proceedings

MDPI

Proceedings

A real-time snore detector using neural networks and selected sound features†

Stelios A. Mitilineos [1], Nicolas-Alexander Tatlas [1], Georgia Korompili [1], Lampros Kokkalas [1], and Stelios M. Potirakis [1,*]

[1] University of West Attica, Department of Electrical and Electronic Engineering, 12243, Aigaleo, GREECE; smitil@uniwa.gr; ntatlas@uniwa.gr; gkorompili@uniwa.gr; lkokkalas@uniwa.gr; spoti@uniwa.gr

* Correspondence: spoti@uniwa.gr

† Presented at 2nd International Electronic Conference on Applied Sciences, 15–31 October 2021; Available online: https://asec2021.sciforum.net/.

Abstract: Obstructive Sleep Apnea Hypopnea Syndrome (OSAHS) is a widespread chronic disease that mostly remains undetected, mainly due to the fact that it is diagnosed via polysomnography which is a time and resource-intensive procedure. Screening the disease's symptoms at home could be used as an alternative approach in order to alert individuals that potentially suffer from OSAHS without compromising their everyday routine. Since snoring is usually linked to OSAHS, developing a snore detector is appealing as an enabling technology for screening OSAHS at home using ubiquitous equipment like commodity microphones (included in, e.g., smartphones). In this context, we developed a snore detection tool and herein present our approach and selection of specific sound features that discriminate snoring vs. environmental sounds, as well as the performance of the proposed tool. Furthermore, a Real-Time Snore Detector (RTSD) is built upon the snore detection tool and employed in whole-night sleep sound recordings resulting to a large dataset of snoring sound excerpts that are made freely available to the public. The RTSD may be used either as a stand-alone tool that offers insight to an individual's sleep quality or as an independent component of OSAHS screening applications in future developments.

Keywords: obstructive sleep apnea hypopnea syndrome; apnea screening; snoring detection; machine learning; neural networks.

Citation: Mitilineos, S. A.; Tatlas, N. A.; Korompili, G.; Potirakis, S. M. A real-time snore detector using neural networks and selected sound features. *Proceedings* **2021**, *68*, x. https://doi.org/10.3390/xxxxx

Published: date

1. Introduction

Obstructive Sleep Apnea-Hypopnea Syndrome (OSAHS) is a chronic condition held responsible for a number of well-documented effects on patients' health. It is linked to increased cardiovascular morbidity and mortality, including sudden heart death [1], while an estimated 4% and 2% of the male and female population respectively suffer from OSAHS. Interestingly enough, an estimated 85% of patients remain undiagnosed [2]. This underestimation poses an increased risk for individuals and society as a whole and is mainly due to polysomnography being the only method for OSAHS diagnosis currently trusted by doctors. Polysomnography is a time and resource-consuming procedure that monitors sleep with a multitude of specialized sensors and equipment and is performed in dedicated sleep laboratories or hospital care clinics. As such, most of the suffering population remains unscreened and, hence, undiagnosed.

The APNEA research project aims at accurately and cost-efficiently screening patients at home, using sound recordings via the users' smartphone during sleep [3]. In an ongoing measurements campaign, the APNEA project is collecting polysomnography data together with time-synchronized and high quality tracheal and ambient microphone recordings. The data are collected during sleep studies that are performed by project partners following the relevant medical protocols, and are of a duration of about 8-hour each.

Insofar, the acquired database consists of more than 200 complete polysomnography studies and our respective findings are reported in [4]. In parallel, and inspired by literature findings linking snoring to OSAHS episodes (e.g., see Refs. [5-7]), APNEA aims at developing a Real-Time Snore Detector (RTSD) in order to use it for pre-screening of microphone recordings at home. The RTSD is intended to be either used as a stand-alone tool for apnea screening or integrated within more sophisticated apnea detection solutions by allowing to the latter to focus on timeslots of increased OSAHS probability.

As long as snore classifiers are concerned, we have focused on neural networks. They have been used in the literature for snoring detection with substantial classification accuracy, usually in the order of 90 % or larger [8-10]. However, neural networks are usually trained using a relatively small dataset or a fragment of whole night sleep sound recordings. On the contrary, RTSDs – neural based or other- are meant to be employed in much larger datasets (see whole night recordings of multiple patients), while larger datasets are typically related to reduced accuracy performance. In this respect, our contribution lies in (i) our approach and findings about which sound features are more promising and should be used for snoring classification, (ii) the training of a successful neural network for snoring detection with superior classification accuracy despite been trained using a much larger dataset compared to those used in the literature, ~~–~~ (iii) the development of a RTSD tool, and (iv) the availability of a large body of annotated snoring sound excerpts (upon which the neural network training was implemented) together with an extremely large body of snoring sound excerpts that correspond to the output of the RTSD upon a large subset of whole-night sleep sound recordings. We present the architecture of the proposed classification tool in section 2, while we report our findings regarding feature selection in sub-section 3.1. Numerical results on the performance of the proposed neural network and RTSD are demonstrated in sub-section 3.2 and section 4 concludes the paper and includes a discussion regarding future work for RTSD improvement.

2. Architecture of the proposed classification tool and real-time snore detector

The architecture of the proposed classification tool is based on a neural network classifier and is illustrated in Fig. 1. Sound excerpts are used as input to the classifier. Each sound excerpt is de-noised using wavelet filtering and then is normalized with respect to its average energy. Selected features are calculated for each sound excerpt (sampled at 48 kHz, 24-bit), while a Gaussian Mixture Model (GMM) is also calculated for the Time-Frequency (TF) features. Finally, a neural-network classifier is employed in order to infer whether the input sound excerpt is a snore or not. Keeping in mind the big picture of a RTSD that will ultimately run in smartphones at home, we selected the implementation of a shallow neural network classifier, with one hidden layer. Given that we experimented with different options for the implemented features, the number of nodes of the network hidden layer was equal to (rounded) 2.5 times the number of nodes of the input layer. A detailed discussion on the features that we implemented and used for this work is provided in Section 3, while details on the architecture of the neural network per se as well as the implementation of wavelet de-noising, energy normalization, GMM and neural network training are provided in our previous work [11].

Furthermore, the architecture of the proposed RTSD is illustrated in Figure 2. The RTSD is designed to be used in real-time but its operation is herein emulated using whole-night sleep recordings as its input. As such, the input sound recording is parsed with a sliding window of duration 6 s and a sliding step of 2 s (i.e., there is an overlap of around 66.7 % between adjacent windows). The window duration of 6 s was selected because we have observed that a typical breathe-in-breathe-out cycle is about 4 s, so we opted for a guard interval of 1 s before and after. The sliding length is then equal to the sum of these guard intervals. The sound within each window is captured and the proposed classification tool of Figure 1 is employed in order to infer whether the specific time window corresponds to snoring or not. If it is classified as snoring, then we record the sound excerpt within the specific window to a separate .wav file for further processing or else we proceed to the next time window according to the predefined time-step. The procedure is repeated until the end of the whole-night sound recording or, in a real-life scenario, until the user aborts the application in her/his smartphone.

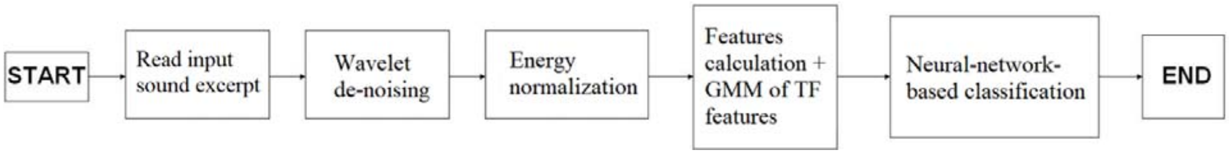

Figure 1. Architecture of the proposed classification tool and neural network.

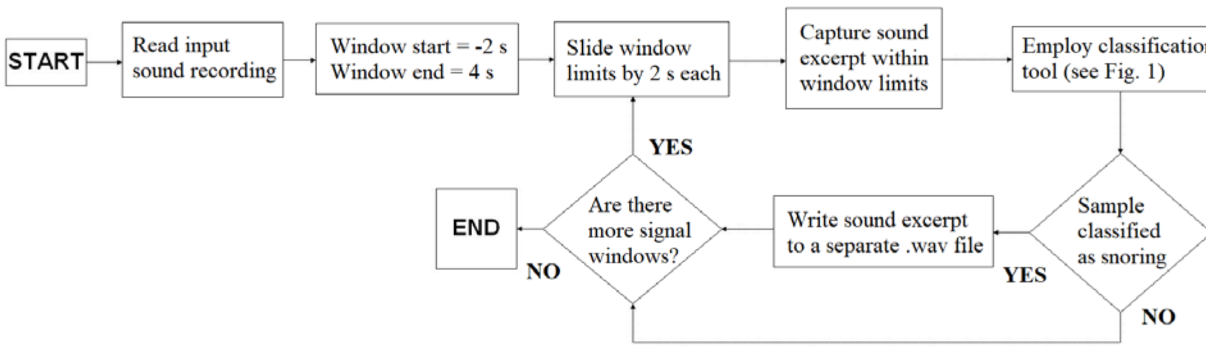

Figure 2. Architecture of the proposed Real-Time Snore Detector.

3. Numerical results

3.1. Features selection and performance of the proposed neural network snore detection tool

In the literature, sound classification is performed using carefully selected features that are broadly categorized in time-domain (such as zero-crossing-rate (ZCR), energy, volume, etc.) and frequency-domain features (pitch, bandwidth, mel-frequency cepstral coefficients (MFCCs), etc.). However, such "static" features fail to capture the time evolution of the signal. Time-frequency (TF) features are therefore proposed and consist in crafting a sequence of static features calculated on a time window that is sliding over the entire sound signal. With such an approach, the temporal evolution of the signal is captured, however, the resulting feature space is usually huge and therefore needs to be reduced by the means of, e.g., a Gaussian mixture modeling in the case of shallow neural networks [11-12] (or repeated convolutional layers in the case of convolutional or deep neural networks).

Most of the aforementioned sound features are also used for snore detection in the literature [13]. On top of these, features that are used for snore detection include low-level descriptors and functional-based features that are reported in [14], positive/negative amplitude ratio, sampling entropy and 500 Hz power ratio reported in [15], local dual octat pattern reported in [16], and many more. Nonetheless, there is not yet a clear consensus on what should be considered an appropriate feature selection when it comes to snore detection in whole night sleep studies [13]. In this respect, we performed a preliminary study about selecting a set of well-performing sound features.

The first features subset that we opted to compare consists of scalar features including (i) the ZCR, pitch, bandwidth, volume and intensity of the signal, (ii) a set of entropy metrics, specifically the Shannon, Tsallis, wavelet and permutation entropy, and (iii) a few statistical metrics, namely the median, average, variance, skewness, kurtosis of the signal amplitude. The second features' subset includes the MFCCs of the sound signal; more specifically, 13 MFCCs are calculated over the frequency range between 20 Hz and 6 kHz of the recorded signal. Implementation details for scalar features and MFCCs are provided in [11-12], while both are calculated over the entire signal portion that corresponds to the relative position of the sliding window described in Section 2.

Furthermore, inspired by studies reporting that snoring frequencies are mostly centered on specific and narrow ranges [17-18], we developed a modified spectrogram of the input signal to be used as a sound feature suitable for snore detection. More specifically, we calculate the spectrogram of each sliding window; each sound excerpt is down-sam~~m~~pled to 12 kHz, hence the resulting spectrogram ranges from 0 up to 6 kHz. Then, we calculate the average spectral coefficients in adjacent, non-overlapping windows of length 100 Hz each, resulting to the so-called Modified Spectral Coefficients (MSC). Finally, we extract the normalized MSC values in order to capture the energy concentration within specific frequency ranges. As an example, Figure 3 compares the normalized MSC between a snoring and a non-snoring sound excerpt. In this case, snoring sound energy exhibits a peak at around 170 Hz that complies with the snoring frequencies reported in [17]. On the contrary, the non-snoring excerpt exhibits a smoother distribution of energy vs. frequency. Following multiple similar by-visual-inspection comparisons, we considered that the normalized MSC can be successful in discriminating snoring events and we herein report numerical results that justify this approach.

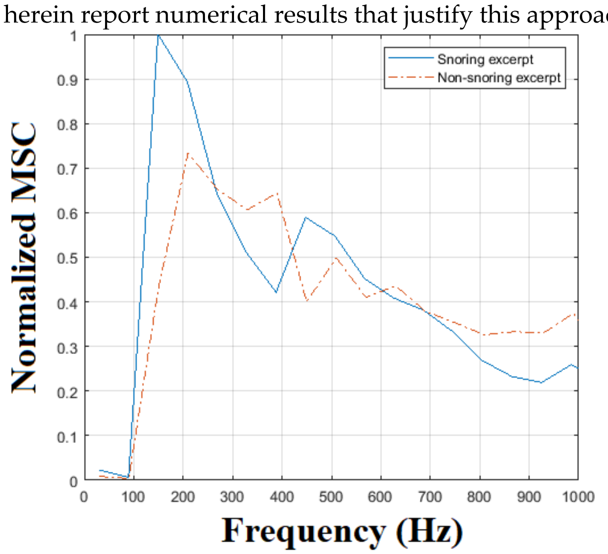

Figure 3. Modified spectral coefficients for a snoring and a non-snoring sound excerpt (solid and dash-dotted curve, respectively).

Then, in order to infer about which combination of the implemented features provides the best accuracy with respect to snoring classification, we executed multiple training sessions of the proposed neural network using different features and feature combinations. More specifically, we selected fifty different whole-night sound recordings from fifty different patients. For each one of them, we manually selected and isolated fifty snoring sound excerpts and fifty non-snoring sound excerpts, of duration 6 s each. This results to a snoring and non-snoring database of 2500 + 2500 sound excerpts respectively (a total of 5000 excerpts), with a total duration of about 30000 s (15000 s of snoring and 15000 s of non-snoring); this database of manually annotated sound excerpts is freely available upon request and the interested reader is referred to the Data Availability section below.

We then selected 70 % of this dataset for network training and the remaining 30 % for testing. The resulting classification accuracy *of the test set* vs. selected features combinations is tabulated in Tables I and II. Classification accuracy is defined as the ratio of correct classification events (snoring excerpt classified as snoring plus non-snoring excerpt classified as non-snoring) vs. the total number of classification attempts (which is equal to the number of available excerpts, i.e., 5000 sound excerpts). According to Table I, the normalized MSC exhibit better accuracy compared to MFCC or the set of scalar features. On the other hand, the normalized MSC exhibit similar accuracy when combined with either MFCC or the set of scalar features. Given that scalar features are computationally less intensive than MFCC to calculate, and that the proposed tool is envisioned to be used in portable devices, we selected the combination of normalized MSC plus scalar features to be used from now on in our experiments.

Table 1. Test set classification accuracy per feature class.

Scalar features	MFCC	Normalized MSC
93.4 %	95.7 %	97.7 %

Table 2. Test set classification accuracy per feature classes' combination.

	Scalar features	MFCC	Normalized MSC
Scalar features	-	96.0 %	98.6 %
MFCC	-	-	98.7 %
Normalized MSC	-	-	-
All feature classes		97.3 %	

Furthermore, the *Precision* and *Recall* of the selected combination of normalized MSC and scalar features are calculated. Precision corresponds to the proportion of positive identifications that was actually correct and is calculated as the ratio of true positives vs. the sum of true positives plus false positives. Recall corresponds to the proportion of actual positives that were identified as such and is calculated as the ratio of true positives vs. the sum of true positives plus false negatives. For the aforementioned features combination and test set, the resulting precision is equal to 99.59 % while the recall is equal to 98.32 %. Taking into account these performance metrics together with the reported overall accuracy of 98.6 %, we consider that the proposed classification tool is eligible to be used as a building block of a RTSD.

3.2. Application of the real-time snore detector

Following the training and testing of the proposed classification tool, we employed it within the proposed RTSD scheme illustrated by Figure 2. Then, we selected a set of twenty-five whole-night sound recordings and applied the RTSD upon them. The total duration of the whole-night recordings that were employed is equal to 51 hours, 45 minutes and 13 seconds. Among these, a total of 12090 different sound excerpts of duration 6 s each are classified as snoring by the RTSD, corresponding to a total duration of 20 hours and 9 minutes. These sound excerpts are freely available to the interested reader upon request (please see the Data Availability section below).

4. Conclusions and future work

We report herein a snoring classification tool with substantial performance (estimated accuracy equal to 98.6 %), as well as the availability of a small dataset of annotated snoring and non-snoring excerpts together with a large dataset of non-annotated excerpts classified as snoring. In the immediate future, we intend to fully annotate the latter and offer a large, freely available database of annotated snoring excerpts. We also intend to use this full annotation in order to train a cascaded neural network that will have as input only the positive output of the classification tool proposed herein, as shown in Figure 4. The cascaded neural network will be trained with the aim of discriminating between true and false positives, thus providing a new classification output that is expected to be much more accurate than that of the first neural network alone.

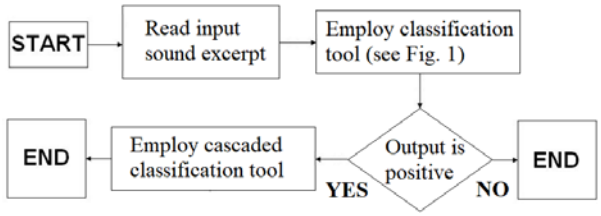

Figure 4. Architecture of a cascaded neural-network classification tool.

Author Contributions: Mitilineos, S. A.: conceptualization, methodology, software, validation, writing—original draft preparation, Tatlas, N.-A., Korompili, G., Kokkalas, L., and Potirakis, S. M.: writing—review and editing, S.M.P.: conceptualization and supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This research was co-financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH—CREATE—INNOVATE (project code: T1EDK-03957_Automatic Pre-Hospital, In-Home, Sleep Apnea Examination).

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethics Committee of Sismanoglio – Amalia Fleming General Hospital of Athens on March 16, 2017, with protocol number 05/16.03.2017.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Two sets of data samples are provided free of charge to the interested reader(s). These include (i) the database of manually annotated sound excerpts upon which the proposed neural network is trained, including snoring and non-snoring sounds (2500 + 2500 sound excerpts respectively, summing up to a total of 5000 excerpts and a total duration of about 30000 s), and (ii) the database the RTSD output, including 12090 snoring sound samples of duration 6 s each, corresponding to a total duration of 20 hours and 9 minutes of snoring sounds. These data are provided upon request at spoti@uniwa.gr

Acknowledgments: The authors acknowledge Mr. M. Kouvaras for his help during the manual annotation of the neural network training dataset.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Qaseem, A., et al., "Management of obstructive sleep apnea in adults: a clinical practice guideline from the American College of Physicians", October **2013**, DOI: 10.7326/0003-4819-159-7-201310010-00704.
2. Pack, A. I., "Advances in sleep-disordered breathing", January **2006**, DOI: 10.1164/rccm.200509-1478OE.
3. APNEA Project: "Automatic pre-hospital and in-home screening of sleep apnea", Operational Programme "Competitiveness, Entrepreneurship and Innovation", website: http://apnoia-project.gr/ (accessed on April, 01, **2021**).
4. Korompili, G., et al., "Poly-SleepRec: a scored polysomnography dataset with simultaneous audio recordings for sleep apnea studies", *Nature Scientific Data*, **2021**, accepted for publication.
5. Karunajeewa, A. S., et al., "Multi-feature snore sound analysis in obstructive sleep apnea-hypopnea syndrome", January **2010**; DOI: 10.1088/0967-3334/32/1/006.
6. Ben-Israel, B., et al., "Obstructive apnea hypopnea index estimation by analysis of nocturnal snoring signals in adults", September **2012**; DOI: 10.5665/sleep.2092.
7. Azarbarzin, A., Moussavi, Z., "Snoring sounds variability as a signature of obstructive sleep apnea", April **2013**; DOI: 10.1016/j.medengphy.2012.06.013.
8. Arsenali, B., et al., "Recurrent neural network for classification of snoring and non-snoring sound events", July **2018**; DOI: 10.1109/EMBC.2018.8512251.
9. Emoto, T., et al., "Neural networks for snore sound modeling in sleep apnea", July **2005**; DOI: 10.1109/CIMSA.2005.1522887.
10. Kang, B., et al., ""Snoring and apnea detection based on hybrid neural networks", December **2017**; DOI: 10.1109/ICOT.2017.8336088.

11. Mitilineos, S. A., et al., "Neural network fusion and selection techniques for noise-efficient sound classification", January **2019**; DOI: 10.17743/jaes.2018.0071.
12. Mitilineos, S. A., et al., "A two-level sound classification platform for environmental monitoring", June **2018**; DOI: 10.1155/2018/5828074.
13. Dafna, E., et al., "Automatic detection of whole night snoring events using non-contact microphone", December **2013**; DOI: 10.1371/journal.pone.0084139.
14. Zhang, Z., et al., "Snore-GANs: improving automatic snore sound classification with synthesized data", **2018**; DOI: 10.1109/JBHI.2019.2907286.
15. Wang, C., et al., "Automatic snoring sounds detection from sleep sounds via multifeatures analysis", **2017**; DOI: 10.1007/s13246-016-0507-1.
16. Tuncer, T., et al., "An automated snoring sound classification method based on local dual octal pattern and iterative hybrid feature selector", *Biomedical* **2021**; DOI: 10.1016/j.bspc.2020.102173.
17. Agrawal, S., et al., "Sound frequency analysis and the site of snoring in natural and induced sleep", *Clinical Otolaryngology*, Vol. 27, pp. 162-166, **2002**; DOI: 10.1046/j.1365-2273.2002.00554.x.
18. Saunders, N. C.,et al., "Is acoustic analysis of snoring an alternative to sleep nasendoscopy?", *Clinical Otolaryngology*, Vol. 29, pp. 242-246, **2004**; DOI: 10.1111/j.1365-2273.2004.00800.x.