AIRFLOW ESTIMATION FROM RESPIRATORY SOUNDS

by

İlhan Yıldırım

B.S., Electrical & Electronics Engineering, Boğaziçi University, 2013

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Electrical & Electronics Engineering
Boğaziçi University
2017

AIRFLOW ESTIMATION FROM RESPIRATORY SOUNDS

APPROVED BY:

Prof. Yasemin Z. Kahya ....
(Thesis Supervisor)

Assist. Prof. İpek Şen ....
(Thesis Co-supervisor)

Prof. Emin Anarım ....

Assoc. Prof. Ali Emre Pusane ....

DATE OF APPROVAL: 01.01.2017

ACKNOWLEDGEMENTS

Acknowledgements come here...

ABSTRACT

AIRFLOW ESTIMATION FROM RESPIRATORY SOUNDS

One page abstract will come here.

ÖZET

SOLUNUM SESLERİNDEN SOLUK AKIŞI KESTİRİMİ

Bir sayfa uzunluğunda özet gelecektir.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

a_{ij}	Description of a_{ij}
$\mathbf{A}$	State transition matrix of a hidden Markov model
α	Blending parameter *or* scale
$\beta_t(i)$	Backward variable
Θ	Parameter set

LIST OF ACRONYMS/ABBREVIATIONS

AR	Autoregressive Model
FFT	Fast Fourier Transform
TVAR	Time Varying Autoregressive Model
LPF	Low Pass Filter
LMS	Least Mean Squares
LS	Least Squares
RLS	Recursive Least Squares
HPF	High Pass Filter
STFT	Short Time Fourier Transform
PSD	Power Spectral Density

1. INTRODUCTION

Start with an introduction...

2. EXPERIMENTAL SETUP AND DATA

Start with an introduction...

3. AIRFLOW CURVE ESTIMATION

3.1. Literature Review

3.2. Autoregressive Modeling

In this chapter, we will analyze the respiratory sounds as if they are time varying autoregressive processes and try to find a correlation between time varying autoregressive coefficients and airflow over the mouth. We will look at three different methods for time varying autoregressive modeling of a signal after we give the description of autoregressive processes. Three methods are

- Windowing Based Autoregressive Modeling
- Time Varying Autoregressive Modeling with Basis Functions
- Time Varying Autoregressive Modeling with Particle Filters.

3.2.1. Description of Autoregressive Processes

Autoregressive process is a class of discrete random processes whose output at a time can be written with a fixed combination of number of past values plus a sample from a white Gaussian noise. It can be formulated as in 3.1. The a_i's are the AR coefficients, N is the order of process and e is the noise term which has a constant variance.

$$x(n) = \sum_{i=1}^{N} a_i x(n-i) + e(n) \tag{3.1}$$

An autoregressive process is equivalent to a infinite impulse response filter whose input is white Gaussian noise. Autoregressive processes are widely used in many areas, including economics, statistics and signal processing. In general, and in this thesis too, the processes used for analysis purposes are assumed to be stationarity.

<u>3.2.1.1. Mean, Variance and Autocorrelation Of Autoregressive Processes.</u> Expectation is a linear operator, then we can write the equation 3.2. Recalling that e is white Gaussian i.e. $\mu_e = 0$ at least one of the followings must hold $\sum_{i=1}^{N} a_i = 1$ or $\mu_X = 0$. For variance, one can write the 3.3. Since white noise is uncorrelated with the values of process, the last term is zero. However, the middle term is a finite, nonzero value. This implies $\sum_{i=1}^{N} a_i < 1$.

$$\mu_X = \sum_{i=1}^{N} a_i \mu_X + \mu_e \tag{3.2}$$

$$\sigma_X^2 = \sum_{i=1}^{N} a_i \sigma_X^2 + \sigma_e^2 + 2\sigma_{Xe} \tag{3.3}$$

If we reorganize what we have from deductions we had while calculating mean and variance, we can list as following:

- $\sum_{i=1}^{N} a_i = 1$ or $\mu_X = 0$, from mean
- $\sum_{i=1}^{N} a_i < 1$, from variance

Now we can say that $\sum_{i=1}^{N} a_i < 1$ is needed for an autoregressive process to be at least wide sense stationary and the mean of an autoregressive process must be zero. In order to look at autocorrelation, let's first rewrite the variance equation as in 3.4, 3.5 and 3.6.

$$\sigma_X^2 = E(X(n)X(n)) - \mu_X^2 \tag{3.4}$$

$$\sigma_X^2 = \sum_{i=1}^{N} a_i E(X_n X_{n-i}) + E(X_n e_n) - \mu_X^2 \tag{3.5}$$

$$\sigma_X^2 = \sum_{i=1}^{N} a_i R_{XX}(i) + \sigma_e^2 \tag{3.6}$$

One of the descriptive functions of random processes is autocorrelation function given in 3.7. If we rewrite X(n) as in 3.1 we can conclude in a recursive equation for autocorrelation function as in 3.8. The equations in 3.8 are called "Yule-Walker Equations".

$$r_{XX}(k) = \frac{E[X(n-k)X(n)]}{\sigma_X^2} \tag{3.7}$$

$$r_{XX}(0) = 1, \qquad r_{XX}(k) = \sum_{i=1}^{N} a_i r_{XX}(k-i), \qquad k \geq 1 \tag{3.8}$$

This leads to a difference equation which can be expressed with vector equations as in 3.9.

$$\begin{pmatrix} r_{XX}(1) \\ r_{XX}(2) \\ \vdots \\ r_{XX}(N-1) \\ r_{XX}(N) \end{pmatrix} = \begin{pmatrix} 1 & r_{XX}(1) & ... & r_{XX}(N-2) & r_{XX}(N-1) \\ r_{XX}(1) & 1 & ... & ... & r_{XX}(N-2) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ r_{XX}(N-2) & ... & ... & 1 & r_{XX}(1) \\ r_{XX}(N-1) & r_{XX}(N-2) & ... & r_{XX}(1) & 1 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_{N-1} \\ a_N \end{pmatrix} \tag{3.9}$$

3.2.2. Description Of Time Varying Autoregressive Processes

Time varying autoregressive process is defined in 3.10. In this model, AR coefficients are dependent on time. This model is very useful for applications where signals are not stationary.

$$x(n) = \sum_{i=1}^{N} a_i(n) x(n-i) + e(n) \tag{3.10}$$

3.2.3. Autoregressive Modeling With Windows

This method divides the signal into parts and finds the autoregressive coefficients by a well known method such as Yule-Walker or Burg or direct matrix inversion. The important thing to note is the divided parts are assumed to be stationary. The method is given in algorithm 1.

Algorithm 1 AR Coefficients Estimation With Overlapping Windows

procedure WINDOWEDAR(signal, order, windowLength, overlap)
    $windowStart \leftarrow 1$
    $L \leftarrow Length(signal)$
    $lastWindowStart \leftarrow L - windowLength$
    while $windowStart \leq lastWindowStart$ **do**
        $temp \leftarrow signal(windowStart : (windowStart + windowLength))$
        $AR(:, i) \leftarrow EstimateAR(temp, N)$
        $windowStart \leftarrow windowStart + windowLength$
    end while
    return AR
end procedure

There are several methods to estimate the AR coefficients of given signal. Yule Walker Method, Burg's Method, Least Squares are the well known methods. Usually Yule Walker approach is used, however it is reported that Yule-Walker method fails where the autocovariance matrix is poorly conditioned and least squares approach doesn't guarantee the estimated autoregressive model to be stable, we used Burg's method to estimate autoregressive coefficients [Wharton Statistics].

To test the relation between ar coefficients and flow, we looked at the correlation between the AR coefficients of overlapping windows and down sampled version of flow. We also looked at the correlation between AR coefficients and absolute value of flow. The absolute value of correlation between downsampled version of flow and array of estimated AR coefficients for different model and coefficient orders is given in figures

3.1 and 3.2. Third channel is chosen randomly for this analysis.

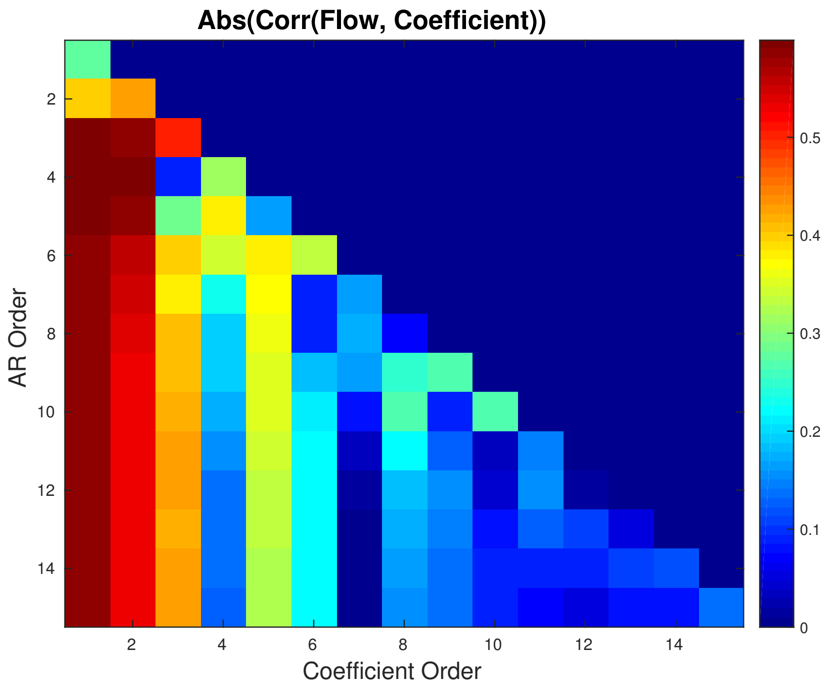

Figure 3.1. Absolute Correlation Between Flow and AR Coefficients for Different Model and Coefficient Orders

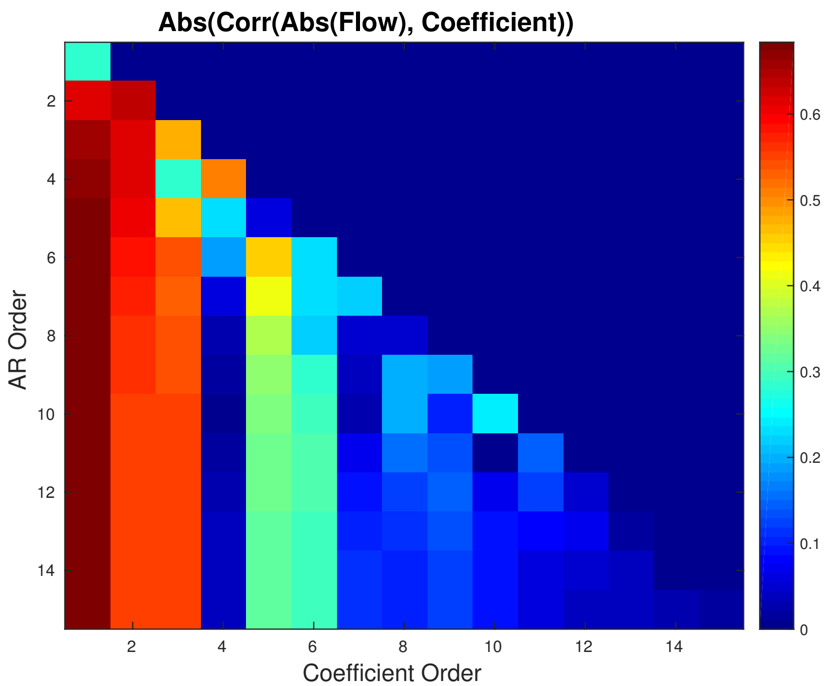

Figure 3.2. Absolute Correlation Between Absolute Flow and AR Coefficients for Different Model and Coefficient Orders

It can be concluded that the greatest correlation is observed for the first AR coefficient in both cases and the correlation with absolute value of flow is greater. It is seen that the correlation increases with model order however it doesn't increase significantly after a point. To find this point, we plot the correlation of absolute flow with first AR coefficient for different model orders in figure 3.3. In the boxplot graph, each box consists of 14 values, each is the average of corresponding channel calculated over 25 subjects. The graph showing the ratio of mean of correlation to variance of correlation is given in By looking at figures 3.3 and 3.4, we decided to use 6 as our model order for further analysis.

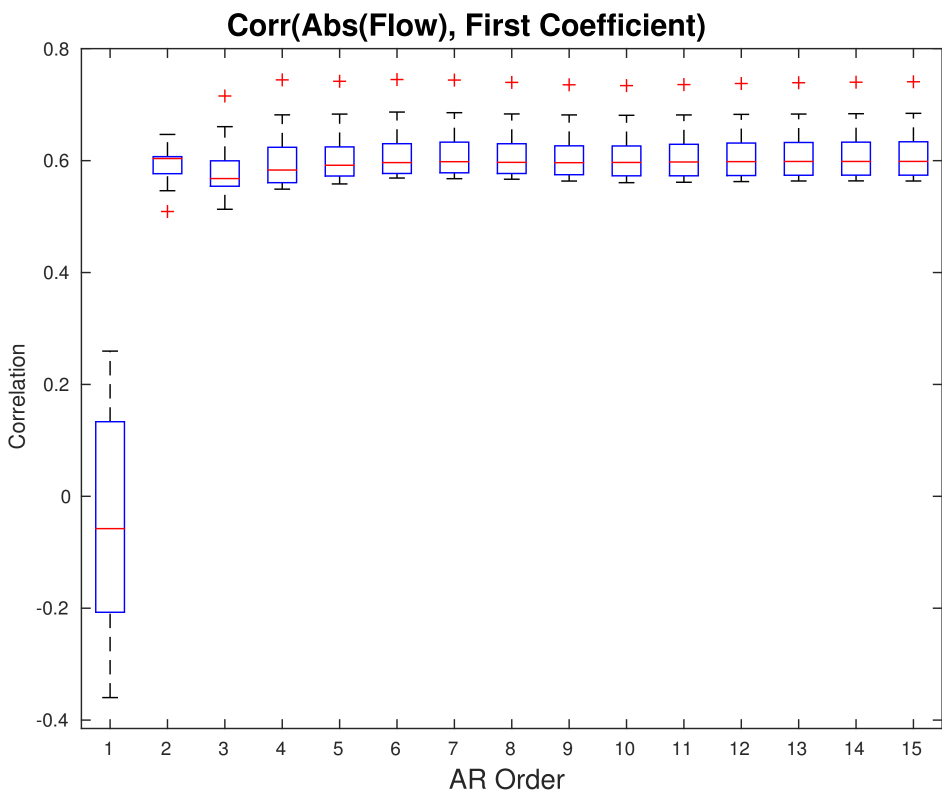

Figure 3.3. Average Correlation Between Absolute Flow and First AR Coefficient for Different Model Orders

Figure 3.4. Ratio Of Correlation Mean To Correlation Variance for Different Orders

After setting model order to 6, we run tests to choose the window length and overlap ratio pair giving the best correlation. The resulting graph is given in figure 3.5. From the figure, one can say that window length must be greater than 400 and must not exceed 1200 and overlap is not making a lot of difference. This makes sense because the correlation is calculated with the decimated versions of flow, in other words increasing the overlap ratio increases the resolution.

After deciding AR order, window length and overlap, we run tests to find the optimum post filtering setup. To test this, we passed the estimation through butterworth filters with different cutoff frequency and orders. We reached the figure 3.6 for different orders and cutoff frequencies.

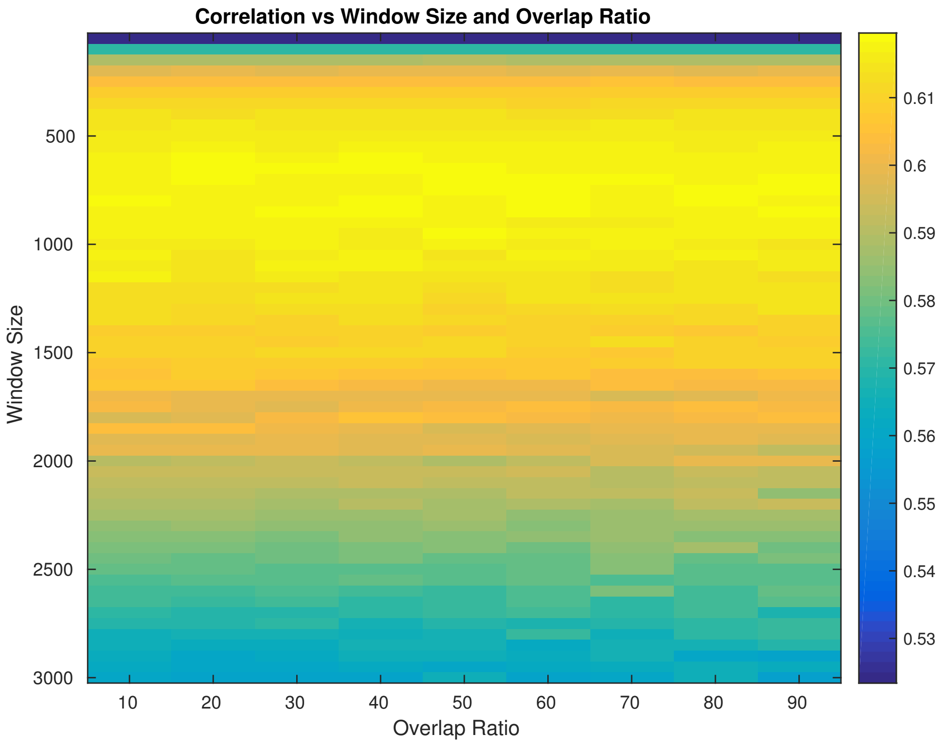

Figure 3.5. Average Correlation Between Absolute Flow and First AR Coefficient for Different Window Lengths and Overlaps

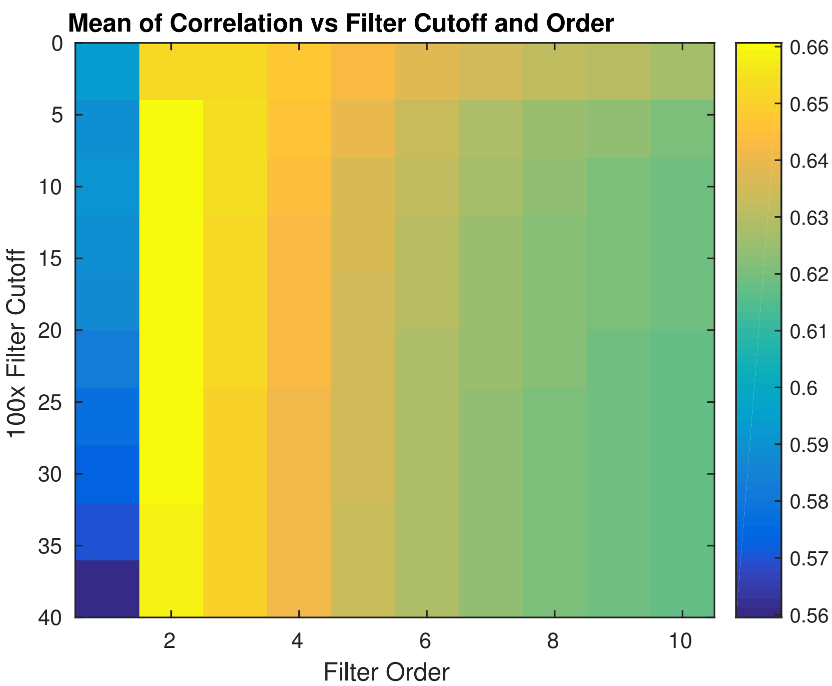

Figure 3.6. Average Correlation For Different Filter Cutoff and Orders

3.3. Time Varying Autoregressive Modeling

Time varying auto regressive (TVAR) modelling is widely used in nonstationary signal analysis. TVAR is also used to analyze respiratory sounds. It is reported in [Koray, SİU] a good correlation is achieved with first TVAR coefficients and absolute value of flow curve. In this thesis, we will go over two approaches for this modeling. The first one is using combinations of Fourier basis functions for coefficients, the second one is using particle filters. The equation defining a TVAR process is given in (3.10). As the name suggests, AR coefficients are changing over time.

3.3.1. Fourier Basis Functions

One of the methods for TVAR coefficient estimation is using Fourier basis functions. This method rely on the assumption that the coefficients to be estimated have some periodicity. Using Fourier basis functions for airflow is based on the assumption that airflows are mostly periodic and one can approximate them by a linear combination of sine and cosine waves.

3.3.1.1. Method. The assumption can be formulated as in (3.11). Each coefficient is a linear combination of K basis functions with constant weights b_j

$$a_i(n) = \sum_{j=1}^{K} a_{ij} u_j(n-i) \tag{3.11}$$

By rewriting (3.10) using (3.11) for coefficients, we can end up in

$$x(n) = \sum_{i=1}^{N} \sum_{j=1}^{K} a_{ij} u_j(n-i) x(n-i) + e(n) \tag{3.12}$$

3.3.1.2. Tests and Results.

3.3.2. Particle Filtering

3.3.2.1. Method.

3.3.2.2. Tests and Results.

3.4. Time - Frequency Analysis

3.4.1. Method

3.4.2. Tests and Results

3.5. Unifying Estimations

3.5.1. Method

3.5.2. Tests and Results

4. AIRFLOW PHASE DETECTION

Start with an introduction...

4.1. Estimation of Periods

4.2. Estimation of Change Point Locations

4.3. Feature Extraction

4.4. Results

5. CONCLUSION

The conclusions of the thesis should come here.

REFERENCES

APPENDIX A: APPLICATION

The appendices start here.