CHEST EXPANSION RECONSTRUCTION FROM RESPIRATION SOUND BY USING ARTIFICIAL NEURAL NETWORKS

A. Bonarini*, M. Matteucci*, S. Tognetti*

*Politecnico di Milano, Italy, IIT-Unit, Dip. Elettronica ed Informazione

Keywords: Affective computing, chest expansion, sound, Artificial Neural Network.

Abstract

Affective Computing is a growing area in which researchers are focusing on the recognition of emotions through the analysis of biomedical signals. Emotion recognition is useful when it is done during real life activities; this is possible only by the use of devices that can be easily worn by the subject and that do not affect his/her activities. In this work, we present a way to reconstruct the chest expansion signal, usually measured by an uncomfortable belt around the chest, from the analysis of respiration sound gathered with a microphone placed on the upper part of the neck. We will show that it is possible to reconstruct the respiration spectrum with an error lower than 0.06Hz in the frequence that characterizes it.

1 Introduction

The goal of affecting computing is to recognize emotions during real life activities and, in such case, it is not confortable to have too many sensors and cables around the body. Since many years, researchers have focused on the possibility to collect physiological data in different ways [4, 1] so that a sensor influences as less as possible the normal activity of the subject; having small and easy-to-wear devices is important to maintain the natural behavior of people that are not conditioned by their use.

One of the most significant signals is the chest expansion (or respiration signal) because from its analysis it is possible to get a lot of information about the subject state. For instance, we can assume that having only one dominant frequency means that the subject is doing something difficult to do because the respiration pattern repeats constantly with a high frequency. On the other hand, if the spectrum is spread over all the frequencies, the activity can be interpreted as easy. Short time measures of the chest expansion are also interesting; for instance, if the breath stop suddenly, it could be possible that an unexpected or fearing situation is just happened.

Hult et al. [2] proposed a method for respiration monitoring by the use a bio-acoustic signal. The method is based on the recognition of respiratory pause using the sum of the sound spectrum into a specific band. They also introduced a way to discriminate between inspiratory and expiratory phases. The method is able to give information on the time when each phase starts/ends and it relays on a professional microphone. This method gave us a first idea on the way it is possible to perform respiration reconstruction, but it does not address completely the problem since the obtained result was just a square wave that identifies each respiratory phase.

In this work, a new technique for chest expansion reconstruction is presented, based on the analysis of the respiration sound obtained by a microphone placed on the upper part of the subject neck. This method uses a particular signal processing technique, based on artificial neural networks, to reconstruct chest expansion. This work aims at spectrum reconstruction instead of signal reconstruction, the spectrum gives most of the information needed for emotion recognition. Moreover, we want to provide a method than can be used with commercial microphones to have a broad range of applications.

In the next sections, we firstly introduce the method, then we discuss the results of this work, and finally we draw some conclusions.

2 Chest expansion reconstruction

When a person is breathing, a sound is produced by the air going inside and outside the trachea. The intensity of the sound produced is related to the flow of air that is going in/out the lung. The integral of such flow is a measure of how much the chest has been contracted or expanded.

The elaboration process starts from the signal acquisition. In Figure 1 there is an example of the chest expansion signal, as detected by the respiration Sensor SA9311M (Thought Technology), to which we will refer as $resp(t)$. In the same figure the corresponding sound is also reported.

Firstly, the sound is sampled at a frequency of 1024 sample/sec and then the spectrogram is computed (256 samples window with 224 overlapping elements). In Figure 2, a short segment of chest expansion and the corresponding sound spectrogram are presented.

It is possible to see that in the 0-190Hz band there is a constant noise probably due to the low quality microphone. In the remaining frequency range there is an activity, over the time, that could be related to the respiration. The spectrum is integrated over the band 190-512Hz and the resulting signal $x_s(t)$ is related to the flow of the air that is going through the trachea. The louder is the signal, the higher is the flow. From $x_s(t)$, it is not possible to understand which is the direction of the flow.

Even if we do not know the real flow direction, we can relate $x_s(t)$ to the absolute value of the respiration gradient. We introduce a technique to smooth $x_s(t)$ from the noise

Authorized licensed use limited to: Politecnico di Milano. Downloaded on October 27, 2009 at 20:09 from IEEE Xplore. Restrictions apply.

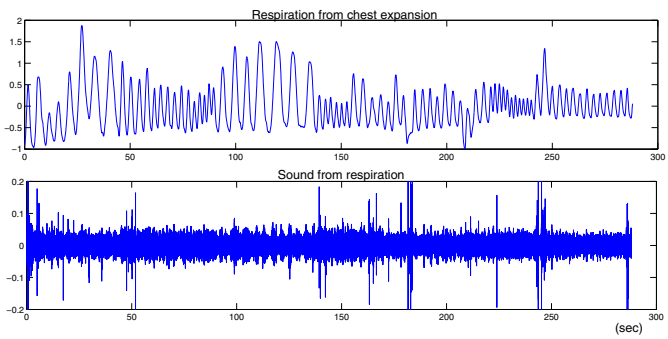

Figure 1: Example of respiration and corresponding sound.

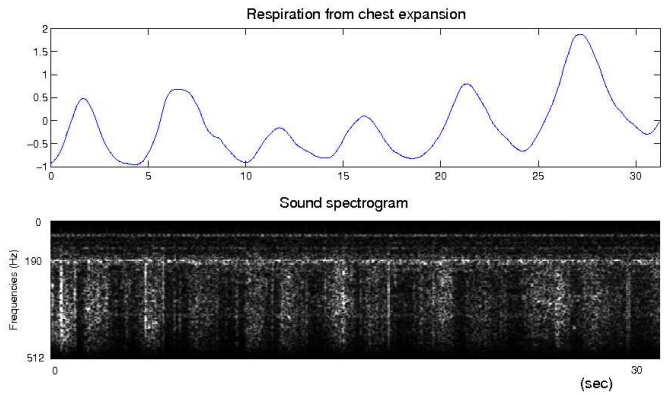

Figure 2: Example of respiration signal and the sound spectrogram.

that affects the spectrogram, from which it is derived and to eliminate spikes resulting from casual environmental sounds. This is done by applying a median filter to the spectrogram in the time dimension (16 samples) and summing again into the 190-512Hz band. We will refer to the resulting signal as $xm_s(t)$.

In Figure 3, $|\frac{d}{dt}resp(t)| = d_{resp}(t)$, $x_s(t)$ and $xm_s(t)$ are presented. It is easy to see from Figure 3 that $d_{resp}(t)$ and $xm_s(t)$ are highly correlated. This confirms our first observations that the sound produced by the air going through the trachea is related to the flow that is going inside it and also that the sound gives us information about how much the chest is expanding/contracting. The higher is the value of $xm_s(t)$, the higher is the flow. Furthermore, if we know the inspiration time window, by integrating $xm_s(t)$ over it, we can estimate how much air is gone inside the lung and so we know how much the chest has expanded.

Consider the sequence of time steps for which $d_{resp}(t) = 0$ namely: $\{t_1, t_2, .., t_k, .., t_n\}$. These are situations in which the respiration signal is in a local maximum or local minimum. Knowing, for instance, that at the initial condition the chest is expanding (or contracting), it is possible to reconstruct $resp(t)$ from $d_{resp}(t)$:

$$resp(t)|_{t\in(t_k,t_{k+1}]} = \int_{t_{k-1}}^{t_k} d_{resp}(\tau)d\tau - \int_{t_k}^{t} d_{resp}(\tau)d\tau \quad (1)$$

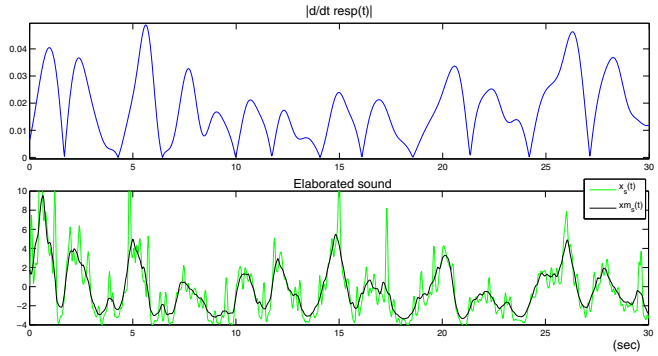

Figure 3: Absolute value of respiration signal gradient and the elaborated sound. In the bottom graph the solid line represents $xm_s(t)$, the dotted one $x_s(t)$

By the use of this consideration one could expect that it is also possible to reconstruct the respiration signal ($resp(t)$) from the elaborated sound ($xm_s(t)$).

The reconstruction process starts from the elaborated sound and gives as output the related respiration signal. This is done by the use of a dynamical filter applied to the elaborated sound $xm_s(t)$ to get a signal that is much more similar to $d_{resp}(t)$. This filter is made up of an artificial neural network that maps $xm_s(t)$ to $d_{resp}(t)$. These two signals are first resampled at a frequency of 8Hz because they have a lower frequency range and $d_{resp}(t)$ is also normalized into the 0-1 range. The artificial neural network (ANN) we chose has a maximum of 64 inputs that, at 8Hz, are used to map 8 seconds of $xm_s(t)$ signal to one value of $d_{resp}(t)$. If we reflect the 8-sec input time window on the target signal $d_{resp}(t)$, we have 64 possibles positions P* to which the mapping can be done (we exclude positions outside this window). The ANN is used, indeed, as a non-linear filter in which we can both select the prediction position P* and inputs to use. The network has two hidden layers, the first one with 32 neurons and the second one with 16 neurons. These layers have a tan-sigmoid transfer function. The output layer has, instead, a log-sigmoid transfer function.

After one possible input/output configuration is selected the network is trained and the network output is taken as best approximation of our target signal $d_{resp}(t)$. We use a 5-fold cross-validation technique; inputs are divided into 5 parts: four are used for training a the last one is used half for testing and half for early stopping. We will refer to the network output as $\widehat{d}_{resp}(t)$.

The last operation is the reconstruction of respiration signal from the network output. This operation is performed finding the time steps for which $\widehat{d}_{resp}(t)$ has a local minimum that should also be time steps for which $d_{resp}(t) = 0$. This is not true in general because the sound is not perfect and the elaboration process produces some false minima that cause a local wrong reconstruction. This is a big problem that produces a displacement in the reconstructed signal, but the spectrum is practically preserved.

We have also applied a linear regressor to check the need of the neural network, but even if the reconstruction is

Authorized licensed use limited to: Politecnico di Milano. Downloaded on October 27, 2009 at 20:09 from IEEE Xplore. Restrictions apply.

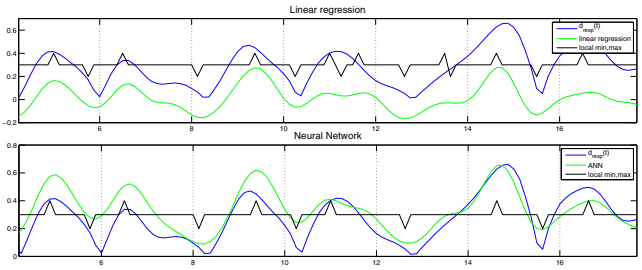

Figure 4: Linear regression and Neural network output with local minima/maxima computation

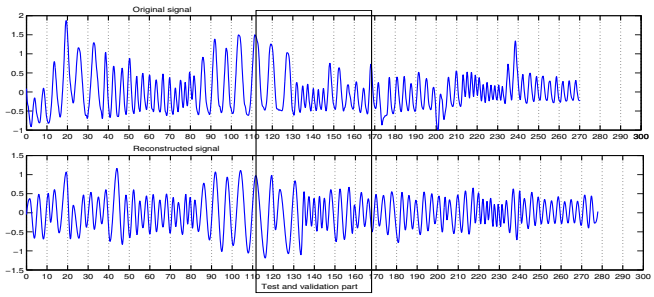

Figure 5: Example of respiration reconstruction

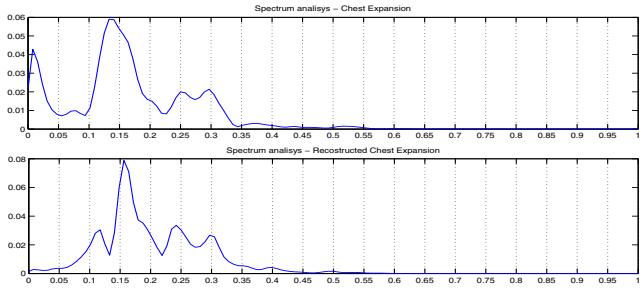

Figure 6: Spectrum of original and reconstructed respiration

apparently good, the performance is lower than the one obtained by the ANN. In Figure 4, there is an example of linear regressor and neural network output with local minima computation. A sequence of impulses is obtained in which the positive value represents a maximum and the negative a minimum. It is possible to see that most of the times the minimum of the ANN output and linear regressor output are the same, but there are also wrong minima that do not correspond to zero values of $d_{resp}(t)$. Non linear model gives a better signal which has less false-positive minima. We have not implemented any false-positive detection mechanism and this could be an improvement of the work. After the sequence of impulses is computed, we chose all the time steps $\{t_1, t_2, \ldots, t_k\} \cup \{t_{min}, t_{max}\}$ for which we have a local minimum; these are, according to what $\widehat{d}_{resp}(t)$ represents, possible stationary points of $resp(t)$. If we start form the hypothesis that in t_{min} the chest begins to expand, in t_1 the expansion is finished and the contraction lasts until t_2. The reconstructed respiration in $[t_1, t_2]$ is obtained interpolating one point in t_1 whose value is $\int_{t_{min}}^{t_1} \widehat{d}_{resp}(\tau)d\tau$ and one point whose value is $-\int_{t_1}^{t_2} \widehat{d}_{resp}(\tau)d\tau$ according to Equation 1. The reconstruction continues by considering following time steps until the end of the sequence.

3 Results

The signal acquisition is done with an ADC converter (Procomp Infinity, Thought Technology) to which both the chest expansion sensor (Respiration Sensor SA9311M, Thought Technology) and the microphone, with its pre-amplifier circuit, are connected. The ADC converter acquires signals at a frequency of 2048 sample/sec and we use it to acquire both signals to get perfect synchronization.

The experiments we made are based on five minutes of registrations performed on four subjects in a not too noisy environment (i.e. office desk). The results are presented with a graphical analysis showing a typical reconstructed signal and with a numerical analysis in which the reconstruction performance is evaluated with respect to three indexes namely: frequency of maximum power (FP), mean frequency of power spectra (MPF) and highest frequency at which the power in the spectrum is at least 10 percent of the maximum power (FM). These indexes are the same used in [3].

In Figure 5, an example of the chest expansion signal and the reconstructed one by the use of ANN are presented. This plot gives us a first idea on the quality of the reconstruction. The first element to notice is that there is no perfect synchronization between the two signals. This is due to the reconstruction phase which is affected by the presence of false-positive minima. Even if we have such problems, the shape of the signal is preserved and the resulting signal follows quite accurately the original one. This is also confirmed by the spectrum of the two signals that are shown in Figure 6. It is possible to see that very low frequencies are filtered (under 0.05Hz). These frequencies represent variations in the air that remains into the lung that are not reconstructed during the process. There are also some unexpected frequencies that have been introduced by the non-perfect reconstruction, but the shape of the spectrum is also preserved. We want to point out that the reconstruction of the spectrum is our first objective and we will perform some validations to discover if we can truly replace the respiration sensor with the microphone.

In the numerical analysis we have reconstructed 7 signals (5 minutes each) coming from four subjects in a low-noise environment. For each performance index, we have computed the difference between the original signal and the reconstructed one. The result of this difference is the frequency error made in the reconstruction. In Figure 7, we show a box and whiskers plot of the results for the artificial neural network. Each index is represented by the use of median, first quantile, third quantile, minimum value and maximum value. Possible outliers are represented by "plus". This graph gives a better picture of the performances. The first index is the most important because it tells us the main frequency of the respiration. FP has a median error of 0.0155Hz, that is, in the 0-1Hz band, the

Authorized licensed use limited to: Politecnico di Milano. Downloaded on October 27, 2009 at 20:09 from IEEE Xplore. Restrictions apply.

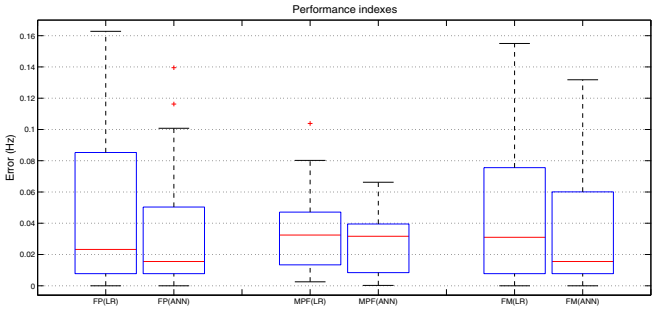

Figure 7: Numerical evaluation of the reconstruction process with comparison between the linear regressor (LR) and the artificial neural network(ANN)

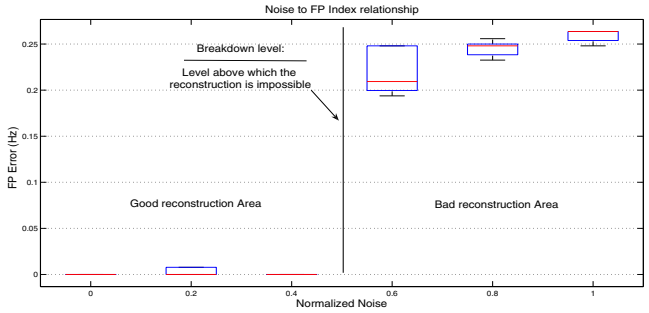

Figure 8: Noise to Performance relationship. As the noise increases, the performances do not decrease until we reach a breakdown noise value above which the reconstruction is impossible

1.5% of the frequency range. The third quantile tells us that the most probable error is less than 0.0503Hz (5% of the range). We have of course some outliers that have an error up to 0.1395Hz that are cases in which the reconstructed FP frequency became one of the others dominant frequencies in the original signal. The MPF shows the lowest error, meaning that the mean frequency of the spectrum is always reconstructed. This confirms the previous observations made for FP. FM gives us information on the weaker frequencies. Also in this case, we have good performance since the most probable errors fall in the range 0.0078-0.06Hz.

We have also compared the performance of the ANN with the performance of the linear regressor. For each performance index the ANN have a lower error meaning that a non-linear model is better for the reconstruction of the respiration. In Figure 7 there is a box and whiskers plot of the results for the linear regressor.

Other kind of experiments have been performed in order to establish the relationship between noise and reconstruction performance. In these experiments a subject is sitting 1 meter far from a computer speaker that is playing a music. We have recorded 6 sessions of 5 minutes for one subject in order to make results comparable. In each session the same music was playing at different volume levels; we have indeed 6 values of noise that range from 0 (no music) to 1 (high volume music). In Figure 8 there is a box and whiskers plot of the results. It is possible to see that as the noise increases the considered performance index (we present only the FP index) do not decrease until we reach a breakdown noise value above which the reconstruction is impossible.

4 Conclusions

In this work, we have presented a way to reconstruct the respiration signal from the analysis of respiration sound. We have shown that the process can be done even with the use of commercial microphones cheap and easy to plug directly into the computer audio card. We have presented the errors made in the prediction of three major frequencies of the spectrum namely: peak frequency (FP), mean frequency (MPF) and weaker frequency (FM). The most probable errors for the ANN are under the 5% of the considered range(0-1Hz). The median value of the error is under the 4% and we believe that with some improvements this error could be further reduced. On the contrary, the linear regressor has lower performances since most probable errors are under the 8%.

We have also shown that the method is quite robust to noise because, if the noise is under a threshold, it does not affect the performances. As expected, if the noise increases too much the method is totally unuseful.

This is a good result especially for application in affective computing. The new sensor can now be applied in affective computing applications, if environmental noise conditions are good, to find how the respiration signal we obtain is related to a specific emotional state.

References

[1] D. De Rossi, F. Carpi, F. Lorussi, A. Mazzoldi, R. Paradiso, E.P. Scilingo, and A. Tognetti. Electroactive Fabrics and Wearable Biomonitoring Devices. *AUTEX Research Journal*, 3(4):180–185, 2003.

[2] P. Hult, T. Fjallbrant, S. Dahle, P. Danielsson, and P. Ask. A method for respiration monitoring by the use of a bioacousticsignal. *Advances in Medical Signal and Information Processing, 2000. First International Conference on (IEE Conf. Publ. No. 476)*, pages 22–25, 2000.

[3] C.S. Lessard, W.C. Wong, and B. Prograni. Correlation of Constant Flow Rate with Frequency Spectrum of Respiratory Sounds When Measured at the Trachea. *Biomedical Engineering, IEEE Transactions on*, pages 461–463, 1986.

[4] R. Paradiso, G. Loriga, and N. Taccini. A wearable health care system based on knitted integrated sensors. *Information Technology in Biomedicine, IEEE Transactions on*, 9(3):337–344, 2005.

Authorized licensed use limited to: Politecnico di Milano. Downloaded on October 27, 2009 at 20:09 from IEEE Xplore. Restrictions apply.