Speech Enhancement using Kalman and Wiener Filtering

P.Taraka Raghavendra Sai, P.Manisha Reddy, R.Sindhuja

Department of Electronics and Communication Engineering, Geethanjali College of Engineering and Technology, Hyderabad, India

Abstract: The term "Quality of Speech" in Speech Denoising techniques is associated with Clarity and Intelligibility. Till now due to the variable nature and characteristics of noise with time and process to process, Speech Denoising is a difficult problem in Noisy environment. In this paper, we proposed a method to improve the quality of speech based on the combination of Digital Audio Effects with Improved Adaptive Kalman Filter when only corrupted speech is available. In this approach to enhance the Speech content in the Noisy speech signal, Digital audio effects are used. A Digital Expander generates an audio effect which operates on a low signal level and creates more likely sound characteristics. And further, noise is removed by Auto Regressive modelled improved adaptive Kalman filter. The performance of the proposed method with additive color noise is found to be better compared to other spectral subtraction, Wiener and Kalman filter methods in terms of Signal-to-Noise ratio and intelligibility.

Check for updates

DOI of the Article: https://doi.org/10.46501/IJMTST0707029

Available online at: http://www.ijmtst.com/vol7issue07.html

As per UGC guidelines an electronic bar code is provided to seure your paper

To Cite this Article: P.Taraka Raghavendra Sai, P.Manisha Reddy, R.Sindhuja. Speech Enhancement using Kalman and Wiener Filtering. *International Journal for Modern Trends in Sceicen and Technology* 2021, *7*, 0707060, pp. 179-184. https://doi.org/10.46501/IJMTST0707029

Article Info. Received: 24 June 2021; Accepted: 14 July 2021; Published: 19 July 2021

INTRODUCTION:

Speech plays a vital role in our daily communication and also for human machine interfacing. Therefore, production and perception of speech have become an interesting part of the research since decades. But the quality and intelligibility of the speech are significantly degraded by the presence of background noise, which affects the ability in understanding others speech, causes error in Human Machine Interfacing, etc. In this digital world, it's really hard for any signal in a real-time environment to escape from noise. This hits us really hard when it comes to delivering a message from one place to another and there is a need for cleaning up or enhancing the message signal but at the same time, not giving up any intelligibility of the message (content, not just clarity). Since speech messages have been the mode of communication everywhere, need for speech denoising is required whenever the signal comes in contact with the real-time environment. Modelling of the human speech production process helps in enhancing the speech. But, as speech is a highly non-stationary signal, it is difficult to model the human speech production process. Though speech is a highly non-stationary signal, it is stationary for a very short period of time. Based on this fact, Classical speech denoising techniques are considered for speech segment models for short time, but these short time models do not include the effects of the noise as noise has long term characteristics. On the other hand, such long-term characteristics are naturally taken care of in the autoregressive approach as speech signals are not modelled on a short-time basis but as a whole. The AR model is also known to be good for representing unvoiced speech. However, it is not quite appropriate for voiced speech , since voiced speech is often quite periodic in nature. This has motivated us to look into speech models which can satisfactorily describe both voiced and unvoiced speech, and allow for exploitation of the long-term characteristics of noise.

TECHNICAL APPROACH

Speech is the most primary human communication. For that reason, there exists a big trend to increase and improve telecommunications. Nowadays, all the people use the communication devices almost as a primary good: telephones, mobiles, internet...and the customers demand a high coverage and quality. However, the background noise is an important handicap. If it is joined with other distortions, it can seriously damage the service quality. Added to this human-human interaction, there also exists a human- machine interaction based on a graphical user interface. However, still today the computers have a lack of human abilities like speaking, listening, understanding and learning. We live in a noisy world. In all applications (telecommunications, hands-free communications, recording, human-machine interfaces, etc) that require at least one microphone, the signal of interest is usually contaminated by background noise and reverberation. As a result, the microphone signal has to be “cleaned” with digital signal processing tools before it is played out, transmitted, or stored. Speech processing is the study of speech signals and the processing methods of these signals. The signals are usually processed in a digital representation whereby speech processing can be seen as the intersection of digital signal processing and natural language processing. Speech processing can be divided in the following categories: Speech recognition, which deals with analysis of the linguistic content of a speech signal. Speaker recognition, where the aim is to recognize the identity of the speaker. Denoising of speech signals (this is the area of this paper) Speech coding, a specialized form of data compression, which is important in the telecommunication area. Voice analysis for medical purposes, such as analysis of vocal loading and dysfunction of the vocal cords. Speech synthesis: the artificial synthesis of speech, which usually means computer generated speech. Speech denoising: enhancing the perceptual quality of speech signal by removing the destructive effects of noise, limited capacity recording equipment, impairments, etc. Speech processing has a lot of applications; one of them could be a tickets sales system by phone, where, without the necessity of an operator, a customer can buy tickets with different characteristics and options thanks to the word recognition systems.

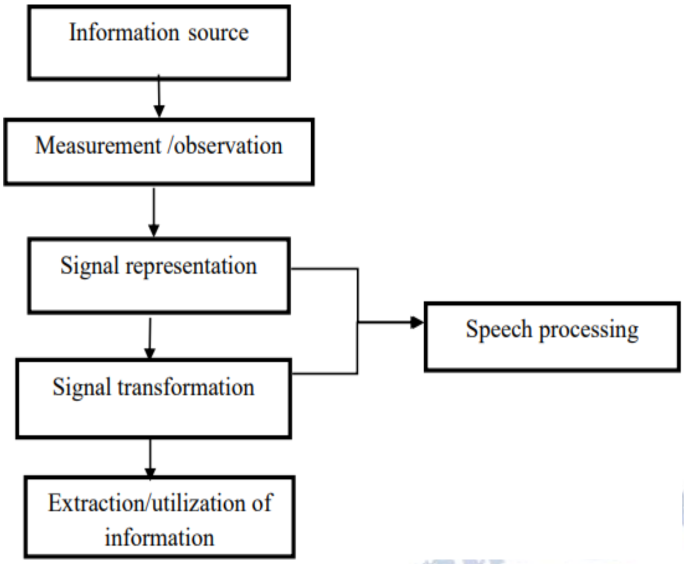


Fig 1 : Speech processing

SPEECH DENOISING

The modelling of speech studies how humans produce the voice. Nowadays we have a lot of devices which "speak" to us and this voice should be as similar as possible to a real human voice. For that reason, a lot of researchis aimed to find a good model of speech production.

All this activity starts in the speaker's mind with a message to be transmitted to the listener via speech. This is the linguistic stage. After a message is created, the next step is to convert the message into a sequence of words. Each word consists of a sequence of phonemes that correspond to the pronunciation of the words. The spoken signal appears when the air crosses the trachea from the lungs. This air crosses the vocal cords, situated in the larynx, which have two functions:

- With the voiced sounds the vocal cords are in tension and they vibrate when the air goes across them.
- With the unvoiced sounds the vocal cords are relaxed and the air can cross them freely. In the next step, the brain sends the information to the vocal tract, where the air takes the characteristics of each formant. This is the physiological stage. After that, when the speaker starts to speak we are in the physic-acoustic stage. The sounds are materialized but, at the same time, there is feedback because the ear of the speaker can hear what is he saying and the brain analyses the meaning. And the process in the speaker's brain that starts is the same as when it is the listener who speaks and the speaker who listens. The sounds travel by the air in the transmission stage from the mouth of the speaker to the ear of the listener. In this stage, when the sounds cross the channel, the noise and other distortions are added. Finally, we have the same stages but on the listener side in reverse order. First the message is passed to the cochlea in the inner ear, which performs frequency analysis as a filter bank. A neural transduction converts the spectral signal into activity signal on the auditory nerve. Currently, it is unclear how the neural activity is converted into the language system and how the brain can achieve the comprehension of the message. Speech signals are composed of analog units, which are the symbolic representation of the spoken language: phonemes, syllables and words.

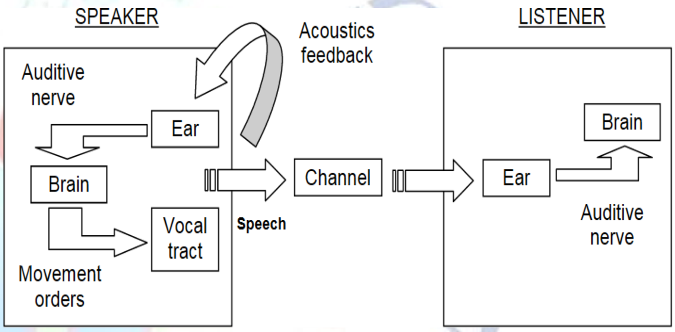


Fig 2 : Speech generation and speech understanding

KALMAN AND WIENER FILTERING

The filter has its origin where it is described as a recursive solution for the linear filtering problem for discrete data. The research was in a wide context of state – space models, where the point is the estimation through the recursive least squares. Since that moment, due to the development of digital calculation, Kalman filter has been researched and applied, particularly in self and assisted navigation, missiles search and economy. The study of the Kalman filter is based on the Wiener filter. The Wiener filter is the precursor of the Kalman filter. The goal of Wiener filter is to remove the noise from a corrupted signal.

. In general there are two processes which affect the signal that we want to measure:

First of all, it is a fact that every device introduces an error in the output when a signal is measured. If our original signal is xk and the response of the device is hk our signal in the output is:

$$y_k = x_k * h_k \leftrightarrow Y_j = X_j \cdot H_j$$

Secondly, the signal outside has noise added due to the process.

$$\hat{y}_k = y_k + n_k$$

To solve this equation, if we don't have noise and we know the response, then the solution is easy to find:

$$X_j = \frac{Y_j \cdot W_j}{H_j}$$

For that, we should find the optimal Wiener filter. This kind of filter was proposed. To reduce the amount of noise in the corrupted signal this filter is based on a statistical approach. Normally, the filters are designed for a specific frequency, but in Wiener filters, first of all, we have to have knowledge about the spectral properties of the original signal and noise, and after that, we have to find a LTI filter whose output would be as close as possible to the original signal. The Wiener filters are characterized by the following concepts:

- Assumption: signal and (additive) noise are stationary linear stochastic processes with known spectral characteristics or known autocorrelation and cross-correlation.
- Requirement: the filter must be physically realizable, i.e. causal (this requirement can be dropped, resulting in a non-causal solution).
- Performance criteria: minimum mean-square error. 3.2 Kalman filter The filter is a mathematical procedure which operates through a prediction and correction mechanism. In essence, this algorithm predicts a new state from its previous estimation by adding a correction term proportional to the predicted error. In this way, this error is statistically minimized. This filter is the main algorithm to estimate dynamic systems specified in state-space form. A Kalman filter is simply an optimal recursive data processing algorithm. If we focus on the word optimal, its definition depends on the criteria chosen to evaluate. A feature is called optimum if the Kalman filter incorporates all the information provided. It processes all the measurements available, regardless the precision, to estimate the current value of the interest variables, using:
- Knowledge of the system and the measurement devices.
- Statistic description of the system noises, measurements of errors and the uncertainty of the dynamics models.

  Any information available about the initial conditions of the variables under study. A Kalman filter would be built to combine all these data and with the knowledge of some dynamic systems to generate the best estimation of the interest variable. If, on the contrary, we focus on the word recursive, this means that the Kalman filter doesn't require storing all the previous samples and it neither needs to reprocess them on each new measurement taken. This feature is very important to the filter practicality. We say that this is a data processing algorithm because it is just a computer program in a processing center. The complete estimation procedure is as follows: The model is formulated on state-space and for an initial set of parameters given, the model prediction errors are generated by the filter. These are used recursively to evaluate the probability function until its maximization. As a summary, we can say that the 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.

SPEECH DENOISING USING MATLAB

5.1 Developed MATLAB functions for speech denoising

5.1.1 FAST ICA

Fast ICA is an efficient and popular algorithm for independent component analysis invented by AapoHyvärinen at Helsinki University of Technology. Like most ICA algorithms, Fast ICA seeks an orthogonal rotation of prewhitened data, through a fixed-point iteration scheme, that maximizes a measure of non-gaussianity of the rotated components. Non-gaussianity serves as a proxy for statistical independence which is a very strong condition and requires infinite data to verify. Fast ICA can also be alternatively derived as an approximative Newton iteration.

5.1.2 *Independent Component Analysis (ICA)*

Independent component analysis is a computational method for separating a multivariate signal into additive subcomponents. This is done by assuming that the subcomponents are non-Gaussian signals and that they are statistically independent from each other.

The ICA separation of mixed signals gives very good results based on two assumptions and three effects of mixing source signals. Two assumptions:

1. The source signals are independent of each other.
2. The values in each source signal have non-Gaussian distributions.

ICA finds the independent components (also called factors, latent variables or sources) by maximizing the statistical independence of the estimated components. We may choose one of many ways to define a proxy for independence, and this choice governs the form of the ICA algorithm.

5.1.3 *Kernel-independent component analysis (KICA)*

Kernel-independent component analysis (Kernel ICA) is an efficient algorithm for independent component analysis which estimates source components by optimizing a generalized variance contrast function, which is based on representations in reproducing Kernel Hilbert space.

Kernel ICA is based on the idea that correlations between two random variables can be represented in a reproducing kernel Hilbert space.

5.1.4 *Principal Component Analysis (PCA)*

Principal component analysis is the process of computing the principal components and using them to perform a change of basis on the data, sometimes using only the first few principal components and ignoring the rest.

PCA is used in exploratory data analysis and for making predictive models. It is commonly used for dimensionality reduction by projecting each data point onto only the first few principal components to obtain lower-dimensional data while preserving as much of the data's variation as possible. The first principal component can equivalently be defined as a direction that maximizes the variance of the projected data.

RESULTS AND DISCUSSION

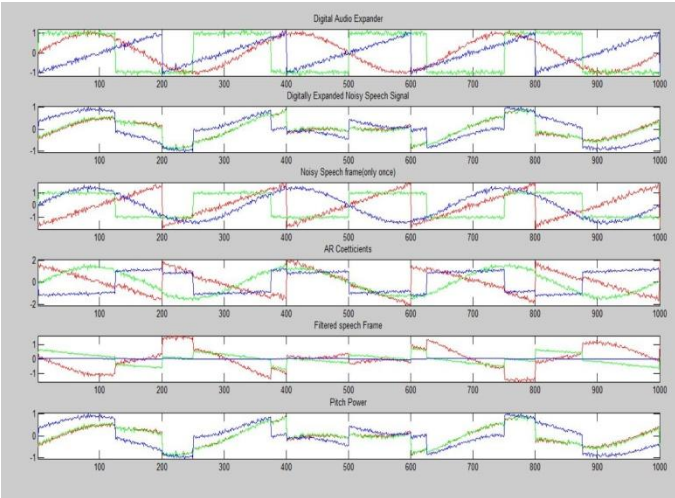
Figure shows the block diagram for combination of Digital Audio Effect with modified adaptive Kalman filter based speech denoising method. Based on it Matlab code is developed. A noisy speech is generated using a Clean Speech, which is taken from the Noizeus Database and random values (random noise or colour noise) are added to the clean speech. Later it passed through a Digital Expander. And the output of Digital Expander is shown in Figure. Digital Expander expanding factor value is set to 0.5. It is further applied to the Iterative modified Kalman filter to suppress the noise. We set the Kalman filter AR model order to P=20. These 20 AR coefficients are updated for every time frame of 25ms duration

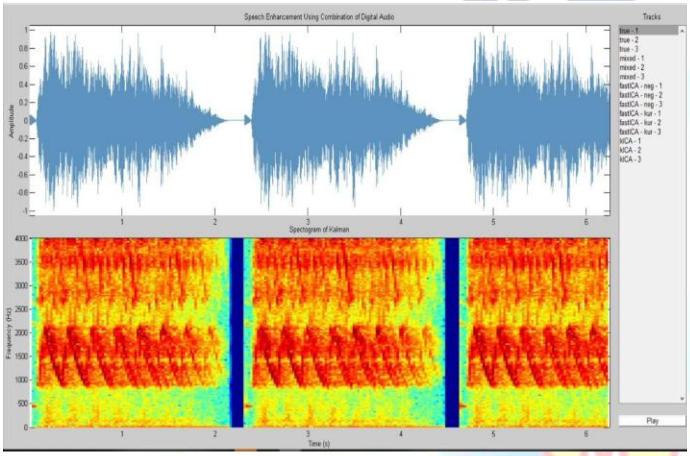
which is chopped by the Hanning window and analysed using the linear prediction analysis method (LPC). The additive measurement noise is assumed to be stationary during each small frame. LPC coefficient estimation order is taken as 13 for both noisy speech and noise

signals. Number of iterations is set to be 7. Real time noisy signals (NOIZEUS database) of 0dB, 5dB, 10dB and 15dB are considered for performance analysis, with the Hanning window.

We have observed and tabulated the results of basic Spectral Subtraction, Wiener Filter, Kalman filter methods and compared with Digital Audio Effect based Kalman filtering method. Compared to all these methods, the proposed algorithm gives better improvement in terms of SNR as well as intelligibility. The corresponding waveforms are shown below.

Experimental results show that the proposed technique is effective for speech denoising compared to conventional Kalman filters. Iterative Kalman filter and proposed method results and waveforms are placed below





CONCLUSION

A speech enhancing method based on an improved SS algorithm was introduced. For effective noise reduction with minimal distortion the proposed algorithm takes in account perceptual aspects of the human ear. It can be seen from the experimental results that the proposed method effectively reduces background noise in comparison with commonly used SS algorithms. Proposed method results in greater improvement of SNR and considerably improvement of perceptual speech quality in comparison to conventional spectral subtraction Method.

In the present study, an improved method for speech denoising by combining Digital Audio Effecting techniques with improved Adaptive Kalman filter technique is proposed. In this paper, we discussed the drawbacks of basic methods such as speech denoising with spectral subtraction and wiener filter methods. Even though other Kalman filter approach-based speech denoising methods are giving better results than a conventional Kalman filter, more complexity is involved. It leadsto a moretime-consuming process. In this paper, we proposed a method to overcome the disadvantages of earlier methods in terms of performance and Speed. All these methods are simulated using MATLAB and input output SNR values of respective methods are compared. Performance of the Proposed method is analyzed with different Input SNR noise levels. It is observed that the proposed method gives better output SNR values and its performance is comparatively superior for both stationary and non-stationary signals

ACKNOWLEDGEMENT

We would like to express gratitude to Dr.V. S. Srinivas, ECE Dept., GCET for his valuable and constructive suggestions during the planning and development of this work

REFERENCES

1. M. Berouti, R. Schwartz, and J. Makhoul, "Denoising of speech corrupted by acoustic noise" in Processing of international Conference of Acoustic, Speech and Signal Processing,1979, pp.208-211.
2. T. Esch, P. Vary, "Efficient musical noise suppression for speech denoising system" IEEE International Conference on Acoustics, Speech and Signal Processing, Taipei, Taiwan, 2009,pp. 4409 - 4412.
3. E. Dong, X. Pu, "Speech denoising based on perceptual weighting filter" 9th International Conference on Signal Processing ,Leipzig, Germany, 2008, pp 705-708.
4. V. Prasad, R. Sangwan et al., "Comparison of voice activity detection algorithms for VoIP", proc. of the Seventh International Symposium on Computers and Communications, Taormina, Italy, 2002, pp. 530-532.
5. K.Sakhnov, E.Verteletskaya, B. Šimák, "Dynamical Energy-Based Speech/Silence Detector for Speech Denoising Applications". In World Congress of Engeneering 2009 Proceedings. Hong Kong:, 2009, pp.801-806.
6. S. Ogata, T.Shimamura, "Reinforced spectral subtraction method to enhance speech signal", Proceedings of IEEE International Conference on Electrical and Electronic Technology, 2001, vol 1, pp 242 – 245.
7. P. Pollák, "Speech signals database creation for speech recognition and speech enchancement applications" [associate professor innagural dissertation] CTU in Prague, FEE, Prague, 2002.
8. "Speech Denoising" by J. Benesty, J. Chen, Eds., and S. Makino,Springer, Berlin, 2005.