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

Stelios A. Mitilineos, Nikolas-Alexander Tatlas, Georgia Korompili, Labros Kokkalas, and Stelios M. Potirakis

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 in-creased cardiovascular morbidity and mortality, while an estimated 4% and 2% of the male and female population respectively suffer from OSAHS. The APNEA research project aims at accurately and cost-efficiently screening patients at home, using sound recordings via the users' smartphone during sleep. To achieve this goal, we are collecting polysomnography data together with time-synchronized and high quality tracheal and ambient microphone recordings in a large number of patients (insofar, the acquired database consists of more than 200 complete polysomnography studies and synchronized microphone recordings). In this context, and inspired by literature findings that link snoring to OSAHS episodes we developed and herein present a Real-Time Snore Detector (RTSD) in order to use it for pre-screening of micro-phone 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 al-lowing to the latter to focus on timeslots of increased OSAHS probability.

SNORE CLASSIFIERS AND OUR CONTRIBUTION

As long as snore classifiers are concerned, we are focusing 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. 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 while 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 together with an extremely large body of sound excerpts that correspond to the output of the RTSD upon labelled as "snoring" – both datasets are available upon request at spoti@uniwa.gr.

ARCHITECTURE OF THE PROPOSED SNORE CLASSIFIER AND REAL-TIME SNORE DETECTOR (RTSD)

Figure 1 depicts the proposed snore classifier architecture:

- Each sound excerpt is de-noised using wavelet filtering and then is normalized with respect to its average energy.

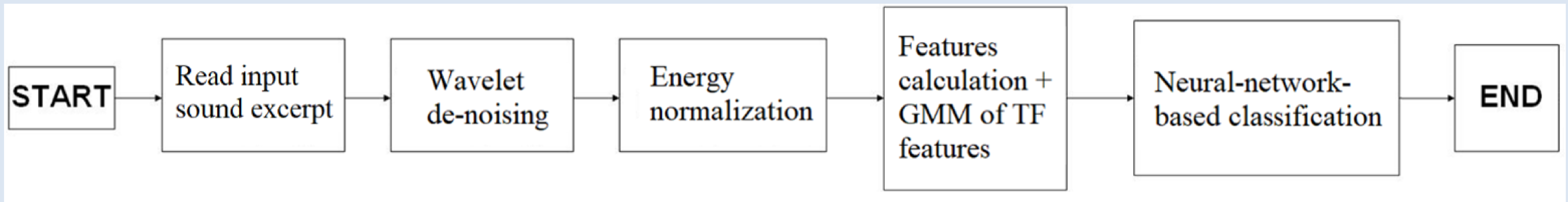

Figure 1. Architecture of the proposed classification tool and neural network.

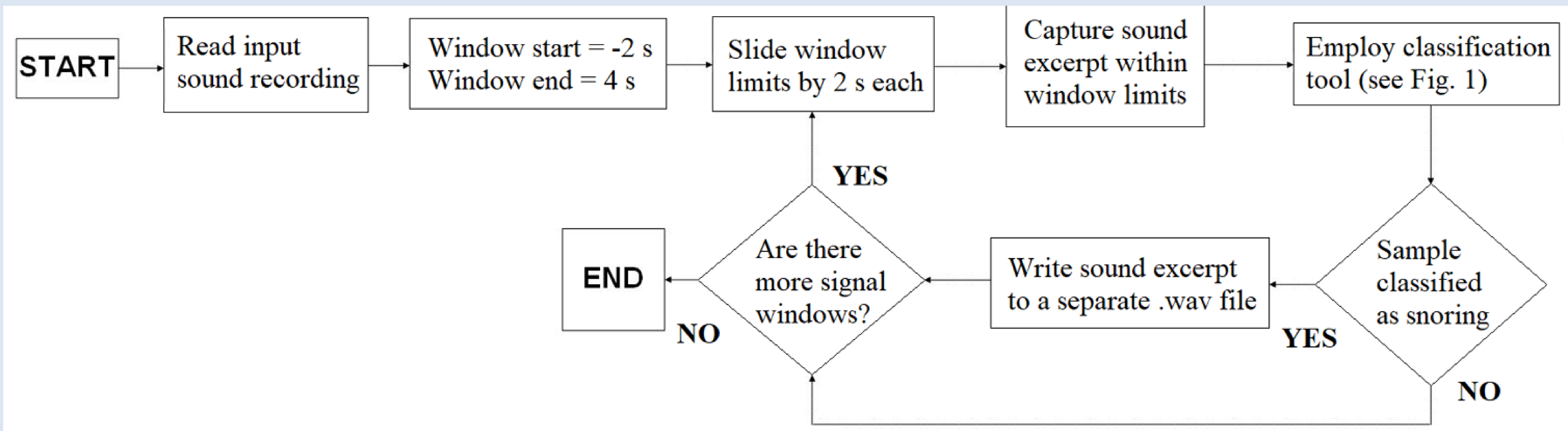

Figure 2. Architecture of the proposed Real-Time Snore Detector.

- Selected features are calculated for each sound excerpt
- A neural-network classifier is employed in order to infer whether the input sound excerpt is a snore or not

Figure 2 depicts the propsed RTSD architecture and block diagram:

- The input sound recording is parsed with a sliding window of duration 6 s and a sliding step of 2 s
- The sound within each window is classified using the neural network of Figure 1
- . If it is classified as snoring then we record the sound excerpt within the specific window to a separate .wav file for further processing.

SOUND FEATURES SELECTION

We use a large variety of sound features that are proposed in the literature and coded in-house. These include:

- Time and frequency domain static features (zero-crossing-rate (ZCR), energy, volume, pitch and bandwidth
- Time-frequency features (MFCC, Spectrogram, and a newly proposed feature that is a modified spectrogram coefficients feature (MSC for brevity)
- A set of entropy and statistics metrics (Shannon, Tsallis, wavelet and permutation entropy, and the median, average, variance, skewness and kurtosis

In order to select the optimal features combination, we performed extensive testing with different network parameters and input features. Tables 1 and 2 summarize our findings that conclude in that (i) the proposed MSC performs best among the proposed features, and (ii) a combination of MFCC and the MSC is the optimal set of sound features for snore detection in our case.

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

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

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 %	

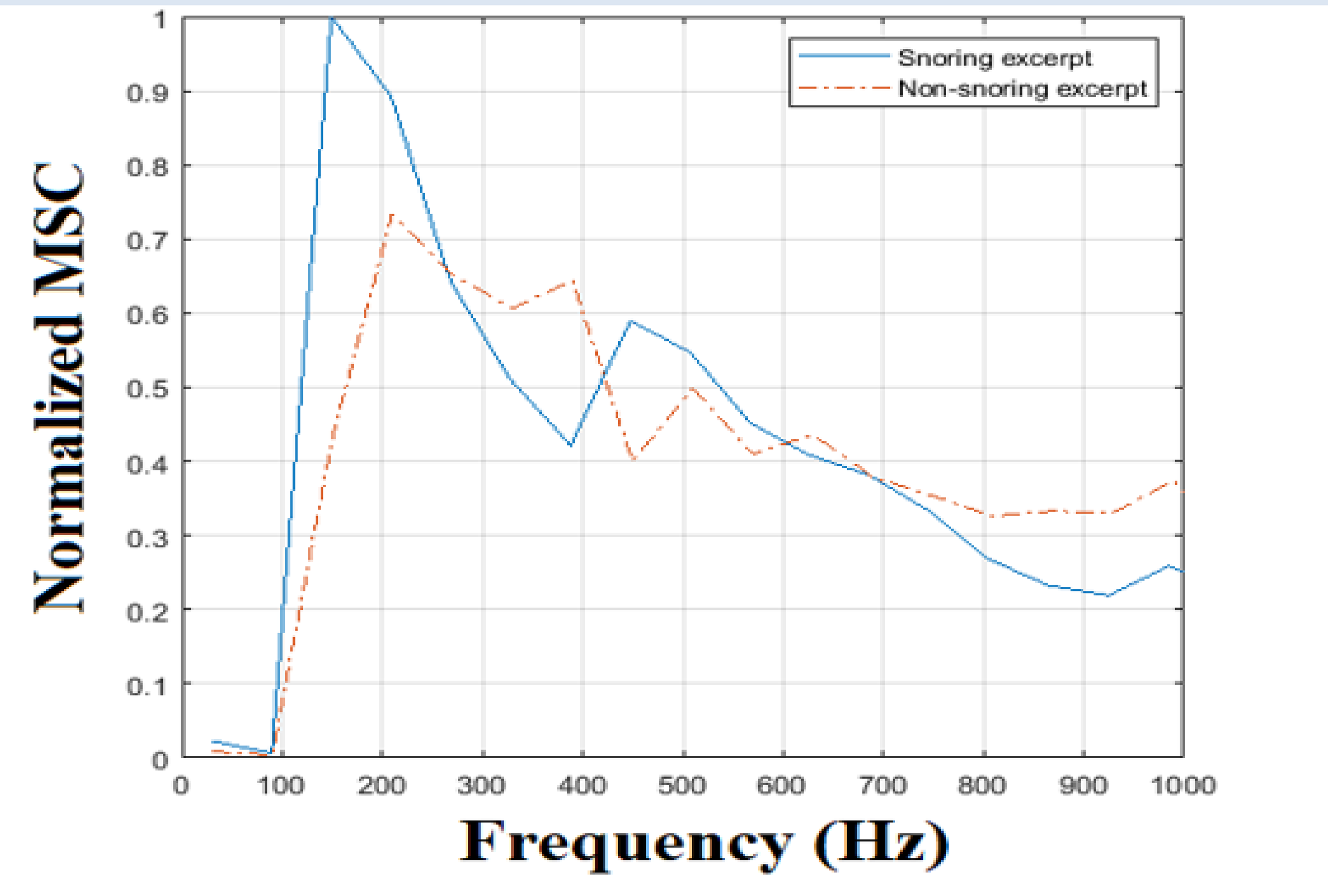

Figure 3. Modified spectral coefficients for a snoring and a non-snoring sound excerpt (solid and dash-dotted curve, respectively).

MODIFIED SPECTROGRAM COEFFICIENTS FEATURE (MSC)

We analyze the proposed MSC feature a little bit more herein. To implement this, we first calculate the spectrogram of each sliding window that is denoted in Figure 2. Then, we calculate the average spectral coefficients in adjacent, non-overlapping frequency ranges 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 the literature. On the contrary, the non-snoring excerpt exhibits a smoother distribution of energy vs. frequency.

CONCLUSIONS

A snoring classification neural network is reported with substantial performance (test set accuracy 98.6 %). Further, a modified sound feature that substantially increases performance is introduced and two large datasets are made available upon request. A RTSD is built upon the neural classifier for efficient OSAHS screening at home.