Please cite as: Kevin Curran, Peter Yuan and Damian Coyle (2011) *Using Acoustic Sensors to Discriminate between Nasal and Mouth Breathing.* International Journal of Bioinformatics Research and Applications, Vol. 7, No. 4, pp: 382–396, Sept-Dec 2011, ISSN: 1744-5485, Inderscience, https://doi.org/10.1504/IJBRA.2012.049623

Using Acoustic Sensors to Discriminate between Nasal and Mouth Breathing

Kevin Curran, Peng Yuan, Damian Coyle

School of Computing and Intelligent Systems
Faculty of Computing and Engineering,
University of Ulster, Northern Ireland, UK
Email: kj.curran@ulster.ac.uk

***Abstract* - The recommendation to change breathing patterns from the mouth to the nose can have a significantly positive impact upon the general well being of the individual. We classify nasal and mouth breathing by using an acoustic sensor and intelligent signal processing techniques. The overall purpose is to investigate the possibility of identifying the differences in patterns between nasal and mouth breathing in order to integrate this information into a decision support system which will form the basis of a patient monitoring and motivational feedback system to recommend the change from mouth to nasal breathing. Our findings show that the breath pattern can be discriminated in certain places of the body both by visual spectrum analysis and with a Back Propagation neural network classifier. The sound file recoded from the sensor placed on the hollow in the neck shows the most promising accuracy which is as high as 90%.**

***Index Terms* —** pervasive health, respiratory monitoring, sensors

1. Introduction

In the 1990's Buteyko identified the need for asthmatics to breathe through their nose to prevent against over-breathing caused by mouth breathing (McHugh et al., 2003). There are currently a number of clinical studies being undertaken to prove the clinical significance of the Buteyko breathing technique. It is also well recognised that nasal breathing warms, humidifies and releases endogenous nitric oxide into inhaled air. These all play an important role in conditioning the lungs. Unfortunately breaking the habit of mouth breathing is difficult, and asthmatics practising the Buteyko breathing technique will often revert to mouth breathing automatically. Asthmatics attempting to adopt this technique to gain better control over their condition need to be constantly reminded to breathe correctly. It is well known that normal lung sounds show interpersonal variations. In addition to this, it has to be taken into account that both a same-day variability and a between-day variability exist in lung sounds (Mahagna and Gavriely, 1994).

On the basis of these large variations, concrete changes in nasal and mouth will be seen only with investigation of a larger number of subjects. The purpose of computer-supported analysis of breathing sounds is for objective understanding and archiving. Because the sensitivity of human hearing is reduced, particularly in the lower frequency ranges, an objective electronic recording for recognizing deviations within these ranges could be helpful. Since the procedure for doing this is not costly, invasive, or particularly intensive, it would be suitable as an examination method for high-risk groups such as pneumonia patients. A daily or even more frequent analysis of lung-sound spectra could help to identify patients with say, incipient pneumonia before the appearance of any radiologic abnormality

The ability to electronically record and analyze biological sounds, such as of heart and respiratory sounds, first developed in the 1950s (McKusick et al., 1955) and it was soon adopted by others in the 1960s and 1970s (Rozenblat et al., 1968). Since that time, it has been possible to describe lung sounds in terms of their timing in the respiratory cycle, duration, waveform, and frequency components. The names used for different respiratory sounds took on definitions based in part on their objective features revealed by sound analysis (Loudon and Murphy, 1984). Lung sounds are frequently recorded for teaching purposes and analyzed for research (Kraman et al., 2006). The implicit goal of such research is to extract from the lung sound signal, qualitative or quantitative information that relates to important physiological or pathological processes. The fact that these sounds may be easily and noninvasively recorded adds to the attraction of the approach (Kraman et al., 2006). Chest auscultation is an important part of patient assessment for the detection of respiratory disease. In clinical practice, lung disease may be diagnosed when adventitious sounds are present, or when an individual's breath sounds are perceived as having a frequency content and intensity that differ from normal. This latter process is critically dependent on the listener's knowledge of the range of frequencies and intensities that can be found in normal breath sounds. Thus, clinicians need to have a clear perception of what

these normal ranges of frequency and intensity are, to avoid systematic errors in auscultation. This may be one reason why agreement on auscultatory signs of lung disease among different observers is reported to be modest (Smylle et al., 1965; Godfrey et al., 1969; Pasterkamp et al., 1987; Gavriely et al., 1995).
The frequency range of normal lung sounds in adults extends from 50 Hz to 500 Hz, and the shape of the linear amplitude frequency spectrum (AFS) of these sounds follows an exponential decay pattern. The studies of normal breath sounds in children have been more limited but suggest that the frequency content of their breath sounds differs from that of adults ((Schreiber et al., 1981; Hidalgo et al., 1991). (Forgacs et al., 1971) demonstrated that the intensity of breath sounds at the mouth is related to the forced expiratory volume in one second (FEV) in chronic bronchitis and asthma. The development of modern signal processing has greatly increased the possibilities to investigate the relationship between respiratory sound signal and ventilatory function. The main interest has been focused on wheezing sounds and on variables reflecting the frequency distribution of breath sounds in lung sound spectra. The aim of the study by (Malmberg et al., 1994) was to investigate the relationship between ventilatory function and frequency spectrum variables of breath sounds in asthmatic patients and in healthy control subjects. The changes in ventilatory function were spirometrically determined during a graded histamine challenge test; airflow standardized breath sound recording was used. The trachea and the chest were compared as recording sites. They found that in asthmatics, the breath sound frequency distribution in terms of median frequency reflected acute changes in airways obstruction with high sensitivity and specificity (Malmberg et al., 1994). Vital signs information influences the patient care decisions of emergency services providers and there is a sound clinical basis in striving for comprehensive and accurate recordings (Brown & Prasad, 1997). The respiratory rate for instance is an accurate reflection of severity of illness and if carefully measured is a sensitive marker of accurate respiratory and metabolic dysfunction, especially in the critical care setting) (Gravelyn & Weg, 1980). Primary clinical courses emphasise that subtle changes in the rate of breathing are highly significant. The need for continuous, non-invasive and reliable respiratory monitoring has long been recognized.

In a study by (Baird & Neuman, 1992) a nasal-oral air temperature sensor for measuring breathing in infants was developed using microelectronic technology and studies on infants showed that 80% of them demonstrated periods of oral, along with nasal, breathing. The sensor is quite invasive however and had to be taped to the upper lip of the infant being studied so that the side with the two temperature sensors was positioned with each sensor lying just outside of the infant's nares. The temperature sensor on the opposite side of the structure was then positioned over the infant's mouth. (Yu et al., 2008) used a Respiratory Inductive Plethysmography (RIP) to detect a user's breathing status. It can detect the breathing condition according to a natural expansion or a natural shrinkage of the user's body. The breathing condition at least includes breathing mode (breast breathing or abdominal breathing), breathing rate, and breathing depth. The monitor displays a relative breathing condition picture according to the detected breathing condition. As such, the user can understand the practical breathing condition by viewing the breathing condition pictures or sounds, so as to be instructed to learn to use a breathing pattern suitable for him however it does not attempt to discriminate between nasal and mouth breathing.

We investigate the principles of automated discrimination of breathing patterns using an acoustic sensor, to examine if the two breathing types can be classified with high accuracy for certain locations in order to later form the basis of a patient monitoring and motivational feedback system to recommend the change from mouth to nasal breathing.

2. Sound Analysis and Feature Extraction

Based on characteristics of the human voice and hearing, a number of theoretical models of sound signal analysis have been developed, such as Short Time Fourier Transform, Filter Analysis, LPC Analysis and Wavelet Analysis. These theories have been widely used for Sound Coding, Sound Synthesis and Feature extraction of the Sound. In Linear Predictive Coding (LPC), the feature of the sound could be extracted by calculating the coefficients in different order of the linear predictive; The Filter Analysis Theory first filters out the frequency of the sound signal by using a bandpass filter, then extracts the frequency feature based on simulating the function of the hearing system of the biological nerve cells.

The main part of the Filter Analysis is a bandpass filter which is used to separate and extract the different and useful frequency bands of the signal. However, a complete and perfect Filter Analysis Model should be a bandpass filter that is followed by non-linear processing, a low-pass filter, resampling by a lower sampling rate and compression of the signal's amplitude process. The common function for non-linear processing is the Sin function and Triangular windowing function. In order to smooth the sudden changing parts of the signal, the signal should pass through a low-pass filter after the non-linear processing. The Alternative process which is to re-sample the signal by a lower sampling rate or compress the amplitude aims at reducing the calculations at a later stage. The fundamental problem of the sound signal recognition lies on what and how to choose the reasonable features and characteristics of the signal. A sound signal is a typical time-varying signal, however if we zoom in to observe it at a millisecond level, the sound signal shows certain periods that seems to be a stable signal to some extents. Therefore, these features are extracted to represents the original signal. The characteristic parameters of a sound signal fall into two types. One is the features in the time domain and the other is the features in the frequency domain after transformation of the

Please cite as: Kevin Curran, Peter Yuan and Damian Coyle (2011) *Using Acoustic Sensors to Discriminate between Nasal and Mouth Breathing.* International Journal of Bioinformatics Research and Applications, Vol. 7, No. 4, pp: 382–396, Sept-Dec 2011, ISSN: 1744-5485, Inderscience, https://doi.org/10.1504/IJBRA.2012.049623

signal. Usually, just the sampling values in one frame, such as the average amplitude of short-time or the average zero crossing rate of short-time, could constitute the characteristic parameters in the time domain. Another type of feature can be obtained by transforming the original signal into the frequency domain Fast Fourier Transform, for example can retrieve the Linear Prediction Cepstral Coefficients (LPCC) or Mel Frequency Cepstral Coefficients (MFCC) features of the signal. This has the advantage of simpler calculations but has a large dimensions of feature parameters and is not suitable for representing the amplitude spectrum features. On the contrary, the latter type has a quite complex calculation of transforming, but could characterise the amplitude spectrum of the signal from several different angles.

The short-time energy of the sound signal reflects the characteristics of amplitude over the time. Linear Predictive Coding (LPC) analysis is based on the theory that the signal in this moment could be approximately figured out by the linear combination several signals before. By minimising the average variance between the actual sampling value and the linear predictive sampling value, the LPC parameters could be obtained. In sound recognition, the LPC parameters are seldom used directly but rather the Linear Predictive Cepstrum Coefficients (LPCC) derived by the LPC. The LPC parameters are the acoustic feature derived by research on the voice mechanism of humans. The Mel Frequency Cepstrum Coefficients (MFCC) is from research on human hearing. The theory is that when two tones of similar frequency appear at the same moment, only one tone can be heard by humans. The critical bandwidth is the bandwidth boundary where the human feels the sudden change. When the frequency of the tone is less than the bandwidth boundary, people usually mistake hearing the two tones as one, and this is called the shielding effect. Mel calibration is one of the methods to measure the critical bandwidth, and the calculation of the MFCC is based on the Mel frequency.

In subjects with healthy lungs, the frequency range of the vesicular breathing sounds extends to 1,000 Hz, whereas the majority of the power within this range is found between 60 Hz and 600 Hz ((Pasterkamp et al., 1997). Other sounds, such as wheezing or stridor, can sometimes appear at frequencies above 2,000 Hz ((Gross et al., 2000). The normal classification of lung sounds in frequency bands involves low (100 to 300 Hz)-, middle (300 to 600 Hz)-, and high (600 to 1,200 Hz)- frequency bands (Pasterkamp et al., 1997). We concentrated on the 1 Hz to 1000 Hz range. The main interest in our study was focused on frequencies below 600 Hz, and our coupler depth of 6 mm was therefore sufficient. In order to investigate whether there were differences in sounds at different locations, we recorded the breathing sounds in five positions in the neck region. The microphone coupler was secured with double-sided tape in these fixed positions. The controlled breathing needed for this was practiced before data recording. The subject attained but did not exceed a certain flow value (1.7 L/s), which resulted in the airflow being approximately the same in all records. By using standardized breathing, we assumed that the influence of sounds emanating from the mouth or pneumotachograph tube were the same.

3. Discriminating Breathing Sounds

The Axelis acoustic sensor comprised a substrate with a layer of piezoelectric film disposed on the curved portion of the substrate second surface and a layer of pliable material disposed on a second surface of the piezoelectric film opposite the first surface for contacting with an object to be sensed. This electronic stethoscope operates as a vibration sensing element contacted to a body part such as neck or throat so as to detect sound propagating through tissue, muscle, tendon, ligament and bone. We connected the sensor to a recorder to record the sound directly onto a flash drive. We used MATLAB for analysis of the audio

To determine the feasibility of automatically distinguishing between nasal and mouth breathing based on acoustic sound a range of tests on various recordings were carried out. Recordings were made on a number of subjects where the subjects performed Nasal and Mouth breathing whilst the acoustic sensor is placed at different locations close to the throat. These included directly under the chin, on the windpipe, on the hollow point directly above suprasternal notch (also known as juglar notch) and on the right and left hand side of the neck. Research shows the frequency of lung sound lies mainly below 1000 (Hz) and this project focus the frequency under 1200 (Hz). Different frequency bands are used for different function for this experiment, the much lower frequency under 100 (Hz) is cut off to extract acoustic features and the higher frequency above 1200 (Hz) carry a lot of noise which has been filtered out at the first stage, the rest in between is used for end-point detection. So the original signal should pass through several bandpass filters to cut out the specified frequency band. However, unlike speech signals coming from lips that has a attenuation of 6dB/oct, the pre-recorded sound signals do not have to be pre-emphasized as they came from the acoustic sensor attached to the hollow. We created a filter built on the Matlab function 'butter':

$$[v, u] = \text{butter}(\text{order}, Wn, \text{function})$$

where parameter 'order' is the order of the filter that results in a better filter effect when use a larger order but also brings in a larger quantity of calculations, and the length (L) of the parameter vectors 'u' and 'v' have a relationship with parameter 'order' that:

$$L_{u,v} = \text{order} + 1$$

Parameter 'Wn' is the normalization value of the frequency that to be filtered. When the sample frequency is expressed as fs, as the highest frequency that could be processed is fs/2, if it is the frequency (f) 2000 that aims to be filtered out then:

$$Wn = f/(\frac{fs}{2})$$

Parameter 'function' is a string that indicates the specified function of the filter. E.g. function = 'low' means a Low Pass Filter, function = 'high' represents a High Pass Filter. Based on the formula Lu,v = order + 1, the higher order of the frequency was more effective. The filter with the parameter vectors 'u' and 'v' having larger values requires more complex calculations. However, we found that the filter was increasingly effective when applying a larger order gradually from 1 to 8. We used short-time average energy to describe the amplitude features of the sound signal. In an acoustic signal, most of the energy lies on the higher frequency band, and high frequency means a higher zero crossing rate. The energy should have some relationship with the zero crossing rate but the breathing sound is unlike normal speech signals. Firstly, a large part of the recorded file is noise as the equipment used contains a sensitive sensor which records the breathing sound inside the body as well as the noise from the skin and airflow through the skin. In fact, sometimes the noise is much larger than the useful breathing sound. Secondly most of the energy lies on the frequency band below 100 (Hz) which is a frequency band the human beings can barely hear. This frequency band is also the most useful band for feature extraction. So it is uncommon to see the much lower frequency band having a higher zero crossing rate as the changing rate is larger in that particular band.

4. Our Method

The first stage was to take recordings of mouth breathing only, nose breathing only and a mix of breath patterns with mouth and nose with the sensor placed in the hollow of the neck (which was discovered to be most accurate location). All recordings were done for 60 seconds each whilst subjects sat a in a quiet room. The pre-processing filtered out certain frequency bands and added a window to smooth the signal as well as apply a Fast Fourier Transform. The pre-processing filtered out the noise in the higher frequency above 1200. Then the signal was cut into smaller frames and windowed for each one. The Fourier Transform was applied to the windowed frame until we got a spectrum map by pseudo-color mapping. We use neural networks as they have been used in the medical field for a long time for their effectiveness in pattern recognition. Other classification methods such Support Vector Machine (SVM), K-Nearest Neighbour (KNN) may also be potentially suitable for this task. Characteristic extraction followed which involves End-point Detection and passing through a band-pass filter. Next, we applied the Back-propagation neural network with training data from mouth and nose breathing in order to be trained and weight adjustment. Finally, for the recognition stage, test data is input to the neural network using the weights obtained in the training process.

The spectrum analysis of the signal is based on Short Time Fourier Transform (STFT) analysis of discrete time domain. Discrete time domain sampling signal can be expressed as x(n) where n = 0, 1, … , N-1 means the sampling point number and N is the signal length. In the process of the digital signal people usually frame the signal by adding window on it, then x(n) could be expressed as $X_m(n)$ where n = 0, 1, … , N-1 and 'm' means the number of the frame, 'n' is the time number of the synchronous frame, N is the sampling points within one frame known as the frame length. the Discrete Time domain Fourier Transform (DTFT) of windowed signal $X_m(n)$ could be illustrated as below:

$$X(m, e^{j\omega}) = \sum_{n=0}^{N-1} \omega_m(n) \bullet x_m(n) \bullet e^{-j\omega n}$$

in order to simplify the discrete calculation, the Discrete Fourier Transform (DFT) of $w_m(n) * x_m(n)$, we use:

$$X(m,k) = \sum_{n=0}^{N-1} w_m(n) \bullet x_m(n) \bullet e^{-j2\pi nk/2N}, k = 0, \cdots, N-1$$

The |X(m, k)| is then the estimated value of short-time amplitude in terms of one frame Xm(n). Take m as the time variable, k as the frequency variable then |X(m, k)| is the dynamic spectrum of signal x(n). Since the Decibel (dB) could be calculated as:

$$DB(x(n)) = 20*\log10(|X(m, k)|$$

Please cite as: Kevin Curran, Peter Yuan and Damian Coyle (2011) *Using Acoustic Sensors to Discriminate between Nasal and Mouth Breathing.* International Journal of Bioinformatics Research and Applications, Vol. 7, No. 4, pp: 382–396, Sept-Dec 2011, ISSN: 1744-5485, Inderscience, https://doi.org/10.1504/IJBRA.2012.049623

We can then get the dynamic spectrum of the signal displayed by DB. Again simplify the calculation of the |X(m, k)| by Fast Fourier Transform (FFT) (Cooley et al., 1965). Take 'm' as the abscissa, 'k' as the ordinate and the value of |X(m, k)| as the pseudo-color mapping on the two-dimensional plane, we get the dynamic spectrum of the signal x(n). Mapping the value of |X(m, k)| to the pseudo-color enables better resolution and visual effects of the dynamic spectrum as well as the improvement of the diagram's readability. The method is firstly mapping the minimum value (X_{min}) of |X(m, k)| to the normalized zero, the maximum value (X_{max}) of |X(m, k)| to the normalized 1 and the rest of them to the C_i between 0 and 1 linearly. Secondly, display the the C_i by the mapped color on the monitor. In order to make full use of the dynamic range of the color space, the appreciated base spectrum value should be chosen. The value that less than the base is limited on the base and that greater than the base then be normalized linearly. The color value matrix is expressed as C = {c(m, k)} then the mapping from |X(m, k)| to c(m, k) is illustrated mathematically as below:

$$c(m,k) = \frac{B(m,k) - Base}{[Max(B(m,k) \forall (m,k))] - Base}$$

where:

$$B(m,k) = \begin{cases} |X(m,k), & |X(m,k)| > Base \\ Base, & |X(m,k)| \leq Base \end{cases}$$

When using a Discrete Fourier Transform (DFT), the frequency resolution of the spectrum refers to the interval between the discrete frequencies. This means the frequency interval (f_0) represented by variable 'k' in the expression X(m, k). The value depends on the frame length N and the sampling frequency of the signal f_s. Based on the Nyquist sampling theorem, f_0, f_s and N fall into the relationship as below:

$$f_0 = f_s / N$$

As the formula suggests, the frequency interval (f_0) has nothing to do with the frequency that the signal contains. As long as the sampling frequency is a constant, increase the frame length (N) will result in the higher resolution of the spectrum or the smaller bandwidth that represented by the 'k' in the expression X(m, k), in that case the spectrum will tend to be a Narrow-band one, otherwise it will be a Broad-band spectrum. Increasing the resolution in frequency domain by using a larger value of N will result in a lower resolution in time domain of the spectrum. The way to resolve the contradiction is to introduce the sub-frame by frame shift (N_1, $N_1 < N$) while choosing a larger but appropriate frame length (N), in this way a spectrum that with balanced resolution in frequency domain and time domain will be obtained, the sub-frame shift could be illustrated as below:

$$x_m(n) = x(n + N_1 * m) \qquad n = 0, 1, \ldots, N - 1, \quad N_1 < N$$

The aim of End-point Detection (EPD) is to find out the starting point and ending point of the digital signal which is meaningful. ZRO means the amount of zero-crossing value points within one frame, in general the zero-crossing rate of the voice sound is more or less larger than that of the silence sound (under the conditional of already got rid of the noise), and thus is used to detect the starting-point and ending-point within this project. When calculate the zero-crossing rate, the importance of the exactly zero value should not be ignored. Because the sample values retrieved from the audio file have been normalized, in order to avoid the possibility of increasing the ZRO by using the float number to calculate the bias, they should be unnormalized by multiplying the bit resolution to get back the original values. In an Arithmetic Progression $a_1, a_2, \ldots, a_n, \ldots$ where $a_{n+1} = a_n + d$, n = 1, 2, ... and 'd' so called common difference is a constant. In a general series expressed as y = y(t)

$$\Delta y(t) = y(t+1) - y(t)$$

where Δy(t) is called the first order difference of y(t) at the point t, so

$$\Delta y_t = y_{t+1} - y_t$$

is defined as the first order difference of the expression y(t) where Δ is the Difference Operator. Combining the zero crossing rate with the High-order Difference (HOD) operation enables the end point detection to achieve a high precision level. Here the order should be adjusted to decide whether first-order or second-order or even third-order should be chosen to attain the best performance.

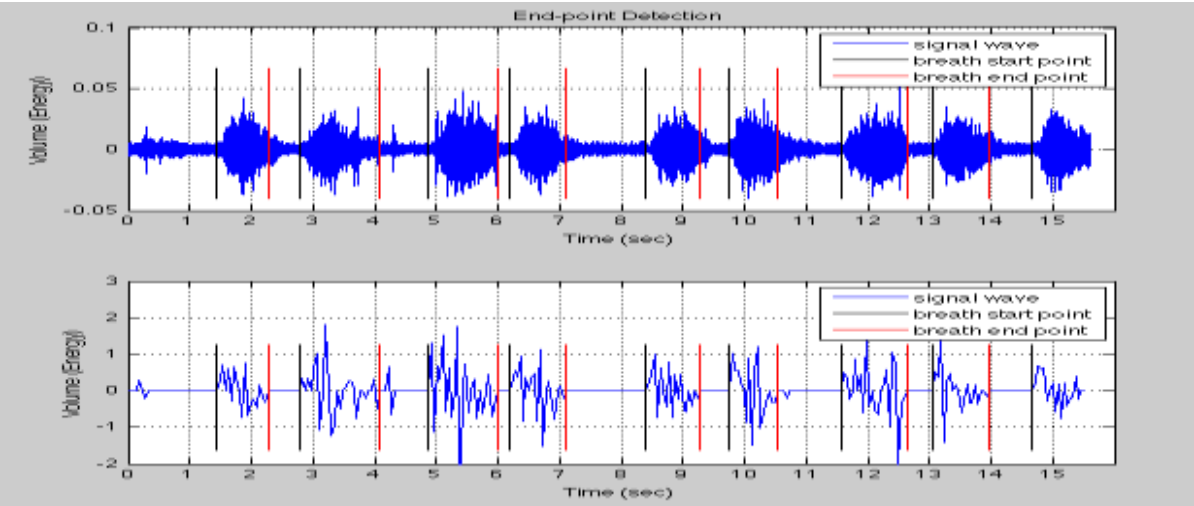

Figure 1: End-point Detection using the ZCR and HOD.

Figure 1 shows End-point Detection after filtering out the noise in the high frequency (1200 Hz). The other information contained a large amount of energy in a low frequency (110 Hz)

4.1 Artificial Neural Network

The back-propagation (BP) learning algorithm is a supervised multi-layered feed-forward neural network algorithm. In a single ANN without a hidden layer, the δ learning algorithm could be applied to the input and output sampling data to train the network. It is important that the back propagation refers to the output errors but does not feedback the result of the output to the hidden layers or even to the input layer. The network itself does not have the feedback function but back propagates the output errors to adjust the connection weights of the hidden layers and output layer, so the BP network could not be regarded as nonlinear dynamic systems but a nonlinear mapping system. If we consider a two layer neural network, that introduces one hidden layer between the input and output layers, then the direction of the arrows indicates the way that the information flow through the network. The node pointed to is called the low layer of the arrow and the node in the arrow tail is the upper layer of the arrow, then the output of the j node in a given training samples could be expressed as:

$$\text{net}_j = \sum o_i * w_{ij}$$

where o_i is the output of the i node in the upper layer, w_{ij} is the connection weight between the i node in the upper layer and the j node in current layer, as for the input layer the input is always equals to the output at any node. The output (o_j) of the j node is the transformation of its input by the expression given below:

$$o_j = f_s(net_j) = \frac{1}{1+e^{-net_j}}$$

where the output (o_j) is taken as the input of nodes in the lower layers. We can abstract the above to get the function:

$$f_s(\alpha) = \frac{1}{1+e^{-\alpha}}$$

then the differential expression of the output (o_j) is given as:

$$f_s'(\alpha) = \frac{-1}{(1+e^{-\alpha})^2}.(-e^{-\alpha}) = f_s(\alpha)[1-f_s(\alpha)]$$

Please cite as: Kevin Curran, Peter Yuan and Damian Coyle (2011) *Using Acoustic Sensors to Discriminate between Nasal and Mouth Breathing.* International Journal of Bioinformatics Research and Applications, Vol. 7, No. 4, pp: 382–396, Sept-Dec 2011, ISSN: 1744-5485, Inderscience, https://doi.org/10.1504/IJBRA.2012.049623

If we set the target output of the j node in the output layer as t_j, the output error is then obtained as $t_j - o_j$. We back propagate this error value from the output layer to the hidden layers and continually adjust the weights according to the principle of the amendment to decrease the errors. The error function for the network is:

$$e = \frac{1}{2}\sum_j (t_j - o_j)^2$$

In order to have the error (e), a decrease trend, the amendment of the weights should follow the gradient descent of the error function, that is:

$$\Delta\omega_{ij} = -\eta \frac{\partial e}{\partial \omega_{ij}}$$

where η is a gain coefficient that greater than zero.

The number of the input neurons depends on the dimension of the features extracted after the pre-processing stage. The number of output neurons is two, representing mouth breathing and nasal breathing respectively. The number of the hidden layer neurons is twice the input neurons typically but it is still adjustable to achieve the best performance. The neural network is illustrated in figure 2.

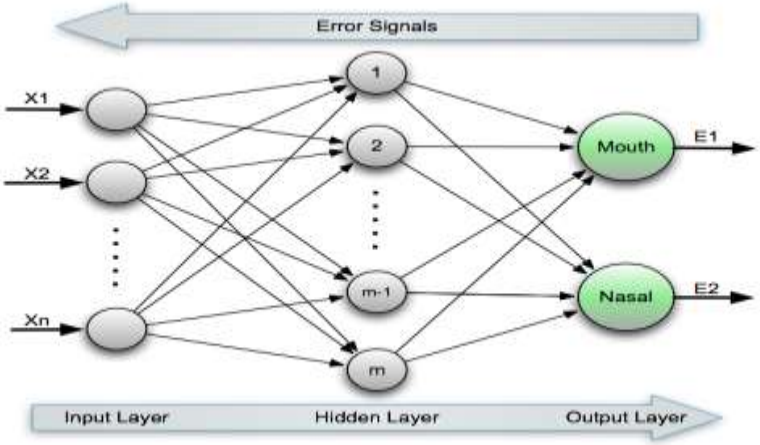

Figure 2: Design of twp-layer artificial back-propagation neural network

We initialize the hidden layer and assign connection weights with random numbers between 0 and 1 for the hidden layer and output layers. We normalize the feature values extracted to form a vector of n where $n = (x_1, x_2, \ldots, x_n)$ and the two types of target output are $t_1 = (1, 0)$ which represent mouth breathing and $t_2 = (0, 1)$ for nasal breathing, so each training sample falls into $I_1 = (x_1, x_2, \ldots, x_n; 1, 0)$ or $I_2 = (x_1, x_2, \ldots, x_n; 0, 1)$. The non-linear function:

$$y_j = \left[1 + \exp(-\sum_i \omega_{ij} x_i)\right]^{-1}$$

is used to calculate the output of each node layer by layer without the input layer. We finally get the output as:

$$O = (o_1, o_2, \ldots, o_m)$$

We adjust the weights from the output layer to the hidden by using the following equation:

$$w_{ij}(N+1) = w_{ij}(N) - \eta\delta_j o_i$$

where o_i is the output of the i node in the upper layer. If j is a node in the output layer, then

$$\delta_j = o_j (1 - o_j) (o_j - t_j)$$

if j is a node in the hidden layer, then

$$\delta_j = o_j (1 - o_j) \sum \delta_k w_{jk}$$

where k represents all the nodes in the lower layer of layer where j node located. The non-linear function with a 'S' shape used for BP Neural Network algorithm is:

$$f(x) = \frac{1}{1 + e^{-x}}$$

We adjust the weights for the hidden and output layers respectively using the gradient descent method, as well as the threshold which is amended in each loop. The adjusting value for the weights and threshold between output layer and hidden layer is:

$$\Delta W_{kj}^{p} = \vartheta(y_{pk} - d_{pk}) \cdot f'(net_{pk}) \cdot O_{pj} = \vartheta \sigma_{pk} O_{pj}$$

$$\Delta \theta_{k}^{p} = \vartheta \sigma_{pk}$$

5. Evaluation of Results

A frequency analysis was carried out on the signals recorded from each of the locations. For each location, each type of breathing was performed for approximately 60 seconds. Every 100Hz range up to 1000Hz was analysed using a short time Fourier transform (STFT) with a window length of 5 seconds and 50% overlap. The signal data which acts as the input for the classification method is the raw data recorded by the acoustic sensor. This wav audio file data is not modified. The first aspect performed is the pre-processing (described in section 4) which is end point detection. We then extract features in each breathe (mainly under the frequency of 110 Hz) and finally, we use those features as the input for the BP Neural Network.

Average Accuracy [%] 1-100Hz					
Window[s]	**Under Chin**	**Windpipe**	**Hollow**	**RHS neck**	**LHS neck**
5	46.0	58.7	100.0	68.8	54.2
7.5	50.0	45.5	100.0	70.8	50.0
10	31.3	50.0	100.0	75.0	50.0
12.5	75.0	40.0	100.0	75.0	75.0
15	30.0	37.5	100.0	87.5	50.0

Table 1: 1-100Hz average accuracy

The spectrogram plots for nasal and mouth breathing for each location and for each frequency band displayed visually discernable differences in certain frequency bands of the acoustic signals produced for each type of breathing for this subject. The differentiation between the two types of breathing (see Table 1 and Figure 3) is more discernable for certain recording locations such as the hollow of the neck (Curran et al., 2010).

The interface is shown in Figure 4, where the upper left part is the control, the upper right is the results, the lower left shows the spectrum graph, and bottom right shows the signal wave. The control part has the 'Choose File' button to choose an audio file in the 'wav' format. The 'Detect Breath' button to the right shows the end points in the signal file and extracts the feature of the signal at a certain frequency band under 110 (Hz).

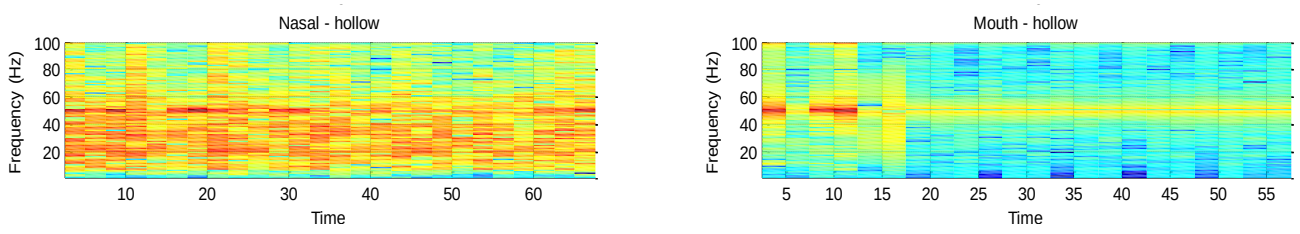

Figure 3: Spectrogram for nasal and mouth (1-100Hz) in hollow

Please cite as: Kevin Curran, Peter Yuan and Damian Coyle (2011) *Using Acoustic Sensors to Discriminate between Nasal and Mouth Breathing.* International Journal of Bioinformatics Research and Applications, Vol. 7, No. 4, pp: 382–396, Sept-Dec 2011, ISSN: 1744-5485, Inderscience,
https://doi.org/10.1504/IJBRA.2012.049623

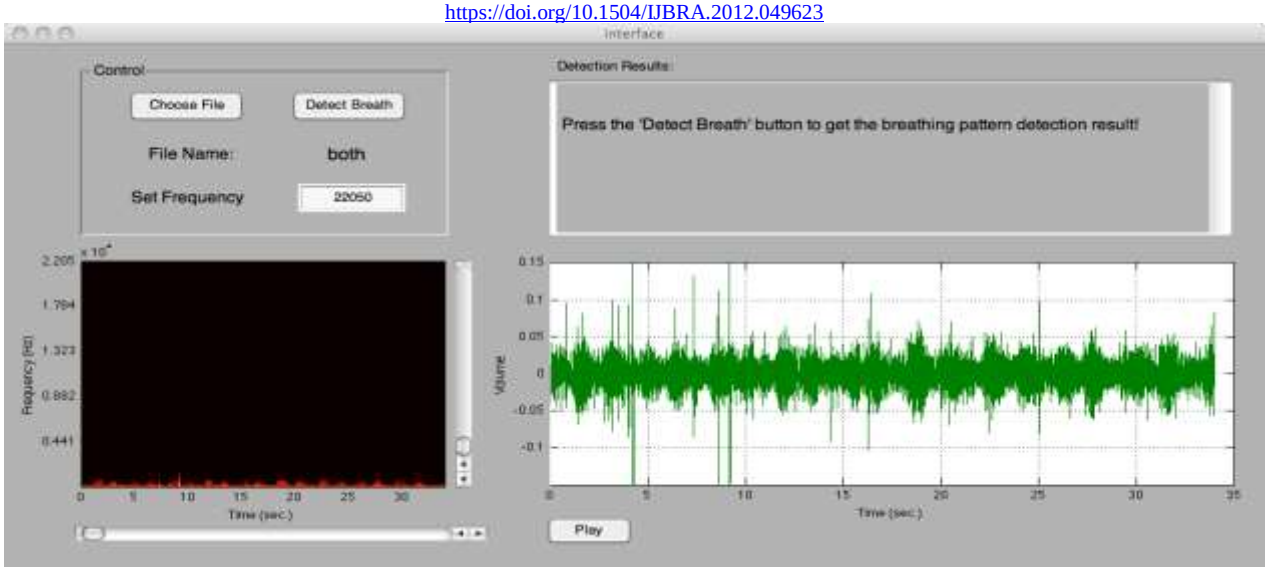

Figure 4: System GUI with a sound file opened

The system then passes the features to the Back-propagation Neural Network to detect the breath patterns. The slider button to the left allows jumping to a specific time to hear the audio. The slider to the right adjusts the frequency on the y-coordinate that actually enables the user to zoom in or out the spectrum to get an overview or a more detailed view of the spectrum range from the sampling frequency to a low of 20 (Hz). The text box is used for displaying results. The plotting area at the lower right corner shows the signal wave after pre-processing. The end-point detection process happens after the 'Detect Breath' button has been pushed. Figure 5 shows the end-point detection function for the mouth breath signal. The blue line indicate the starting point where one breath procedure begins and the dark red line tells the ending point that one breath has finished.

Figure 6 shows accurate breath pattern detection for the nasal breath. Only one mistake occurs in the second breath cycle. The classification rate is 90% correct in this instance. End-point detection result is not as good as for mixed mode breathing as compared to nasal or mouth alone. For the mixed breath pattern sounds, the end-point detection function only detects a little more than half the breath cycles.

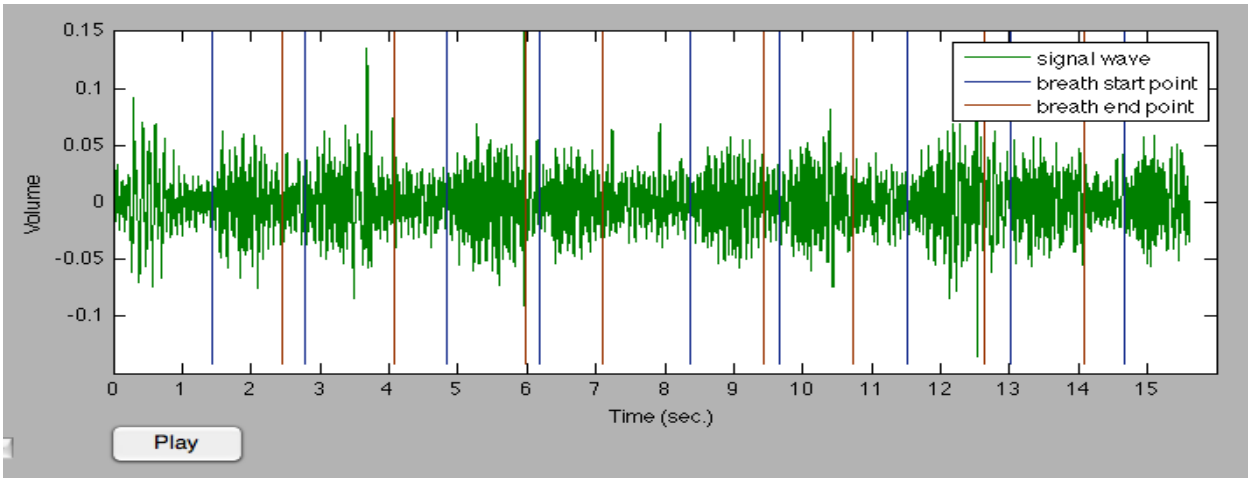

Figure 5 Mouth breathing End-point Detection

The reason for this is that when a person breathes with their mouth for a period and then changes the pattern to nasal breathing, the breath sound is usually smaller than a single breath pattern. If the threshold for the end-point detection is too small, it will detect more false end-points. However, if the threshold is larger, a lot of true end-points will not be detected. Therefore the threshold has to be adjusted to fit each situation.

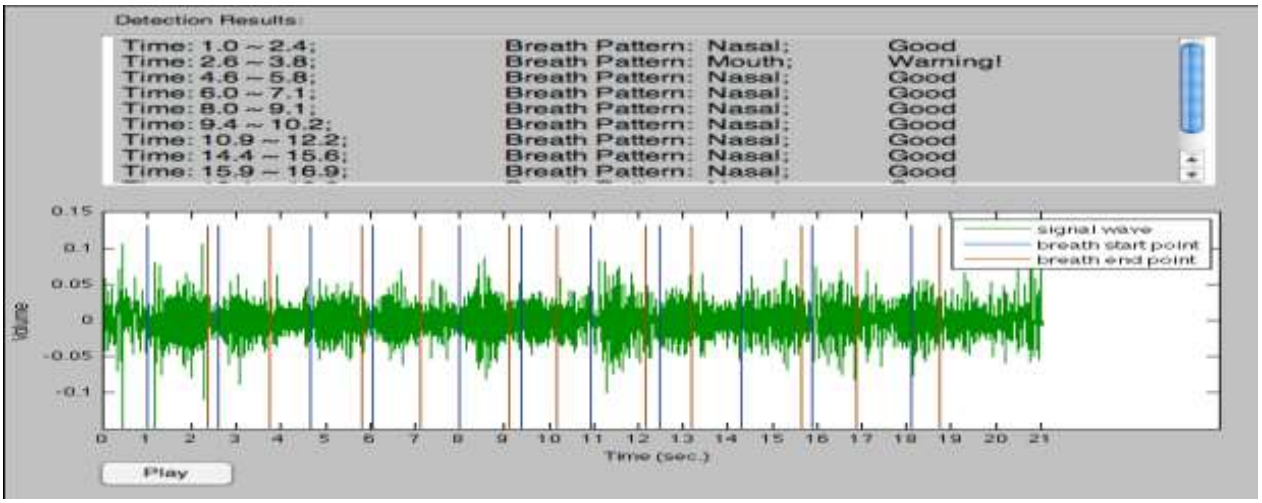

Figure 6 Nasal breath only breathing pattern detection

6. Conclusion

It is well known that the breathing pattern changing from mouth to nose impacts on patients respiratory disease, even in a healthy person. We investigate here whether the necessary discriminatory information of nasal versus mouth breathing can be obtained from acoustic sensors placed at various positions on the body. The experiment result shows that the difference between nasal and mouth breath can be discriminated successfully with a high enough accuracy and therefore integrated into a on-the-body acoustic sensor to try to give appropriate feedback to end-users. End point detection for mouth and nasal breathing is detected with 90% accuracy, however mixed breath patterns yield lesser results.

In addition to this, it has to be taken into account that both a same-day variability and a between-day variability exist in lung sounds. On the basis of these large variations, concrete changes in nasal and mouth will be seen only with investigation of a larger number of subjects. The purpose of computer-supported analysis of breathing sounds is for objective understanding and archiving. Because the sensitivity of human hearing is reduced, particularly in the lower frequency ranges, an objective electronic recording for recognizing deviations within these ranges could be helpful. Since the procedure for doing this is not costly, invasive, or particularly intensive, it would be suitable as an examination method for high-risk groups such as pneumonia patients. A daily or even more frequent analysis of lung-sound spectra could help to identify patients with say, incipient pneumonia before the appearance of any radiologic abnormality. The results presented in this work are from a preliminary analysis in which the experimental protocol was investigated for performing the breathing and also recording locations were tested. There are a number of additional measures which would be incorporated in further studies such as those mentioned above and a more in depth analysis of each recording site or multiple recording sites could be developed to determine if this could improve detection rates for subjects who may not show significant differences between the two breathing types. Further investigations into the physiological findings of the study are ongoing to explain why the acoustics sounds are more prominent in some frequencies and some areas for each breathing type. Further work can be conducted by extracting information from the neural network so as to convince the medical doctor to use it for classification purposes. There may also be merit in building on the work of (Kwong et al., 2009) in extracting useful information (associate rules) from the neural networks.

References

Baird, T., Neuman, M. (1992) A Thin Film Temperature Sensor For Measuring Nasal And Oral Breathing In Neonates, Engineering in Medicine and Biology Society, 1992. Vol.14. Proc of Annual International Conference of the IEEE, Volume 6, Issue , 29 Oct-1 Nov 1992 Page(s):2511 – 2512

Brown L., Prasad N. (1997) Effect of vital signs on advanced life support interventions for pre-hospital patients. Prehosp.Emerg.Care 1997 Jul-Sep; 1(3):145-8.

Cooley J W, Tudey J W. (1965) An algorithm for the machine computation of complex Fourier Series[J]. Mathematical Computation, 1965, 19: 296 ~ 302.

Curran, K., Yuan, P. and Coyle, D. (2010) Discriminating between Nasal and Mouth Breathing via Non-Invasive Acoustic Sensors. IERIC 2010 - Intel European Research and Innovation Conference 2010, pp: 167, Intel Ireland Campus, Leixlip, Co Kildare, 12-14th October 2010

Forgacs P, Nathoo A, Richardson H. (1971) Breath sounds. Thorax 1971; 26:288-95

Godfrey S, Edwards B, Campbell E, Armitage P., Oppenheimer E. (1969) Repeatability of physical signs in airways obstruclion. Thorax 1969; 24:4-9

Gravelyn T., Weg J. (1980). Respiratory rate as an indication of acute respiratory dysfunction. JAMA 1980, Sep;244(10):1123-5.
Gavriely, N, Nissan, M, Rubin, A. (1995). Spectral characteristics of chest wall breath sound in normal subjects. Thorax 50:1292–1300.

Gross, V., Dittmar, A., Penzel, T., Schüttler, F., and Von Wichert, P. (2000) Relationship between Normal Lung Sounds, Age, and Gender, A. J. of Respiratory Critical Care Medicine, Vol. 162. pp 905–909, 2000

Hidalgo, H., Wegmann, M. and Waring, W. (1991) Frequency spectra of normal breath sounds in childhood, *Chest* 1991;100;999-1002

Kwong, C.K., Wong, T.C. and Chan, K.Y. (2009) A methodology of generating customer satisfaction models for new product development using a neuro-fuzzy approach, Expert Systems with Applications, Vol. 36, No. 8, pp. 11262-11270

Kraman, S., Wodicka, G., Pressler, G. and Pasterkamp, H. (2006) Comparison of lung sound transducers using a bioacoustic transducer testing system, J Applied Physiol 101:469-476, 2006. April 2006

Loudon R and Murphy R. (1984) Lung sounds (Review). Am Rev Respir Dis 130: 663–673, 1984.

Malmberg, L. Sovzjdrvi, A., Paajanen, E., Piirild, P., Haahtela, T. and Katila, T. (1994) Changes in Frequency Spectra of Breath Sounds During Histamine Challenge Test in Adult Asthmatics and Healthy Control Subjects, Chest, Vol. 105, No. 1, pp: 122-133

Mahagna, M., and Gavriely, N. (1994). Repeatability of measurements of normal lung sounds. Am. J. Respir. Crit. Care Med. 149:477–481.

McKusick V, Jenkins J, and Webb G. (1955) Acoustic basis of the chest examination; sound spectrography studies. Am Rev Tuberc 72: 12–34, 1955.

Pasterkamp H, Wiebicke W, Fenton R. (1987) Subjective assessment vs computer analysis of wheezing in asthma. Chest 1987; 91:376-81

Pasterkamp H, Consunji-Araneta R, Oh Y, and Holbrow J. (1997) Chest surface mapping of lung sounds during methacholine challenge. Pediatr Pulmonol 23: 21–30, 1997.

Rozenblat V, Likhacheva E, Stolbun B, Sazhina T, Forshtadt V, and Ganiushkina S. (1968) The frequency spectrum of respiratory sounds in normal subjects and in pneumoconiosis. Klin Med (Mosk) 46:135–139, 1968.

Schreiber J, Anderson W, Wegmann M, Waring W. (1981) Frequency analysis of breath sounds. Med Instrum 1981; 15:331-34

Smylle H., Blendis L., Armitage P. (1965) Observer disagreement in physical signs ofthe respiratory system. Lancet 1965; 2:412-13

Yu, M., Ko, J., Lin, C., Chang, C, Yang , Y., Lin, S., Chen, J., Chang, K., Kuo, S., Hsu, S., Hung, Y. (2008) Multimedia feedback for improving breathing habits, Ubi-Media Computing, 2008 First IEEE International Conference, July 31 2008-Aug. 1 2008, pp:267-272