DETECTION AND IDENTIFICATION OF BEEHIVE PIPING AUDIO SIGNALS

Dominique Fourer and Agnieszka Orlowska

IBISC (EA 4526) - University of Evry / Paris-Saclay*
Evry-Courcouronnes, France
dominique.fourer@univ-evry.fr

ABSTRACT

Piping signals are particular sounds emitted by honey bees during the swarming season or sometimes when bees are exposed to specific factors during the life of the colony. Such sounds are of interest for beekeepers for predicting an imminent swarming of a beehive. The present study introduces a novel publicly available dataset made of several honey bee piping recordings allowing for the evaluation of future audio-based detection and recognition methods. First, we propose an analysis of the most relevant timbre features for discriminating between tooting and quacking sounds which are two distinct types of piping signals. Second, we comparatively assess several machine-learning-based methods designed for the detection and the identification of piping signals through a beehive-independent 3-fold cross-validation methodology.

Index Terms— bees piping signals, quacking, tooting, audio signal recognition, smart beekeeping

1. INTRODUCTION

Nowadays, smart beekeeping is gaining interest since it aims at developing innovative methods for enhancing the monitoring of beehives using AI techniques. To this end, the audio-based approach [1, 2] is promising since it allows to use low-cost sensors for monitoring a bee colony. Recent work pioneered the bee sound analysis problem through a machine learning approach to predict the different health states of a beehive. For example, the task of predicting the bee queen presence is investigated in [3, 4] and could help beekeepers to reduce the number of inspections which are stressful for a beehive. The prediction of colony swarming from audio signal is investigated [5, 6] and can be related to specific sounds emitted by the bees. Several studies analyze different piping sounds and show their interest for beekeepers [7, 5, 6]. Other studies explain that piping signals can also have other functions for synchronizing the colony activity [7, 8]. Such particular sounds can respectively be emitted by bee workers or by a queen and can easily be distinguished from classical background beehive sounds. A more recent study [9] proposes an acoustic analysis of piping signals which can be segregated into two classes with specific audio signatures: tooting and quacking. Both tooting and quacking signals can occur about 1 day before swarming and their occurrences can increase every 10 minutes during approximately 6 hours.

The present study pursues the piping sounds investigation with an analysis of the most relevant audio features using a machine learning-based methodology. Our contributions are manifold. First, we introduce a new publicly available audio piping dataset made of several recordings collected from various beekeepers which were manually segmented and annotated as tooting or quacking. Second, we present an acoustic analysis through timbre features to discriminate between the tooting and the quacking signals. Finally, we assess several methods for a supervised detection and classification of audio field recordings of beehive sounds. This paper is organized as follows. In Section 2, we explain the differences between piping signals and we introduce our new proposed dataset. In Section 3, we perform an acoustic analysis of piping sounds using timbre features. Section 4 presents our audio detection and classification results using several proposed methods. Finally, the paper is concluded by a discussion with future work directions in Section 5.

2. MATERIALS

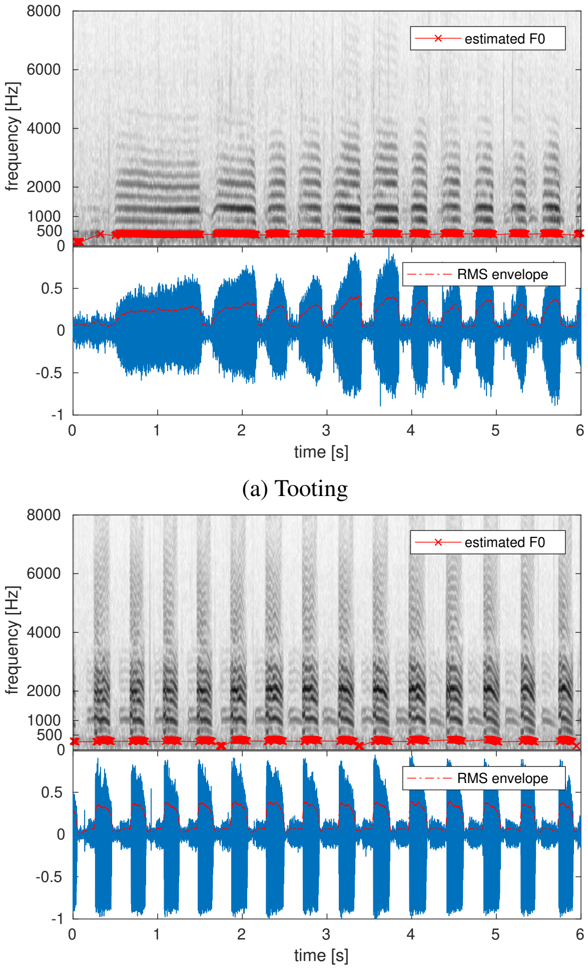
2.1. Tooting and Quacking

Piping sounds (cf. Fig. 1) are among the most noticeable signs of swarming. *Tooting* corresponds to the sound emitted by a virgin queen bee who announces her presence by releasing pheromones and by tooting. Tooting corresponds to a series of pulsed, high-pitched sounds produced by pressing her thorax and operating her wing-beating mechanism without spreading her wings [10]. Mature queens still confined within their queen cells answer the tooting with a distinct piping sound, called *Quacking*. A chorus of synchronized quacking follows each tooting, and those specific swarming sounds are broadcasting in the bee nest as vibrations of the combs and perceived by vibration detectors in the workers' tarsi [11]. Toots and quacks are made of different varying pulses: during the process of tooting, the queen produces a one-second-long pipe immediately followed by several bursts of less than half a second. The fundamental frequency increases with the age of queens, ranging from 200 to 550 Hz, and is usually observed around 400 Hz [12]. Quacks are made of several short pulses which are typically less than 0.2 seconds at a lower fundamental frequency around 350 Hz [13]. Piping sounds are not only emitted by queens but also by workers in queenless colonies: laying-workers and guarding-workers [14]. More recent studies show that workers could emit piping sounds to prepare a synchronized liftoff [7]. This prompts a conclusion that workers pipe in a variety of circumstances, while queens pipe only in the context of colony reproduction [15]. The queens' toots and quacks last several seconds and are broken up into syllables [12]. Piping sounds emitted by workers come from several sources and have a duration below one second. It often consists of a single pulse [14].

2.2. New Proposed Piping Dataset

We introduce a novel dataset of natural honey bee piping audio signals which was built by collecting 44 different recordings pub-

*This work is partly supported by the French ANR ASCETE project (ANR-19-CE48-0001).



(a) Tooting

(b) Quacking

Figure 1: Spectrograms with highlighted F_0 and waveforms with RMS envelope of two distinct piping signals.

lished on the YouTube platform by various beekeepers around the world.These audio recordings were obtained in field conditions using various non-professional microphones located close to the beehive when a piping signal is emitted. Each recording has a duration varying from 2 to 13 seconds and is annotated according to the beekeeper comment respectively as *Tooting* or *Quacking*. We extracted and segmented the audio from 14 distinct videos from which the signal is recorded without a loss of quality into WAVE files with a sampling frequency of $F_s = 22.05$ kHz and a sample precision of 16 bits. After manually removing the silent and spurious frames, the resulting dataset contains 36 tooting signals and 8 quacking signals which correspond to a duration of 145 seconds for tooting and 60 seconds for quacking (total 205 seconds). To avoid possible copyright issues, we only made publicly available the Short-Time Fourier Transform (STFT) matrices and the timbre descriptors computed using a matlab implementation of the timbre toolbox [16] from the post-processed signals used in our experiments. We propose a more detailed description of the dataset containing the links of the original Youtube videos with our matlab loader codes published on IEEE DataPort [17].

3. ACOUSTIC ANALYSIS

3.1. Signal analysis

We present in Fig. 1 the waveform of a tooting and of a quacking signal both extracted from our proposed dataset (Toot1 and Quack1) with almost the same duration of about 6 seconds. Colored in red, we plot the Root Mean Square (RMS) envelope computed for a window length of 23ms. We also display the spectrograms of the same signals where the fundamental frequency (F_0) estimated using the SWIPE method [18] is highlighted. From these observations, one can notice that tooting and quacking are both harmonic signals but with very different temporal and spectral structures. The tooting signal contains longer pulses with a higher F_0 (mean value of $\mu_T = 382.97$Hz with a standard deviation $\sigma_T = 61.45$ Hz) and a slightly lower number of pulses for the same observation duration. For the comparison, the quacking signal contains more pulses with a lower F_0 ($\mu_Q = 306.60$ Hz, $\sigma_Q = 23.98$ Hz). We also notice that the F_0 decreases at the end of each pulse for both tooting and quacking signals.

Table 1: Top-10 most relevant timbre descriptors selected using a mutual information criterion.

	Timbre feature	Relevance score
1	*ERB-gammatone Spectral Centroid*	0.428
2	*ERB-gammatone Spectral Kurtosis*	0.419
3	*ERB-fft Spectral Kurtosis*	0.402
4	*ERB-gamatone Spectral Skewness*	0.373
5	*ERB- fft Spectral Skewness*	0.373
6	*ERB-fft Spectral Centroid*	0.371
7	*ERB-fft Spectral Spread*	0.334
8	*Zero-crossing rate*	0.321
9	*STFT Spectral Kurtosis*	0.314
10	*STFT Spectral Roll-Off*	0.311

3.2. Timbre Feature Selection

The timbre toolbox proposed by Peeters et al. [16] proposes a large set of hand-crafted audio features used in various audio recognition tasks. These features are expected to convey information about the perceived timbre of an arbitrary sound. They include temporal, spectral, harmonic and perceptual descriptors which are directly computed from the waveform and from the time-frequency representation of the analyzed signal. In this study, we investigate a total of 164 timbre features (cf. [19] Table. 2 for details) summarized by median and Inter Quartile Range (IQR) statistics related to the signal acoustic parameters. In Table 1, we present the top-10 most relevant features sorted by descending order of relevance according to the mutual information (MI) criterion [20] by considering the tooting/quacking classification problem. Our computation uses the scikit-learn MI python implementation which shows that perceptual-based Equivalent-Rectangular-Bandwith (ERB) spectral features appear to be the most relevant. Fig. 2a plots in 3 dimensions the whole dataset where each individual corresponds to a one-second-long frame where the axes correspond to the top-3 most relevant features. This figure shows that the components can almost be separated into two distinct clusters corresponding to tooting and quacking signals (plotted with different colors) using only 3 relevant features. In Fig. 2b, we plot a whole dataset projection using Principal Component Analysis (PCA) which is a dimension reduction method reducing the redundancy between the features while preserving original data inertia. This second projection shows that the separation between tooting and quacking sounds is not trivial despite each cluster seem located in a different area. Finally, we perform a Linear Discriminant Analysis (LDA) [21] which can be viewed as a supervised PCA providing the optimal linear projection of the dataset which maximizes the Euclidean distance between individuals of different classes while minimizing the distance between individuals of the same class. Fig. 2c shows that there exists a linear combination of the original timbre features enabling to perfectly separate tooting and quacking sounds. This result paves the way of a supervised classification investigated in Section 4.

4. DETECTION AND CLASSIFICATION RESULTS

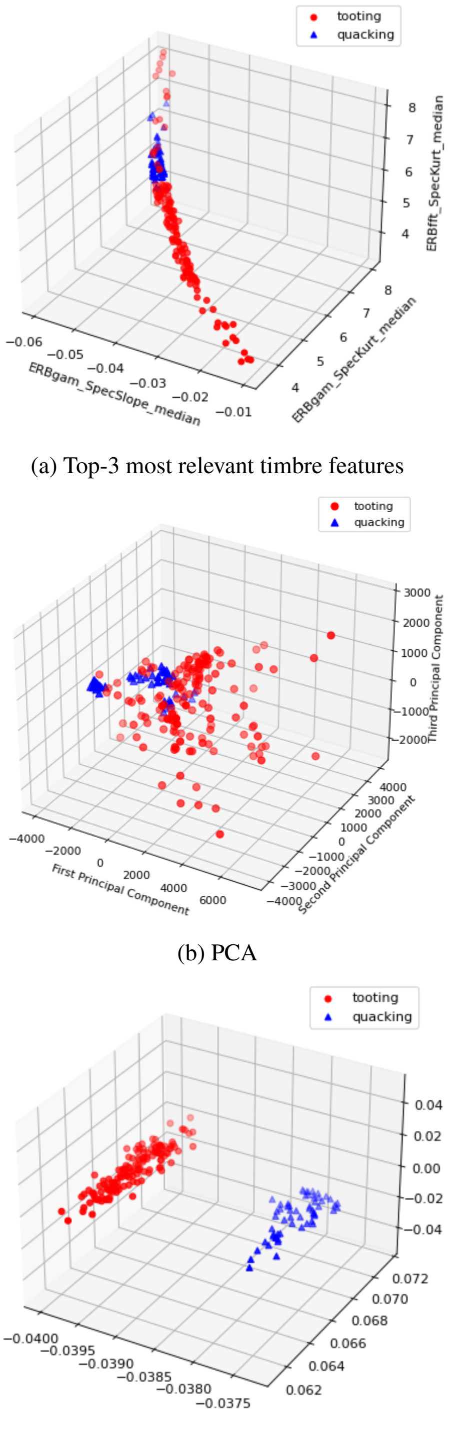
4.1. Experimental Setup

We focus on two distinct tasks which consist of the detection of piping signals and the discrimination between tooting and quacking piping signals. To this end, we consider three distinct experiments. **Experiment 1** focuses on the detection of piping signals from beehives recordings. We address this problem through a binary classification problem involving samples from our proposed dataset and beehive recordings from the OSBH dataset[1] made of several beehives sounds. **Experiment 2** focuses on the binary piping audio classification problem which consists in identifying respectively tooting and quacking signals where 145 recordings are labeled as *tooting* and 60 recordings as *quacking*. **Experiment 3** considers both the detection and the classification problem that is addressed through a 3-label supervised classification approach consisting in predicting if a signal is a *tooting*, a *quacking* or a *non-piping* signal. For each experiment, datasets are preprocessed by splitting signals into one-second-long chunks sampled at $F_s = 22.05$ kHz. Each signal is centered by subtracting the mean and the amplitude is normalized by dividing each sample by $\max(|x|)$. Our evaluation uses a 3-fold cross-validation methodology (2 training folds and 1 testing fold) where the recordings are beehive-independent to avoid overfitting and to assess over the whole dataset the generalizing capability of the trained models. Hence, all recordings from the same Youtube video are only present into a unique fold and cannot simultaneously appear in both the training and testing sets. In experiments 1 and 3 involving *non-piping* signals, we randomly add bee signals from the OSBH dataset to obtain the same number of *piping* and *non-piping* signals in each fold.

4.2. Methods

4.2.1. Classification

We comparatively assess four distinct supervised classification methods suitable for beehive audio signals. The **TTB+SVM** method uses the 164 timbre descriptors investigated in Section 3 combined with a support vector machines (SVM) classifier with a Gaussian radial basis function kernel [22]. The proposed **1D-CNN** method uses the modulus of the discrete Fourier transform of the signal as input of a 1D-convolutional neural network (CNN) with residual connections. This architecture (total: 7,684,226 trainable parameters) is made of 4 residual blocks with a different number of kernel filters (sequentially: 16, 32, 64, 128). Each residual block is made of 3 one-dimensional convolutional layers interspersed by the addition of the input followed by a Rectified Linear Unit (ReLU) activation and a max-pooling. Output of the last residual block is average-pooled and connected to 3 fully-connected (FC) layers including flatten and with ReLU and softmax activation for the final output. The **MFCC+CNN** and the **STFT+CNN** are based on the same 2D-CNN architecture (total: 404,770 trainable



(a) Top-3 most relevant timbre features

(b) PCA

(c) LDA

Figure 2: Three-dimensional projections of our proposed piping dataset where each point corresponds to a one-second-long excerpt.

[1] https://zenodo.org/record/1321278

parameters) with 2 distinct inputs: Mel-Frequency Cepstral Coefficients (MFCC) and the spectrogram defined as the squared modulus of the short-time Fourier transform (STFT). The proposed 2D-CNN architecture is inspired from [2] and consists of 4 convolutional blocks containing 16 kernel filters of size 3×3, a 2×2 max-pooling layer and a 25% dropout layer. The output is connected to a 3 FC layers including 2 dropout layers of respectively 25% and 50% followed by a softmax activation function to compute the output predicted label. Convolutional and FC layers both use a LeakyReLU activation function defined as $LeakyRELU(x) = \max(\alpha x, x)$, with $\alpha = 0.1$.

4.2.2. Detection

For detecting piping in an arbitrary audio signal as proposed in **Experiment 1**, we also consider the 4 proposed classification methods using a binary *piping/non-piping* taxonomy. We also consider two additional methods based on the stochastic modeling of the estimated F_0 distribution respectively for piping and non-piping signals. This later approach is motivated by the harmonic property of piping signals described in Section 3. The **F0 Gaussian model** estimates the parameters $\theta = [\mu, \sigma^2]$ of a Gaussian probability distribution used to model respectively piping and non-piping signals. Thus, given the estimated F_0 denoted f_x of a signal, the decision to detect a piping signal is made when $p(f_x|\theta_{piping}) > p(f_x|\theta_{non-piping})$. The **F0 kernel model** is a variant of the **F0 Gaussian model** where $p(f_x|\theta_{piping})$ is estimated using the empirical distribution (i.e. histogram) of the estimated F_0 smoothed by a convolution product using a Gaussian kernel [23]. Our experiments used the SWIPE F_0 estimator [18] for which the median function is used to summarize a frame of signal with an arbitrary length.

4.3. Implementation details

The 17 first cepstral coefficients of the **MFCC+CNN** method are computed each 20 ms. The STFT is computed using a Hann analysis window with a 50% overlap and a FFT size of $M = 1025$ (i.e. 512 positive frequency bins). The input of the **1D-CNN** is set to half of the frequency sampling due to the Fourier transform Hermitian symmetry of a real signal (i.e. 11,025 real-valued coefficients). During the testing of each of the 3 folds, we use data augmentation (DA) [24] to artificially increase the number of training recordings by generating new samples from the original ones by the addition of a white Gaussian noise (SNR $=$ 25dB) and by the application of temporal random circular shifts. The results reported in Tables 2, 3 and 4 correspond to the best ones obtained after several iterations (no significant improvement is shown by data augmentation). The training of our CNN methods is configured for a constant number of 25 epochs for the **1D-CNN** and 50 epochs for the **2D-CNN**, with a batch size of 16. The overall evaluation framework and the **TTB+SVM** method are implemented in matlab. The deep learning methods are implemented in Python using Keras with Tensorflow frameworks. Our codes are freely available online[2] for the sake of reproducible research.

4.4. Comparative results

According to Table 2, the best detection results in terms of accuracy for **Experiment 1** are obtained using the **TTB+SVM** and the **MFCC+CNN** method which both obtain 94%. The best piping classification (**Experiment 2**) results (cf. Table 3) are obtained using the **STFT+CNN** method with an overall accuracy of 95%, followed from far by the **MFCC+CNN** method which obtains an accuracy of 78%. Despite efforts, the two other techniques fail to identify quacking sounds and obtain poorer results with a quacking F-measure below 0.5. These poor quacking recognition results are confirmed in **Experiment 3** (cf. Table 4) where the best method remain **STFT+CNN** for which the results are poorer than in **Experiment 2**. This suggests the best pipeline which detects piping signals using **MFCC+CNN** or **TTB+SVM** before attempting to discriminate between tooting and quacking signals using **STFT+CNN**.

Table 2: Experiment 1: Piping signals detection comparative results.

Method	Feat. dimension	Label	Recall	Precision	F - score	Accuracy
F0 kern. model	1	Piping	0.68	0.96	0.79	0.84
		Non-piping	0.97	0.78	0.87	
F0 Gauss. model	1	Piping	0.69	0.99	0.81	0.85
		Non-piping	1	0.79	0.88	
TTB+SVM	164	Piping	**0.91**	0.96	**0.94**	**0.94**
		Non-piping	0.97	**0.93**	**0.95**	
1D-CNN	11,025	Piping	0.84	**1.00**	0.91	0.93
		Non-piping	**1.00**	0.88	0.93	
MFCC+CNN	17×47	Piping	0.87	**1.00**	0.93	**0.94**
		Non-piping	**1.00**	0.90	0.94	
STFT+CNN	512×42	Piping	0.86	0.96	0.91	0.92
		Non-piping	0.97	0.89	0.93	

Table 3: Experiment 2: Piping signals binary classification comparative results.

Method	Feat. dimension	Label	Recall	Precision	F - score	Accuracy
TTB+SVM	164	Tooting	0.78	0.85	0.71	0.66
		Quacking	0.24	0.18	0.38	
1D-CNN	11,025	Tooting	**0.97**	0.72	0.82	0.71
		Quacking	0.08	0.50	0.14	
MFCC+CNN	17×47	Tooting	0.93	0.79	0.86	0.78
		Quacking	0.42	0.71	0.53	
STFT+CNN	512×42	Tooting	0.94	**0.98**	**0.96**	**0.95**
		Quacking	**0.96**	**0.87**	**0.92**	

Table 4: Experiment 3: Simultaneously detection and classification comparative results.

Method	Feat. dimension	Label	Recall	Precision	F - score	Accuracy
TTB+SVM	164	Tooting	0.88	0.78	0.83	0.82
		Quacking	0.03	0.12	0.05	
		Non-piping	0.99	0.89	0.94	
1D-CNN	11,025	Tooting	0.93	0.84	0.88	0.85
		Quacking	0.10	0.54	0.16	
		Non-piping	0.99	0.86	0.92	
MFCC+CNN	17×47	Tooting	0.88	0.81	0.84	0.84
		Quacking	0.18	0.45	0.26	
		Non-piping	**0.99**	**0.90**	**0.95**	
STFT+CNN	512×42	Tooting	**0.94**	**0.97**	**0.95**	**0.91**
		Quacking	**0.50**	**0.76**	**0.60**	
		Non-piping	**0.99**	0.89	0.94	

5. CONCLUSION

We introduced a new dataset made of beehive piping sounds designed for identifying tooting and quacking signals emitted by bees. The most relevant timbre features were presented and reveal a link with perceptual spectral features. Our numerical experiments involving several state-of-the-art approaches show that a time-frequency represention combined with a 2D-CNN is currently the most promising approach for addressing the tooting/quacking binary classification problem and can obtain an accuracy above 85%. Future work consists in evaluating new methods in more realistic application scenarios involving embedded systems.

[2] `https://fourer.fr/dcase22`

6. REFERENCES

[1] S. Cecchi, A. Terenzi, S. Orcioni, P. Riolo, S. Ruschioni, and N. Isidoro, “A preliminary study of sounds emitted by honey bees in a beehive,” in *Audio Engineering Society Convention 144*, Milan, Italy, May 2018.

[2] S. Cecchi, A. Terenzi, S. Orcioni, and F. Piazza, “Analysis of the sound emitted by honey bees in a beehive,” in *Audio Engineering Society Convention 147*, 2019.

[3] I. Nolasco and E. Benetos, “To bee or not to bee: Investigating machine learning approaches for beehive sound recognition,” in *Proc. DCASE*, Nov. 2018.

[4] I. Nolasco, A. Terenzi, S. Cecchi, S. Orcioni, H. L. Bear, and E. Benetos, “Audio-based identification of beehive states,” in *Proc. IEEE ICASSP*, 2019, pp. 8256–8260.

[5] A. Zgank, “Bee swarm activity acoustic classification for an iot-based farm service,” *Sensors*, vol. 20, no. 1, p. 21, 2020.

[6] M.-T. Ramsey, M. Bencsik, M. I. Newton, M. Reyes, M. Pioz, D. Crauser, N. S. Delso, and Y. Le Conte, “The prediction of swarming in honeybee colonies using vibrational spectra,” *Scientific reports*, vol. 10, no. 1, pp. 1–17, 2020.

[7] T. D. Seeley and J. Tautz, “Worker piping in honey bee swarms and its role in preparing for liftoff,” *Journal of Comparative Physiology A*, vol. 187, no. 8, pp. 667–676, 2001.

[8] C. Thom, D. C. Gilley, and J. Tautz, “Worker piping in honey bees (apis mellifera): the behavior of piping nectar foragers,” *Behavioral Ecology and Sociobiology*, vol. 53, no. 4, pp. 199–205, 2003.

[9] T. Yamamoto, M. Sugahara, R. Okada, and H. Ikeno, “Differences between queen piping temporal structures of two honeybee species, apis cerana and apis mellifera,” *Apidologie*, vol. 52, no. 2, pp. 524–534, 2021.

[10] J. Simpson, “The mechanism of honey-bee queen piping,” *Zeitschrift für vergleichende Physiologie*, vol. 48, no. 3, pp. 277–282, 1964.

[11] J. Simpson and S. M. Cherry, “Queen confinement, queen piping and swarming in apis mellifera colonies,” *Animal Behaviour*, vol. 17, pp. 271–278, 1969.

[12] A. Michelsen, W. H. Kirchner, B. B. Andersen, and M. Lindauer, “The tooting and quacking vibration signals of honeybee queens: a quantitative analysis,” *Journal of Comparative Physiology A*, vol. 158, no. 5, pp. 605–611, 1986.

[13] W. Kirchner, “Acoustical communication in honeybees,” *Apidologie*, vol. 24, no. 3, pp. 297–307, 1993.

[14] T. Ohtani and T. Kamada, “’worker piping’: The piping sounds produced by laying and guarding worker honeybees,” *Journal of Apicultural Research*, vol. 19, no. 3, pp. 154–163, 1980.

[15] S. Pratt, S. Kühnholz, T. D. Seeley, and A. Weidenmüller, “Worker piping associated with foraging in undisturbed queenright colonies of honey bees,” *Apidologie*, vol. 27, no. 1, pp. 13–20, 1996.

[16] G. Peeters, B. L. Giordano, P. Susini, N. Misdariis, and S. McAdams, “The timbre toolbox: Extracting audio descriptors from musical signals,” *The Journal of the Acoustical Society of America*, vol. 130, no. 5, pp. 2902–2916, 2011.

[17] A. Orlowska and D. Fourer, “Identification of beehive piping audio signals,” in *IEEE Dataport (doi:10.21227/53mq-g936)*, 2021. [Online]. Available: https://dx.doi.org/10.21227/53mq-g936

[18] A. Camacho and J. G. Harris, “A sawtooth waveform inspired pitch estimator for speech and music,” *The Journal of the Acoustical Society of America*, vol. 124, no. 3, pp. 1638–1652, 2008.

[19] D. Fourer, J.-L. Rouas, P. Hanna, and M. Robine, “Automatic timbre classification of ethnomusicological audio recordings,” in *Proc. ISMIR*, Taipei, Taiwan, Oct. 2014.

[20] N. Hoque, D. K. Bhattacharyya, and J. K. Kalita, “Mifs-nd: A mutual information-based feature selection method,” *Expert Systems with Applications*, vol. 41, no. 14, pp. 6371–6385, 2014.

[21] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*. New York, USA: Wiley-Blackwell, 1958.

[22] A. J. Smola and B. Schölkopf, “A tutorial on support vector regression,” *Statistics and computing*, vol. 14, no. 3, pp. 199–222, 2004.

[23] A. W. Bowman and A. Azzalini, *Applied smoothing techniques for data analysis: the kernel approach with S-Plus illustrations*. OUP Oxford, 1997, vol. 18.

[24] D. A. Van D. and X.-L. Meng, “The art of data augmentation,” *Journal of Computational and Graphical Statistics*, vol. 10, no. 1, pp. 1–50, 2001.