# A Real-Time Snore Detector Using Neural Networks and Selected Sound Features †

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

Department of Electrical and Electronic Engineering, University of West Attica, 12243 Aigaleo, Greece; smitil@uniwa.gr (S.A.M.); ntatlas@uniwa.gr (N.-A.T.); gkorompili@uniwa.gr (G.K.); lkokkalas@uniwa.gr (L.K.)
* 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 in 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 concerning 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

## 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 measurement 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-h each. 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 the 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 otherwise, 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 Section 3.1. Numerical results on the performance of the proposed neural network and RTSD are demonstrated in Section 3.2. 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 illustrated in Figure 1. Sound excerpts are used as input to the classification tool. Each sound excerpt is de-noised using wavelet filtering and then normalized with respect to its average energy. Selected features are calculated for each sound excerpt (sampled at 48 kHz, 24-bit), including temporal (time-domain), spectral (frequency-domain) and time-frequency features (for a complete discussion please refer to the subsequent Sections of this paper). Due to the high dimensionality of the available features, a Gaussian mixture model (GMM) is calculated for the time-frequency features. Afterwards, a neural-network classifier is employed in order to infer whether the input sound excerpt is a snore or not based on the calculated features and GMM models. 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. After extensive trial-and-error, the number of nodes of the network hidden layer was selected to be equal to (rounded) 2.5 times the number of nodes of the network input layer. A detailed discussion on the features that we implemented and used for this work is provided in Sections 3 and 4, 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 [11,12].

As far as the neural network classifier is concerned, a schematic of the internal architecture of the neural network that we used is illustrated in Figure 2. We consider only feed-forward artificial neural networks (FANNs) with the training function being an error back-propagation variant. The input layer of the network is used for data entry and weighting and consists of a number of nodes that is equal to the number of the independent feature inputs. The weights that multiply each data entry are subject to the network's training that is performed off-line and prior to classification. The weighted input features are then forwarded to an intermediate layer of neurons. The middle layer's number of nodes is tuned around the empirical rule-of-thumb value of two and a half

times the number of input layer nodes. These neurons sum up all the weighted features and, essentially, configure all possible convex classes of data in the feature space. The output of the intermediate layer is then forwarded to two output neurons. These neurons at the output layer are essentially combining convex classes in order to configure non-convex classes to classify the input data. Each output is taking a value of "+1" or "−1" that corresponds to "true" or "false" state respectively. The usage of two output nodes instead of one with a true/false implementation is empirically proven to add robustness to the network and improve performance. As such, a decision for snoring is formed in the case where $(Z_1, Z_2) = (1, -1)$ while a decision for non-snoring is formed in the case where $(Z_1, Z_2) = (-1, 1)$.
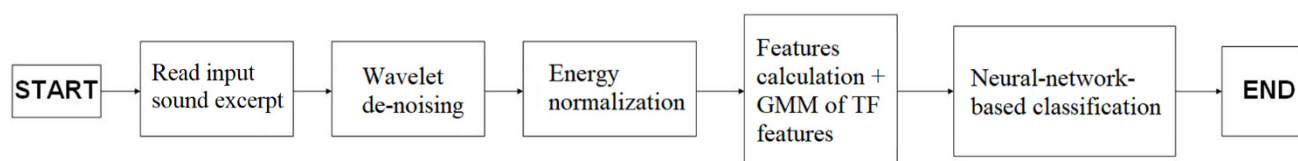


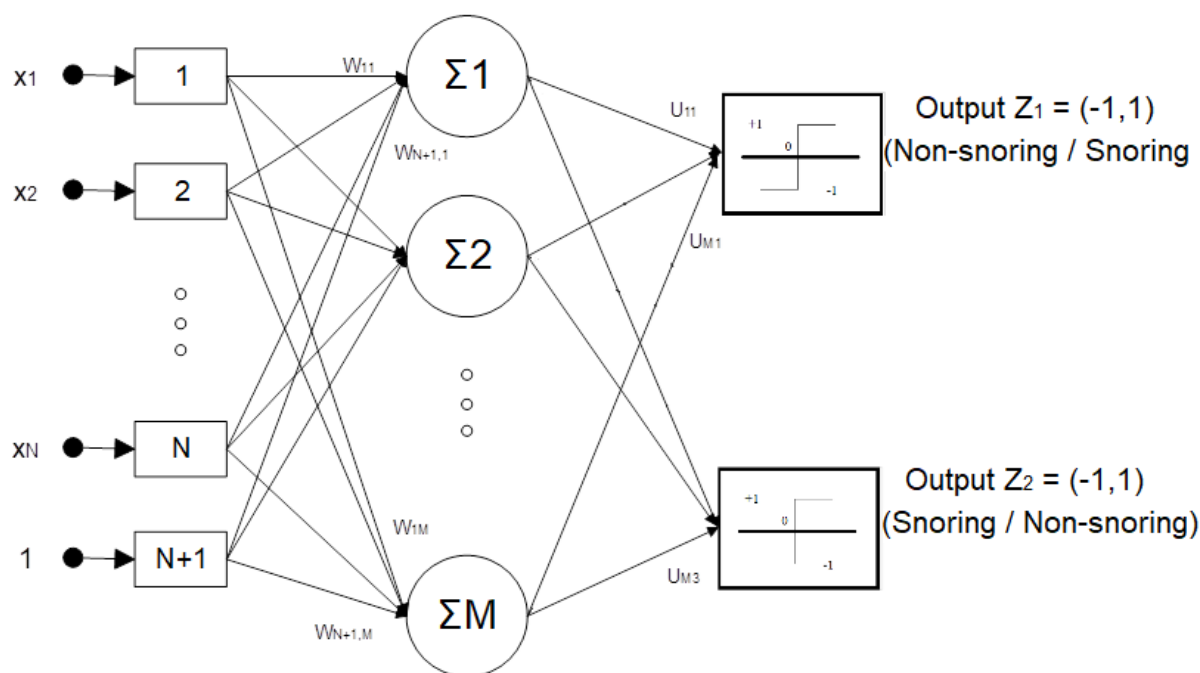**Figure 1.** Architecture of the proposed classification tool and neural network.



**Figure 2.** Neural network classifier for high−level snoring/non-snoring classification.

Furthermore, the architecture of the proposed RTSD is illustrated in Figure 3. 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 this is the case, 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.
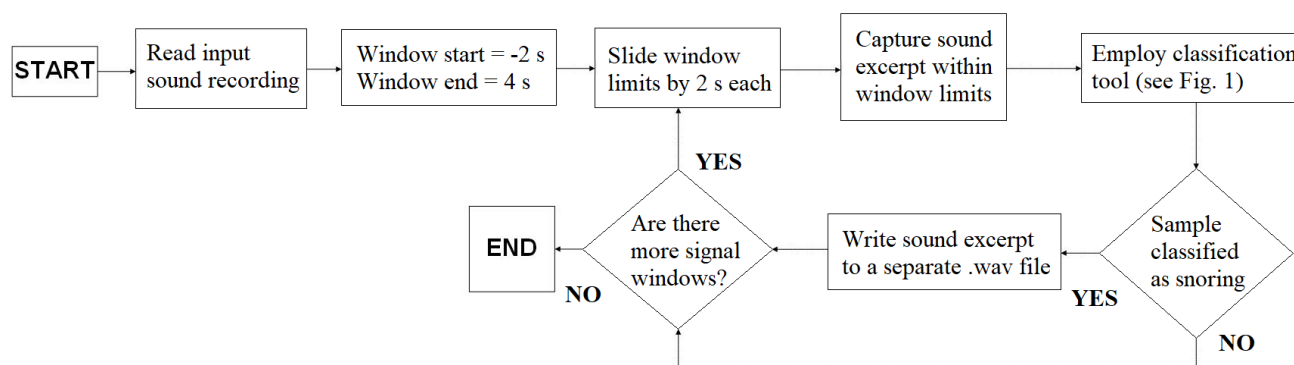
**Figure 3.** Architecture of the proposed Real-Time Snore Detector.

## 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, and 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-sampled 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 4 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.
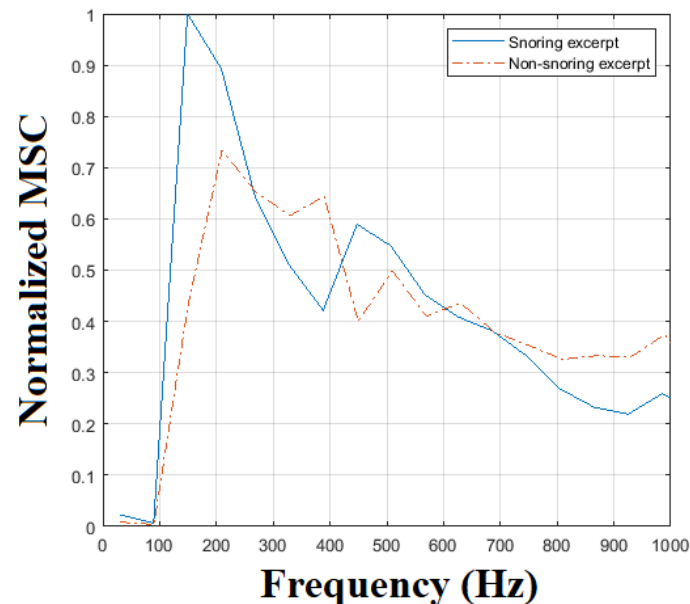


**Figure 4.** Modified spectral coefficients for a snoring and a non−snoring sound excerpt (solid and dash-dotted curve, respectively).

Then, in order to infer 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 50 snoring sound excerpts and 50 non-snoring sound excerpts, of 6 s duration each. This results constitute a snoring and non-snoring database of 2500 + 2500 sound excerpts respectively (a total of 5000 excerpts), with a total duration of about 30,000 s (15,000 s of snoring and 15,000 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 1 and 2. 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 1, 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 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 1. Then, we used a set of twenty-five whole-night sound recordings, other than those used for the training and testing of the classification tool, in order to test the proposed RTSD. The total duration of the whole-night recordings that were tested is equal to 51 h, 45 min and 13 s. A total of 12,090 different sound excerpts of duration 6 s each are classified as snoring by the RTSD, corresponding to a total duration of 20 h and 9 min. These snoring sound excerpts are freely available to the interested reader upon request (please see the Data Availability section below).

In order to assess the performance of the proposed RTSD, a scoring campaign of experts listening to the extracted excerpts is required. For a complete evaluation, experts should also listen to the parts of the recordings that were classified as non-snoring, in order to evaluate with respect to false negatives. In order to provide a quick performance estimate, we opted to score 100 excerpts that were classified as snoring from each whole-night recording. Statistics of the resulting sound excerpts classified as snoring are tabulated in Table 3. Out of 2500 tested excerpts, a total of 2032 (or about 80%) were scored as snoring. However, this figure should not be translated to accuracy or TPR, since accuracy and TPR assume that the denominator (i.e., total number of true snoring excerpts and true non-snoring excerpts) of the ratios is known. Notwithstanding, 80% is an encouraging ratio in the case where we are solely interested in providing a database of actual snoring excerpts. Furthermore, it is also expected that during a whole-night recording there are large periods of non-snoring (actually, this should be the default state). As such, even a small percentage of FPR shall produce a large number of non-snoring excerpts that are classified as snoring.

**Table 3.** Performance results of the proposed real-time snore detector.

| Number of Sound Excerpts Classified as Snoring | Subset of Sound Excerpts That Were Manually Scored by Experts | True Positive Ratio within Subset of Sound Excerpts (Number of True Positive Excerpts) |
|---|---|---|
| 12,090 (total duration equal to 20 h, 9 min) | 2500 (100 excerpts from each whole-night sound recording) | ~80% (2032 excerpts out of 2500) |

## 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 5. 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 5.** Architecture of a cascaded neural-network classification tool.

## References

1. Qaseem, A.; Holty, J.-E.C.; Owens, D.K.; Dallas, P.; Starkey, M.; Shekelle, P. Management of Obstructive Sleep Apnea in Adults: A Clinical Practice Guideline From the American College of Physicians. *Ann. Intern. Med.* **2013**, *159*, 471–483. [CrossRef] [PubMed]
2. Pack, A.I. Advances in Sleep-disordered Breathing. *Am. J. Respir. Crit. Care Med.* **2006**, *173*, 7–15. [CrossRef] [PubMed]
3. APNEA Project: "Automatic Pre-Hospital and In-Home Screening of Sleep Apnea", Operational Programme "Competitiveness, Entrepreneurship and Innovation. Available online: http://apnoia-project.gr/ (accessed on 1 April 2021).
4. Korompili, G.; Amfilochiou, A.; Kokkalas, L.; Mitilineos, S.A.; Tatlas, N.-A.; Kouvaras, M.; Kastanakis, E.; Maniou, C.; Potirakis, S.M. Poly-SleepRec: A scored polysomnography dataset with simultaneous audio recordings for sleep apnea studies. *Nat. Sci. Data* **2021**, *8*, 1–13.
5. Karunajeewa, A.S.; Abeyratne, U.; Hukins, C. Multi-feature snore sound analysis in obstructive sleep apnea–hypopnea syndrome. *Physiol. Meas.* **2010**, *32*, 83–97. [CrossRef] [PubMed]
6. Ben-Israel, N.; Tarasiuk, A.; Zigel, Y. Obstructive Apnea Hypopnea Index Estimation by Analysis of Nocturnal Snoring Signals in Adults. *Sleep* **2012**, *35*, 1299–1305. [CrossRef] [PubMed]
7. Azarbarzin, A.; Moussavi, Z. Snoring sounds variability as a signature of obstructive sleep apnea. *Med. Eng. Phys.* **2013**, *35*, 479–485. [CrossRef] [PubMed]
8. Arsenali, B.; van Dijk, J.; Ouweltjes, O.; Brinker, B.D.; Pevernagie, D.; Krijn, R.; Van Gilst, M.; Overeem, S. Recurrent Neural Network for Classification of Snoring and Non-Snoring Sound Events. In Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 18–21 July 2018. [CrossRef]
9. Emoto, T.; Abeyratne, U.R.; Akutagawa, M.; Nagashino, H.; Kinouchi, Y.; Karunajeewa, S. Neural networks for snore sound modeling in sleep apnea. In Proceedings of the 2005 IEEE International Conference on Computational Intelligence for Measurement Systems and Applications, Messian, Italy, 20–22 July 2005. [CrossRef]
10. Kang, B.; Dang, X.; Wei, R. Snoring and apnea detection based on hybrid neural networks. In Proceedings of the 2017 International Conference on Orange Technologies (ICOT), Singapore, 8–10 December 2017. [CrossRef]
11. Mitilineos, S.A.; Tatlas, N.-A.; Potirakis, S.M.; Rangoussi, M. Neural Network Fusion and Selection Techniques for Noise-Efficient Sound Classification. *J. Audio Eng. Soc.* **2019**, *67*, 27–37. [CrossRef]
12. Mitilineos, S.A.; Potirakis, S.M.; Tatlas, N.-A.; Rangoussi, M. A Two-Level Sound Classification Platform for Environmental Monitoring. *J. Sens.* **2018**, *2018*, 1–13. [CrossRef]
13. Dafna, E.; Tarasiuk, A.; Zigel, Y. Automatic Detection of Whole Night Snoring Events Using Non-Contact Microphone. *PLoS ONE* **2013**, *8*, e84139. [CrossRef] [PubMed]
14. Zhang, Z.; Han, J.; Qian, K.; Janott, C.; Guo, Y.; Schuller, B. Snore-GANs: Improving Automatic Snore Sound Classification with Synthesized Data. *IEEE J. Biomed. Health Inform.* **2019**, *24*, 300–310. [CrossRef] [PubMed]
15. Wang, C.; Peng, J.; Song, L.; Zhang, X. Automatic snoring sounds detection from sleep sounds via multi-features analysis. *Australas. Phys. Eng. Sci. Med.* **2016**, *40*, 127–135. [CrossRef] [PubMed]
16. Tuncer, T.; Akbal, E.; Dogan, S. An automated snoring sound classification method based on local dual octal pattern and iterative hybrid feature selector. *Biomed. Signal Process. Control.* **2020**, *63*, 102173. [CrossRef] [PubMed]
17. Agrawal, S.; Stone, P.; McGuinness, K.; Morris, J.; Camilleri, A. Sound frequency analysis and the site of snoring in natural and induced sleep. *Clin. Otolaryngol.* **2002**, *27*, 162–166. [CrossRef] [PubMed]
18. Saunders, N.; Tassone, P.; Wood, G.; Norris, A.; Harries, M.; Kotecha, B. Is acoustic analysis of snoring an alternative to sleep nasendoscopy? *Clin. Otolaryngol.* **2004**, *29*, 242–246. [CrossRef] [PubMed]