Classification of Sounds Indicative of Respiratory Diseases

Stavros Ntalampiras[1] and Ilyas Potamitis[2](✉)

[1] University of Milan, Milan, Italy
stavros.ntalampiras@unimi.it
[2] Technological Educational Institute of Crete, Rethymno, Greece
potamitis@staff.teicrete.gr

Abstract. This work presents a system achieving classification of respiratory sounds directly related to various diseases of the human respiratory system, such as asthma, COPD, and pneumonia. We designed a feature set based on wavelet packet analysis characterizing data coming from four sound classes, i.e. *crack*, *wheeze*, *normal*, *crack+wheeze*. Subsequently, the captured temporal patterns are learned by hidden Markov models (HMMs). Finally, classification is achieved via a directed acyclic graph scheme limiting the problem space while based on decisions made by the available HMMs. Thorough experiments following a well-established protocol demonstrate the efficacy of the proposed solution.

Keywords: Respiratory sound classification · Acoustic signal processing · Respiratory diseases

1 Introduction

It is widely accepted that diseases of the human respiratory system, such as asthma, COPD, and pneumonia are associated with distinctive acoustic patterns [13]. This is due to the abnormalities they cause in the airway path. Typically, a medical expert is able to correctly identify such patterns (e.g. by means of a stethoscope) and subsequently propose the corresponding treatment. However, this process relies both on the availability of an expert as well as their degree of expertise. Thus, the need to automatize the diagnosis process has arose in the last years igniting the development of such algorithms.

Even though during the last decade there has been a significant amount of research in this direction, a standardized way to compare the existing solutions is yet to appear. Systematic reviews of the state of the art are available in [13,17]. A great variety of temporal, spectral and wavelet features along with generative and non-generative classifiers have been employed in the literature. Such a review is beyond the scope of the present article; we rather focus on a recent standardized attempt approaching this problem. More in detail, the challenge organized within the International Conference on Biomedical Health Informatics in 2017 provides

© Springer Nature Switzerland AG 2019
J. Macintyre et al. (Eds.): EANN 2019, CCIS 1000, pp. 93–103, 2019.
https://doi.org/10.1007/978-3-030-20257-6_8

a dataset characterizing the properties of the classes of interest as well as an experimental protocol allowing the extraction of comparable results.

So far, two solutions employing the challenge's experimental protocol stand out. The first one [4] uses hidden Markov models fed with mel-frequency cepstral coefficients. The second [19] employs non-linear spectral features along with a support vector machine with a radial basis function kernel.

This work builds on the existing findings and proposes the usage of features derived from the wavelet domain, the distribution of which is learned by a directed acyclic graph scheme composed of hidden Markov models. More precisely, we designed a three-level wavelet packet band-based analysis component able to capture the behavior of the involved sound events within various spectral bands. Subsequently, we construct a directed acyclic graph limiting the problem space into a series of binary classifiers, each one relying on a pair of hidden Markov models. On top of that, we provide a solution to the topological ordering of such a graph. Experimental results using the protocol of the ICBHI 2017 challenge demonstrate the efficacy of the proposed approach.

The rest of this work is organized as follows: Sect. 2 describes the wavelet packet feature extraction module, while Sect. 3 explains the graph-based classification scheme. Section 4 details the employed dataset, the parameterization of the proposed approach as well as the achieved results and how these compare to existing solutions. Finally, in Sect. 5 we draw our conclusions.

2 The Feature Set

This section introduces the usage of band-based multiresolution analysis for automated respiratory sound classification. Lately, digital signal processing using wavelets has become a common tool in various research areas with heterogeneous needs. Such cases refer to enhancement of biological signals [11,15], geophysics applications like tropical convection, dispersion of ocean waves, etc. [21], speech/music discrimination [12], emotion prediction [10], farm monitoring [7], voice activity detection [2], moving vehicle classification [8], audio fingerprinting [1], generalized sound recognition [6], to name but a few.

The fundamental property of the Fourier transform is the usage of sinusoids with infinite duration. Whereas sinusoids functions are smooth and predictable, wavelets tend to be irregular and asymmetric. The main advantage of the wavelet transform is that it can analyze at many different frequencies time series characterized by non-stationary power. They comprise a dynamic windowing technique which can treat with different precision low and high frequency information. The first step of the wavelet packet analysis is the choice of the original (or mother) wavelet and by utilizing this function, the transformation breaks up the signal into shifted and scaled versions of it. In this paper we utilized Daubechies 1 (or Haar) function as the original wavelet while its optimal choice will be a subject of our future work. Unlike discrete wavelet transform (DWT), when wavelet packets (WP) are employed both low and high frequencies coefficients are kept. In our case the DWT is applied three subsequent times and consists of three-stage filtering of the audio signals as we can see in Fig. 1.

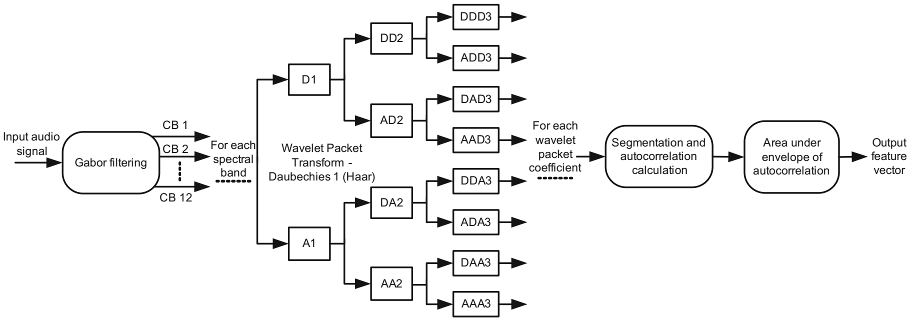

Input audio signal
Gabor filtering
CB 1
CB 2
CB 12
For each spectral band
Wavelet Packet Transform - Daubechies 1 (Haar)
D1
A1
DD2
AD2
DA2
AA2
DDD3
ADD3
DAD3
AAD3
DDA3
ADA3
DAA3
AAA3
For each wavelet packet coefficient
Segmentation and autocorrelation calculation
Area under envelope of autocorrelation
Output feature vector

Fig. 1. The block diagram of proposed feature extraction module. Audio signals are filtered and each spectral band is analyzed by a 3-level wavelet packet transform. After segmenting and computing the area under the autocorrelation envelope, we obtain the feature vector.

Table 1. The frequency limits used for perceptual wavelet packet integration analysis.

Band ID	Lower (Hz)	Center (Hz)	Upper (Hz)
1	0	125	250
2	250	375	500
3	500	625	750
4	750	875	1000
5	1000	1125	1250
6	1250	1375	1500
7	1500	1625	1750
8	1750	1875	2000
9	2000	2250	2500
10	2500	2750	3000
11	3000	3250	3500
12	3500	3750	4000

The idea behind the specific set is the production of a vector that provides a complete analysis of the audio signal across different spectral areas while they are approximated by WP. We should also take into account that respiratory signals do not distribute their energy across the spectrum in a homogeneous way. Thus, a fine partitioning of the spectrum could offer relevant distinctive information. Based on this observation, we designed a band-based signal analysis with the frequency ranges denoted in Table 1. Such a division is achieved by Gabor bandpass filters. Subsequently, three-level wavelet packets are extracted out of each spectral band. The specific level is able to provide detailed information regarding the signal characteristics at a specific band. Downsampling is applied on each coefficient at each stage in order not to end up having the double amount of data, as Nyquist theorem requests. The wavelet coefficients are then segmented and the autocorrelation envelope area is computed and normalized by half the segment size. M normalized integration parameters are calculated for each frame, where M is the total number of the frequency bands multiplied by the number of the wavelet coefficients. This series of parameters comprises the WP-Integration feature vector and the entire calculation process is depicted in Fig. 1.

WP-Integration parameters reflect upon the degree of variability of a specific wavelet coefficient within a frequency band. Since the audio signals we try to classify exhibit great differences among these bands, we decided to utilize the normalized autocorrelation envelope area.

3 The Classification Scheme

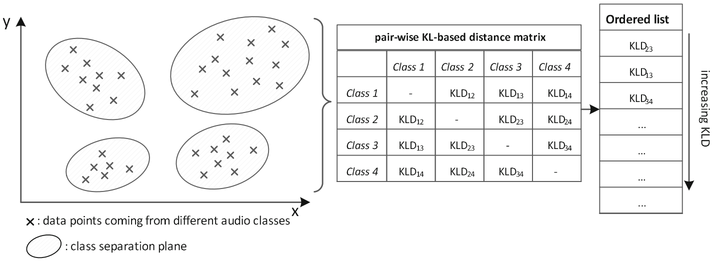
The proposed framework relies on the Directed Acyclic Graph (DAG) logic [9], i.e. the classification scheme is a graph denoted as $\mathcal{G} = \{N, L\}$, where $N = \{n_1, \ldots, n_m\}$ represents the nodes and $L = \{l_1, \ldots, l_p\}$ the links associating the nodes. Each node in N is responsible for a binary classification task conducted via a set of HMMs which fit well the specifications of audio pattern recognition tasks, thus the DAG-HMM notation.

The motivation behind creating such a graph-based classification system is that in this way, one is able to limit the problem space and design classification algorithms for two mutually-exclusive classes than having to deal with the entirety of the different classes at the same time. Essentially, the proposed methodology breaks down any m-class classification problem into a series of 2-class classification problems.

DAGs can be seen as a generalization of the class of Decision Trees, while the redundancies and repetitions that may occur in different branches of the tree can be observed more efficiently since different decision paths might be merged. In addition, DAGs are able to collect and conduct a series of tasks in an ordered manner, subject to constraints that certain tasks must be performed earlier than others. The sequential execution of tasks is particularly important and directly related to the efficacy with which the overall task is addressed [22].

The DAG-HMM architecture used in this paper includes $m(m-1)/2$ nodes (m being the total number of classes) each one associated with a two-class classification problem. The connections between the different nodes in $\mathcal{G}$ have only one orientation without any kind of loop(s). As a result, each node of a such a so-called *rooted* DAG has either 0 or 2 leaving arcs.

The principal issue associated with the design of every DAG is the *topological ordering*, i.e. ordering the nodes in a way that the starting endpoints of every edge occur earlier than the corresponding ending endpoints. In the following, we describe how such a topological ordering is discovered based on the Kullback-Leibler divergence.


y
x
× : data points coming from different audio classes
◯ : class separation plane
Ordered list
increasing KLD

pair-wise KL-based distance matrix

	Class 1	Class 2	Class 3	Class 4
Class 1	-	KLD_{12}	KLD_{13}	KLD_{14}
Class 2	KLD_{12}	-	KLD_{23}	KLD_{24}
Class 3	KLD_{13}	KLD_{23}	-	KLD_{34}
Class 4	KLD_{14}	KLD_{24}	KLD_{34}	-

Ordered list
KLD_{23}
KLD_{13}
KLD_{34}
...
...
...
...

Fig. 2. Determination of the topological ordering.

3.1 Determining the Topological Ordering of the DAG-HMM

Naturally, one would expect that the performance of the DAG-HMM depends on the order in which the different classification tasks are conducted. This was also evident from early experiments. This observation motivated the construction of the DAG-HMM so that "simple" tasks are executed earlier in the graph. In other words, these are placed in the top nodes of the DAG-HMM, in a way that classes responsible for a high amount of misclassifications are discarded early in the graph operation. In order to get an early indication of the degree of difficulty of a classification task, we employed the metric representing the distance of the involved classes in the probabilistic space, i.e. the Kullback-Leibler Divergence (KLD) between per-class GMMs in the feature space. The basic motivation is to place early in the DAG-HMM tasks concerning the classification of classes with large KLD, as they could be completed with high accuracy. The scheme determining the topological ordering is illustrated in Fig. 2.

The KLD between two J-dimensional probability distributions A and B is defined as [20]:

$$D(A||B) = \int_{R^J} p(X|A) log \frac{p(X|A)}{p(X|B)} dx \tag{1}$$

KLD provides an indication of how distant two models are in the probabilistic space. It is important to note that KLD as given in Eq. 1 comprises an asymmetric quantity. The symmetrical form can be inferred by simply adding the integrals in both directions, i.e.

$$D_s(A||B) = D(A||B) + D(B||A) \tag{2}$$

In the special case where both A and B are Gaussian mixture models KLD can be defined as follows:

$$KLD(A||B) = \int A(x) log \frac{B(x)}{A(x)} dx \tag{3}$$

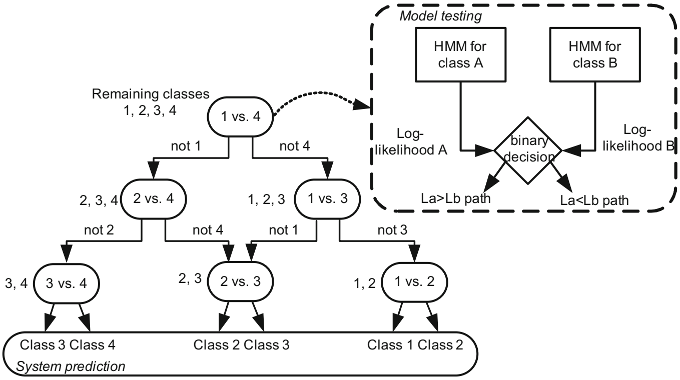


Fig. 3. An example of a DAG-HMM addressing a problem with four classes.

Unfortunately, there is not a closed-form solution for Eq. 3, thus we employed the empirical mean as follows

$$KLD(A||B) \approx \frac{1}{n} \sum_{i=1}^{n} log \frac{B(x_i)}{A(x_i)} \tag{4}$$

given that the number of Monte Carlo draws is sufficiently large. During our experiments we set $n = 2000$.

It should be noted the KLD between HMMs was not used since computing distances between HMMs of unequal lengths, which might be common in this work as HMMs representing different classes might have different number of states, can be significantly more computationally demanding without a corresponding gain in modeling accuracy [5,23].

After computing the KLD for the different pairs of classes, i.e. reach the second stage depicted in Fig. 2, the KLD distances are sorted in a decreasing manner. This way the topological ordering of the DAG-HMM is revealed, placing the classification tasks of low difficulty on its top. Each node removes a class from the candidate list until there is only one class left, which comprises the DAG-HMM prediction. The distance matrix elements could be seen as early performance indicators of the task carried out by the corresponding node. The proposed topological ordering places tasks likely to produce misclassifications at the bottom of the graph. This process outputs a *unique* solution for the topological sorting problem, as it is usually met in the graph theory literature [3].

3.2 The DAG-HMM Operation

The operation of the proposed DAG-HMM scheme is the following: after extracting the features of the unknown audio signal, the first/root node is activated. More precisely, the feature sequence is fed to the HMMs, which produce two log-likelihoods showing the degree of resemblance between the training data of each HMM and the unknown one. These are compared and the graph flow continues on the larger log-likelihood path. It should be stressed out that the HMMs are optimized (in terms of number of states and Gaussian components) so that they address the task of each node optimally. That said, it is possible that a specific class is represented by HMMs with different parameters when it comes to different nodes of the DAG-HMM.

An example of a DAG-HMM addressing a problem with four classes is illustrated in Fig. 3. The remaining classes for testing are mentioned beside each node. Digging inside each node, Fig. 3 also shows the HMM-based sound classifier responsible for activating the path of the maximum log-likelihood.

The operation of the DAG-HMM may be parallelized with that of investigating a list of classes, where each level eliminates one class from the list. More in detail, in the beginning the list includes all the potential audio classes. At each node the feature sequence is matched against the respective HMMs and the model with the lowest log-likelihood is erased from the list, while the DAG-HMM proceeds to the part of the topology without the discarded class. This process terminates when only one class remains in the list, which comprises the system's prediction. Hence, in case the problem deals with m different classes, the DAG's decision will be made after the evaluation of $m - 1$ nodes.

4 Experiments

This section explains *(a)* the dataset, *(b)* the parameterization of the proposed solution for classification of respiratory sounds, and *(c)* finally presents and analyses the achieved results.

4.1 Dataset

The respiratory sound database comes from the challenge organized within the International Conference on Biomedical Health Informatics in 2017 and it is publicly available[1]. The recordings span over several years. The database has a total duration of 5.5 h and contains 6898 respiratory cycles, of which 1864 contain crackles, 886 contain wheezes, and 506 contain both crackles and wheezes, in 920 annotated audio samples coming from 126 subjects.

The cycles were annotated by respiratory experts as including *crackles*, *wheezes*, a *combination* of them, or *no adventitious* respiratory sounds. The recordings were collected using heterogeneous equipment and their duration ranges from 10 s to 90 s. In addition, noise levels in some respiration cycles is high, representing very well, real life conditions. Finally, training and testing data are already defined by the challenge organization committee. More information regarding the dataset is available in [18].

[1] https://bhichallenge.med.auth.gr/.

Table 2. The recognition rates for the proposed and contrasted methods.

Approach	Recognition rate (%)
HMMs+MFCCs [4]	39.5
Non-linear spectral features+SVM [19]	49.8
DAG-HMM + WP-Integration	**50.1**

4.2 System Parameterization

The low-level feature extraction window is 30 ms with 10 ms overlap between subsequent windows, while the Daubechies mother wavelet was selected. The HMMs of each node are optimized in terms of number of states and nodes following the Expectation-Maximization and Baum Welch algorithms [14]. As the considered sound events are characterized by a distinct time evolution, we employed HMMs with left-right topology, i.e. only left to right states transitions are permitted. Moreover, the distribution of each state is approximated by a Gaussian mixture model of diagonal covariance, which may be equally effective to a full one at a much lower computational cost [16].

The maximum number of k-means iterations for cluster initialization was set to 50 while the Baum-Welch algorithm used to estimate the transition matrix was bounded to 25 iterations with a threshold of 0.001 between subsequent iterations. The number of explored states ranges from 3 to 7 while the number of Gaussian components used to build the GMM belongs to the {2, 4, 8, 16, 32, 64, 128, 256, and 512} set. The final parameters were selected based on the maximum recognition rate criterion. The machine learning package Torch[2] was used to construct and evaluate GMMs and HMMs.

4.3 Results

Table 2 depicts the rates achieved by two contrasted approaches as well as the proposed one. We observe that the solution based on DAG-HMM fed with PWP-Integration feature set achieved the highest recognition rate which is equal to 50.1%. Interestingly the inferred topological order suggested the execution of classification tasks with the following order: *(a)* crack+wheeze vs. normal, *(b)* normal vs. wheeze, *(c)* crack vs. crack+wheeze, *(d)* crack vs. wheeze, *(e)* crack+wheeze vs. wheeze, and *(f)* crack vs. normal.

Towards a more detailed picture of its classification capabilities, Table 3 tabulates the confusion matrix. As we can see, the class identified with the highest accuracy is the *wheeze* one with 64.5% and second is the *normal* one with 63%. On the contrary, *crack* sound events were the most misclassified ones with the respective rate being 36.7%.

Even though the achieved rate is the highest one reported in the literature, it is still far from satisfactory. Interestingly, when samples from *crack*,

[2] Freely available at http://torch.ch/.

Table 3. The confusion matrix (in %) achieved by the proposed approach. The average recognition rate is 50.1%.

Presented	Responded			
	crack	*crack+wheeze*	*normal*	*wheeze*
crack	36.7	3.1	58.5	1.6
crack+wheeze	3.1	38.4	57.9	0.6
normal	32.4	3.3	63	1.3
wheeze	0.5	3.2	31.8	64.5

crack+wheeze, and *wheeze* are misclassified, they are identified as *normal* at most cases. This indicates that the patterns exhibited by the data coming form the *normal* class are similar to all other classes. A potential solution to this problem is collecting more data form the *non-normal* classes, such that the differences are highlighted.

5 Conclusions

This work explained a graph-based classification scheme encompassing wavelet analysis and temporal modeling. Such an approach was able to surpass solutions existing in the literature. Nonetheless, the achieved classification rates highlight the fact that automated respiratory sound analysis systems are not yet ready to assist medical experts. Towards improving the current performance, in the future we intent to pursue the following directions: *(a)* augment the *non-normal* part of the dataset, *(b)* employ a combination of spectral and wavelet features, and *(c)* include a discriminative classifier, possibly forming a synergistic framework.

Acknowledgment. This research was funded by the ELKE TEI Crete funds related to the domestic project: Bioacoustic applications, number 80680.

References

1. Baluja, S., Covell, M.: Waveprint: efficient wavelet-based audio fingerprinting. Pattern Recogn. **41**(11), 3467–3480 (2008). https://doi.org/10.1016/j.patcog.2008.05.006. http://www.sciencedirect.com/science/article/pii/S0031320308001702
2. Chen, S.H., Wu, H.T., Chen, C.H., Ruan, J.C., Truong, T.: Robust voice activity detection algorithm based on the perceptual wavelet packet transform. In: 2005 International Symposium on Intelligent Signal Processing and Communication Systems. IEEE (2005). https://doi.org/10.1109/ispacs.2005.1595342
3. Cook, S.A.: A taxonomy of problems with fast parallel algorithms. Inf. Control **64**(1), 2–22 (1985). https://doi.org/10.1016/S0019-9958(85)80041-3. http://www.sciencedirect.com/science/article/pii/S0019995885800413. International Conference on Foundations of Computation Theory

4. Jakovljević, N., Lončar-Turukalo, T.: Hidden Markov model based respiratory sound classification. In: Maglaveras, N., Chouvarda, I., de Carvalho, P. (eds.) Precision Medicine Powered by pHealth and Connected Health. IP, vol. 66, pp. 39–43. Springer, Singapore (2018). https://doi.org/10.1007/978-981-10-7419-6_7
5. Liu, P., Soong, F.K., Zhou, J.L.: Divergence-based similarity measure for spoken document retrieval. In: 2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP 2007, vol. 4, pp. IV-89–IV-92, April 2007. https://doi.org/10.1109/ICASSP.2007.367170
6. Ntalampiras, S.: A novel holistic modeling approach for generalized sound recognition. IEEE Sig. Process. Lett. **20**(2), 185–188 (2013). https://doi.org/10.1109/LSP.2013.2237902
7. Ntalampiras, S.: A classification scheme based on directed acyclic graphs for acoustic farm monitoring. In: 2018 23rd Conference of Open Innovations Association (FRUCT), pp. 276–282, November 2018. https://doi.org/10.23919/FRUCT.2018.8588077
8. Ntalampiras, S.: Moving vehicle classification using wireless acoustic sensor networks. IEEE Trans. Merg. Top. Comput. Intell. **2**(2), 129–138 (2018). https://doi.org/10.1109/TETCI.2017.2783340
9. Ntalampiras, S.: Directed acyclic graphs for content based sound, musical genre, and speech emotion classification. J. New Music Res. **43**(2), 173–182 (2014). https://doi.org/10.1080/09298215.2013.859709
10. Ntalampiras, S.: A transfer learning framework for predicting the emotional content of generalized sound events. J. Acoust. Soc. Am. **141**(3), 1694–1701 (2017). https://doi.org/10.1121/1.4977749
11. Ntalampiras, S.: Bird species identification via transfer learning from music genres. Ecol. Inf. **44**, 76–81 (2018). https://doi.org/10.1016/j.ecoinf.2018.01.006. https://www.sciencedirect.com/science/article/pii/S1574954117302467
12. Ntalampiras, S., Fakotakis, N.: Speech/music discrimination based on discrete wavelet transform. In: Darzentas, J., Vouros, G.A., Vosinakis, S., Arnellos, A. (eds.) SETN 2008. LNCS (LNAI), vol. 5138, pp. 205–211. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-87881-0_19
13. Pramono, R.X.A., Bowyer, S., Rodriguez-Villegas, E.: Automatic adventitious respiratory sound analysis: a systematic review. PloS One **12**(5), e0177926 (2017). https://doi.org/10.1371/journal.pone.0177926
14. Rabiner, L.R.: A tutorial on hidden Markov models and selected applications in speech recognition. Proc. IEEE **77**(2), 257–286 (1989). https://doi.org/10.1109/5.18626
15. Ren, Y., Johnson, M.T., Tao, J.: Perceptually motivated wavelet packet transform for bioacoustic signal enhancement. J. Acoust. Soc. Am. **124**(1), 316–327 (2008). https://doi.org/10.1121/1.2932070
16. Reynolds, D.A., Rose, R.C.: Robust text-independent speaker identification using Gaussian mixture speaker models. IEEE Trans. Speech Audio Process. **3**(1), 72–83 (1995). https://doi.org/10.1109/89.365379
17. Rizal, A., Hidayat, R., Nugroho, H.A.: Signal domain in respiratory sound analysis: methods, application and future development. J. Comput. Sci. **11**(10), 1005–1016 (2015). https://doi.org/10.3844/jcssp.2015.1005.1016
18. Rocha, B.M., et al.: A respiratory sound database for the development of automated classiffication. In: Maglaveras, N., Chouvarda, I., de Carvalho, P. (eds.) Precision Medicine Powered by pHealth and Connected Health. IP, vol. 66, pp. 33–37. Springer, Singapore (2018). https://doi.org/10.1007/978-981-10-7419-6_6

19. Serbes, G., Ulukaya, S., Kahya, Y.P.: An automated lung sound preprocessing and classification system based onspectral analysis methods. In: Maglaveras, N., Chouvarda, I., de Carvalho, P. (eds.) Precision Medicine Powered by pHealth and Connected Health. IP, vol. 66, pp. 45–49. Springer, Singapore (2018). https://doi.org/10.1007/978-981-10-7419-6_8
20. Taylor, P.: The target cost formulation in unit selection speech synthesis. In: Ninth International Conference on Spoken Language Processing, ICSLP, INTERSPEECH 2006, Pittsburgh, PA, USA, 17–21 September 2006 (2006). http://www.isca-speech.org/archive/interspeech_2006/i06_1455.html
21. Torrence, C., Compo, G.P.: A practical guide to wavelet analysis. Bull. Am. Meteorol. Soc. **79**, 61–78 (1998)
22. VanderWeele, T.J., Robins, J.M.: Signed directed acyclic graphs for causal inference. J. Roy. Stat. Soc.: Ser. B (Stat. Method.) **72**(1), 111–127 (2010). https://doi.org/10.1111/j.1467-9868.2009.00728.x
23. Zhao, Y., Zhang, C., Soong, F.K., Chu, M., Xiao, X.: Measuring attribute dissimilarity with HMM KL-divergence for speech synthesis, 6 p. (2007)