INSTITUTE OF PHYSICS PUBLISHING PHYSIOLOGICAL MEASUREMENT

Physiol. Meas. **27** (2006) 1047–1056 doi:10.1088/0967-3334/27/10/010

Automatic detection, segmentation and assessment of snoring from ambient acoustic data

W D Duckitt[1], **S K Tuomi**[2] **and T R Niesler**[1]

[1] Department of Electronic Engineering, University of Stellenbosch, Stellenbosch, South Africa
[2] Department of Speech, Language and Hearing Therapy, University of Stellenbosch, Stellenbosch, South Africa

E-mail: trn@dsp.sun.ac.za

Received 16 May 2006, accepted for publication 8 August 2006
Published 1 September 2006
Online at stacks.iop.org/PM/27/1047

Abstract

Snoring is a prevalent condition with a variety of negative social effects and associated health problems. Treatments, both surgical and therapeutic, have been developed, but the objective non-invasive monitoring of their success remains problematic. We present a method which allows the automatic monitoring of snoring characteristics, such as intensity and frequency, from audio data captured via a freestanding microphone. This represents a simple and portable diagnostic alternative to polysomnography. Our system is based on methods that have proved effective in the field of speech recognition. Hidden Markov models (HMMs) were employed as basic elements with which to model different types of sound by means of spectrally based features. This allows periods of snoring to be identified, while rejecting silence, breathing and other sounds. Training and test data were gathered from six subjects, and annotated appropriately. The system was tested by requiring it to automatically classify snoring sounds in new audio recordings and then comparing the result with manually obtained annotations. We found that our system was able to correctly identify snores with 82–89% accuracy, despite the small size of the training set. We could further demonstrate how this segmentation can be used to measure the snoring intensity, snoring frequency and snoring index. We conclude that a system based on hidden Markov models and spectrally based features is effective in the automatic detection and monitoring of snoring from audio data.

Keywords: snoring, hidden Markov models, polysomnography

(Some figures in this article are in colour only in the electronic version)

1. Introduction

Snoring can be defined as a respiratory noise that is generated during sleep when breathing is obstructed by a collapse in the upper airway. Studies have shown that it affects over 60% of adult men and 44% of women over the age of 40 (Lugaressi *et al* 1980, Dalmasso and Prota 1996). Hence, it is a highly prevalent condition and affects a substantial part of the population.

The impact of snoring varies from slight irritation and daytime drowsiness to social disharmony and in severe cases potentially life-threatening obstructive sleep apnea (OSA). Both surgical (Ikematsu 1964, Fujita *et al* 1981, Wedman and Harald 2002) and therapeutic (Tuomi 2002) treatments for heavy snoring have been developed. However, objective assessment of the condition's severity and the success of treatments remains problematic. The ultimate authority is usually considered to be polysomnography, a procedure in which a number of physiological parameters, such as oesophageal pressure, gastric pressure, EEG and EMG, are measured during sleep by specialized equipment. However, this method is expensive and intrusive, requiring overnight stays in a suitable facility. Furthermore, the discomfort caused by sensors that are physically attached to the body, as well as the unfamiliarity of the surroundings, has been shown to lead to deviations from normal sleeping patterns, drawing the relevance of polysomnographic results into question (Osman *et al* 1998). A second option often employed to gauge the severity of snoring is feedback from a sleeping partner. However, due to its intrinsic subjective nature, this type of assessment is often unreliable.

We propose the analysis of snoring activity based only on audio data gathered via a microphone. The advantage of this approach is its unobtrusiveness, portability and lower cost. Ultimately such algorithms could be integrated into bedside devices installed at the subject's home, making monitoring of the condition both more convenient as well as potentially more accurate, due to the absence of extenuating factors that can disturb normal sleep patterns. The benefits of this approach have already been pointed out by Lee *et al* (1999), using a portable audio data logger to record snoring sounds for later offline visualization and analysis. We propose to integrate a much greater processing effort into the device to allow for continuous automatic assessment.

Several authors have proposed the characterization of snoring based on its acoustic properties (Beck *et al* 1995, Dalmasso and Prota 1996, Agrawal *et al* 2002, Saunders *et al* 2004). In particular, it has been established that the site of snoring (be it tongue based or palatal) can be identified to a certain extent from the spectral properties of the audio signal. We extend these insights by incorporating spectrally based acoustic features into a statistical classification system designed to detect individual snores and reject other ambient noises. This approach allows the automatic estimation of quantities that have been suggested to be of pathological importance, such as the snoring intensity, the number of snores per hour of sleep (known as the snoring index) and the number of snores per minute of snoring (snoring frequency).

Severe snoring may be associated with obstructive sleep apnea (OSA). This condition occurs in response to a complete obstruction of the airway. Breathing ceases completely for periods exceeding approximately 10 s, followed by a sudden gasp for air and possible partial or full awakening. OSA has been found to lead to irregular heartbeats and an associated increased risk of cardiac arrest and stroke. A successful attempt to automatically determine the occurrence of OSA using sleep laboratory audio recordings has recently been reported by Abeyratne *et al* (2005). In this system, a pitch-based feature is calculated and used to determine the onset of OSA. Although the detection of apnea is a possible application of our methods, we focus in this study on the accurate detection of individual snores with the aim of automatically monitoring the snoring severity.

2. Methods

2.1. Subjects

Our experiments are based on overnight audio recordings (each approximately 8 h in length) made of six subjects, two female and four male, between the ages of 43 and 75. These six subjects are a convenience sample comprising individuals with an acknowledged snoring habit. Although the existence of apnea cannot be excluded, no subject or sleeping partner reported concern over long duration airway collapse. None of the subjects were taking medication at the time when the recordings were made, and all consented to be part of this study.

2.2. Instrumentation

A Carol Sigma Plus 5 condenser microphone with a 50–18 000 Hz frequency response was interfaced to a notebook computer via an Edirol UA-1X external USB-based 16 bit audio interface sampling at 44.1 kHz. The external interface was used to ensure a good signal-to-noise ratio, which is not possible via the PC's electrically noisy internal audio interface. The microphone was placed on the bedside table in the subject's bedroom in order to be as unobtrusive as practically possible. The PC was located in a separate room to avoid corruption of our recordings by fan or other computer-generated noise.

2.3. Data extraction

Manual screening of the recordings identified the following classes of sounds in the recordings:

(i) snoring,
(ii) breathing,
(iii) duvet noise,
(iv) silence,
(v) other noise (includes car noises, barking dogs, etc).

Periods of silence were annotated since they will be modelled explicitly by our system. Duvet noises are the sounds made by the bedlinnen when the subject moves during sleep. In future there is scope to expand this set of sounds, to include for example speech (to allow for sleep talking), a greater variety of 'other' sounds and possibly different types of snore (e.g., tongue based and palatal). For this study however we have restricted ourselves to this limited set, in order to determine the feasibility of the overall method using the restricted collection of audio recordings at our disposal.

From each of the six full 8 h recordings, the following portions were extracted for further consideration:

(a) One hour of audio containing frequent snoring, extracted from the first half of the 8 h recording.
(b) Half an hour of audio data containing frequent snoring, extracted from the second half of the 8 h recording.

The one-hour portions were intended as training data and the half-hour portions as testing data. The approximately 9 h of data extracted in this way were then manually time aligned and annotated with one of the five sound classes listed above. All labelling was carried out by the same individual. Annotation was sometimes difficult, for example when labelling very soft breathing noises or loud breathing noises that were gradually developing into snoring noises. Table 1 indicates the number of snores, breaths, duvet noises and other noises that were hand labelled in this way.

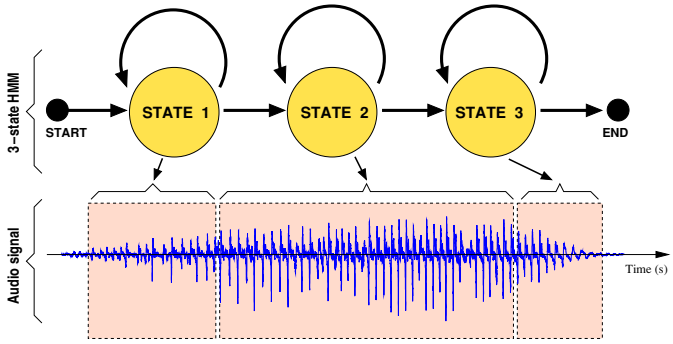

Figure 1. A three-state left-to-right hidden Markov model (HMM), as used to model snoring sounds. A possible segmentation of a single snoring sound into onset, body and end is indicated.

Table 1. Number of snores, breaths, duvet noises and other noises that were detected and annotated in the training and test data.

Sound class	Training data	Testing data
Snores	4006	1554
Breaths	3144	1046
Duvet noises	71	62
Periods of silence	6681	2740
Other noises	132	89

For ease of processing, the training and testing material was split up into files each approximately 30 s in length. The audio data were then parameterized as Mel-frequency cepstral coefficients (MFCCs) which have been found to be very effective in automatic speech recognition systems (Haeb-Umbach and Loog 1999). These parameters encode the sound in a way that mimics the function and capabilities of the human ear. Furthermore, MFCCs are fairly invariant to pitch changes, which makes them suitable for the development of a snorer-independent system. Finally, a post-processing step known as cepstral mean normalization (CMN) can be applied, which further increases the systems robustness to changes in room acoustics, microphone and snorer identity (Moreno *et al* 1995).

A set of 12 MFCCs was calculated every 10 ms using a 30 ms window of audio data. The instantaneous signal energy was appended to this 12-dimensional feature vector, after which both first and second differentials were added. This resulted in a 39-dimensional feature vector representing the signal spectrum once every 10 ms.

2.4. Data analysis

We have used hidden Markov models (HMMs) to model different types of sounds in our classification system (Rabiner 1989). A hidden Markov model is a statistical technique well suited to the modelling, classification and segmentation of sampled time series. HMMs have been employed with great success in the field of automatic speech recognition to segment audio signals into linguistic units such as phones or words. Since this has much in common with our goal of isolating snoring sounds in audio data, we have chosen HMMs as an appropriate basis for our system. Figure 1 illustrates a three-state HMM, which was used to model the sound of a single snore.

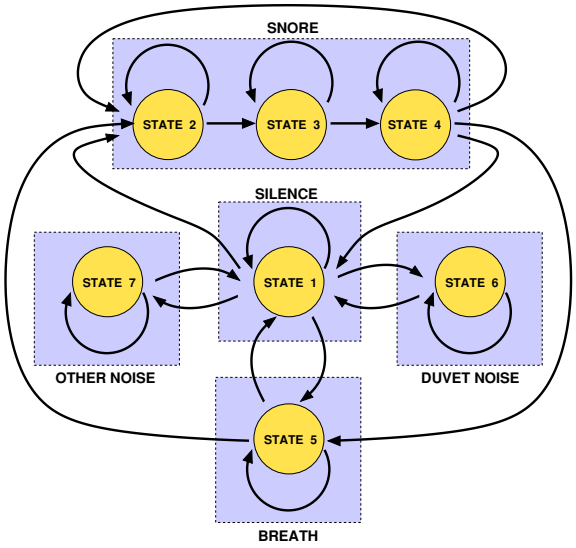

Figure 2. Hidden Markov model (HMM) network used to segment audio signals into periods of snoring, silence, breathing, duvet noise, and other noise.

The arrows denote allowable state-to-state transitions. At each new time instant (every 10 ms in our case), a path through the model is extended by following one of the links exiting the current state. There are many possible paths from the start to finish (to account for time series of varying length), and the left-to-right topology ensures a sequential progression without the possibility of skipping back in time. Each model state is associated with a probability distribution function, which describes the average qualities of a particular portion of the audio signal. For the HMMs we used, these probability density functions (PDFs) describe the typical spectral characteristics of the signal, as captured by the MFCC vectors. Hence, when a three-state left-to-right HMM is used to model the sound of a single snore, the PDF of the first, second and third states naturally describe the spectral qualities of the onset, body and end of the snore, respectively, as indicated in figure 1. Precisely how much of the snore would be ascribed to each of the three states is determined automatically during HMM training.

We chose to model snoring sounds by a three-state HMM and each of the remaining sounds identified in table 1 by single-state HMMs. Whether these are optimal choices for our application remains the subject of future work. Each HMM state employed an eight-component full covariance Gaussian mixture model (GMM) as probability density function, using the 39-dimensional MFCC-based feature vectors. The GMMs were initialized using the time-aligned training transcriptions and the k-means clustering algorithm. The model parameters were then trained by two iterations of Viterbi-based embedded reestimation (Rabiner 1989).

In order to test the ability of the system to segment new data into snores, periods of silence and the other sound classes, the five HMMs were interconnected to form a network as shown in figure 2.

This topology allows for the following eventualities, which were observed in the training data during annotation:

(i) A snore may fade to become a breath. Hence, a snore is not always followed by a silence.
(ii) A breath may develop into a snore. Hence, a snore is not always preceded by a silence.
(iii) When snoring occurs while breathing in as well as out, a snore may be followed immediately by another, with a possible separating silence.

Table 2. Classification performance for system 6/6.

	Percentage classified as				
Test sound	Silence	Snore	Breath	Duvet	Other
Silence	69.0	5.3	12.7	9.0	3.8
Snore	2.9	89.0	6.7	0.3	1.1
Breath	7.1	13.6	73.1	4.3	1.9
Duvet noise	23.3	3.9	10.7	61.2	1.0
Other noise	12.0	18.5	13.9	25.0	30.5

Table 3. Classification performance for system 3/3.

	Percentage classified as				
Test sound	Silence	Snore	Breath	Duvet	Other
Silence	48.7	9.3	14.4	25.3	2.2
Snore	4.4	82.2	11.9	1.1	0.2
Breath	18.3	7.5	42.3	29.8	1.9
Duvet noise	30.4	8.6	15.2	41.3	4.3
Other noise	20.0	35.0	13.8	18.5	12.3

Finally, the test-set audio data were segmented into periods of snoring, silence, breathing, duvet noise and other noise using the network shown in figure 2 and the Viterbi algorithm. This resulted in a time-aligned segmentation of the audio data in terms of these sounds.

3. Results

Since data from only six subjects were available, we chose to develop and test our system in two different ways:

(a) *System 6/6*. Train the HMMs on the training data from all six subjects and test on the test data of all six subjects. Although the testing and training data are taken from different portions of the recording, this approach uses the same snorers in the training as well as the test set. However, it makes best use of the limited available data.

(b) *System 3/3*. Train using the training material of three subjects and test using the testing material of the other three subjects. This guarantees complete independence of training and testing material, but fragments already limited data.

The performance of the two systems was determined by segmenting the test sets using the network of HMMs shown in figure 2 and comparing the result with the manually produced transcription of the same data. The classification results are summarized in tables 2 and 3. Each row represents a true sound class, as identified by the manual annotations of the test set. The columns indicate the system's classification decisions. For example, the first row of table 2 indicates that 69% of all silences were classified as silences and that 5.3% were incorrectly classified as snores.

From tables 2 and 3 it is apparent that 89% of snores are identified correctly by system 6/6 and 82.2% by system 3/3. The superior performance of system 6/6 indicates that better performance is possible when more training data are available and also when data from the snorer in question are available.

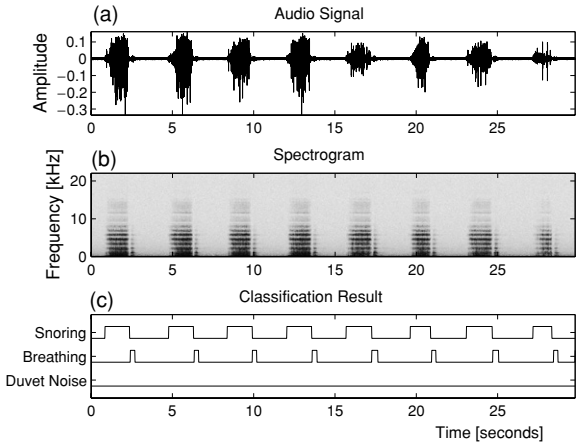

Figure 3. Segmentation of a 30 s audio signal into snores, breaths and duvet noises. The audio signal is shown in (a), the associated spectrogram in (b) and the classification decisions in (c).

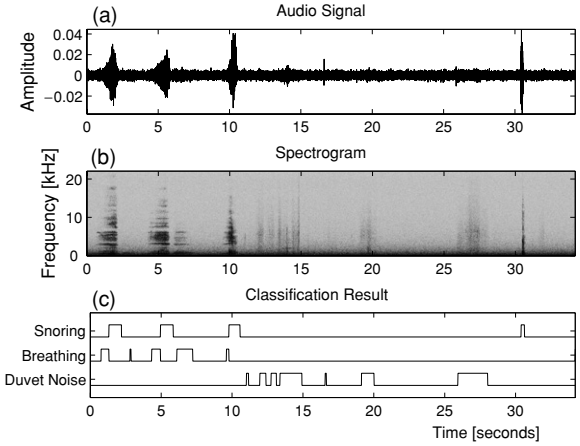

Figure 4. Segmentation of a second 30 s audio signal, indicating correctly identified duvet noises. The audio signal is shown in (a), the associated spectrogram in (b) and the classification decisions in (c).

Figure 3 illustrates the operation of the system for a particular 30 s audio signal consisting of a regular succession of snores, breaths and periods of silence. Snoring is pronounced and occurs during inspiration, while the breathing noises are much softer and heard during expiration. No duvet noises are detected in this audio segment.

Figure 4 shows a 30 s audio signal from a different subject. In this case, snoring is much softer than it was in figure 3, accounting for the higher amplitude of the background noise. Furthermore, for this subject an initial breathing noise develops into a snoring sound during inspiration, after which a softer breathing is heard during expiration. The exhaled breath is barely audible after the first snore in the figure, but louder after the second. After the third snore the subject changes sleeping position, leading to a cessation of snoring and breathing

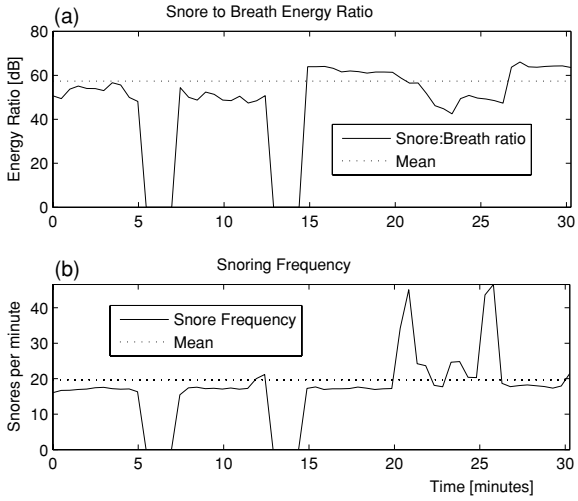

Figure 5. Snoring loudness (a) and snoring frequency (b) calculated over a 30 min interval.

noises, and the production of duvet noise, which is in this case correctly identified by the system. Finally, snoring resumes at approximately 31 s.

4. Discussion

Although system 6/6 exhibits the better accuracy, the good performance of system 3/3 is very encouraging in the light of the small number of snorers in the training set. From tables 2 and 3 it is evident that snores were most commonly misclassified as breaths. As already pointed out, it is sometimes very difficult to decide where loud breathing ends and snoring starts, and therefore some of these misclassifications may be borderline cases. Periods of silence are most often misclassified by the system as breaths or duvet noises. Similarly, duvet noises are often misclassified as silence or breath noise. This is due to the similarity of soft breathing noises, soft duvet noises and periods that have been labelled as silence. However, a comparison between tables 2 and 3 shows that these misclassifications are reduced when the amount of training data are increased.

System 3/3 was less effective at discerning between silence, duvet noise and breathing sounds than system 6/6. These accuracies may be expected to improve if more training data become available. Finally, the category 'other noise' was not well recognized by either system, although again system 6/6 showed an improvement over system 3/3. Future overall performance may benefit from additional categories, such as car noise, animal noises, etc.

As a last step, the segmentation of the audio signal into the five sound classes was used to calculate the following two indicators of the severity of snoring:

(a) The loudness of the snore measured relative to the loudness of a breath, calculated using a 1 min sliding window at 30 s intervals. The relative measure makes this quantity less affected by changes in overall loudness caused, for example, by changes in the subject's sleeping position.
(b) The frequency of snores, again calculated over a 1 min sliding window at 30 s intervals. Both the snoring frequency and the snoring index can be calculated from this signal.

In figure 5, these two quantities are plotted for one of the 30 min test sets.

Periods of snoring can be clearly identified from both the snoring loudness and frequency plots. The peaks in the snoring frequency indicate periods during which the subject snored during exhalation as well as inhalation. The snoring frequency can be used to determine the snoring index, by determining the total number of snores per hour of recording. The average loudness as well as snoring frequency or index, measured regularly over an extended time period, could be a useful indicator of snoring severity and be used to judge the effectiveness of a treatment or therapy.

5. Conclusion

We have demonstrated that hidden Markov models and acoustic parameterizations which are effective in speech recognition systems are able to identify snores in audio recordings with 82–89% accuracy. This good performance was achieved despite relying on a rather small data set with which to train the parameters of the system. We have further shown how quantities such as the snoring index, frequency and intensity can be calculated from these classifications. We conclude that these are promising methods with which to develop unobtrusive and cost effective automatic snore monitoring systems that can be used as diagnostic tools to form reliable and objective opinions about the severity of snoring and effectiveness of treatment. Moreover, since our method does not retain the audio signal, but only calculates statistics and averages, patient confidentiality concerns are addressed.

Acknowledgment

This work was supported by the South African National Research Foundation (NRF) under grant number FA2005022300010.

References

Abeyratne U, Wakwella A and Hukins C 2005 Pitch jump probability measures for the analysis of snoring sounds in apnea *Physiol. Meas.* **26** 779–98

Agrawal S, Stone P, McGuinness K, Morris J and Camilleri A 2002 Sound frequency analysis and the site of snoring in natural and induced sleep *Clin. Otolaryngol.* **27** 162–6

Beck R, Odeh M, Oliven A and Gavriely N 1995 The acoustic properties of snores *Eur. Respir. J.* **8** 2120–8

Dalmasso F and Prota R 1996 Snoring: analysis, measurement, clinical implications and applications *Eur. Respir. J.* **9** 146–59

Fujita S, Conway W, Zorick F and Roth T 1981 Surgical correction of anatomic abnormalities in obstructive sleep apnea syndrome: uvulopalatopharyngoplasty *Otolaryngol. Head Neck Surg.* **89** 923–4

Haeb-Umbach R and Loog M 1999 An investigation of cepstral parameterisations for large vocabulary speech recognition *Proc. Eurospeech (Budapest, Hungary)* pp 1323–6

Ikematsu T 1964 Study of snoring. IVth report therapy *J. Japan. Otorhinolaryngol.* **64** 334–5

Lee B, Hill P, Osbourne J and Osman E 1999 A simple audio data logger for objective assessment of snoring in the home *Physiol. Meas.* **20** 119–27

Lugaressi E, Cirignotta F, Coccagna G and Piana C 1980 Some epidemiological data on snoring and cardiocirculatory disturbances *Sleep* **3** 221–4

Moreno P, Raj B, Gauvea E and Stern R 1995 Multivariate Gaussian-based cepstral normalisation for robust speech recognition *Proc. ICASSP (Detroit, USA)* pp 137–40

Osman E, Osbourne J, Hill P and Lee B 1998 Snoring assessment: do home studies and hospital studies give different results? *Clin. Otolaryngol.* **23** 524–7

Rabiner L 1989 A tutorial on hidden Markov models and selected applications in speech recognition *Proc. IEEE* **77** 267–96

Saunders N, Tassone P, Wood G, Norris A, Harries M and Kotecha B 2004 Is acoustic analysis of snoring an alternative to sleep nasendoscopy? *Clin. Otolaryngol.* **29** 242–6

Tuomi S 2002 Snoring therapy—a new area of expertise for speech-language pathologists *Annual Convention of American Speech-Language-Hearing Association (Atlanta, USA)*

Wedman J and Harald M 2002 Treatment of simple snoring using radio waves for ablation of uvula and soft palate: a day-case surgery procedure *Laryngoscope* **112** 1256–9