Detection of crackle events using a multi-feature approach

L. Mendes, I. M. Vogiatzis, E. Perantoni, E. Kaimakamis
I. Chouvarda, N. Maglaveras, J. Henriques, P. Carvalho, R. P. Paiva

Abstract— **The automatic detection of adventitious lung sounds is a valuable tool to monitor respiratory diseases like chronic obstructive pulmonary disease. Crackles are adventitious and explosive respiratory sounds that are usually associated with the inflammation or infection of the small bronchi, bronchioles and alveoli. In this study a multi-feature approach is proposed for the detection of events, in the frame space, that contain one or more crackles. The performance of thirty-five features was tested. These features include thirty-one features usually used in the context of Music Information Retrieval, a wavelet based feature as well as the Teager energy and the entropy. The classification was done using a logistic regression classifier.**

Data from seventeen patients with manifestations of adventitious sounds and three healthy volunteers were used to evaluate the performance of the proposed method. The dataset includes crackles, wheezes and normal lung sounds. The optimal detection parameters, such as the number of features, were chosen based on a grid search. The performance of the detection was studied taking into account the sensitivity and the positive predictive value. For the conditions tested, the best results were obtained for the frame size equal to 128 ms and twenty-seven features.

I. INTRODUCTION

Crackles are short explosive respiratory sounds that are usually associated with the inflammation or infection of the small bronchi, bronchioles and alveoli, quite common in chronic obstructive pulmonary disease. The automatic detection of these sounds (additional respiratory sounds superimposed on breath sounds) is a valuable tool to follow-up respiratory diseases.

These adventitious sounds seem to result from an abrupt opening or closing of the airways [1]. Several methods have been proposed for automatic detection of crackles such as 1) wavelets based method [2][3], 2) empirical mode decomposition method with Katz fractal dimension filter [4], 3) diffuse systems [5] and 4) autoregressive models [6].

In this study we have developed a multi-feature approach to detect crackles. Since crackles can appear individually or in group, the algorithm aims to detect crackle events, i.e., intervals where at least one crackle is present. The performance of thirty-five features was evaluated. Thirty-one of the features tested to detect these high pitch sounds are usually used in the context of musical information retrieval. A wavelet based feature proposed by Bahoura and Lu [2] to detect crackles was also included in the set of tested features. Another feature tested was the entropy. Signals with high dispersion have high entropy. Since crackles are sounds with high frequency and intensity the Teager energy was also tested.

Patients with manifestations of crackles can also present manifestation of other adventitious sounds. The dataset used to test the proposed method includes patients with manifestation of crackles and wheezes. A feature specially designed to detect wheezes [7], wheeze signatory in the spectrogram space (WS-SS), was also included in the set of features tested.

The performance of the sequential combination of the features was studied taking into account the measured values of the sensitivity and of the positive predictive value. In order to optimize the detection parameters a grid search was done. The detection of crackle events was done in the frame space.

II. THEORY

A. Fractal dimension of the filter WPST–NST

For the automatic detection of crackles Bahoura and Lu [3] proposed a wavelet based method (WPST–NST) using the Daubechies wavelet 8th with 5 levels of decomposition. As done in [2] a filter was applied to the non-stationary part of the signal and the Katz fractal dimension computed. For each frame the maximum of the Katz fractal dimension was used as feature. The values used in this study for the first and second threshold (see [2]) were equal to 0.85 and 1.25, respectively. The value of the filter threshold was equal to 0.6.

B. Detection of the wheezes signature in the spectrogram space (WS-SS)

The dataset used in this study includes patients with manifestation of crackles and wheezes. To improve the robustness of the method against the presence of wheezes, a feature that aims to detect the wheezes signature in the spectrogram space was tested [7].

C. Teager energy operator

For each frame the maximum of the Teager energy of the normalized signal was computed. The Teager Energy Operator, $\psi(.)$, for discrete signals, x[n], is given by

$$\psi(\mathrm{x[n]}) = \mathrm{x}^2[\mathrm{n}] - \mathrm{x}[\mathrm{n}-1]\mathrm{x}[\mathrm{n}+1] , \qquad (1)$$

This work was supported by projects LINK (H2020 - 692023) and WELCOME (FP7 - 611223).
L. Mendes, J. Henriques, P. Carvalho, R. P. Paiva and are with CISUC, Department of Informatics Engineering, University of Coimbra, Portugal (e-mail: {lgmendes, jh, ruipedro, carvalho,} @dei.uc.pt).
I. Vogiatzis is with General Hospital of Imathia- Health Unit of Naoussa, Greece (email: ioanvogiatzis@gmail.com).
E. Perantoni and E. Kaimakamis are with the General Hospital of Thessaloniki ‘G. Papanikolaou’, Greece. (email: elperantoni@gmail.com, vkaimak@med.auth.gr).
N. Maglaveras and I. Chouvarda are with the Laboratory of Medical Informatics, Medical School, Aristotle University of Thessaloniki, Thessaloniki, Greece. (email: {nicmag, ioanna}@med.auth.gr).

with $n \in \mathbb{Z}$.

D. Musical features

Crackles are high pitch sounds. Thirty-one features computed using the MIR toolbox [8] were included in the set of tested features. Table 1 presents the names and the description of the musical features used in this work.

E. Entropy

The information entropy, H, is a measurement of the disorder of a system. The entropy of a segment of a signal quantized into V levels is given by [9]

$$\widehat{\mathrm{H}} = -\sum_{v=1}^{\mathrm{V}} \frac{\mathrm{g}_v}{\mathrm{N}} \log \frac{\mathrm{g}_v}{\mathrm{N}}, \qquad (2)$$

where g_v is the number of times that the vth level appears in the segment of the signal and N is the size of the segment.

The maximum of the entropy of each frame was also used as feature. The normalized signal was quantized into 5 levels and the neighborhood size used to compute the local entropy was equal to 51 samples.

TABLE 1. FEATURES AND THE CORRESPONDENT LABELS USED IN THIS STUDY. THE FEATURES NOT COMPUTED USING THE MIR TOOLBOX [8] ARE HIGHLIGHTED.

Feature	Description	Label
RMS	Root-mean square energy of the frame	1
Spec. Centroid	Geometric center (centroid) of the spectral distribution	2
Spec. Brightness	Amount of energy of the frame above 500 Hz	3
Spec. Spread	Variance of the spectral distribution	4
Skewness	Coefficient of skewness of the spectral distribution	5
Spec. Kurtosis	Excess kurtosis of the spectral distribution	6
Spec. Rolloff 85	Frequency such that a 85% of the total energy is contained below that frequency	7
Spec. Rolloff 75	Frequency such that a 75% of the total energy is contained below that frequency	8
Spec. Flatness	Ratio between the geometric mean (of the spectral distribution) and the arithmetic mean	9
Roughness	Average of all the dissonance between all possible pairs of spectrogram frame peaks	10
Spec. Irregularity	Degree of variation of the successive peaks of the spectrum	11
Chromagram centroid	Centroid of the redistribution of the spectrum energy along the different pitches	12
Chromagram peak	Peak of the redistribution of the spectrum energy along the different pitches	13
Zerocross	Number of times the signal change the sign	14
Keyclarity	Key strength associated to the best key	15
Mode	Estimation of the modality (major mode vs minor mode)	16
13 Mel-frequency cepstral coefficients	Compact description of the shape of the spectral envelope of an audio signal	[17,29]
WS-SS	See section II.B	30
Pitch	Existence of pitches	31
Inharmonicity	Amount of partials that are not multiples of the fundamental frequency	32
FD of WPST–NST	See section II.A	33
Teager energy	See section II.C	34
Entropy	See section II.E	35

III. MATERIAL AND METHODS

A. Data

Lung sounds of twenty volunteers, seventeen patients with manifestations of adventitious sounds and three lung healthy subjects, were acquired mainly at the General Hospital of Thessaloniki 'G. Papanikolaou' and at the General Hospital of Imathia (Health Unit of Naoussa), Greece. The dataset includes crackles, wheezes and normal lung sounds. Fifteen patients exhibit manifestations of crackles or manifestations of crackle in conjugation with other adventitious sounds. The respiratory sounds of two patients contain adventitious sounds that are not crackles.

The acquisitions were performed using a 3M Littman electronic stethoscope (model 3200), at 4000 Hz, which complies with the EMC requirements of the IEC 60601-1-2.

The auscultation positions used in this study were selected among the six possible positions presented in Fig. 1. For each volunteer it was selected the data acquired from the two positions where the adventitious/normal sounds were better heard. A total of forty sound files, with approximated 30 seconds each, were used in this study. Each sound file was normalized.

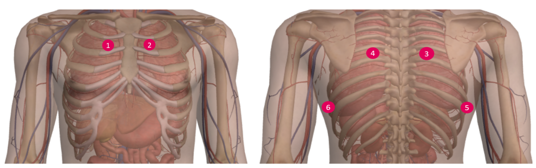

Figure 1. Potential positions for the acquisition of sounds (red). For each volunteer the data acquired from the two positions where the adventitious/normal sounds were better heard were selected.

Four hundred crackle events with duration of 587 ± 492 ms (mean ± std) were annotated by doctors in the temporal space.

The detection of the crackle events was done in the frame space. Frames were automatically annotated as containing or not containing crackles. Depending of the frame size the total number of the events considered as true events change. Neighborhood frames, with a maximum frame distance of 5 frames, were grouped and considered to belong to the same event. The ethical committee of the General Hospital of Thessaloniki 'G. Papanikolaou' authorized the data acquisition.

B. Detection of crackle events

In this study, the performance of a multi-feature approach to detect crackle events was studied. Thirty-five features were tested (see section II): the fractal dimension of the filter WPST–NST, WS-SS, thirty-one musical features, the Teager energy and the entropy of the sound.

For the automatic classification of the frames, as containing or not containing crackles, the logistic regression classifier was used. After the classification, neighborhood frames marked as containing crackles within a maximum frame distance were grouped and considered belonging to the same crackle event. After that, groups of frames with duration below or equal to a pre-defined threshold, the minimum group length, were discarded. Groups of frames with duration superior to 3 seconds were also discarded.

In this work the optimization of the detection parameters was also done. It was studied the influence of the frame size, the maximum frame distance (MFD), the minimum group length (MGL) and the decision threshold (DC) of the classifier as a function of the number of features (NF) used to classifier the data. A grid search (see Table 2) was done as a function of the number of features (NF). In this study the number of events considered as true events was equal to 357 and 326 for the frame size (FS) equal to 64 ms and 128 ms, respectively.

For the optimization of the number of features it was only considered the sequential combination of the thirty-five features. The rank of the features was done using the sequential feature selection in the forward direction taking into account the Matthews correlation coefficient. Using the data of all patients, a stratified 10-fold cross-validation approach with ten Monte Carlo repetitions was used. Although the datasets used in the training and testing are different they may contain frames acquired from the same patient. Each frame was classified as containing or not containing crackles.

TABLE 2. VALUES OF THE FRAME SIZE, THE MAXIMUM FRAME DISTANCE (MFD) , MINIMUM GROUP LENGTH (MGL) AND DECISION THRESHOLD (DT) USED IN THE GRID SEARCH DONE TO OPTIMIZE THE CLASSIFICATION PARAMETERS.

Frame size [ms]	MFD [frames]	MGL [frames]	DT [a.u.]
128	5	2	0.5
64	7	3	0.475
-	9	4	0.45
-	10	-	-

C. *Performance criteria*

A leave-one-out (volunteer) cross-validation approach was used to test the performance of the detector, i.e., data from nineteen volunteers were used to train the model and the data of the remaining volunteer was used to test the model. The cost function utilized to optimize the method parameters, the balanced F-score, is expressed by

$$c\{\mathrm{FS,NF,MFD,MGL,DC}\} = \tag{3}$$

$$< 2\frac{\mathrm{sen_E}\{\mathrm{p_c}\}\mathrm{ppv_E}\{\mathrm{p_c}\}}{\mathrm{sen_E}\{\mathrm{p_c}\}+\mathrm{ppv_E}\{\mathrm{p_c}\}} >_{\{\mathrm{FS,NF,MFD,MGL,DC}\}},\ \mathrm{p_c}=\{1,2,\ldots15\}$$

with,

$$\mathrm{sen_E} = \frac{\mathrm{TP_E}}{\mathrm{E}}, \tag{4}$$

$$\mathrm{ppv_E} = \frac{\mathrm{TP_E}}{\mathrm{TP_E}+\mathrm{FP_E}}, \tag{5}$$

where $\mathrm{TP_E}$ is the number of the true positives events, $\mathrm{FP_E}$ the number of false positives events and E the number of events for a given frame size. The $\mathrm{sen_E}$ and the $\mathrm{ppv_E}$ correspond to the sensitivity and to the positive predict value measured in the detection of crackle events, respectively. The $<a>$ notation stands for the mean value of the array of values a. An event was considered detected (true positive event) if at least a group of frames, that contains at least a part of the event, was classified as containing a crackle. If a crackle event was detected by n groups of frames, n-1 false positives events were count. If a group of frames classified as crackles contains more than one crackle event, only one crackle event was considered detected.

For evaluate the performance of the detection method for the volunteers without crackles it was measured the specificity given by

$$\mathrm{spe_F} = \frac{\mathrm{TN}}{\mathrm{TN+FP}}, \tag{6}$$

where the TN is the number of frames true negative and the FP the number of frames false positives.

IV. RESULTS

In Table 3 and Table 4 are presented the features ranked by importance when the frame size was equal to 128 ms and 64 ms, respectively. The feature number 32, inharmonicity, was discarded from this study due to the high number of frames with the value equal to Not-a-Number.

TABLE 3. RANK OF THE FEATURES SELECTED BY THE SEQUENTIAL FEATURE SELECTION IN THE FORWARD DIRECTION WHEN THE FRAME SIZE WAS EQUAL TO 128 MS.

#	1	2	3	4	5	6	7	8	9	10
Label	21	8	23	31	11	5	17	1	14	29
#	11	12	13	14	15	16	17	18	19	20
Label	15	7	33	30	19	25	9	35	16	2
#	21	22	23	24	25	26	27	28	29	30
Label	10	12	22	6	3	4	34	13	27	18
#	31	32	33	34	35	-	-	-	-	-
Label	28	26	24	20	32	-	-	-	-	-

In the top of Fig. 2 are presented the mean values of the $\mathrm{ppv_E}$ and of the $\mathrm{sen_E}$ as a function of the number of features, when the frame duration was equal to 128 ms. For a given number of features (NF) used in the classification, the detection parameterization which maximizes the criterion express by (3), with FS equal to 128 ms, was used to measure the performance criteria. The correspond value of the cost function and the mean of the $\mathrm{spe_F}$ measured in the data of the volunteers without crackles manifestations are also presented. In the bottom of the same figure is presented similar information when the frame size was equal to 64 ms. In the top of Fig. 3 is presented the standard deviation values (std) of the $\mathrm{ppv_E}$, $\mathrm{sen_E}$ and $\mathrm{spe_F}$ measured for the same detection results presented in the Fig. 2. In Fig. 4 and Fig. 5 are presented the spectrogram and the corresponding detection results of four acquisitions included in the dataset, respectively.

V. DISCUSSION AND CONCLUSION

The objective of this study was the development of a multi-feature method to detect crackle events. Instead of trying to detect individual crackles, the developed method aims to detect crackle events in the frame space. The performance of the sequential combination of thirty-five features were evaluated. Thirty-one of the features tested, for the detection of these high pitch sounds, are usually used in the context of music information retrieval. Due to the non-stationary nature of these sounds was also tested a wavelet based feature, a Teager energy based feature and an entropy-based feature.

A grid search was used to optimize the parameters of the detection method. For the conditions tested, the best results were obtained for the frame size equal to 128 ms (see Fig. 2 and Fig. 3) and twenty-seven features. Using the optimal configuration a ppv_E and a sen_E equals to 0.77±0.22 (mean ± std) and 0.76±0.23 were measured, respectively. For the same configuration the spe_F (measured in the data of the volunteers without crackles) was equal to 0.91±0.10. The great values measured for the standard deviation may be explained due to the small size of the dataset, the variability of the adventitious sounds between patients (see Fig. 4 and Fig. 5) and the presence of involuntary movement artifacts. The low values measured for the mean values of the sensitivity and of the positive predictive values may also be related to the variability of the crackles sounds between patients.

In the future, improvements of the developed method will be explored such as the introduction of a segmentation phase, to be done before the features extraction, and the exploration of non-linear classifiers. Since the vest that is being development under the WELCOME project [10] will allow the simultaneous acquisition of sound and Electrical Impedance Tomography (EIT), the integration of extra features extracted from the EIT, e.g., respiratory phase and the relative airflow, will be investigated.

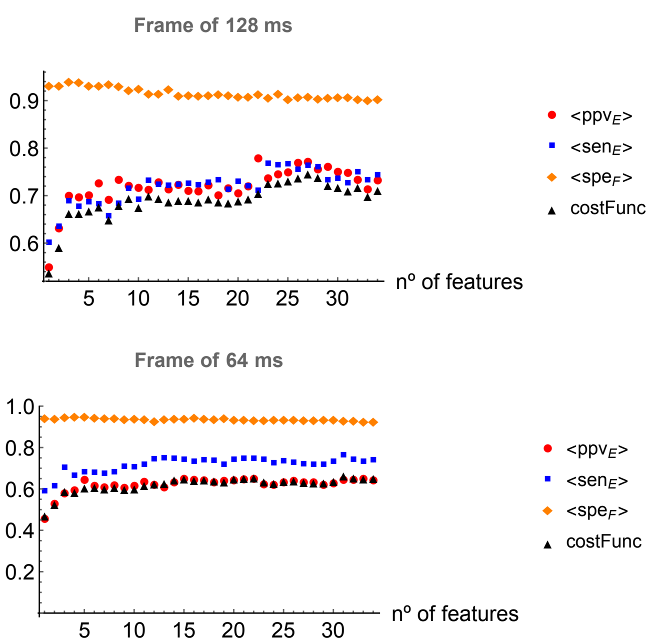

Figure 2. Performance measurements obtained for the different frame sizes tested. In the top is presented the results when the size frame was equals to 128 ms and in the bottom when was equal to 64 ms. The mean values of the ppv_E and of the sen_E, as a function of the number of features, are presented in red and blue, respectively. The mean values of the spe_F measured in the volunteers without crackles manifestation are also presented in orange. The detection parameterization which maximizes the cost function express by (3) was used to measure the performance criteria. The value of the cost function as a function of the number of iterations is presented in black.

TABLE 4. RANK OF THE FEATURES SELECTED BY THE SEQUENTIAL FEATURE SELECTION IN THE FORWARD DIRECTION WHEN THE FRAME SIZE WAS EQUAL TO 64 MS.

#	1	2	3	4	5	6	7	8	9	10
Label	21	8	23	29	4	31	11	16	9	33
#	11	12	13	14	15	16	17	18	19	20
Label	20	1	17	3	12	35	6	14	30	19
#	21	22	23	24	25	26	27	28	29	30
Label	2	18	28	7	24	13	15	10	34	22
#	31	32	33	34	35	-	-	-	-	-
Label	27	26	25	5	32	-	-	-	-	-

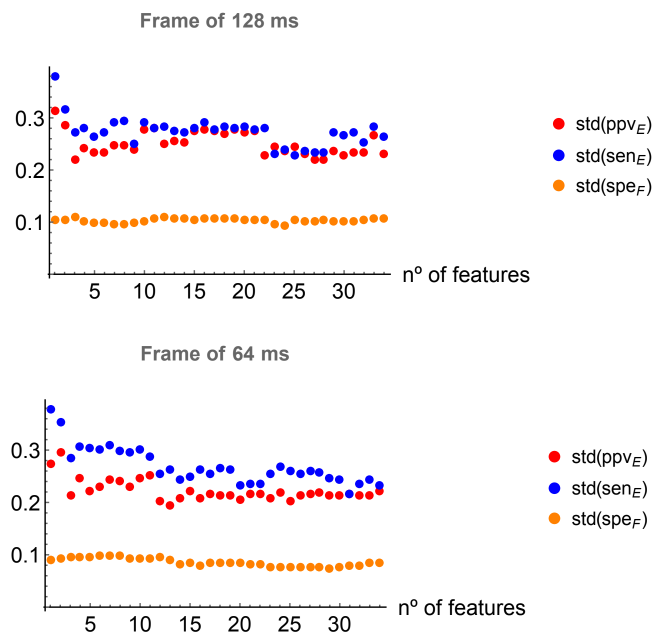

Figure 3. Performance measurements obtained when the frame size was equal to 128 ms (top) and 64 ms (bottom). The standart deviation values (std) of the the ppv_E and the sen_E, as a function of the number of features, are presented in red and blue, respectively. The std of the spe_F measured in the volunteers without crackles manifestation is also presented in orange. The detection parameterization which maximizes the criterion express by (3) was used to measure the performance criteria.

REFERENCES

[1] P. Piirilä and A. R. A. Sovijärvi, "Crackles: recording, analysis and clinical significance," *Eur. Respir. J.*, vol. 8, no. 12, Dec. 1995.

[2] X. Lu and M. Bahoura, "An integrated automated system for crackles extraction and classification," *Biomed. Signal Process. Control*, vol. 3, Jul. 2008.

[3] M. Bahoura and X. Lu, "Separation of crackles from vesicular sounds using wavelets packet transform," in *2006 Proc. Int. Conf. on Acoustics, Speech, and Signal*.

[4] L. J. Hadjileontiadis, "Empirical Mode Decomposition and Fractal Dimension Filter," *IEEE Eng. Med. Biol. Mag.*, Feb. 2007.

[5] P. A. Mastorocostas and J. B. Theocharis, "A dynamic fuzzy neural filter for separation of discontinuous adventitious sounds from vesicular sounds," *Comput. Biol. Med.*, vol. 37, Jan. 2007.

[6] S. Charleston-Villalobos, G. Martinez-Hernandez, R. Gonzalez-Camarena, G. Chi-Lem, J. G. Carrillo, and T. Aljama-Corrales, "Assessment of multichannel lung sounds parameterization for two-class classification in interstitial lung disease patients," *Comput. Biol. Med.*, vol. 41, Jul. 2011.

[7] L. Mendes et. al, "Detection of wheezes using their signature in the spectrogram space and musical features", 37th Annual International

Conference of the IEEE Engineering in Medicine and Biology Society, 2015.

[8] O. Lartillot and P. Toiviainen, "MIR in Matlab (II): A Toolbox for Musical Feature Extraction From Audio," *in Int. Conf. Mus. Inf. Ret.*, Vienna, 2007.

[9] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, Digital Image Processing Using Matlab. Gatesmark Publishing, 2004.

[10] "WELCOME Project" [Online]. Available: http://www.welcome-project.eu/about-the-project.aspx. [Accessed: 7-March-2016].

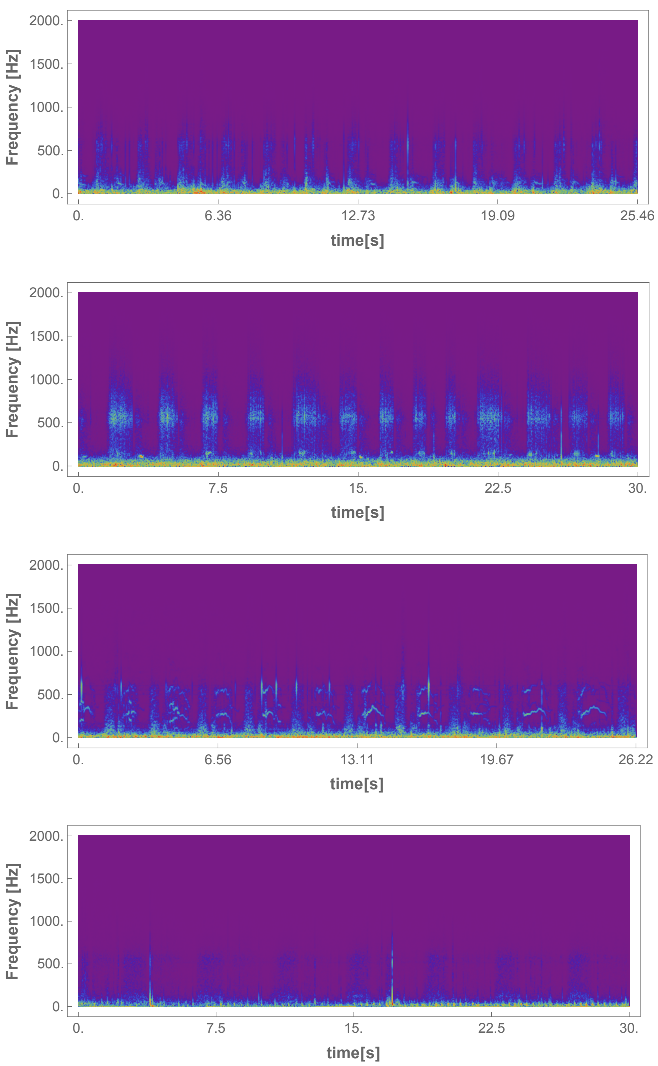

Figure 4. Spectrograms of four acquisitions included in the dataset.

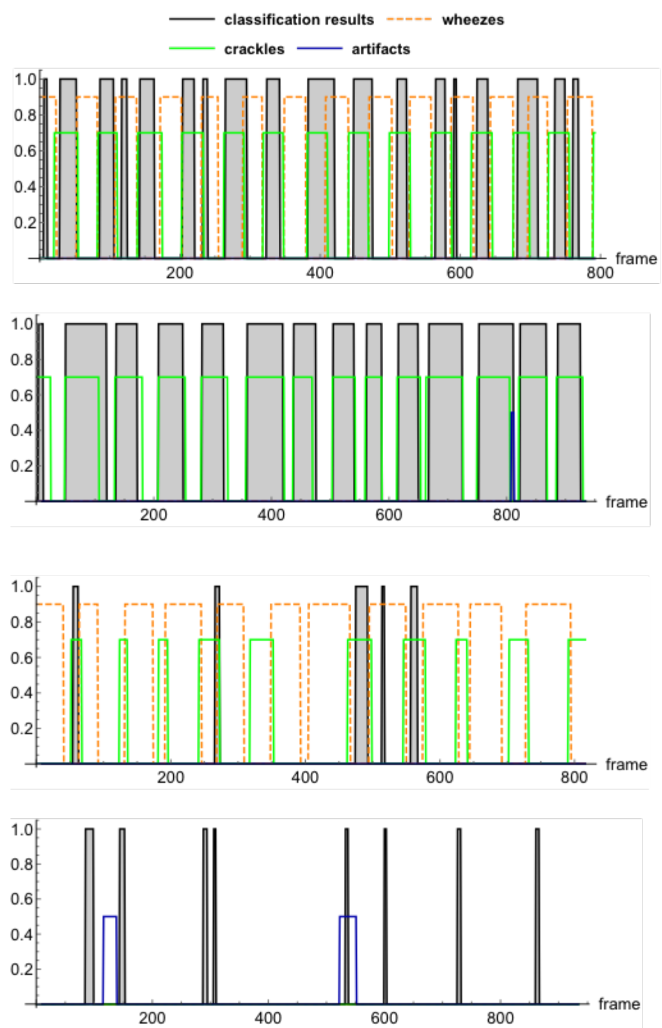

Figure 5. Four detection results obtained using the proposed method. The corresponding spectrograms are presented in Fig. 4. The crackle, wheeze and artifact events are presented in green, orange and blue, respectively. The gray bars correspond to group of frames classified as containing crackles.