Robust features for detection of crackles: an exploratory study

L. Mendes, P. Carvalho, C. A. Teixeira, R. P. Paiva and J. Henriques

***Abstract*— Crackles are adventitious and explosive respiratory sounds that can be classified as fine or coarse. These sounds are usually associated with cardiopulmonary diseases such as the chronic obstructive pulmonary disease. In this work seven different features were tested with the objective to identify the best subset of features that allows a robust detection of coarse crackles. Some of the features used in this study are new, namely those based on the local entropy, on the Teager energy and on the residual fit of a Generalized Autoregressive Conditional Heteroskedasticity process.**

The best features as a function of the number of features used in classification were identified having into account the Matthews correlation coefficient. The best individual feature was based on the local entropy. A significant improvement in the performance was obtained by using the feature based on local entropy and the feature based on the wavelet packed stationary transform – no stationary transform. The addition of more features only allows a smaller improvement.

I. INTRODUCTION

The analysis of the respiratory sounds is a valuable diagnostic tool for the detection and follow-up of respiratory diseases such as chronic obstructive pulmonary disease (COPD). The respiratory sounds can be classified as breath sounds, abnormal breath sounds and adventitious sounds [1]. Adventitious sounds refer to additional respiratory sounds superimposed on breath sounds. These sounds include wheezes (continuous sounds), stridors, squawks and crackles (discontinuous sounds). Crackles are short explosive sounds that seem to result from an abrupt opening or closing of the airways[2]. Crackles can usually be classified based on their total duration (2CD) as fine (<10 ms) or coarse (>10 ms) [3]. These sounds are associated with cardiopulmonary diseases and typically present a very characteristic waveform. The waveform of the crackle generally begins with a width deflection, followed by deflections with greater amplitude [2]. Several methods have been proposed for automatic detection of crackles: based on wavelets [4][5], on empirical mode decomposition method with Katz fractal dimension filter [6], on adaptive computing methods [7] and on autoregressive models [8].

The objective of this study is to identify the best features that allow the robust detection of crackles. In the future, the identified features will be used to monitor the presence of crackles in patients with COPD in different acquisition environments, namely in a telemonitoring system. This work was done under the FP7 European Project WELCOME [9] that aims to provide continuous monitoring for an integrated care to COPD patients with comorbidities. Some of the features used in this study are new, such as a feature based on the local entropy, a feature based on the Teager energy and a feature based on the Generalized Autoregressive Conditional Heteroskedasticity process. The performance of the feature(s) to discriminate segments of respiratory sounds with coarse crackles was studied by having into account the Matthews correlation coefficient [10]. The best combination of features was found by testing the performance of all the possible combination of features.

This work was supported by iCIS (CENTRO-07-ST24-FEDER-002003) and EU Project WELCOME (FP7 - 611223).

L. Mendes, P. Carvalho, C. Teixeira, R. P. Paiva, J. Henriques are with the Centre for Informatics and Systems of the University of Coimbra (CISUC), Pólo II, Coimbra, Portugal (e-mail: {lgmendes, carvalho, cteixei, ruipedro, jh} @dei.uc.pt).

II. THEORY

A. *Teager energy operator*

The Teager energy operator is a non-linear operator that is frequently used in signal and image processing applications, such as in the demodulation of AM-FM signals. In the continuous case the Teager Energy Operator, $\psi(.)$, [11] for real signals x(t) is defined as

$$\psi(\mathrm{x(t)}) = \left(\frac{\mathrm{dx}}{\mathrm{dt}}\right)^2 - \mathrm{x}\frac{\mathrm{d^2x}}{\mathrm{dt^2}}. \quad (1)$$

The discrete version is given by

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

with $n \in \mathbb{Z}$ and $t \in \mathbb{R}$.

B. WPST–NST

Bahoura and Lu [5] proposed a filter scheme based on the wavelet packed transform to separate the crackles from vesicular sounds. The proposed filter, the wavelet packed stationary transform – no stationary transform (WPST–NST), is a double thresholding non iterative method. In the formulation of the WPST–NST filter two assumptions are made: 1) the wavelet coefficients related with the crackles have larger coefficients when compared with the coefficients related with the vesicular sound, 2) the background related coefficients decrease to zero with the increase of scale. The first threshold is computed as

$$\mathrm{S1}_\mathrm{k}^\mathrm{j} = \mathrm{P_1}\sigma_\mathrm{k}^\mathrm{j}\,, \quad (3)$$

where σ_k^j is the standard deviation of coefficients corresponding to the k^{th} subband of the level j and $P_1 = 0.75$. The second threshold is computed as

$$S2^j = P_2 \frac{1}{N/2^j}\sum_{m=1}^{N/2^j}\sum_{k=1}^{2^j} M_k^j[m]\,, \quad (4)$$

with,

$$M_k^j[m] = \begin{cases} 1 & \text{if} \left|w_k^j[m]\right| \geq S1_k^j \\ 0 & \text{else} \end{cases}, \quad (5)$$

where $w_k^j[m]$ is the wavelet coefficient m of the k^{th} subband of the level j, N the signal length and $P_2 = 2$. The wavelet coefficients of the non-stationary part of the signal are the ones that satisfy the condition

$$\left|w_k^j[m]\right| \geq S1_k^j \text{ and } \sum_{k=1}^{2^j} M_k^j[m] \geq S2^j . \quad (6)$$

C. Fractal dimension

The fraction dimension (or capacity dimension of a fractal or Hausdorff dimension) is a measurement of the complexity of a given waveform. This measurement can be calculated by

$$D = \frac{\log(L)}{\log(d)}, \quad (7)$$

where L is the total length of the curve [12] and d the diameter of the curve. To avoid the dependence on the particular units of measure (in the discrete case) Katz [12] proposes a modification of (7). A study done by Raghavendra and Dutt [13] shows that the performance of the method proposed by Katz in the estimation of the fractal dimension of waveforms is poor. A more robust method was proposed by Higuchi [14]. Let $\{x[1], x[2], \dots, x[N]\}$ be a time series composed by N points. The first step to calculate the Higuchi FD is to construct u new time series, x_i^u, as

$$x_i^u = \left\{x[i], x[i+u], x[i+2u], \dots, x\left[i + \left\lfloor \frac{N-i}{u} \right\rfloor u\right]\right\}, \quad (8)$$

with $i = 1,2,\dots,u$, and $u = 1,2,\dots,u_{MAX}$, where i and u are integers indicating the initial time value and the discrete time interval between points, respectively. The u_{MAX} is the maximum value of u and $[a]$ is the integer part of a**.** Then for each time series x_i^u, the average length, $L_i[u]$, is computed as

$$L_i[u] = \frac{\sum_{h=1}^{\lfloor (N-i)/u \rfloor} |x[i+hu] - x[i+(h-1)u]|(N-1)}{\left\lfloor \frac{N-i}{u} \right\rfloor u}. \quad (9)$$

After that, the length of the time series for the time interval u, L[u], is computed as

$$L[u] = \sum_{i=1}^{u} L_i[u] . \quad (10)$$

Finally the FD Higuchi corresponds to slope of the curve $\ln(L[u])$ versus $\ln(1/u)$ estimated by a linear least squares fitting.

D. Empirical Mode Decomposition

The EMD method was developed by Huang *et al.* [15] and allows to adaptively decompose a (non-stationary) signal into a finite sum of S curves called intrinsic mode functions (IMF), i.e.,

$$x[n] = \sum_{f=1}^{S} c_f[n] + r_S , \quad (11)$$

where c_f denotes the f intrinsic mode function and r_S the residual. The IMF has two main characteristics: 1) the number of extremes and the number of zero-crossing must differ by at most one, 2) the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero at every point.

E. Entropy

The information entropy, H, is a measurement of the disorder of a system. A discrete random variable X with V possibles outcomes $\{a_1, a_2, \dots, a_V\}$ and associated probabilities $\{P(a_1), P(a_2), \dots, P(a_V)\}$ has an entropy equal to [16]

$$H = -\sum_{v=1}^{V} P(a_v) \log P(a_v). \quad (12)$$

The entropy of a signal quantized into V levels is given by [16]

$$\hat{H} = -\sum_{v=1}^{V} \frac{g_v}{N} \log \frac{g_v}{N}, \quad (13)$$

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

F. Generalized Autoregressive Conditional Heteroskedasticity (GARCH) Process

Time series models can be used to try to model the respiratory sounds. By definition, a process $\{X_n, n = 1,2,\dots\}$ is said to be an GARCH(p,q) if it is stationary and if [17]

$$X_n = \sigma_n Z_n , \qquad \{Z_n\} \sim \text{IID } N(0,1) , \quad (14)$$

where

$$\sigma_n^2 = \alpha_0 + \alpha_1 X_{n-1}^2 + \dots + \alpha_p X_{n-p}^2 + \beta_1 \sigma_{n-1}^2 + \dots + \beta_q \sigma_{n-q}^2 \quad (15)$$

and $\alpha_0 > 0$, $\alpha_j \geq 0$ for $j = 1, \dots, p$, $\beta_k \geq 0$ for $k = 1, \dots, q$ and if Z_n and $X_{n-1}, X_{n-2}, \dots$ are independent for all n. *IID* stands for independent and identical distributed and $N(0,1)$ refers to a normal distribution with zero mean and variance equal to 1.

III. Material and Methods

A. Data

In this study two datasets available on-line [18] (first channel of the repository "Crackle (a)" and the repository "Crackle (c)") of respiratory sounds containing crackles were used. The two datasets (~10 seconds each) with a sampling rate of 11025 Hz were subdivided, with an overlap factor of 75%, into segments with a length of 2048. The segments were manually classified as containing coarse crackles or not containing coarse crackles. This is the expected type of crackles to be found in patients with COPD [2].

B. Data processing

Different processing algorithms were applied to each segment of data. The first one was the Teager energy operator (ψ). The local entropy (H_L), i.e., the information entropy (see (13)) , was calculated for the values of the signal blocks with length of 131 points centered in each point. A global quantization of the data into 6 levels was applied before the calculation of the H_L. The third one was the local Higuchi FD of non-stationary signal part of the WPNST-ST filter (FD-NST). The Higuchi FD was calculated for the values of the signal blocks with a length of 101 points centered in each point. For the WPNST-ST filter we used a Daubechies-8 wavelet for a 5th level decomposition tree. The fourth and the fifth processing algorithms were, respectively, the local Higuchi FD ($FD-EMD$) and the Teager energy ($\psi-EMD$) of the signal reconstructed using only the first $\mathcal{L}$ IMFs. The number of IMFs , $\mathcal{L}$, were selected using the criterion proposed by Hadjileontiadis [6]:

$$\mathcal{L} = \min\{\lambda: |\eta'_\lambda| > p \wedge |\eta'_{\lambda+1}| \leq p \wedge \eta''_\lambda > 0\} \qquad (16)$$

with

$$\eta_\lambda = 1 - \frac{\sum_{f=1}^{\lambda} E\{{c_f}^2\}}{\sum_{f=1}^{S} E\{{c_f}^2\}}, \qquad \lambda = 1,2,\dots,S, \qquad (17)$$

where η'_λ and η''_λ represent the first and second derivate of η_λ with respect to λ, p=0.05, and $E\{a\}$ denotes the expected value of a. Finally, the local Higuchi FD of the fit residues of best GARCH model (FD-GARCH) that fit the data was also calculated. The order of the model was chosen using the Akaike Information Criterion over a set of 8 candidates. Prior to the estimation of the GARCH model a median filter was applied to the data.

C. Features

For each segment 7 features were computed. We began by subdividing the segment of sound, with an overlap factor of 50%, into $Nsub$ =7 sub-segments with 512 of length. For each sub-segment of a given segment *seg*, x_{seg}^{sub} (with $sub = 1,2,\dots,Nsub$), the most appropriated quartile (the lower ($Q1$) or the upper ($Q3$)) of the result of the different processing algorithms (including the amplitude, Amp) was computed. The features computed for each segment of sound correspond to the appropriated extreme value (maximum (Max) or minimum (Min) value) of the set composed of the quartiles values of the sub-segments. In summary the 7 features computed for the segment *seg* were:

1. $Q3\text{–}\psi_{\mathrm{seg}} =$
$$\mathrm{Max}\{\mathrm{Q3}(\psi_{\mathrm{seg}}^{1}), \mathrm{Q3}(\psi_{\mathrm{seg}}^{2}), \dots, \mathrm{Q3}(\psi_{\mathrm{seg}}^{\mathrm{Nsub}})\}$$
2. $\mathrm{Q1-FD-NST_{seg}} =$
$$\mathrm{Min}\left\{\begin{matrix}\mathrm{Q1}(\mathrm{FD-NST_{seg}^{1}}), \mathrm{Q1}(\mathrm{FD-NST_{seg}^{2}}), \dots, \\ \mathrm{Q1}(\mathrm{FD-NST_{seg}^{Nsub}})\end{matrix}\right\}$$
3. $\mathrm{Q1-FD-EMD_{seg}} =$
$$\mathrm{Min}\left\{\begin{matrix}\mathrm{Q1}(\mathrm{FD-EMD_{seg}^{1}}), \mathrm{Q1}(\mathrm{FD-EMD_{seg}^{2}}), \dots, \\ \mathrm{Q1}(\mathrm{FD-EMD_{seg}^{Nsub}})\end{matrix}\right\}$$
4. $\mathrm{Q3}-\psi-\mathrm{EMD_{seg}} =$
$$\mathrm{Max}\left\{\begin{matrix}\mathrm{Q3}(\psi-\mathrm{EMD_{seg}^{1}}), \mathrm{Q3}(\psi-\mathrm{EMD_{seg}^{2}}), \dots, \\ \mathrm{Q3}(\psi-\mathrm{EMD_{seg}^{Nsub}})\end{matrix}\right\}$$
5. $\mathrm{Q1-FD-GARCH_{seg}} =$
$$\mathrm{Min}\left\{\begin{matrix}\mathrm{Q1}(\mathrm{FD-GARCH_{seg}^{1}}), \mathrm{Q1}(\mathrm{FD-GARCH_{seg}^{2}}), \\ \dots, \mathrm{Q1}(\mathrm{FD-GARCH_{seg}^{Nsub}})\end{matrix}\right\}$$
6. $\mathrm{Q3-H_{L_{seg}}} =$
$$\mathrm{Max}\{\mathrm{Q3}(\mathrm{H_{L_{seg}}}^{1}), \mathrm{Q3}(\mathrm{H_{L_{seg}}}^{2}), \dots, \mathrm{Q3}(\mathrm{H_{L_{seg}}}^{\mathrm{Nsub}})\}$$
7. $\mathrm{Q3-Amp_{seg}} =$
$$\mathrm{Max}\{\mathrm{Q3}(\mathrm{Amp_{seg}^{1}}), \mathrm{Q3}(\mathrm{Amp_{seg}^{2}}), \dots, \mathrm{Q3}(\mathrm{Amp_{seg}^{Nsub}})\}$$

D. Performance criteria

The Matthews correlation coefficient (MCC), measured after classifying the data using the Logistic Regression classifier [19], was used to assert the capacity of the different features (or combination of features) to detect segments with coarse crackles. This coefficient is calculated by

$$\mathrm{MCC} = \frac{\mathrm{TP.TN - FP.FN}}{\sqrt{(\mathrm{TP+FP})(\mathrm{TP+FN})(\mathrm{TN+FP})(\mathrm{TN+FN})}}, \qquad (18)$$

where TP , TN, FP and FN correspond , respectively, to the true positives, true negatives, false positives and false negatives.

Due to the reduced size of the datasets, a leave-one-out cross-validation approach was used to assert the performance of the classification. In order to identify the best subset of features, all possible combinations were tested.

IV. Results

Fig. 1 and Fig. 2 show examples of segments of respiratory sounds used in this study with and without coarse crackles, respectively. Fig. 3 presents a representative example of the result of the application of the different processing algorithms to a segment of sound. Table 1 presents the feature(s) that obtain(s) the higher value of the MCC as a function of the number of features (subset size) used in the classification.

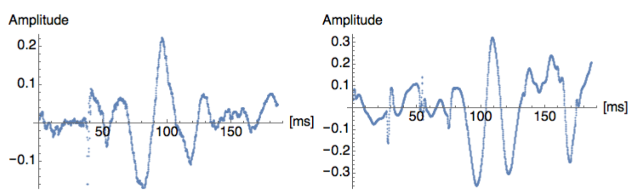

Figure 1. Two examples of segments of the respiratory sound processed that contain coarse crackles.

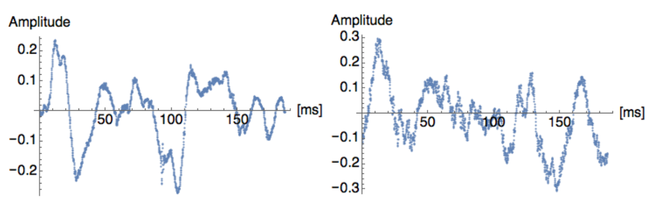

Figure 2. Two examples of segments of the respiratory sound processed without crackles.

V. DISCUSSION AND CONCLUSION

In this study we try to identify the best subset of features that allows a robust detection of coarse crackles. Although the maximum amplitude of the crackles tends to be higher than the maximum amplitude of the breath sounds, we could find several situations where this condition is not satisfied (see Fig. 1 and Fig. 2). As we can see in Fig. 3, the different processing algorithms provide by themselves some ability to discriminate between crackles and breath sounds.

The H_L tends to be higher in the presence of crackles and as we expected the values of the FD-NST and the FD-EMD are lower in this region. Please note that the results presented in [6] use Katz FD instead of the Higuchi FD. In that case we expect higher values for the FD-NST and the FD-EMD in the presence of crackles. GARCH models allow to model stationary signals with non-constant conditional variance. In the presence of breath sounds the fit residues of the GARCH model tends to be lower when compared with the fit residues in presence of coarse crackles. As we can see in Fig. 3 the FD-GARCH is lower in the region with the crackle. In the same figure we can also notice that Teager energy (ψ) and ψ-EMD have higher values in the presence of crackles.

To improve the robustness against outliers and improve the performance of the detection of crackles, we subdivided the segment and we calculated the appropriated quartile for the different processing algorithms. The best individual feature was the Q3-H_L. A significant improvement in the performance of the classification was obtained by using two features (Q3-H_L, Q1-FD-NST). The addition of more features only allows a smaller improvement. When all the features were used, the performance of the classification slightly decreased. Since the number and size of the datasets used in this study was very small, in the near future, under the project WELCOME [9], an extensive amount of respiratory sounds will be acquired and used to validate these results.

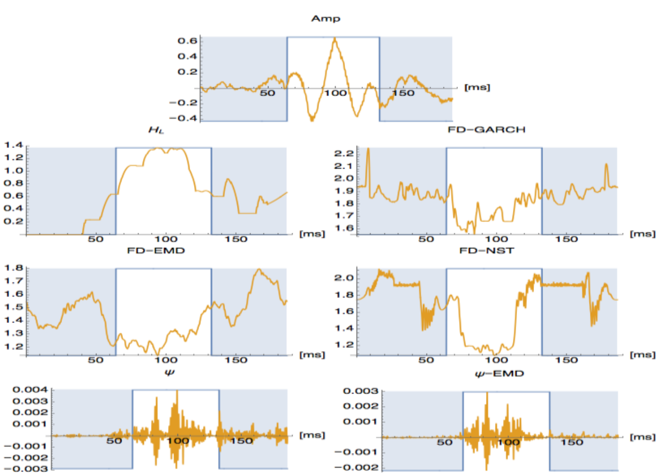

Figure 3. Example of the result of the different processing algorithms that were applied to each segment of sound. The subfigures a,b,c,d,e,f and g correspond, respectively, to the amplitude, local entropy, Higutich FD of the fit residuals of the GARCH model, Higutich FD of the EMD filter, Higutich FD of the non-stationary signal of the WPNST-ST, Teager energy and Teager energy of the EMD filter.

TABLE I. BEST FEATURES AND THE RESPECTIVE VALUE OF MCC AS A FUNCTION OF THE NUMBER OF FEATURES USED IN THE CLASSIFICATION.

Nº of features	*Best Feature(s)*	*MCC*
1	$Q3-H_L$	0.59
2	$\{Q3-H_L,\ Q1-FD-NST\}$	0.75
3	$\{Q3-H_L,\ Q1-FD-NST,$ $Q1-FD-GARCH\}$	0.77
4	$\{Q3-H_L,\ Q1-FD-NST,$ $Q1-FD-EMD,$ $Q3-\psi-EMD\}$	0.79
5	$\{Q3-H_L, Q1-FD-NST,$ $Q1-FD-GARCH,\ Q3-\psi,$ $Q1-FD-EMD\}$	0.79
6	$\{Q3-H_L,\ Q1-FD-GARCH,$ $Q1-FD-NST,$ $Q3-\psi-EMD,$ $Q3-\psi,\ Q1-FD-EMD\}$	0.80
7	All	0.78

REFERENCES

[1] A. R. A. Sovijärvi, L. P. Malmberg, G. Charbonneau, J. Vanderschoot, F. Dalmasso, C. Sacco, M. Rossi, and J. E. Earis, “Characteristics of breath sounds and adventitious respiratory sounds,” *Eur. Respir. Rev*., vol. 10, no. 77, 2000.

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

[3] A. R. A. Sovijärvi, F. Dalmasso, J. Vanderschoot, L. P. Malmberg, and G. Righini, “Definition of terms for applications of respiratory sounds,” *Eur. Respir. Rev.*, vol. 10, no. 77, 2000.

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

[5] 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*.

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

[7] 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.

[8] 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.

[9] “WELCOME Project” [Online]. Available: http://www.welcome-project.eu/about-the-project.aspx. [Accessed: 10-Mar-2014].

[10] G. Jurman, S. Riccadonna, and C. Furlanello, “A comparison of MCC and CEN error measures in multi-class prediction,” *PLoS One*, vol. 7, 2012.

[11] E. Kvedalen, “Signal processing using the Teager energy operator and other nonlinear operators,” Master Thesis, Dep. Informatics, Univ. Oslo, Norway, 2003.

[12] M. J. Katz, “Fractals and the analysis of waveforms.,” *Comput. Biol. Med.*, vol. 18, no. 3, Jan. 1988.

[13] Signals using Multiresolution Box-counting Method,” *Int. J. Inf. Mat. Sci.*, vol. 6, 2010.

[14] T. Higuchi, “Approach to an irregular time series on the basis of the fractal theory”, *Phys. D*, vol 31, June 1988.

[15] N. E. Huang, Z. Shen, S. R. Long, M C. Wu, H. H. Shih, Q. Zheng, N. Yen, C. C Tung and H. H. Liu, “The empirical mode decomposition and the Hilbert spectrum for non-linear and non-stationary time series analysis," *Proc. Roy. Soc. London*, vol. 454 , Mar. 1998.

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

[17] “Time series analysis.” [Online]. Available: http://www.math.kth.se/matstat/gru/sf2943/ts.pdf. [Accessed: 07-Mar-2014].

[18] “R.A.L.E. Repository” [Online]. Available: http://www.rale.ca. [Accessed: 10-Mar-2014].

[19] I. H. Witten and E. Frank, *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann Publishers, 2005.