# ASNORE: Assessing snore health through classification of whole-night audio data

**João Ribeiro**[1] **and Pedro Constantino**[1]

[1]Instituto Superior Técnico, Lisbon, Portugal

**Assessing whole-night snoring through audio-only signal processing has grown in popularity due to advantages over traditional polysomnography (PSG). However, it is still impossible to accurately make prognosis about snoring health using only the whole-night audio recording, since no robust characterization of snores has been attained yet. Here, we present a MATLAB based solution for automatic extraction of relevant features from whole-night recordings, which takes into account both time and spectral information.. Ten whole-night recordings were analyzed, mined and then classified. Small sample size prevents meaningful testing from being done. However, classification results appear to be non-random and show some possibility of distinguishing between people that had different apnea and hypopnea indexes.**

**Obstructive sleep apnea | Respiratory sounds | Whole-night | Spectral features | Classification | Machine learning | Prognosis**

**Problem and motivation.** Snoring is estimated to affect 20% to 40% of the world population (1). It is a result of loss of muscle tonus in the upper airway and is a percursor of more severe conditions, such as obstructive sleep apnea (OSA). Obstructive sleep apnea has been associated with many cardiovascular diseases, like hypertension and arrhythmias, and has also been identified as a risk factor for stroke and all-cause mortality. (2) There is a growing need to diagnose OSA. Polysomnography (PSG) is the standard procedure, but it requires many resources and is impractical for long-term and home utilization. We believe that if a specialized physician were to stay besides a sleeping person for a full night, he could probably provide accurate prognosis regarding this person's respiratory sleeping health, based solely on his understanding of respiratory sounds. Moreover, we believe that it should be possible to replicate both 1) the physician's listening task and 2) prognosis production, recurring respectively to automatic digital processing of the whole-night audio recording and a machine learning tool. We intend to achieve these tasks using a fully automated system which only uses whole-night audio recordings using MATLAB.

To achieve 1), it is necessary to extract relevant features from the recording. Besides being able to extract time-based features (e.g.number of snores per hour, time between snores, average snore duration,etc), this system should also be able to extract relevant spectral features. This is of relevance both because we don't see relevant spectral snore description being done and used in any sort of classification/prognosis, and because, naturally, specialized physicians must use qualitative spectral information of the snores to successfully assess patients. Without spectral information, sounds would lose their quality. Regarding the spectral description, this automated system should carefully listen to each snore sound recorded and be able to integrate each and every one's spectral characteristics into a more general model that would succinctly describe this aspect of the patient - spectral feature extraction.

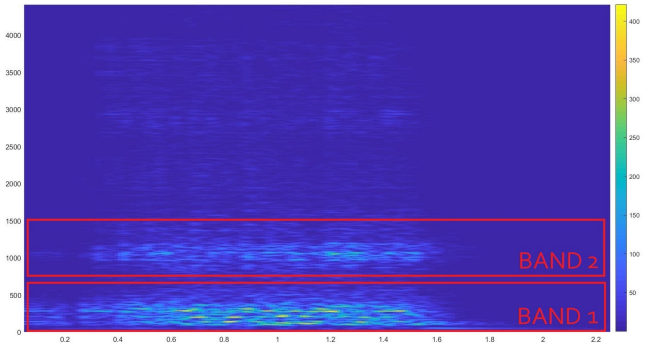


**Fig. 1.** Short-time Fourier transform of a snore sound reveals the two ocurring bands that may be quantitatively described.

We looked, one by one, at the short-time Fourier transforms (STFT) of several hundreds of snore-sounds and it looks like all snore-sounds have two fundamental frequency bands that can be readily described by accurate description of their shape in the spectrum (Fig. 1). This simple description could be done for all snores of a patient and, in the end, we could extract, for example, average width of both bands, average length of both bands, average maximum frequency in both bands, etc.

This simple spectral description together with time-based features and personal information could finally unlock accurate prognosis that time-based features alone can't achieve. Putting it all into perspective, if this tool were to initially provide meaningful prognosis, long-run tweaking of its parameters should probably outperform the specialized physician both in time and accuracy and we would have created an automatic prognostic tool ready to be distributed.

**Background and related work.** This project encompasses several steps in order to produce a robust full-night feature extraction. Some of these steps are whole topics in specific areas of the signal processing field. Next, we talk about some of the challenges present and which related work is associated with it so far.

***Noise reduction.*** In all applications that are related to sound recording, a certain signal of interest captured by a microphone is always contaminated by noise and interference.

Therefore, signal enhancement algorithms are required in order to clean the noisy signals. Since snoring sounds are produced through the same apparatus as speech sounds, it makes sense that filtration of audio with snore-sounds is made on the same basis as filtration of speech recording. One snore-sound segmentation paper reported using the celebrated Wiener-filter (often used in speech recognition routines), in order to clean the noisy whole-night signals (3). For an up to date review on speech enhancement see (4).

***Snore detection.*** In order to successfully extract snore-sound features, one by one, we must be able to segment each snore sound out of the whole recording to perform further processing. This is, as of today, still an ongoing matter of research and several publications have been made regarding automatic detection of snores in whole-night recordings (5), (6), (3), (7). One of the most promising unsupervised method for snore detection is that described in (7), which accomplishes high accuracy (95.6%) and high sensitivity (96.0%). It accomplishes snore segmentation with three steps: 1) detect snore candidates using V-Box control chart (8), 2) extract Mel-frequency cepstral coefficients (MFCCs) (9) and 3) classify snore candidates through K-harmonic means algorithm (10). Unsupervised snore detection has its advantages over other supervised techniques, since no pre-training is required. This also makes this method more portable.

***Snore sound characterization.*** Nowadays, there exist algorithms that can distinguish from different musical instruments, based on audio features (e.g. (11)). These generally focus on MFCCs features to accomplish such task. In a way, MFCCs can characterize a certain musical instrument which makes then possible for a machine learning algorithm to distinguish between different instruments. Snore-sound characterization however is not so straight-forward. Some works have focused on snore duration; others focused on snore intensity; others in wavelet descriptions (12). All of these have been applied to specific applications which are not always linked to assessing snore health; few focused on the spectral characterization of snore-sounds, which is something we take advantage of in the proposed method.

## Approach and uniqueness.

***Materials.*** Ten stereo whole-night recordings (.wav format), sampled at 44.1kHz, of both male and female patients, with different severities of snoring activity. These recordings were made with non-contact microphones, during polysomnography (PSG) routines. PSG data is also available for each patient (including apnea and hypopnea indices).

***Proposed method.*** In short, the available recordings were analyzed and then classified with supervised learning algorithms (Fig. 2).

Each whole-night recording was analysed using a single MATLAB script ((Fig. 3)). This script iteratively imports 5-minute segments of the full-recording and performs three tasks: **1) pre-processing**, which accomplishes conversion to mono (if needed), resampling and noise reduction; **2)**

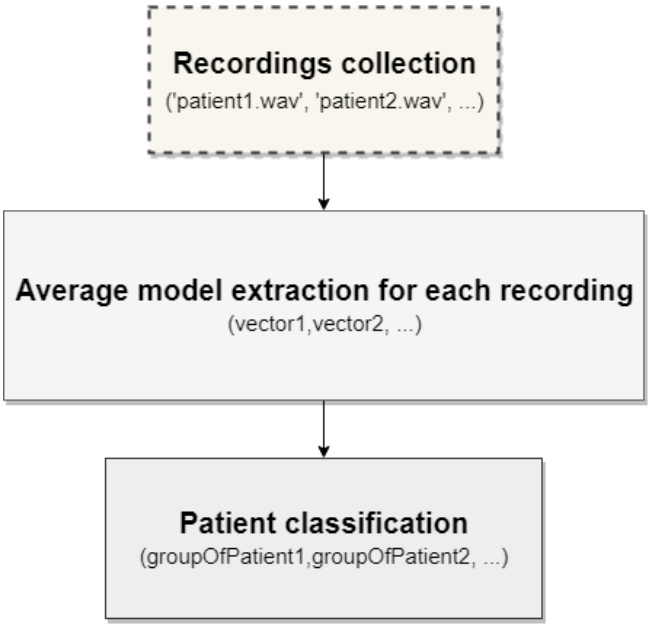


**Fig. 2.** Project routine.

**snore detection** (modified version of algorithm described in (7)) which generates time-stamps for onset and offset of each snoring sound in the segment and **3) spectral features extraction**, which, for each snore sound detected, creates a feature vector which will then be averaged out against all other snore-sounds' feature vectors to generate an average snore model vector of that patient's night. After processing the whole recording, the script further performs **time features extraction**, based on the already available time stamps. In the end, the script outputs a feature vector, which characterizes the full recording, that is ready to join whole-night recording data set and be further classified.

Thus, analysis of a given number of recordings, N, produced N vectors characterizing each record, which were further used as training datasets for supervised learning algorithms, namely Logistic Regression and Naive Bayes classifiers.

***Special topics.*** For **noise reduction**, we used an available MATLAB routine (13), which provides noise reduction through a Wiener filter, as described in (14) and in the style of speech enhancement techniques.

For **snore detection**, the modification to what is described in (7) is: in the V-Box Chart algorithm, the height, $H$, of the box is updated for each new 5-minute entry as:

$$P_{rms} = \frac{1}{N}\sum_{n=1}^{N} x^2(n) \quad \textbf{(1)}$$

$$P = \lfloor \log P_{rms} \rfloor \quad \textbf{(2)}$$

$$H = \frac{P_{rms} \times 10^{-P-2}}{6.5} \quad \textbf{(3)}$$

For **spectral feature extraction**, we used 9 features (3 for lower band + 3 upper band + 1 between bands), taken from STFT of snores sounds: 1) maximum frequency of lower

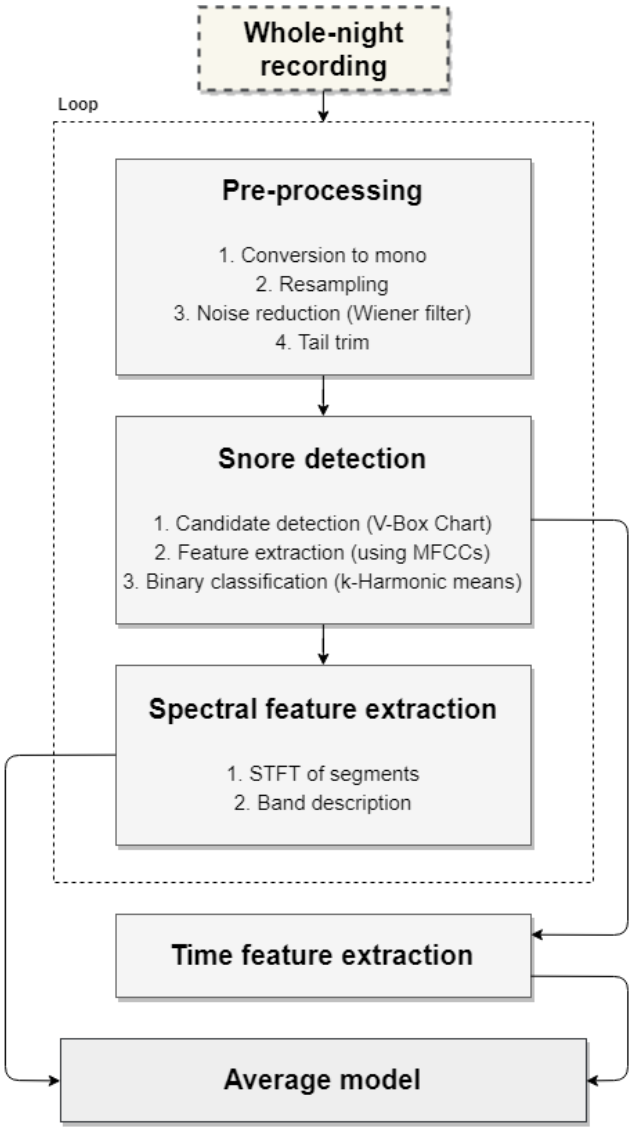


**Fig. 3.** Flow chart describing each whole-night recording analysis routine.

band, 'mf1', 2) maximum average frequency of lower band, 'mavgf1', 3) and 4) width of lower band at frequency bin 'mavgf1', 'w1l' and 'w1u', 5) maximum frequency of upper band, 'mf2', 6) maximum average frequency of upper band, 'mavgf2', 7) and 8) width of upper band at frequency bin 'mavgf2', 'w2' and 9) time difference between maximum average frequencies maximum activation, 'td'. See Fig. 4.

For classification, we exported the features matrix (each row comprising of the feature vector of an individual patient) into a .csv file. To use this as a training set, we defined the class from the apnea-hypopnea index (AHI) of the patient and difining the classes 'severe','medium' and 'mild' as:

1. Mild: $5 < AHI < 15$
2. Moderate: $15 < AHI < 30$
3. Severe: $AHI > 30$

We appended these classes into the feature matrix (the AHI were previsouly known) and used Orange3's software

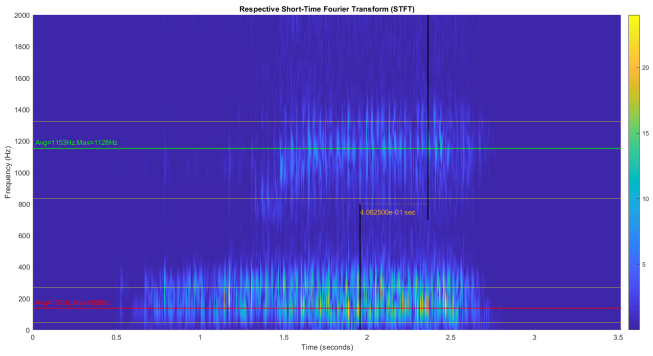


**Fig. 4.** Example of spectral feature extraction. The red line marks 'mavgf1'; contains labels for both 'mavgf1' and 'mf1'. The green line marks 'mavgf2'; contains labels for both 'mavgf2' and 'mf2'. The orange lines mark the recorded widths of bands; from lower to higher: 'w1l', 'w1u', 'w2l' and 'w2u'. The black lines set the points at which there was maximum activation of 'mavgf1' in lower band and 'mavgf2' in upper band; 'td' is the difference in seconds between these black lines.

to train a Naive Bayes (NB) and a Logistic Regresion (LR) classifier.

**Results and contributions.** When tested on the training set, both the Naive Bayes and Logistic Regression classifiers yielded an accuracy of 100%. To validate the models, the Receiver Operating Characteristics (ROC) were computed for each of the classifiers, using random sampling of the data with 100 repetitions and using 70% of the data to train, and the remainder to classify. For this purpose, one of the recordings was excluded, as it was a patient that did not snore at all during the entire night and would introduce significant errors since the data set is not diverse enough to accurately classify this outlier. The ROCs for each of the classes are displayed in figures 5-7. The NB classifier shows an average precision of 52.4% and sensitivity of 36%, while the LR classifier has an average precision of 42.9% and sensitivity of 29%. Both models behave better at classifying the 'severe' patients, but considerably worse at classifying the remaining ones, especially the LR classifier. The accuracy of these classifiers are way below the necessary for clinical applications, but not necessarily because the model isn't accurate, as discussed next.

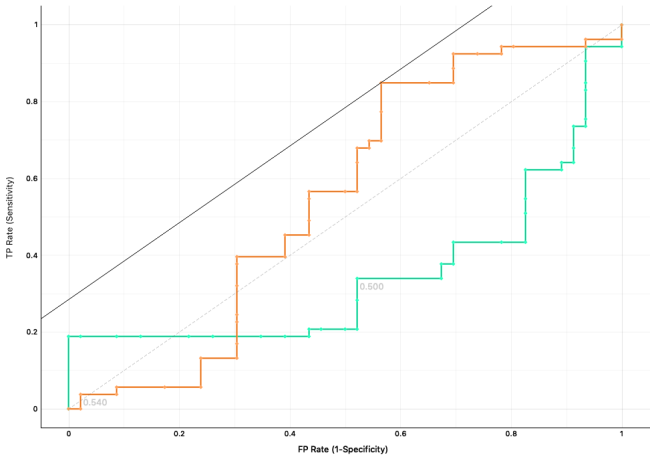


**Fig. 5.** ROC curves for the 'mild' class. Orange curve — NB; Green curve — LR

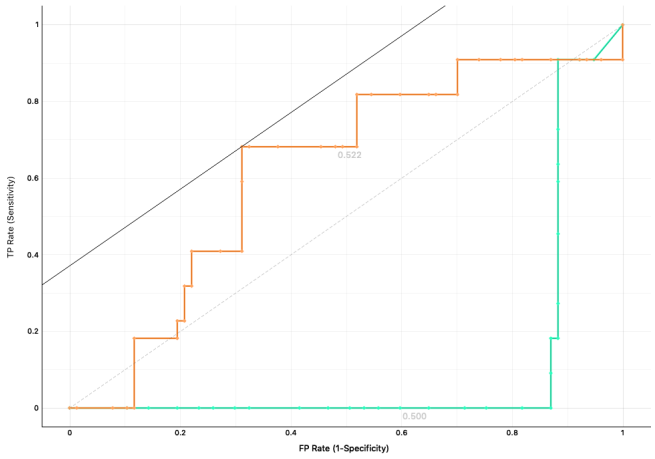


**Fig. 6.** ROC curves for the 'moderate' class. Orange curve — NB; Green curve — LR

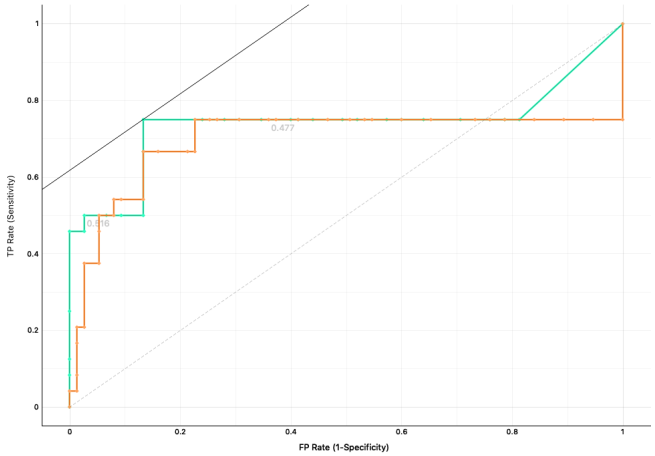


**Fig. 7.** ROC curves for the 'severe' class. Orange curve — NB; Green curve — LR

**Discussion, Conclusions and Future Perspectives.** Despite being systematically acquired, the audio recordings came with different signal-to-noise ratios (SNRs), which made snore detection and feature extraction algorithms produce a wide range of different responses. If for example 'audio1.wav' has low SNR, then the snore detection algorithm will fail to detect some snore candidates (because they are hidden in the background noise) and will also identify as candidates small perturbations that might just be noisy artifacts and not true snore candidates, which will further complicate classification into 'yes-snore', 'no-snore'. Furthermore, for extraction of features, the lower SNR recording will yield results with a decreased resolution because there is little difference between maximum and minimum activation of frequencies. It's as if I want to identifies the borders of a face in a drawing, but there is little difference between the color and texture of the paper and the lines I draw with my pencil - like that I can't identify where the nose ends or starts as well as I would be able to see in a drawing with higher SNR. On the other hand, if the 'audio2.wav' has higher SNR, the snore detection algorithm virtually finds all candidates and, for extraction of features, there is now a friendlier, more-contrasted STFT plot to be assessed.

The small data size (only ten recordings) made inspection of results and code-tweaking possibilities harder, since we can't be sure whether some change in the code itself would produce better or worse results in terms of characterization and classification. From a classification perspective, the size and diversity of the database is the most limiting factor. From the database, six patients show 'mild' snoring condition, two 'severe' and two 'morderate'. The specificity and sensitivity of the models don't tell the whole story here, and we believe that without a large and diverse dataset — with many tens of classified patients, showing diverse characteristics — it is impossible to accurately classify, especially in a topic as complex as snoring sounds, where even in simple applications such as this one, 23 features are recognizable. It's like asking a child to distinguish between tree families in a huge unknown forest, by initially showing him only ten trees from that forest. The kid might go and produce a reasonable model that describes the ten trees and is able to distinguish between groups of trees inside those 10 trees environment; however, when going to explore the huge forest, the child will understand that his way of classifying was biased and inelegant compared to the diversity that he now encounters.

Not that the snore detection applied was perfect: modifications made to the original implementation (7) degraded snore sound detection both in accuracy and sensitivity. This has negative impacts on final results, since not detecting a snore-sound, or identifying a clap as snore-sound weakens the final average model. Besides that, unsupervised snore detection has a severe limitation in the case of whole-night recordings: if a person doesn't snore that much, let's say, in an hour of recording, the fact that maybe we will find only one or two snores makes classification of candidates into yes-snore/no-snore harder, because there is not enough data to achieve meaningful clustering. This fact should tap researchers attention to the importance of having robust, good tools for automatic detection of snores in whole-night recordings. Without a robust snore detection, matters complicate further.

The average snore model here proposed might be a beginning to understand some fundamental spectral characteristics about snore sound in general, and also a tool to integrate automatic snore health prognosis. However, snore sound is variable and although the two bands look like almost all the information there is, a better analysis of this assumption has to be carried. It should also be looked into which aspects of the two bands in the STFT plot are meaningful for snore characterization.

If this whole-night recording classification routine would be perfected, and more data would be available, methods such as principal component analysis (PCA) could be used to understand which type of features really distinguish and explain the variance bewteen types of snorers.


#### ACKNOWLEDGEMENTS

We would like to thank Tiago Marçal, who conceded the ten audio recordings and Professor Agostinho Rosa, who made the connection possible.


#### COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

# Bibliography


1. V. Hoffstein. Apnea and snoring: state of the art and future directions. *Acta oto-rhino-laryngologica Belgica*, 56:205–236, 2001.
2. Samy I McFarlane. Obstructive sleep apnea and stroke. *Sleep Medicine and Disorders: International Journal*, 2(5), 2018. doi: 10.15406/smdij.2018.02.00056.
3. Eliran Dafna, Ariel Tarasiuk, and Yaniv Zigel. Automatic detection of whole night snoring events using non-contact microphone. *PLoS ONE*, 8(12), 2013. doi: 10.1371/journal.pone.0084139.
4. Jacob Benesty. *Fundamentals of Speech Enhancement*. Springer International Publishing, 2018.
5. W D Duckitt, S K Tuomi, and T R Niesler. Automatic detection, segmentation and assessment of snoring from ambient acoustic data. *Physiological Measurement*, 27(10): 1047–1056, 2006. doi: 10.1088/0967-3334/27/10/010.
6. A Azarbarzin and Z M K Moussavi. Automatic and unsupervised snore sound extraction from respiratory sound signals. *IEEE Transactions on Biomedical Engineering*, 58(5): 1156–1162, 2011. doi: 10.1109/tbme.2010.2061846.
7. Ganjun Ma, Biao Xue, Hong Hong, Xiaohua Zhu, and Zhiyong Wang. Unsupervised snore detection from respiratory sound signals. *2015 IEEE International Conference on Digital Signal Processing (DSP)*, 2015. doi: 10.1109/icdsp.2015.7251905.
8. Ewaryst Rafajlowicz, Miroslaw Pawlak, and Ansgar Steland. Nonparametric sequential change-point detection by a vertically trimmed box method. *IEEE Transactions on Information Theory*, 56(7):3621–3634, 2010. doi: 10.1109/tit.2010.2048443.
9. Reinhold Haeb-Umbach and Marco Loog. An investigation of cepstral parameterisations for large vocabulary speech recognition. *EUROSPEECH'99*, page 1323–1326, 1999.
10. B. Zhang. Generalized k-harmonic means. *Hewlett-Packard Laboratoris Technical Report*, 2000.
11. Jordi Pons, Olga Slizovskaia, Rong Gong, Emilia Gomez, and Xavier Serra. Timbre analysis of music audio signals with convolutional neural networks. *2017 25th European Signal Processing Conference (EUSIPCO)*, 2017. doi: 10.23919/eusipco.2017.8081710.
12. Kun Qian, Maximilian Schmitt, Christoph Janott, Zixing Zhang, Clemens Heiser, Winfried Hohenhorst, Michael Herzog, Werner Hemmert, and Björn Schuller. A bag of wavelet features for snore sound classification. *Annals of Biomedical Engineering*, 47(4):1000–1011, 2019. doi: 10.1007/s10439-019-02217-0.
13. Yi-Wen Chen. Noise reduction by wiener filter. 2019. doi: https://www.github.com/JarvusChen/MATLAB-Noise-Reduction-by-wiener-filter.
14. P. Scalart and J.v. Filho. Speech enhancement based on a priori signal to noise estimation. *1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings*. doi: 10.1109/icassp.1996.543199.