Climate change effects on population dynamics of three species of amazonian lizards," *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, vol. 236, p. 110530, 2019.

[3] S. J. Ramson and D. J. Moni, "Applications of wireless sensor networks—a survey," in *2017 International Conference on Innovations in Electrical, Electronics, Instrumentation and Media Technology (ICEEIMT)*. IEEE, 2017, pp. 325–329.

[4] H. Tan, Y. Gao, and Z. Ma, "Regularized constraint subspace based method for image set classification," *Pattern Recognition*, vol. 76, pp. 434–448, 2018.

[5] L. S. Souza, B. B. Gatto, and K. Fukui, "Grassmann singular spectrum analysis for bioacoustics classification," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 256–260.

[6] B. B. Gatto, J. G. Colonna, E. M. dos Santos, and E. F. Nakamura, "Mutual singular spectrum analysis for bioacoustics classification," in *2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE, 2017, pp. 1–6.

[7] L. S. Souza, B. B. Gatto, and K. Fukui, "Classification of bioacoustic signals with tangent singular spectrum analysis," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 351–355.

[8] R. Vautard, P. Yiou, and M. Ghil, "Singular-spectrum analysis: A toolkit for short, noisy chaotic signals," *Physica D: Nonlinear Phenomena*, vol. 58, no. 1-4, pp. 95–126, 1992.

[9] S. M. Alessio, "Singular spectrum analysis (ssa)," in *Digital Signal Processing and Spectral Analysis for Scientists*. Springer, 2016, pp. 537–571.

[10] C. Liu, W. Ke, F. Qin, and Q. Ye, "Linear span network for object skeleton detection," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 133–148.

[11] K. Fukui and A. Maki, "Difference subspace and its generalization for subspace-based methods," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 11, pp. 2164–2177, 2015.

[12] K. Fukui, N. Sogi, T. Kobayashi, J.-H. Xue, and A. Maki, "Discriminant analysis based on projection onto generalized difference subspace," *arXiv preprint arXiv:1910.13113*, 2019.

[13] Q. Gu, Z. Li, and J. Han, "Generalized fisher score for feature selection," *arXiv preprint arXiv:1202.3725*, 2012.

[14] X. Pennec, "Intrinsic statistics on riemannian manifolds: Basic tools for geometric measurements," *Journal of Mathematical Imaging and Vision*, vol. 25, no. 1, p. 127, 2006.

[15] P. Turaga, A. Veeraraghavan, A. Srivastava, and R. Chellappa, "Statistical computations on grassmann and stiefel manifolds for image and video-based recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 11, pp. 2273–2286, 2011.

[16] I. Nolasco, A. Terenzi, S. Cecchi, S. Orcioni, H. L. Bear, and E. Benetos, "Audio-based identification of beehive states," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 8256–8260.

[17] H. Mukundarajan, F. J. H. Hol, E. A. Castillo, C. Newby, and M. Prakash, "Using mobile phones as acoustic sensors for high-throughput mosquito surveillance," *Elife*, vol. 6, p. e27854, 2017.

[18] J. G. Colonna, E. F. Nakamura, and O. A. Rosso, "Feature evaluation for unsupervised bioacoustic signal segmentation of anuran calls," *Expert Systems with Applications*, vol. 106, pp. 107–120, 2018.

[19] R. J. Javier and Y. Kim, "Application of linear predictive coding for human activity classification based on micro-doppler signatures," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 10, pp. 1831–1834, 2014.

[20] B. B. Gatto, L. S. Souza, E. M. dos Santos, K. Fukui, W. S. Júnior, and K. V. dos Santos, "A semi-supervised convolutional neural network based on subspace representation for image classification," *EURASIP Journal on Image and Video Processing*, vol. 2020, no. 1, pp. 1–21, 2020.

[21] B. B. Gatto, E. M. dos Santos, K. Fukui, W. S. Júnior, and K. dos Santos, "Fukunaga–koontz convolutional network with applications on character classification," *NEURAL PROCESSING LETTERS*, 2020.