Imperial Journal of Interdisciplinary Research (IJIR)
Vol-2, Issue-5, 2016
ISSN: 2454-1362, http://www.onlinejournal.in

Comparison of Different Speech Enhancement Techniques

HardikPanchmatia, Karan Gaikar&Dharmesh Patel
Department of Electronics & Telecommunications, K. J. Somaiya Institute of Engineering & Information Technology, Mumbai, India

Abstract*:Speech enhancement is concerned with the processing of corrupted or noisy speech signal in order to improve the quality or intelligibility of the signal. Our goal is to enhance speech signal corrupted by noise to obtain a clean signal with higher quality. However, the presence of noise in speech signals will contribute to a high degree of inaccuracy in a system that requires speech processing. For this, we apply here Kalman, Weiner and Linear Predictive Coding filters to the noise corrupted signal and try to determine the filter with the best signal to noise ratio.*

1. INTRODUCTION

The problem of enhancing speech signals degraded by noise received considerable attention by engineers and researchers in the past and a variety of systems have been proposed. Recently, this interest in enhancing the speech signal quality has emerged again due to the rapid advances in hardware technology which allows sophisticated signal processing algorithms to be implemented in real time. In order to improve the quality of communicated speech signals, speech enhancement techniques have to be used. Ultimately, the main objective of speech enhancement is to improve one or more perceptual aspects of the speech signal, such as overall quality, intelligibility, or degree of listener fatigue. Speech enhancement systems have primarily been developed for human-hearing in noisy environments. A less well developed application area involves systems to improve speech comprehension for the hearing impaired in both noisy and quiet environments. Speech enhancement techniques can also be used for improving the quality of coded speech signals by using pre or post enhancement techniques.

2. KALMAN FILTER

The Kalman filter is a mathematical procedure which operates through a prediction and correction mechanism. Kalman filter combines all the available data measured, plus the knowledge of the system and the measurement devices, to produce an estimation of the desired variables in such a manner that the error is statistically minimized. The Kalman filter uses a system's dynamics model (i.e., physical laws of motion), known control inputs to that system, and measurements (such as from sensors) to form an estimate of the system's varying quantities (its state) that is better than the estimate obtained by using any one measurement alone. As such, it is a common sensor fusion algorithm. The use of Kalman Filter for speech enhancement in the form that is presented here was first introduced by Paliwal (1987). This method however is best suitable for reduction of white noise to comply with Kalman assumption. In deriving Kalman equations it normally assumed that the process noise (the additive noise that is observed in the observation vector) is uncorrelated and has a normal distribution. This assumption leads to whiteness character of this noise. There are, however, different methods developed to fit the Kalman approach to coloured noises. It is assumed that speech signal is stationary during each frame, that is, the AR model of speech remains the same across the segment. To fit the one-dimensional speech signal tothe state space model of Kalman filter we introduce the statevector as:

$$x(k) = [x(k-pt1) * (k-pt2) * (k-pt3) \ldots x(k)J] \quad \text{- (1)}$$

where, x(k) is the speech signal at time k. Speech signal is contaminated by additive white noise n(k)

$$Y(k) = x(k) + n(k) \quad \text{- (2)}$$

The speech signal could be modelled with an AR process oforder p

$$x(k) = \sum aix(k-i) + u(k) \quad i = 1 \ldots p \text{- (3)}$$

$$n(k) = \sum bin(k-i) + v(k) \quad i = 1 \ldots q \quad \text{- (4)}$$

where, a_i's are AR (LP) coefficients and u(k) is the predictionerror which is assumed to have a normal distribution ~ N(0,Q). Substituting equation (1) into equation (3) we get,

$$x(k) = A_x x(k-1) + G_x u(k) \text{- (5)}$$

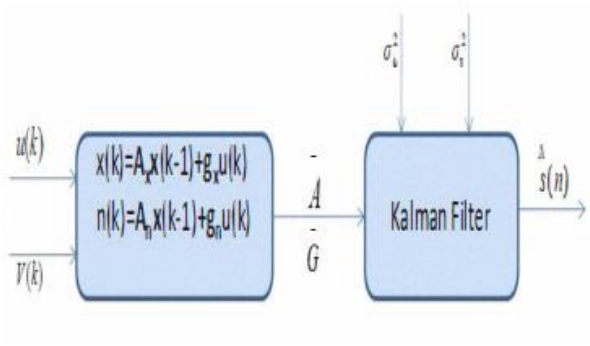
$$n(k) = A_n x(k-1) + G_n u(k) \text{- (6)}$$

where, $G = [0\ 0\ \dots\ 0\ 1]^T$

G has a length of p (LP order) and the observation equation would be

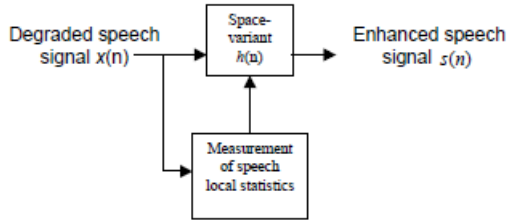
$$Y(k) = H_x(k) + n(k) \qquad \text{- (7)}$$

$$H = GT$$



3. WEINER FILTER

The Wiener filter is a popular technique that has been used in many signal enhancement methods. The basic principle of the Wiener filter is to obtain an estimate of the clean signal from that corrupted by additive noise. This estimate is obtained by minimizing the Mean Square Error (MSE) between the desired signal $s(n)$ and the estimated signal $\hat{s}(n)$.



The frequency domain solution to this optimization problem gives the following filter transfer function:

$$H(w) = \frac{P_s(w)}{P_s(w) + P_v(w)}$$

Where,$P_s(\omega)$ and $P_v(\omega)$ are the power spectral densities of the clean and the noise signals, respectively. This formula can be derived considering the signal s and the noise v as uncorrelated and stationary signals. The

SNR is defined by:

$$SNR = \frac{P_s(w)}{P_v(w)}$$

This definition can be incorporated to the Wiener filter equation asfollows:

$$H(w) = \left[1 + \frac{1}{SNR}\right]^{-1}$$

The drawback of the Wiener filter is the fixed frequency responseat all frequencies and the requirement to estimate the power spectraldensity of the clean signal and noise prior to filtering. In this approach, the estimated speech signal mean$m_x$and variance σ_v^2

are exploited. It is assumed that the additive noise $v(n)$ is of zero mean and has a white nature with variance of σ_v^2. Thus, the power spectrum $Pv(\omega)$ can be approximated by:

$$P_v(w) = \sigma_v^2$$

Consider a small segment of the speech signal, in which the signal $x(n)$ is assumed to be stationary, The signal $x(n)$ can be modeled by:

$$x(n) = m_x + \sigma_x w(n)$$

Where,$m_x$and $\sigma_x$are the local mean and standard deviation of $x(n)$, $w(n)$ is a unit variance noise.

Within this small segment of speech, the Wiener filter transferfunction can be approximated by:

$$H(w) = \frac{P_s(w)}{P_s(w) + P_v(w)} = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_v^2}$$

Because, $H(\omega)$ is constant over this small segment ofspeech, the impulse response of the Wiener filter can be obtained by:

$$h(n) = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_v^2}\delta(n)$$

The enhanced speech signal $s(n)$ in this local segmentcan be expressed as:

$$s(n) = m_x + (x(n) - m_x) * \frac{\sigma_s^2}{\sigma_s^2 + \sigma_v^2}\delta(n)$$
$$= m_x + \frac{\sigma_s^2}{\sigma_s^2 + \sigma_v^2}(x(n) - m_x)$$

If $m_x$and $\sigma_s$are updated at each sample, we can say:

$$s(n) = m_x(n) + \frac{\sigma_s^2}{\sigma_s^2 + \sigma_v^2}(x(n) - m_x(n))$$

the local mean $m_x(n)$ and $(x(n) - m_x(n))$ are modifiedseparately from segment to segment and then the results are combined.

If σ_s^2 is much larger than $\sigma_v^2$the output signal $\hat{s}(n)$ will be primarilydue to $x(n)$ and the input signal $x(n)$ is not attenuated. If $\sigma_s^2$is smallerthan σ_v^2, the filtering effect is performed. Notice that $m_x$is identical to $m_s$when mv is zero. So, we can

estimate$m_x(n)$ in Eq. from $x(n)$ by:

$$m_s(n) = m_x(n) = \frac{1}{(2M+1)}\sum x(k)$$

Where, $(2M+1)$ is the number of samples in the short segment used in the estimation. To measure the local statistics of the speech signal, we need to estimate the signal variance σ_s^2. Since $\sigma_x^2 = \sigma_s^2 + \sigma_v^2$, then $\sigma_s^2(n)$ may be estimated from $x(n)$ as follows:

$$\sigma_s^2(n) = \sigma_x^2(n) - \sigma_v^2$$

Where,

$$\sigma_x^2(n) = \frac{1}{(2M+1)}\sum(x(k) - m_x(n))^2$$

By this method, we guarantee the adaptaion of the filter transfer function from sample to sample based on the local statistics of the speech signal.

4. LINEAR PREDICTIVE CODING FILTER

Linear predictive coding (LPC) is defined as a digital method for encoding an analog signal

in which a particular value is predicted by a linear function of the past values of the signal. It was first proposed as a method for encoding human speech by the United States Department of Defence in federal standard 1015, published in 1984. Human speech is produced in the vocal tract which can be approximated as a variable diameter tube.

The linear predictive coding (LPC) model is based on a mathematical approximation of the vocal tract represented by this tube of a varying diameter. At a particular time, *t*, the speech sample *s(t)* is represented as a linear sum of the *p* previous samples. The most important aspect of LPC is the linear predictive filter which allows the value of the next sample to be determined by a linear combination of previous samples. LPC is another method of separating out the effects of source and filter from a speech signal; similar in intention to cepstral analysis but using quite different methods.

One way of thinking about LPC is as a coding method -- a way of encoding the information in a speech signal into a smaller space for transmission over a restricted channell. LPC encodes a signal by finding a set of weights on earlier signal values that can predict the next signal value:

y[n] = a[1]y[n-1] + a[2]y[n-1] + a[3]y[n-3] + e[n]

If values for a[1..3] can be found such that e[n] is very small for a stretch of speech (say one analysis window), then we can transmit only a[1..3] instead of the signal values in the window. The speech frame can be reconstructed at the other end by using a default e[n] signal and predicting subsequent values from earlier ones. Clearly this relies on being able to find these values of a[1..k] but there are a couple of algorithms which can do this (one is covered in the book). The result of LPC analysis then is a set of coefficients a[1..k] and an error signal e[n], the error signal will be as small as possible and represents the difference between the predicted signal and the original.

There is an obvious parallel between the LPC equation and that of a recursive filter (y*a = x):

y[n] = -a[1]y[n-1] - a[2]y[n-1] - a[3]y[n-3] + ... + x[n]

where we have rearranged the terms as in the text. The LPC coefficients correspond to those of a recursive filter and the error signal corresonds to a source signal. Moreover, the conditions under which the error signal is minimised in LPC analysis mean that the error signal will have a flat spectrum and hence that the error signal will approximate either an impulse train or a white noise signal. This is a very close match to our source filter model of speech production where we excite a vocal tract filter with either a voiced signal (which looks like a series of impulses) or a noise source. So, LPC analysis has the wonderful property of finding the coefficients of a filter which will convert either noise or an impulse train into the original frame of speech.

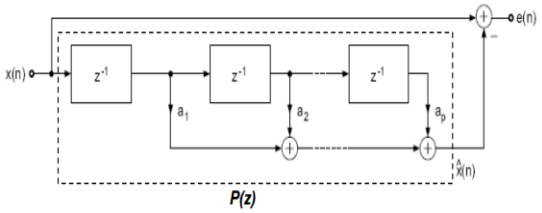
The input sample x(n) is extrapolated, i.e. approximated by a linear combination of past samples of the input signal:

$$x(n) = \sum_{k=1}^{p} a_k x(n-k)$$

Because this is a prediction filter, we will always have an error. This residual error is given by:

$$e(n) = x(n) - x = x(n) - \sum_{k=1}^{p} a_k x(n-k)$$

The prediction error calculation can be implemented by means of a FIR filter:



The Z-transform of the prediction filter is:

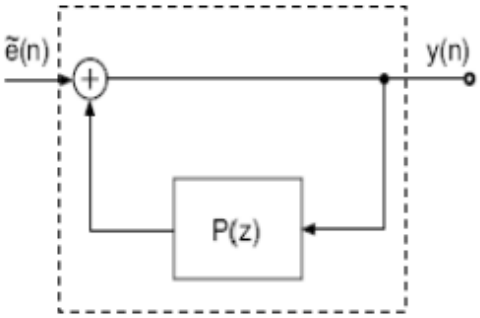
$$P(z) = \sum_{k=1}^{p} a_k z^{-k}$$

Such that, *E(z) = Z(z)[1 – P(z)]*

The Inverse filter can be defined as:

$$A(z) = 1 - P(z) = 1 - \sum_{k=1}^{p} a_k z^{-k} \quad s.t \; E(z) = X(z)A(z)$$

For synthesis we use an approximation of the residual as the excitation used as input to the all-pole (LPC) filter, resulting on the model:



$$Y(z) = E(z)H(z)$$

$$H(z) = \frac{1}{A(z)} = \frac{1}{1 - P(z)}$$

With optimal coefficients, residual energy is minimized. The higher the coefficient order p, the closer the approximation is to X(k).

5. SIMULATION AND RESULTS

We apply the noise contaminated speech signal to each of the three filter models of Kalman, Weiner and Linear Predictive Coding which were constructed using MATLAB™ software. The results will be based upon the Signal to Noise Ratio of the output speech signal at each of the three filters.

Below is the comparison of the Signal to Noise Ratio for the three filters:

Sr. No	Signal to Noise Ratio		
	White Noise (dB)	Random Noise (dB)	Color Noise (dB)
Kalman Filter	0.03225	-4.0962	-16.3584
	17.8923	5.6542	12.7895
Weiner Filter	0.03225	-4.0962	-16.3584
	10.3654	3.1785	9.0123
Linear Predictive Coding	.03225	-4.0962	-16.3584
	18.3169	6.2391	14.6842

As we can see, Linear Predictive Coding has the best Signal to Noise Ratio than both Kalman and Weiner Filters.

6. CONCLUSION

peech enhancement is implemented using three methods namely Linear Predictive Coding, Wiener filter and Kalmanfilter. Weiner Filter has some disadvantages like, it is suitable for stationary signals only but it has problem like musical noise. But in our daily life, the signals are not stationary and will vary randomly. To overcome these problems, Kalman filter and Linear Predictive Coding are used, which are time domain in nature. It has overcome the disadvantages mentioned in earlier method. Each method is implemented in MATLAB™ and Signal to Noise Ratio values of respective methods are compared. It is observed that, among three methods,performance of the Kalman filter and Linear Predictive Coding is comparatively good forboth stationary and non stationary signals.

7. FUTURE SCOPES

Future research directions are given in this section. One extension to the proposed enhancement algorithm would be to investigate sequential techniques of noise parameter estimation. These approaches estimate the unknown parameters to maximise the likelihood at each time slice and thus

allow adaptation to slowly varying noise. Another path for investigation is developing the speech enhancement system based on classification of voiced and unvoiced speech. Performing the speech enhancement by designing the different speech signal estimators based on voiced and unvoiced classification.

8. ACKNOWLEDGEMENT

We acknowledge the efforts and the hardwork by the experts who have contributed greatly towards the development of the different speech enhancement techniques. We take this opportunity to express our profound gratitudeand deep regards to our guide, Prof HarshawardhanAhire for his exemplary guidance, monitoring and constant encouragement throughout the course of this project.

9. REFERENCES

1. S. Kay and S. Marple "Spectrum analysis modern perspective", IEEE Proc., vol. 69, pp.1380 -1418 1981
2. D. O'Shaughnessy Speech Communication, 1987 :Addison-Wesley
3. R. E. Kalman, "A new approach to linear filtering and prediction problems", Trans. ASME&mdash,J. Basic Eng. Automat. Control, vol. 82, no. D, pp.35 -45 1960
4. M. Micheli and M. I. Jordan, "Random sampling of a continiuous-time stochastic dynamical system", 15th Int. Symp. Mathematical Theory of Networks and Systems (MTNS), 200
5. P. S. Maybeck, Stochastic Models, Estimation, and Control, vol. 141, 1979 :Academic
6. J. C. Brailean, R. P. Kleihorst, S. Efstratiadis, A. K. Katsaggelos and R. L. Lagendijk "Noise reduction filters for dynamic image sequences, a review", Proc. IEEE, vol. 83, pp.1272 -92
7. D. T. Kuan, A. A. Sawchuk, T. C. Strand and P. Chavel "Adaptive noise smoothing filter for images with signal-dependent noise", IEEE Trans. PAMI, vol. 7, pp.165 -177 1985