Biomedical Signal Processing and Control 48 (2019) 1–11

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Detecting breathing rates and depth of breath using LPCs and Restricted Boltzmann Machines☆

Eric Ehrhardt Hamke*, Manel Martínez-Ramón, Amir Raeisi Nafchi, Ramiro Jordan

Department of Electrical and Computer Engineering, University of New Mexico, 1 University of New Mexico, Albuquerque, NM 87131-0001, USA

ARTICLE INFO

Article history:
Received 20 February 2018
Received in revised form 10 August 2018
Accepted 27 September 2018
Available online 10 October 2018

Keywords:
First responders
Detection
Prediction of breathing rates and depth or length of breath

ABSTRACT

This paper presents the use of a Restricted Boltzmann Machine to develop an unsupervised machine learning approach to process breathing sounds to predict breathing rates and depth or length of breaths. Breath detection and monitoring has been the subject of several studies involving the health monitoring of patients on respirators. We are proposing to extend the use of non-invasive techniques to provide measures of physical exhaustion or activity. The level of activity or exhaustion could be used to prevent accidents or manage exposure to physically demanding environments such as firefighting or working underwater.

1. Introduction

The use of non-invasive techniques to provide measures of physical exertion or activity can be used to prevent accidents or manage exposure to physically demanding environments such as firefighting or others as working underwater. This work presents two alternative methods for predicting breathing rates and depth or length of breaths. Both methods use a recording of the regulator from the self-contained breathing apparatus (SCBA).

The first method uses the concept that the regulator sound is a colored white noise source. By fitting Linear Predictive Coefficients (LPC) all pole filter to the regulator sound, it is possible to transform a white noise source into a regulator colored sound. Using the poles as zeros (inverting the filter) will amplify frequency elements not used in coloring the white noise and minimally attenuate the spectral peaks of the sound. Using this spectral information and the filter gain (ratio of filtered to unfiltered) it is possible to recognize a regulator sound. The process of recognizing the sound can be automated using an unsupervised classifier based on probability density mixture models and hypothesis testing. The recognition filter is updated using a Levinson–Durbin algorithm on signals identified as a regulator sound. Thus, the filter adapts to better track changes in the systems being used and provides flexibility to allow the system to adapt to different users of the same pieces of equipment.

The second approach uses an unsupervised deep learning approach to classify the respirator sounds. The deep learning method uses a Restricted Boltzmann Machine (RBM) and the Euclidean distance between two adjacent frames. The overlap between frames causes the changes in the distance measure to lag behind the changes on the observed signal. The lag is such that frames with similar trends will have a low distance measure. Frames of a respirator sound (colored noise) will have a greater distance since the new values in the next frame are driven by a random process. This information combined with the normalized spectral power estimates allows for a robust classification. A periodic retraining of the RBM allows for adaption. As stated before, this is essential to account for differences between users and equipment.

2. Previous works

Medical applications have focused in determining breathing rates of patients on respirators [1,2], where the data is collected through a network of strain gauges that measure chest movement, and that may be susceptible of variations due to the body movements. The data is then processed using an Artificial Neural Network (ANN) using a supervised training approach. This learning approach requires a large enough training set to account for all possible variations.

Using capnography [3] is another method in the medical arena. This method uses a nose sensor placed just below the nostril and a series of gas sensors to sense O_2 and CO_2. Inhalations are detected by rapid changes between these two gasses. In a SCBA mask these

☆ This work has been supported by NSF S&CC EAGER grant 1637092.
* Corresponding author.
E-mail address: ehamke@unm.edu (E.E. Hamke).

https://doi.org/10.1016/j.bspc.2018.09.009
1746-8094/

gasses may be difficult to measure without contamination for the firefighting environment. The use of a nose sensor is intrusive, and the gas sensors would need constant recalibration. A primary reason these are not used by the fire departments in general.

The use of a temperature sensor placed at the base of the nose, as discussed in Patel et al. [4] measure breath by looking at changes resulting from expelling air warmed by the body and cooler air inhaled from the environment. The use of a nose sensor mounted in the mask is intrusive. Because of the firefighting environment and the nature of the face mask the temperature differential may not be as pronounced.

Another method, proposed by Güder [5], looks at using a hydrometer to measure the moisture content in the breath. The change in moisture results from the lungs using water to facilitate oxygen exchange. This is intrusive since the hygrometer would have to be placed near the noise and mouth. Further complications of measurement are the confinement of the SCBA mask and the use of fire hoses spraying water into the air.

The method proposed by Li et al. [6] involves speech samples collected from sitting subjects in a quiet environment. Breathing cycles are determined using the gradient of the recorded speech signals to determine the start of breathing cycles. The speech introduces additional quasi-cycles and changes to the nominal breathing rate. Li automated the duration detection using the local maxima and minima of the signal and the time duration between them. The duration of an inhalation is then found as the length of time the noise is present in the recording.

The intention of other speech related applications is to identify breathing artifacts and remove them from the speech using primary features based on cepstral or Mel Frequency Cepstral Coefficients (MFCC). Price et al. [7] automatic breath detection methods of speech signals were based on cepstral coefficients and used the Gaussian Mixture Model (GMM) as the classifier. Price achieved a detection rate of 93% [7], Wightman and Ostendorf algorithm uses a Bayesian classifier and achieved a detection rate of 91.3% [8].

Our intention in this work is to implement and test fast, unsupervised methods that are suitable specifically for this application. Unsupervised classification is usually based on clustering methods, that adjust a model for the data likelihood given a set of latent variables, each one representing each one of the clusters. All these methods can be explained in terms of graphical probabilistic models [9]. The unsupervised classification is performed by maximizing the class conditional likelihood. *K*-means, for example, starts by assigning k mean values to k possible latent variables, and it updates each one of these means by computing the mean value of the samples whose distance to that mean is the minimum. Gaussian Mixture Models (GMM) are a more general method that uses the Expectation Maximization (EM) algorithm in order to iteratively update the mean vector and covariance matrix of all the likelihoods conditional to each latent variable by averaging the samples weighted with the posterior probability of the latent variable given the sample. This can be viewed as a generalization of the *K*-means. GMM is adequate when the data is organized in clusters that can be easily approximated with Gaussian functions, thus they need at least to have some central symmetry around their based on means. In other cases, the Gaussian functions can be changed by others that fit better the shape of non-symmetric conditional distributions for which the parameters can still be inferred by Expectation Maximization. This is the case of the Gamma functions, which have been used to model the distribution of the data used in this paper.

In some cases, the distributions are not easy to approximate with relatively simple functions like Gaussians or Gamma functions. When this is the case and the EM cannot be constructed, variational inference is needed, which results in more computationally complex algorithms, that preclude the real time usability of the procedure [10]. Alternative methods that produce approximations less dependent of a probabilistic class of functions include the use of copulas [11]. These models can also be interpreted as graphical probabilistic models, but their training computational burden is high compared to the previous models.

A way to represent the data in a non-model dependent way is the use of Restricted Boltzmann Machines (RBM) and their Multilayer generalizations, the Deep Boltzmann Machines (DBM). These methods have structures that are identical to those of the standard Multilayer Perceptron, and they can be trained in an unsupervised way. The training is based on gradient descent to optimize the so-called Contrastive Divergence [12,13], and their computational burden is comparable these of GMM if the number of layers and nodes are moderate.

Of the two approaches used in our paper, first one uses a feature extraction based on an LPC process, which seems to have a distribution that fits Gamma functions properly. The second approach uses a different type of feature extraction, for which the RBM produces competitive results with comparable computational burden.

The principal advantages of our methodology with respect to the existing works are the following:

- The final system will process the regulator noise recorded through the SCBA mask's microphone using a small digital signal processing device with a Bluetooth modem.
- Both the MAP and RBM are unsupervised training methods and do not require training sets.
- Firefighters are moving and performing physically demanding tasks. Their exertion will also create other changes in the breathing rate.
- Our recordings for the study included non-speech and breath related sounds, gas powered fan noise, breaching of entryways, water leaving the nozzle, and alarm sounds.

  Like in [6], we will also look at the interval between start of breath of events. The time difference between starts can be likened to an instantaneous measure of breathing rate. The medical community uses an average rate of breath over a 1-min interval.
- The assumption of Gaussian distributions degrades the performance since the energy is non-negative and finite.
- We use a Gamma Mixture Model (GaMM), that provides a measurably better fit to the observed data than GMM. The use of the GaMM enhances the algorithm proposed by Kushner, Harton and Novorita which proposed the use of a fixed threshold chosen in advance. The GaMM allows for an adaptive thresholding process and is a more robust recognition filter [14].

3. Acoustic properties and effects of SCBA masks

The investigation focus on monitoring the breathing rate of the firefighter using a SCBA. The SCBA system is a pressure-demand air delivery system. When a user inhales, negative pressure within the mask causes the regulator valve to open. Pressurized air enters the mask producing a loud broadband hissing noise. This broadband signal is incoherent and it has a mostly flat bandwidth between 500 Hz and 5 KHz. The mask (Fig. 1) is a rigid structure with a clear plastic face plate. The mask includes a flexible rubber seal that contacts the forehead temples, cheeks and chin of the wearer. The SCBA system noise also include low air alarms [14]. Previous measurements in the context of this work show that the aggregated sonises noise have spectral peaks at 2.6 kHz, 0.3 kHz, 4.5 kHz, 3.5 kHz, 0.9 kHz, and 1.8 kHz (with the strongest peak at 2.6 kHz and the weakest at 1.8 kHz).

Fig. 1. External and internal views of a commonly used SCBA mask showing the voicemitter port [14].

4. Methodology

The movement of air in from the tank makes a very distinctive sound. We examined two methodologies for detecting and measuring the intervals between sounds (instantaneous breathing rate) and the length of time (duration) the recorded signal $s[n]$ is present. The first method is referred to as the Feature Extraction using LPC. The second method discussed involves using a RBM for detection.

4.1. LPC-GaMM classifier

4.1.1. LPC model

As stated in the introduction, the regulator noise can be modeled as a colored noise source. The set of LPC determines the characteristics of the transfer function that colors or shapes the white noise to look like the regulator noise. This model supports the use of LPC to represent the noise as a filter. An initial set of filter coefficients can be generated from an initial off-line training data set [14]. As mentioned in the article, these coefficients will be updated each time a breathing event is detected. The following equation expresses the transfer function of the filter model in the z-domain,

$$V(z) = A\frac{1}{\sum_{k=1}^{N} a_k z^{-k}} \tag{1}$$

where $V(z)$ is the transfer function used to color the white noise source. A is the gain of the filter and $\{a_k\}$ is a set of auto-regression coefficients found by LPC algorithms. The upper limit of summation, N, is the order of the all-pole filter. There are many methods to determine the LPC coefficients but the most efficient being the Levinson–Durbin algorithm.

Once the coefficients have been determined by fitting the LPC filter to an offline training set [14], the filter is inverted so that we have a Finite Impulse Response (FIR) or moving average (MA) filter which constitutes the noise "recognition" filter $\Lambda(z)$.

$$\Lambda(z) = \sum_{k=1}^{N} a_k z^{-k} \tag{2}$$

The recognition filter then generates an estimate of the input signal. Ideally, in this application, the gain or relative energy measure of the estimated input to the actual measured output is computed by

$$x_{frame} = \left(\frac{\mathrm{rms}\,(s[n] * \Lambda[n])}{\mathrm{rms}(s[n])}\right)^2\Bigg|_{frame} \tag{3}$$

where operator rms$(\cdot)$ computes the root mean square value of a window of m samples of a given sequence. The relative energy measure can be thresholded to determine both breathing interval and breath duration. In the above mentioned work [14] this threshold was selected based on observation of the data and engineering experience. They indicated that the positive detection events would trigger a moving average filter on the recognition filter coefficients. Instead of using a threshold, we make use of a GaMM classifier.

4.1.2. GaMM detection criteria

It is implemented by using an EM algorithm where the data is modeled by 2 classes of events. The first event class, C_N, is a non-regulator noise event (speech, background noises, and recording noise). The second class of events, C_R, addresses the regulator noise event. Each of these events is associated to data that is modeled with a Gamma distribution. These distributions have two fitting statistics: a shape parameter, k, and a scale parameter, θ, given by

$$\hat{k} = \frac{\bar{x}^2}{\sigma^2} \tag{4}$$

$$\hat{\theta} = \frac{\sigma^2}{\bar{x}} \tag{5}$$

where $\bar{x}$ represents the sample mean and σ^2 is the sample variance of the data. These expressions are manipulations of the expectation functions for the mean, $E[x] = k\theta$, and variance, $Var[x] = k\theta^2$ of the gamma distribution.

The Expectation Maximization algorithm [9] first performs a soft classification of each training data point by finding the probability contributions from each event's distribution. Using Bayes' rule we can write the posteriors

$$\hat{p}(C_N|x) \propto \underset{N}{\hat{\rho}}\Gamma(x|\hat{k}_N, \hat{\theta}_N) \tag{6}$$

$$\hat{p}(C_R|x) \propto \underset{R}{\hat{\rho}}\Gamma(x|\hat{k}_R, \hat{\theta}_R) \tag{7}$$

where $\hat{\rho}_N = p(C_N)$ and $\hat{\rho}_R = p(C_R)$ are the prior probabilities of the events, and $\hat{k}_N, \hat{\theta}_N$ are the fitting parameters for the non-regulator noise class or event and $\hat{k}_R, \hat{\theta}_R$ are the parameters for the regulator noise event.

These probabilities are then normalized to have probability mass function properties. The next phase of the calculation uses normalized soft classification values as an estimate of the probabilities $\hat{p}(C_N|x)$, $\hat{p}(C_R|x)$ of the data point belonging to each class of events. Since, the mean is defined as the sum of the products of the probabilities of occurrence of the data point and the point's value we get,

$$\bar{x}_N = \frac{\sum_{k=0}^{N}\hat{p}(C_N|x_k)x_k}{\sum_{k=0}^{N}\hat{p}(C_N|x_k)} \tag{8}$$

$$\bar{x}_R = \frac{\sum_{k=0}^{N}\hat{p}(C_R|x_k)x_k}{\sum_{k=0}^{N}\hat{p}(C_R|x_k)} \tag{9}$$

The variances are also estimated using the definition of its expectation as follows

$$s_N^2 = \frac{\sum_{k=0}^{N}\hat{p}(C_N|x_k){x_k}^2}{\sum_{k=0}^{N}\hat{p}(C_N|x_k)} - \bar{x}_N^2 \tag{10}$$

$$s_R^2 = \frac{\sum_{k=0}^{N}\hat{p}(C_R|x_k){x_k}^2}{\sum_{k=0}^{N}\hat{p}(C_R|x_k)} - \bar{x}_R^2 \tag{11}$$

The resulting calculation values are then used to compute the estimated fitting statistics using Eqs. (4) and (5) for the next iteration.

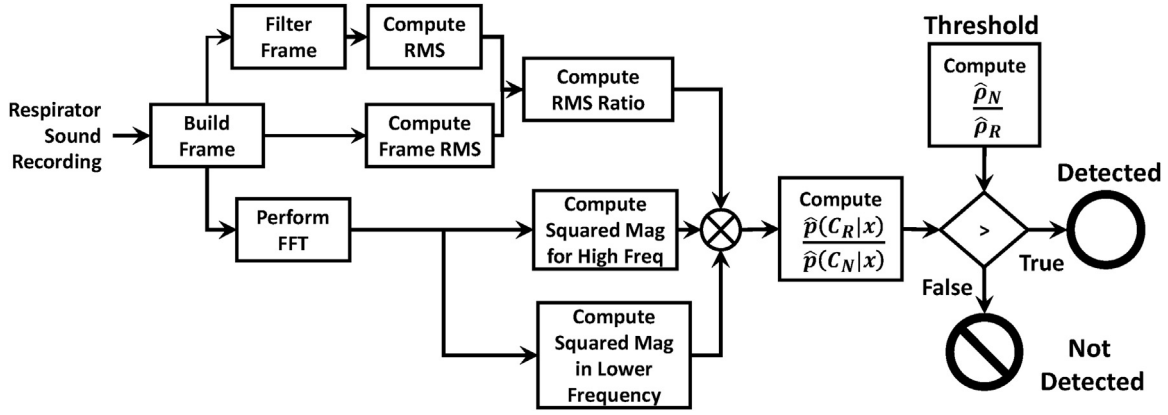

Fig. 2. The recording is broken into 6.25 ms frames with 6.125 ms overlap and processed by computing the measure for the frame and then its probability ratio.

The prior parameters are then updated by marginalizing the posteriors as

$$\hat{\rho}_N = \frac{1}{N}\sum_{k=0}^{N}\hat{p}(C_N|x_k) \tag{12}$$

$$\hat{\rho}_R = \frac{1}{N}\sum_{k=0}^{N}\hat{p}(C_R|x_k) \tag{13}$$

Finally, the difference or change between the last set of estimated values of the fitting parameters and the current set of parameters is computed. When this net change falls below the threshold the iterative process is halted. The resulting parameter sets for each event is then used for the classification of the remaining data not used for training.

The experimental data is classified by choosing the event with the greatest posterior probability as

$$\hat{p}(C_N|x)\underset{N}{\overset{R}{\gtrless}}\hat{p}(C_R|x) \tag{14}$$

The resulting classification rule is then implemented in a data flow as shown in Fig. 2.

4.2. Restricted Boltzmann Machine classifier

4.2.1. Structure of an RBM

A Restricted Boltzmann Machine (RBM) is a pairwise Markov Random Field [9] with layers of hidden nodes $\mathbf{h} \in \mathbb{R}^{d_h}$ and visible nodes $\mathbf{v} \in \mathbb{R}^{d_v}$ [15] restricted so that nodes within the layer are not connected (Fig. 3). In this manner, a joint probability distribution of the states of each node can be factorized, and then the learning task is tractable [16,12].

The most used configuration of the posterior probability distribution $p(h_i|\mathbf{v})$ or $p(v_j|\mathbf{h})$ of a node given the rest is a Bernoulli distribution, which assumes that the states of the nodes are binary. In the context of this work, the hidden nodes are fed with the feature vector of the noise, which is previously normalized so their components are between 0 and 1. The visible nodes are interpreted as the probability that their state is 1. The relationship between the hidden and the visible layers can be written as

$$\begin{aligned}\mathbf{v} &= \mathbf{W}\mathbf{h} + \mathbf{b}\\ \mathbf{h} &= \mathbf{W}^{\top}\mathbf{v} + \mathbf{c}\end{aligned} \tag{15}$$

where matrix $\mathbf{W} \in \mathbb{R}^{d_n \times d_h}$ is called the generative matrix, and its transpose is the recognition matrix, and where $\mathbf{b}$ and $\mathbf{c}$ are bias terms. Thus, the vector of posterior probabilities can be approximated by $p(\mathbf{v}|\mathbf{h}) = \mathrm{sigm}(\mathbf{W}\mathbf{h} + \mathbf{b})$ and $p(\mathbf{h}|\mathbf{v}) = \mathrm{sigm}(\mathbf{W}^{\top}\mathbf{h} + \mathbf{c})$ where

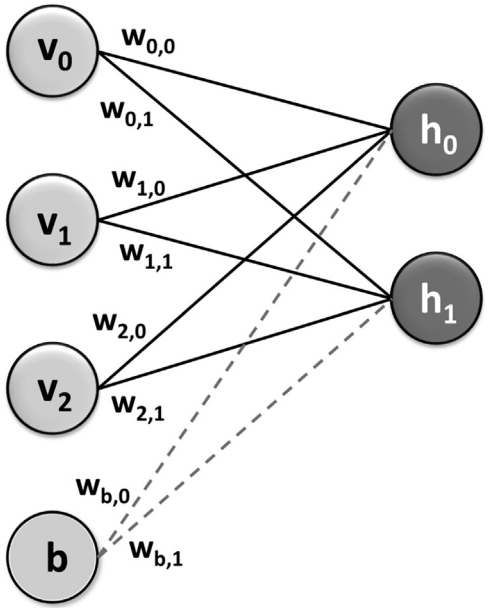

Fig. 3. The visible layer is composed of v_0, v_1, v_2 and a bias node (b). The hidden layer is composed of h_0 and h_1. Each connecting line presents a weighted connection between nodes $w_{v,h}$.

sigm is a sigmoid function. The training method proposed by Hinton [12,13] consists of reducing the so-called Contrastive Divergence between both distributions. Roughly speaking, this can be interpreted as the difference between the cross correlation matrix of the actual values of the visible and hidden nodes and the cross correlation matrix of values randomly sampled from their probability distributions. Assuming a set of normalized input patterns $\mathbf{v}_i$, $1 \le i \le N$, the training consists of computing values $\mathbf{h}_i$ for each input $\mathbf{v}_i$. Then, a set of random values $\mathbf{v}_i'$ and $\mathbf{h}_i'$ are sampled from distribution $p(\mathbf{h}|\mathbf{v}_i)$ and $p(\mathbf{v}|\mathbf{h}_i)$ and the update at iteration k is computed as

$$\begin{aligned}\Delta_{W_k} &= \mathbb{E}(\mathbf{v}\mathbf{h}^{\top}|\mathbf{v}_i) - \mathbb{E}(\mathbf{v}\mathbf{h}^{\top}) \approx \sum_i \mathbf{v}_i\mathbf{h}_i^{\top} - \sum_i \mathbf{v}_i'\mathbf{h}_i'^{\top}\\ \mathbf{W}_k &= \mathbf{W}_{k-1} + \mu\Delta_{W_k}\end{aligned} \tag{16}$$

The operation for $\mathbf{b}$ and $\mathbf{c}$ is analogous. Our implementation includes the use of two stacked RBMs (Fig. 4), which can be trained in a sequential way [17].

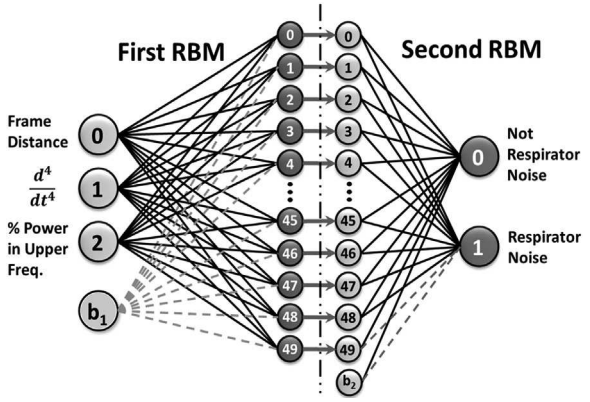

Fig. 4. Structure used to detect the respirator noise events.

4.2.2. RBM feature set

The input has three features: similarity between two adjacent frames, spectral power in the upper frequencies (above 3 kHz), and 4th order finite difference of the signal sequence. The first RBM is used to shatter the three inputs into a higher dimension space. The second RBM reassembles the information into a probability of a respirator sound detected in a frame (breathing event) and a probability of a non-respirator sound detected.

The first visible node for the visible layer is a measure of the Euclidean distance between two consecutive frames. Let the contents of the frame be represented as a vector $\mathbf{f}[n]$ such that

$$\mathbf{f}[n] = [s[k-m]\ldots s[k-p]\ldots s[k]]^T \tag{17}$$

where $s[\cdot]$ are the values recorded, m is the frame size, p denotes the frame overlap and n is an arbitrary index for the frame ranging from instants $k-m$ to k. Thus, index $n-l$ references the frame between instants $k-m-lp$ to $k-lp$, in particular

$$\mathbf{f}[n-1] = [s[k-m-p]\ldots s[k-2p]\ldots s[k-p]]^T \tag{18}$$

Using the sample mean $\tilde{f}[n]$ of each frame, the frames are centered and the square root of the dot product between every two adjacent frames is computed. This measure is intended to detect sudden changes between frames.

The 2nd visible node contains the 4th order finite difference between elements of the signal $s[k]$. This is equivalent to a type II FIR High Pass filter of order 5 with a cutoff frequency at 2 kHz and a group delay of −0.0011 deg/Hz.

The respirator energy is uniformly distributed across the 100–5000 Hz band, but most of the voice, background sounds and audible alarms occupy the region from 500 to 3000 Hz but not uniformly as in the case of the respirator noise. So we chose to use upper spectrum (4134–5000 Hz) as well as the lower spectrum (0–1000 Hz) to indicate the presence of respirator noise. For this reason, the last visible node uses the total power of the frequency range from 4.1 kHz to 5 kHz. This sum is estimated by summing the magnitude of the Fast Fourier Transform (FFT) of the signal frame for the discrete frequencies given in this range.

Fig. 6. Placement of impromptu data collection system on outside of the firefighters SCBA mask [18].

The RBM convergence is sensitive to the variability of the visible node signals taking considerable more time to converge when using un-smoothed data than smoothed data. The data had noise content above the 3 Hz region. We used a low pass filter with a 3 dB bandwidth is 7.694 Hz to smooth the data before classifying it. The cutoff frequency is based on observing that the fastest instantaneous rate observed is 70 breaths per minute. Seventy breaths per minute implies there is 0.8571 s between each start of a breath. This interval is equivalent to 1.1667 Hz.

The RBM is then implemented in a data flow as shown in Fig. 5.

5. Experiments

5.1. Data description

We conducted a series of recordings for six firefighters with ages between 20 and 30 at New Mexico's Fire Department Training Facilities. One recording of 5–30 min from each fire fighter was obtained. The net number of sound data points from all six recording is the following: sound samples for each regulator and non-regulator: 9,386,267 vs. 21,149,815; inhalation vs. non-inhalation intervals: 1292 vs. 1358. The sample rate is 11,025 samples per second.

The microphone was taped to the outside of the SCBA mask in the lower left corner out of the field of view of the firefighter and next to the regulator (Fig. 6). The microphone also recorded the voices of other firefighters, as well as the background sounds such

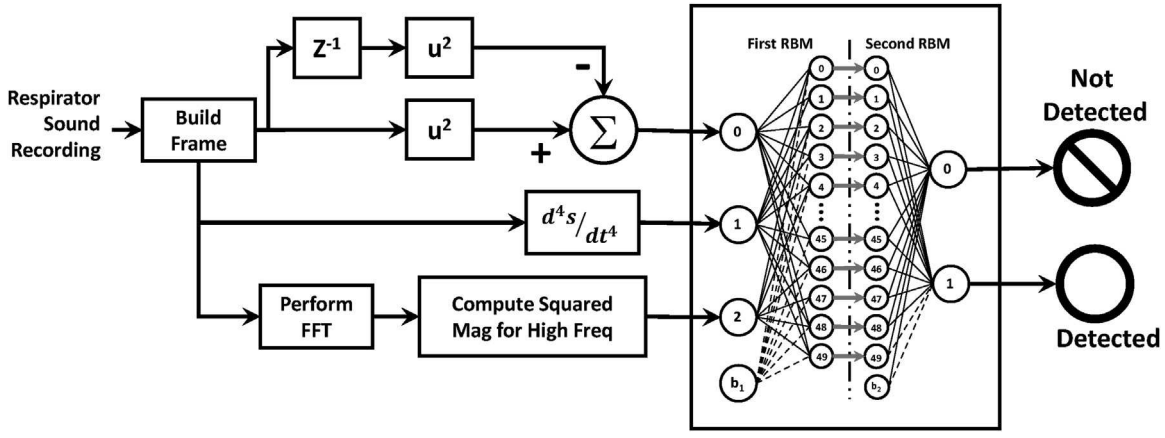

Fig. 5. The recording is broken into 12.5 ms frames with 12.325 ms overlap and classified using the RBM to determine noise or respirator sounds.

as breaching the entryway, the sound of the water nozzle, Personal Alert Safety System (PASS) alarms, ventilation fans used to clear smoke from the rooms, and some electronic interference.

The data was then listened to, and intervals where the regulator's sound is present were annotated using the PRAAT software (a free computer software package for the scientific analysis of speech in phonetics) [19]. Each event in a recording was assigned a unique event identification number. The results of this process were then saved in a textgrid file.

When the classifier runs, the results are inserted into the text grid file as a new tier. The annotation file and the recorded data were then examined in PRAAT. The breathing events detected by the classifier were compared with events from the manually scored tier. The automatically detected events are then manually compared with the had scored events. Automatically scored events are assigned the label of a manually scored event, when the Atutomatically scored event's start and stop times are contained within a manually scored events interval. Rarely, an automatic event detection covered multiple manual events. The event was assigned the label of the first manually scored event. A "?" label was used to identify automated system predictions that do not correspond to any of the manually scored events. A missed event was left unlabeled in the automated detector tier.

In examining the data, it became evident that the data had a distinct banding pattern where the respirator noise would occupy the upper spectrum (4134–5000 Hz) as well as the lower spectrum (0–1000 Hz). The middle part of the spectrum is used by the voice or speech related frequencies and audible alarms such as the PASS alarm (Fig. 7). This division is evident in Fig. 8, where the observed distribution appears to be bimodal.

The noise produced by the hose makes the detection more challenging. The respirator detection measure generally reports a breath duration less than 2 s. The noise generated by the water exiting the fire hose nozzle has features similar to the sound of the compressed air released from the SCBA tank through the respirators valve opening, which produces false respiration detections longer than 2 s. When such an event occurs, the whole sequence is re-filtered with a filter constructed as in Section 4.1, but where the data used to fit the model contains the hose noise. The compared results are shown in Fig. 9.

Another refinement involved updating the auto-regression coefficients each time a breath was detected. The new coefficients were computed using the autocorrelation matrix and

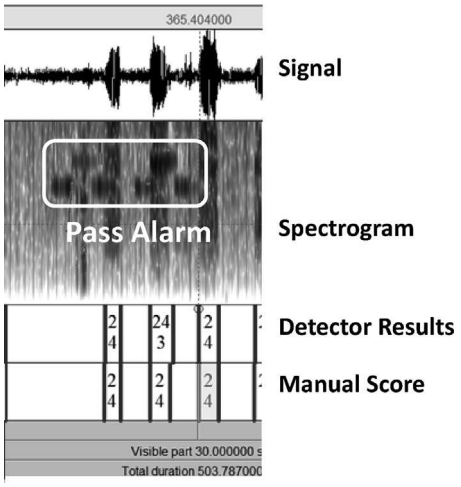

Fig. 7. Two tier PRAAT textgrid with spectrogram and signal displayed [19].

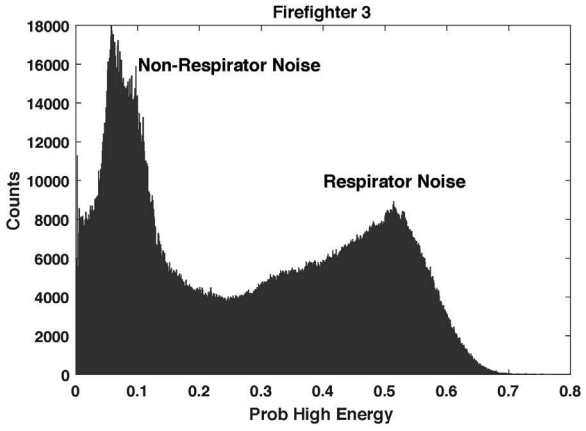

Fig. 8. Typical histogram of percent of power in the (4134–5000 Hz) band.

Levinson–Durbin approach and the set of signal elements just identified. The blending rule is given by

$$a_i^+ = \frac{a_i + a_i^-}{2} \tag{19}$$

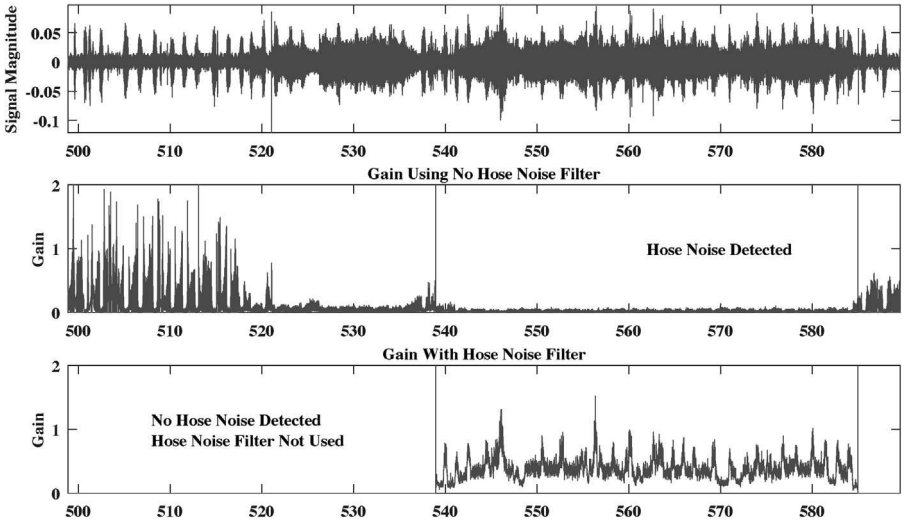

Fig. 9. Example of fire hose noise section.

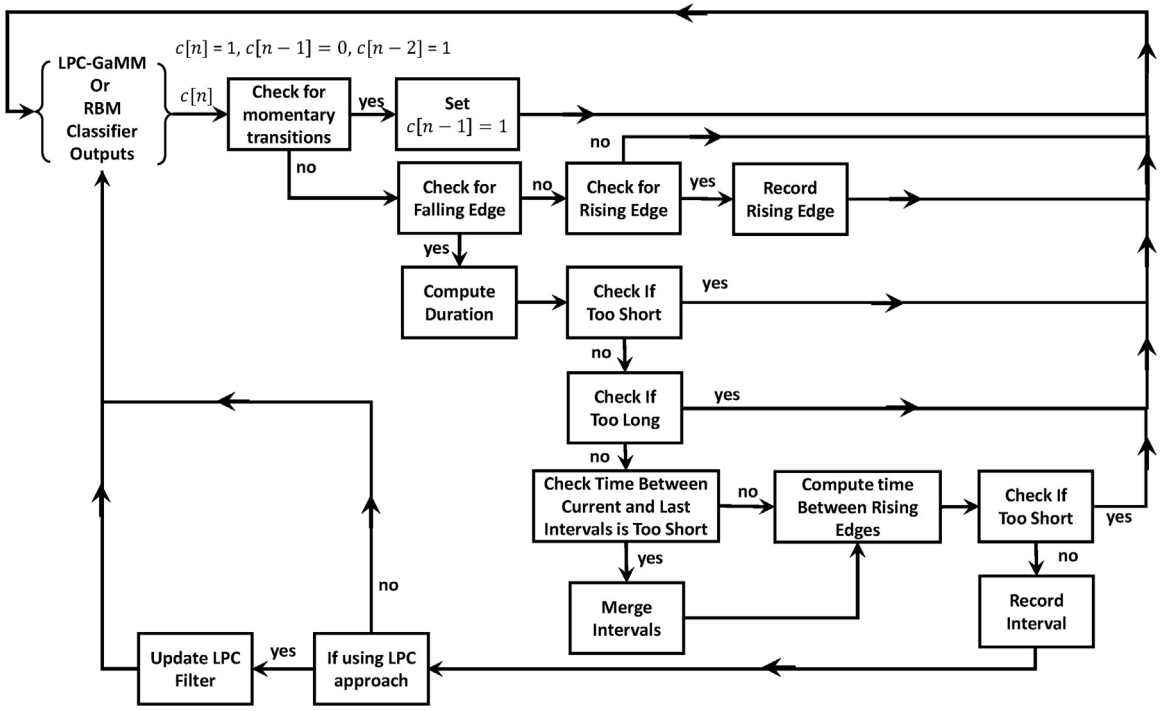

Fig. 10. Physiological processing data flow/overview.

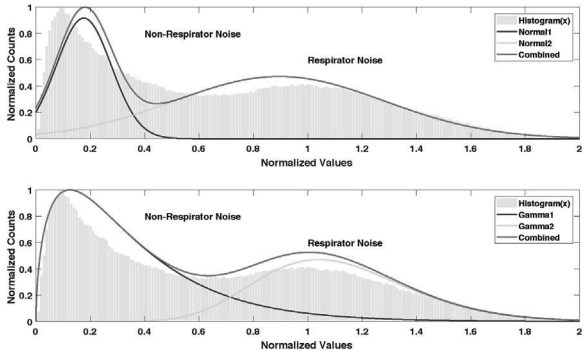

Fig. 11. Relative energy value distributions for the training videos. The upper pane shows the GMM distributions, and the lower panel shows the GaMM result.

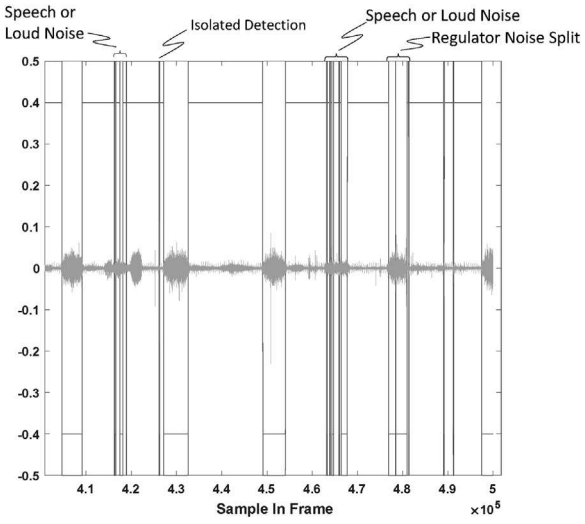

Fig. 12. RBM Classifier Output – the yellow represents the recorded sound for the frame being processed. The red and blue lines represent the outputs of the classifier for both the 1st and 2nd RBM. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

where a_i is the just computed filter coefficient for tap i, a_i^- is the previous coefficient value, and a_i^+ is the blended coefficient value.

5.2. Gamma Mixture Model performance

A GaMM and a GMM have been trained using the sound track from a fire fighter training video. The video focused on firefighters learning how to manage air consumption when the air remaining in the SCBA air cylinder is below acceptable levels (FIRE-GROUND Fire Entrapment - Conserving SCBA Air). A histogram of the values is generated using a bin resolution of 0.01, from 0 to the maximum observed relative value. The approximations made with both models are shown in Fig. 11. The GaMM is a better fit to the observed distribution. This is primarily due to the finite distribution of the Gamma function. The finite tails limit the contribution of each distribution to a bounded interval. Gaussian distributions have infinite tails. These tails influence the summat each mean. Their contributions distort the overall shape of the mixture distribution PDF. Frames consisted of 6.25 ms with an overlapping of 6.125 ms (shifting over one sample).

5.3. Analysis

A majority of the breathing events are shown as long duration intervals without temporary splits. The regulator noise split occurs in an interval where more than one sound is present. From a physiological perspective these intervals are the same breath and need to process as a single event and not as 2 or more extremely short breaths (Fig. 12). The detection algorithms also found isolated instances where the indicator is positive for less than 100 samples. These are sounds that do not indicate a breathing event but might indicate a tool was dropped or some speech activity. A cluster of these short term events usually indicates that the firefighter is talking.

Table 1
Thresholds.

Firefighter	Leave One Out	Half & Half
1	0.185	0.4
2	0.203	0.516
3	0.344	0.294
4	0.344	0.414
5	0.277	0.353
6	0.242	0.684
All Samples	0.266 ± 0.068	0.444 ± 0.139

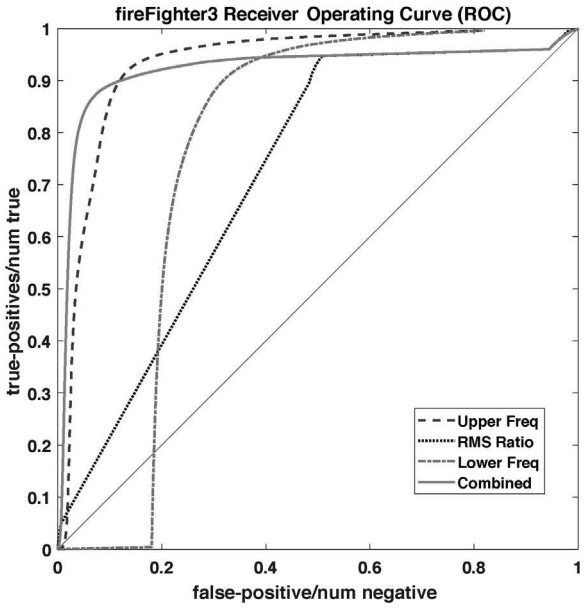

Fig. 13. The upper frequency power alone curve is shown as the blue dashed line. The black dotted line is the RMS alone line. The red dash dot line shows the curve for using the lower frequency probability alone. The solid magenta line is the ROC curve for combined measure.

5.3.1. LPC-GaMM training

In experimenting with the LPC-GaMM approach we needed to develop a threshold based on the ratio of $\hat{\rho}_N$ to $\hat{\rho}_R$. The ratios were determined using two approaches, a Leave One Out Approach which pooled the data from five of the six recordings (1). The resulting ratio was then used to process the recording not included in the pool. The idea being to use as much information as possible. The second approach is a half and half strategy where the first half of the recording was used to develop the ratio (1) for the second half processing. This approach shows more variability, but it can be argued is more indicative of individual firefighter. The leave one out reflects an averaging across multiple individuals. We conducted 3 different training modalities. The first one is a leave one out training where five fire fighter's data is used for training and 1 left for test. The average number of training frames was 35.3 million frames. In the second modality, which is denoted as half and half, the first part of the data of each fire fighter's data is used for training and the rest for test, with and average training size of 3.5 million samples.

The GaMM Classifier uses three separate measures, percent power in upper frequency, percent power in lower frequency, and the RMS ratio are shown in the ROC curves shown in Fig. 13. The combined measure is shown as the solid magenta line. The resulting ROC is better than any of the component measures alone. Note that in the preceding sections we discussed using the RMS ratio alone. This turns out to have a linear ROC curve. By using a product of both the high frequency and low frequency values, we improved the performance.

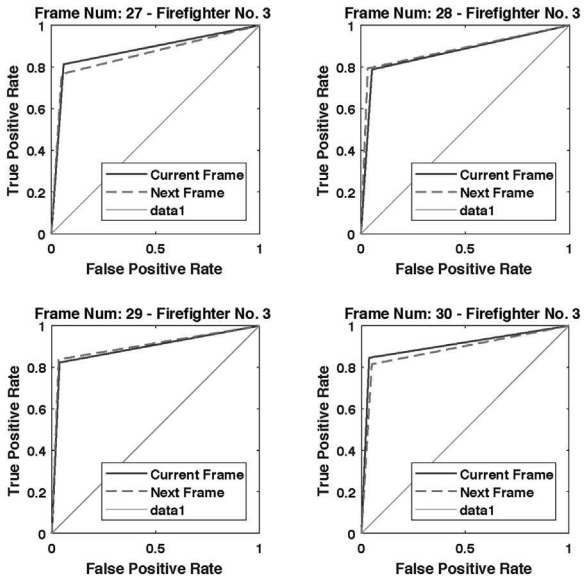

Fig. 14. Four consecutive RBM training instances where the hose noise enters and leaves the recording. The blue line indicates the Receiver Operating Characteristic curve for the current frame using the current trained weights and the red dashed the line the weights used on the next frame.

Table 2
Leave One Out (LOO) and Half and Half (H & H) signal processing overall correct classifications.

Firefighter	RBM	LPC LOO	LPC H & H
1	85.96%	88.50%	89.10%
2	88.76%	95.25%	94.68%
3	88.53%	85.03%	88.51%
4	91.68%	86.17%	89.76%
5	70.42%	56.44%	70.49%
6	79.05%	89.79%	90.24%
All Samples	$84.06 \pm 7.57\%$	$82.28 \pm 14.98\%$	$86.51 \pm 8.44\%$

5.3.2. RBM training

The RBM features are extracted using 12.5 ms frames with a 12.325 ms overlap. The RBM algorithm was trained incrementally using 500,000 feature vectors, each vector corresponding to a frame. The convergence criteria checks for change in the sum of the squared incremental differences of second machine's hidden node values to be less than 0.00125. We also imposed a limit on the number of iterations by using 50 epochs of 10 iterations. In most instances these values converge quickly (4 or 5 epochs or 7 or more seconds) but when the hose noise or a gas powered ventilation fan is present and the training times increase to take the full 50 epochs at 7 or more seconds per epoch. As can be seen in Fig. 14 the noise changes the basic shape of the ROC curve shifting the optimal point to the right. As the noise begins to leave the frame the curve begins to shift back to its previous form. A sequence of 500,000 feature vectors represents 45 s of recorded data and the training time is 350 s or more. We had hoped to be able to implement the training in the same time interval as the frame length and use the weights on the next interval. The structure of the estimator, i.e. the number of layers and the number of hidden nodes was previously validated using the independent training set [14]. The structures tested had 1–3 layers. The number of tested nodes in the hidden layers was 10, 13, 16, 20, 25, 32, 40, 50, 63, 79 and 100. The best results were obtained for one hidden layer with 50 nodes.

5.3.3. Signal processing results

Table 2 shows the results for each firefighter. It should be noted that the recording for firefighter 5 is the poorest due to the micro-

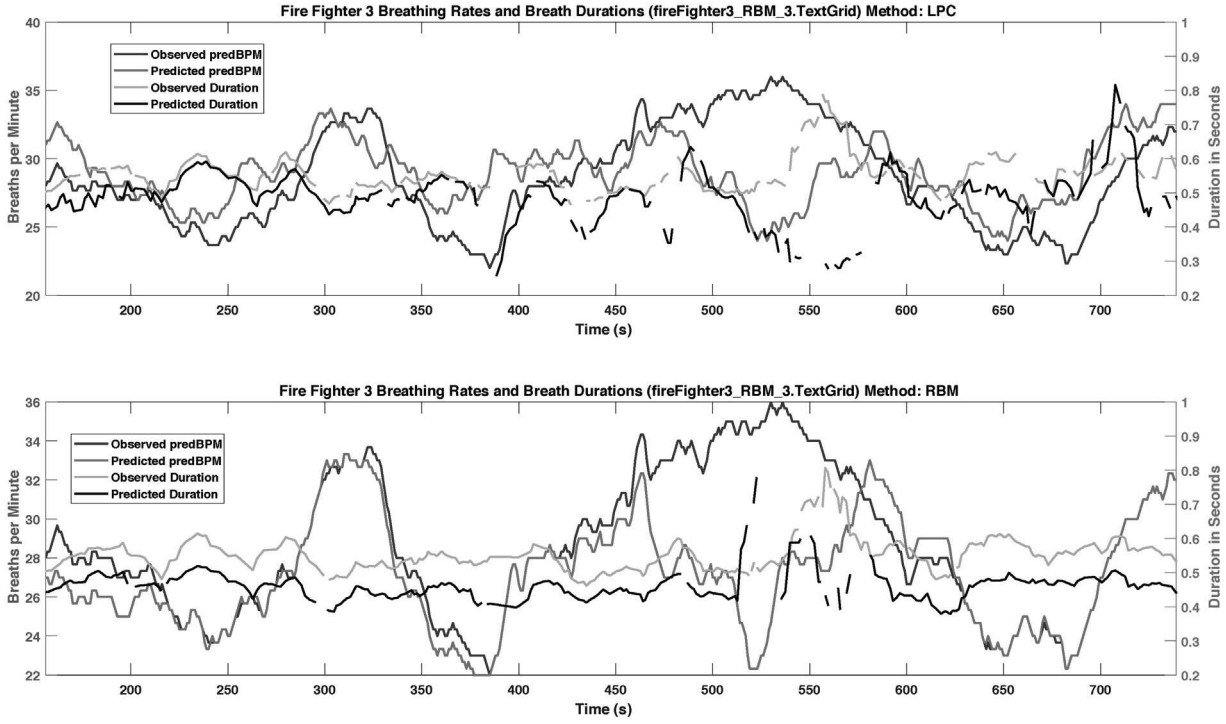

Fig. 15. Example fire hose noise section.

phone working loose during the training exercise. Overall, the accuracy (number of correct classifications) are comparable. The real difference in Table 2 is in their standard deviations, the RBM classifier and the Half and Half training have a significantly lower standard deviation of 7.9% and 8.4% vs. 13.7% for the Leave One Out Approach. This indicates that having one universal threshold for all firefighters and conditions is not possible and that the classifier needs to adapt to each individual and conditions.

5.3.4. Physiological analysis results

The physiological perspective involves correlating the breathing signal classifications into intervals (Fig. 10). Each interval is used to measure depth of breath. The number of intervals in a minute represents the frequency or rate of breathing. It is important to understand that the physiological measurements depend on the accuracy of the signal processing. This process requires that the classifications first be sorted into long (≥2200 samples) and short duration (<2200 samples). The short duration intervals are discarded. The remaining longer intervals are checked for adjacency. The adjacency being defined as the current interval's sample rising edge sample number minus preceding interval's falling edge sample number. The two intervals are merged into one interval when the adjacency values are less than 300 samples. The final interval check eliminates intervals that are too long (e.g. 2 s or more).

The resulting set of intervals are then compared to the manually scored intervals as discussed above. The results are given in Table 3. In both sets of results, the overall averages and their standard deviations (All Samples row) is significantly different. The RBM Classifier is slightly better and noticeably more consistent (lower standard deviation values) than the other LPC-GaMM Classifiers.

Fig. 15 depicts the difference between the LPC and RBM approaches when examined from a physiological perspective. There is a noticeable difference when the data is used to predict breathing rates (bpm) and breath duration. The regions where the red line is above the blue line indicate that the false detections cause the estimate of breaths per minute to be higher than expected. Similarly, when the red line is below the blue line, the processing failed to detect breaths.

Table 3
Leave One Out (LOO) and Half and Half (H & H) physiological processing overall correct classifications.

Firefighter	RBM	LPC LOO	LPC H & H
1	88.31%	85.43%	90.47%
2	96.30%	96.30%	97.78%
3	93.95%	80.90%	89.07%
4	97.68%	87.22%	92.42%
5	77.71%	47.50%	64.17%
6	87.50%	83.08%	84.23%
All Samples	90.21 ± 7.06%	80.07 ± 16.81%	85.49 ± 11.73%

Table 4
Breath per minute RMS error.

Firefighter	RBM	LPC LOO	LPC H & H
1	1.787	2.350	1.487
2	1.514	1.894	1.181
3	3.452	3.637	3.973
4	2.123	5.698	4.886
5	1.028	2.166	1.215
6	2.921	4.585	3.942
All Samples	2.135 ± 0.970	3.388 ± 1.525	2.781 ± 1.667

The comparison of the charts is best summarized in Table 4. The Root Mean Squared (RMS) error the difference between the observed values and the predicted values. The RBM approach is consistently lower as expected given the higher detection percentages (Table 3).

The breath duration RMS Data (Table 5) shows a marked difference between the average predicted duration estimation times for the RBM and the LPC GaMM approaches. The cause of the difference is supported by the inhalation duration times showing a tendency to underestimate the times by as much as 0.14 s vs. 0.04 s. This difference may in part be due to the longer frame size used by the RBM

Table 5
Inhalation duration RMS error.

Firefighter	RBM (s)	LPC LOO	LPC H & H
1	0.251	0.187	0.214
2	0.267	0.126	0.159
3	0.107	0.139	0.123
4	0.113	0.112	0.098
5	0.313	0.318	0.250
6	0.270	0.193	0.151
All Samples	0.220 ± 0.088	0.179 ± 0.075	0.166 ± 0.057

(twice that of the LPC) making detecting changes less accurate. In the future we will use the shorter windows.

6. Conclusion

We have examined three alternative means of predicting breathing rates and depth or length of inhalation times from recordings from an SCBA system's regulator noise.

The LPC based methods use an unsupervised classifier based on probability density mixture models and hypothesis testing. The use of a Gamma Mixture Model improved the fitting of the mixture model by eliminating the infinite tails of the Gaussian distributions. The classification of individual points in the sound recording showed an overall accuracy of 80–85% (Table 2) in detecting the regulator sound. However, when the classifier outputs are used to predict breathing rates and breath duration (Table 3), the number of breathing events detected ranged from 80% to 85%. This is not as accurate as the deep learning RBM approach.

In addition, we believe that the mixture models using two distributions do not have enough degrees of freedom to capture the introduction of additional modes resulting from external noise sources. As discussed in the paper, the presence of the water sound exiting the nozzle of the fire hoses required the need of a second recognition filter. The use of two or more filters introduces an additional requirement to manage the detection process. Therefore, we added an algorithm to switch between filters. Furthermore, we introduced spectral power measurements to assist in the classification process as demonstrated in Fig. 13.

The deep learning classifier uses Euclidean distance similarity measure between two adjacent frames. Adjacent frames sharing similar trends will have a relatively small distance measure due to the overlap between frames. The measured distance is greater in frames containing respirator sound (colored noise). Combining the measure with the normalized spectral power estimates classification yields an overall accuracy of 90.1% (Table 2). Using the classifier outputs to predict breathing rates and breath duration (Table 3) shows an accuracy of 90%. The qualitative difference is depicted in Fig. 15. The distance measure is more robust to external noise sources and eliminated the need of a second filter trained with hose noise. Further, the performance of the RBM with the firefighter 5 recording, where the microphone was not placed as close to the regulator as desired, was still reasonable compared to those of LPC. The classifier out preformed the LPC-GaMM approaches. Its drawbacks are also clearly illustrated when examining the average error in predicting the inhalation duration times. Another concern is that the RBM classifier using 500,000 points takes a significant amount of time to retrain and converge to a stable weight set.

Based on these results, we conclude, that it is possible to monitor respirator sounds to estimate breathing rate and depth of breath using a microphone placed on the regulator of an SCBA system and that our future work will involve optimizing the use of the RBM approach. It should be noted that the study was performed in the firefighting environment and not a clinical setting. Fundamental aspects as the effect of age, gender and other variables were not considered here, but they will need to be studied in the future and, if needed, included in the predictive models. We believe that we can design a microphone mount that places the microphone directly on top of regulator valve and not on the side of the masked near the regulator. This should greatly reduce capture of back ground sounds like the fire hose, other firefighter conversations, ventilation fans. Once he we have a working prototype of a surface mount microphone, we plan to repeat these experiments.

As discussed in the conclusions about the RBM approach, we need to start examining the RBM algorithms convergence criteria and see if there is a balance between the precision of the model and the level convergence. We also need to revisit the use of frame sizes and see if using a smaller frame size would help improve the inhalation duration time predictions. We can validate these modifications in the future experimentation with the surface mount microphone design.

We are hoping to use this data combined with some data being collected in other studies dealing with speech production and heart rate data. This combination can provide a mechanism for monitoring the level of exertion, and possibly predict when the individual is getting tired or becoming physically exhausted.

Acknowledgments

The authors would like to thank the Santa Fe Fire Department's fire fighters, command staff and the training facility. We would especially like to thank Chief Erik Litzenberg and Training Officer Elias Frick for their gracious support and assistance. We would also like to acknowledge Dr. Christine Mermier and Dr. Ann Gibson from the UNM Exercise Physiology Lab of Health, Exercise & Sports Sciences Department for providing guidance and support with the Institutional Review Board and initial contact with the Santa Fe Fire Department.

References

[1] R.C. Sá, Y. Verbandt, Automated breath detection on long-duration signals using feedforward backpropagation artificial neural networks, IEEE Trans. Biomed. Eng. 49 (10) (2002) 1130–1141.
[2] S. Sello, S. kyung Strambi, G.D. Michele, N. Ambrosino, Respiratory sound analysis in healthy and pathological subjects: a wavelet approach, Biomed. Signal Process. Control 3 (3) (2008) 181–191.
[3] F.J. Cereceda-Sánchez, J. Molina-Mula, Capnography as a Tool to Detect Metabolic Changes in Patients Cared for in the Emergency Setting, 2017 (online). Available: http://www.scielo.br/pdf/rlae/v25/0104-1169-rlae-25-e2885.pdf.
[4] P. Rajesh, Extraction of Breathing Pattern Using Temperature Sensor Based on Arduino Board, 2015, http://dx.doi.org/10.1063/1.4917842 (online).
[5] F. Güder, Paper-based Electrical Respiration Sensor, April 2016, http://dx.doi.org/10.1002/anie.201511805 (online).
[6] C. Li, D.F. Parham, Y. Ding, Cycle detection in speech breathing signals, in: Biomedical Sciences and Engineering Conference (BSEC), 2011, IEEE, 2011, pp. 1–3.
[7] P. Price, M. Ostendorf, C. Wightman, Prosody and parsing, in: Proceedings of the Workshop on Speech and Natural Language, Association for Computational Linguistics, 1989, pp. 5–11.
[8] C. Wightman, M. Ostendorf, Automatic recognition of prosodic phrases, in: 1991 International Conference on Acoustics, Speech, and Signal Processing, 1991. ICASSP-91, IEEE, 1991, pp. 321–324.
[9] K.A. Murphy, Machine Learning. A Probabilistic Perspective, The MIT Press, Cambridge, MA, 2012.
[10] M.I. Jordan, Z. Ghahramani, T.S. Jaakkola, L.K. Saul, An introduction to variational methods for graphical models, Mach. Learn. 37 (November (2)) (1999) 183–233.
[11] G. Elidan, Copulas in machine learning, in: P. Jaworski, F. Durante, W.K. Härdle (Eds.), Copulae in Mathematical and Quantitative Finance, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, pp. 39–60.
[12] G.E. Hinton, Training products of experts by minimizing contrastive divergence, Neural Comput. 14 (8) (2002) 1771–1800.
[13] G.E. Hinton, R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks, Science 313 (5786) (2006) 504–507.
[14] W.M. Kushner, S.M. Harton, R.J. Novorita, The distorting effects of SCBA equipment on speech and algorithms for mitigation, in: 2005 13th European Signal Processing Conference, IEEE, 2005, pp. 1–4.
[15] H. Ackley, E. Hinton, J. Sejnowski, A learning algorithm for Boltzmann machines, Cogn. Sci. (1985) 147–169.

[16] P. Smolensky, Information processing in dynamical systems: foundations of harmony theory, in: D.E. Rumelhart, J.L. McClelland, C. PDP Research Group (Eds.), Parallel Distributed Processing: Explorations in the Microstructure of Cognition, vol. 1, MIT Press, Cambridge, MA, USA, 1986, pp. 194–281.
[17] R. Salakhutdinov, G. Hinton, Deep Boltzmann machines, in: D. van Dyk, M. Welling (Eds.), Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics, ser. Proceedings of Machine Learning Research, vol. 5, 16–18 April 2009, PMLR, Hilton Clearwater Beach Resort, Clearwater Beach, FL, USA, 2009, pp. 448–455.
[18] M.S. Appliances, MSA G1 SCBA Integrated Thermal Imaging Camera, 2015, http://dx.doi.org/10.1002/anie.201511805 (online).
[19] P. Boersma, D. Weenink, PRAAT, a system for doing phonetics by computer, Glot Int. 5 (9/10) (2001) 341–345.