Effect of Modulation Domain Coupled Kalman-Spectral Filter on Speech Enhancement over Wireless Voiced Communication System

E. T. Olawole [1], D. O. Akande[2*], Z. K. Adeyemo[2], F. K. Ojo[2] and S. I. Ojo[2]

[1]Computer Engineering Department, University of Ilorin, Ilorin, Kwara State, Nigeria.
[2]Electronic and Electrical Engineering Department, Ladoke Akintola University of Technology, Ogbomoso, Oyo State, Nigeria.

ABSTRACT: Noise suppression in speech signal has received great attention in signal processing research community. However, existing methods of noise suppression such as the time-domain Kalman filter suffers from the psychoacoustic and physiological characteristic feature while the spectral subtraction technique performed in the modulation domain is associated with remnant noise in the enhanced speech signal. Therefore, in this paper, a Modulation-Domain coupled Kalman-Spectral Filtering (MD-KSF) in a single channel system is proposed to address the aforementioned shortcomings in the two existing techniques. The simulation of the proposed coupled MD-KSF was performed in MATLAB software environment. The evaluation of the proposed coupled MD-KSF technique in terms of Spectral waveform, Mean Square Error (MSE) and Log Spectral Distance (LSD) was performed. Furthermore, Perceptual Evaluation of Speech Quality (PESQ) and Short-Time Objective Intelligibility (STOI) tests were employed to validate the quality and intelligibility of the proposed coupled MD-KSF using the NOIZEUS corpus data set. The proposed technique shows significant noise suppression over the existing techniques.

KEYWORDS: Speech enhancement, Modulation-domain, coupled Kalman-Spectral Filter, and Log spectral distance.

[Received Jan. 22, 2024; Revised June 28, 2024; Accepted July 2, 2024] Print ISSN: 0189-9546 | Online ISSN: 2437-2110

I. INTRODUCTION

The primary role of cellular and mobile telecommunication devices is in their ability to enhance received information in which speech signal is of topmost importance. This is because speech information exchange via these devices is subjected to adverse environmental noise effects generated from automobiles, industries, market places just to mention a few (Foth *et al*., 2013; Adeyemo *et al*., 2012). The accumulation of uncontrolled and inevitable noises is mostly present everywhere and mixed with speech conversations. These destructive noises impair the audibility, quality and intelligibility of clean speech or audio information when engaging in conversation via telecommunication means (Santos *et al*., 2019; Olawole *et al*., 2022).

The management of these destructive noise effects is of paramount interest as the integrity and the sensitivity of the original speech or audio information become questionable and mostly results in misinformation to critical decision taking. Noise possesses different statistical properties exhibiting different characteristic effect on speech signals after subjection to enhancement algorithms (Pardede *et al*., 2019). Noise is classified into two namely: stationary and non-stationary noise (Vaseghi, 2000). Stationary noise also known as white noise is characterized with uncorrelated processes and possesses a constant power spectral density. Besides, non-stationary noise known as man-made noise possesses different power spectral density due to the different changes in the rhythm of the generated noise. This kind of noise are very difficult to remove due to the unavailability of the a priori characteristic information of such noise. On a daily basis, speech information exchangers undergo such experience.

Existing speech or audio enhancement algorithm used in combating this noisy speech effect can be classified into two and are discussed as follows: single channel system in which only one input medium or microphone is employed in the collection of both the clean and background noises simultaneously to give a single spectral information (So *et al*., 2010; Upadhyay and Jaiswal, 2016). On the other hand, multi-channel system involves the acquisition of the mixture of the clean as well as that of the noise via multiple input media or microphones to give a unified degraded noisy speech (Upadhyay and Karmakar, 2015). While multi-channel offers better performance, single-channel systems are preferred due to lower cost and easier implementation. This simplicity, however, creates challenges for mobile speech enhancement algorithms, especially regarding real-time processing demands (Loizou, 2013).

Based on existing literature, the primary algorithms used for speech enhancement include: spectral subtraction (Upadhyay and Karmakar, 2015), the Kalman filter (Paliwal and Basu, 1987; Goh *et al*., 1999; Esch and Vary, 2008), the Wiener filter (Upadhyay and Jaiswala, 2015; Bingyin and Bao, 2014), minimum mean square error (Ephraim and Malah,

*Corresponding author: doakande@lautech.edu.ng
doi: http://dx.doi.org/10.4314/njtd.v21i3.2317

1984; Paliwal *et al*., 2012), and model-based methods (Sohn *et al*., 1999; Wang and Brookes, 2018; Chen *et al*., 2012; Chowdhury *et al*., 2018), among others. One of the popular methods that has found usefulness in different area of applications in minimizing errors is the Kalman filter. The Kalman filter algorithm that has been utilized in the area of speech enhancement to solve the problem of degraded speech signal was performed by estimating the linear prediction coefficient of the clean speech before the mixture of corrupt noise (Roy and Paliwal, 2021).

Nonetheless, this process is impracticable in real time applications. This is because both the clean speech and the background noise from different sources are inseparable and at the same time cannot be avoided at the originator side (Sen *et al*., 2019). However, Kalman filter suffers significant drawback in that the algorithm results in instability and nonlinear characteristics due to error observed at the prediction update stage most especially in time domain (Banale *et al*., 2014). This observed error distorts the speech signal and degrades its psychoacoustic and physiological characteristic (So and Paliwal, 2011). In contrast to the Kalman filter, the Spectral subtraction technique, a frequency domain method, is easy to implement, making it attractive to many speech enhancement researchers. The noise spectral in the degraded noisy speech is initially estimated from the silence regions, which are then subtracted from the spectral of the degraded speech. The enhanced speech is then produced using this estimate with less noise characteristics. However, this method gives rise to musical noise remnant and makes the enhanced signals less effective due to noticeable speech distortion and inability to attenuate noise in the silent region (Rao and Kumar, 2016).

Wiener filter is another speech enhancement algorithm based on linear time invariant property and similar to spectral subtraction method in a conceptual manner. This is achieved by replacing the direct subtraction from an optimal statistical estimate of the clean speech spectral (Bingyia and Bao, 2014; Loizou, 2013). The Wiener filter algorithm reduces noise to the barest acceptable minimum level using the Mean Square Error (MSE) of the original clean speech and the estimate of the degraded speech (Venkateswarlu *et al*., 2011). However, this method is limited in performance because it requires a prior knowledge of the properties of both the clean speech and noise to achieve a perfect enhanced speech (Alam, 2010). This makes it to suffer from musical noise at the output signal.

Due to the inability of traditional algorithms to significantly minimize or eliminate noise, extensions of these algorithms have been developed over the years and are discussed therein. Yuhong *et al*. (2013) proposed an improved spectral subtraction speech enhancement algorithm to eliminate the distortions and remnant noise introduced by the traditional spectral subtraction method. This was achieved by using the weighing rule of the simultaneous masking effect of human auditory system. The enhanced speech signal demonstrated significant improvement in the output SNR over the traditional Spectral subtraction algorithm by showing effective reduction in speech distortion but still experienced a minimal low-level residual noise.

Rao and Kumar (2016) proposed an adaptive Kalman filter using time-frequency masking as a post filter to improve the speech signal in a sequence of noisy speech signal. The additive noise was assumed to be an autoregressive process and the time varying estimation parameters as linear prediction coefficient. The result obtained diminished the coloured noise. The method showed better improvement over existing methods but could not give the desired enhanced speech signal.

Pardede *et al*. (2019) proposed a coupled Spectral subtraction and Wiener filter to enhance speech signal in a secured voice communication system which is characterized by different distortion. The method was able to track the changes experienced in the speech signal due to the randomization of the encryption process without over-estimating the noise. The result showed that the method performed better than the existing speech enhancement algorithm on different communication channels.

Roy and Pawali, (2021) proposed a tuning based Kalman filter to tune the gain of the filter in order to improve the single channel speech signal corrupted with stationary noise in real life scenario. The estimate of the noise variance was first computed using speech presence probability method. Whitening filter was constructed to estimate the coefficients of the linear prediction parameter in order to improve the noise effect. The result outperformed traditional algorithms but at the expense of computational complexities.

Olawole *et al*. (2022) proposed a hybrid Spectral-Kalman filter in the time domain for speech enhancement over a wireless communication system. The technique first converted the captured time domain signal via Fast Fourier Transform (FFT) before the spectral subtraction and later performed Inverse Fast Fourier Transform (IFFT) for further processing by the Kalman filter. The result obtained shows that the proposed hybrid Spectral-Kalman filter was superior to the existing Kalman filter and Spectral subtraction technique. However, performing speech enhancement in the modulation domain shows better and significant evidence in psychoacoustic and physiological characteristic compared to speech analysis and processing in the time domain.

In this paper, focus is beamed on the improvement of speech enhancement techniques by coupling Kalman filter and Spectral subtraction in a cascade form. The reason for this is to alleviate the time frame discrepancy generated by prediction update process of the Kalman filter (Leonardro *et al*., 2018) and the subtractive error leading to musical notes which deteriorates the quality of the speech signal (Venkateswarlu *et al.,* 2011). Therefore, in lieu of the aforementioned, the following contributions were made;

A) The coupling of Kalman filter with the Spectral subtraction technique in the modulation domain is proposed for the enhancement of the quality and intelligibility of speech signal over a wireless communication via a single channel system.

B) The mathematical expression for the proposed coupled MD-KSF to minimize the speech compression error generated as a result of the instability experienced in the existing time-domain Kalman filtering and remnant noise present in the spectral subtraction techniques has been derived. This addresses the inadequate presence of psychoacoustic and physiological characteristics and musical note in the enhanced speech signal.

C) Comparison of the performance of the proposed coupled MD-KSF with the existing techniques using Spectral waveform, Mean Square Error, (MSE) and Log Spectral Distance (LSD) has been carried out. Furthermore, quality and intelligibility are carried out using Perceptual Evaluation of Speech Quality (PESQ) and Short-Time Objective Intelligibility (STOI), respectively were employed and validated using NOIZEUS corpus data set.

The remaining parts of the paper are organized as follows. Section II gives the details of the proposed method used. Section III presents the materials and methods. Section IV presents the simulation results. Conclusion is drawn in Section IV.

II. SYSTEM MODEL

Figure 1 depicts the diagram of the proposed coupled MD-KSF speech enhancement technique at the receiver of a wireless communication system. Mixture of the clean speech and noise transmitted over a single channel wireless link is received and then pre-processed. The pre-processing stage comprises of the framing, windowing and FFT. The pre-processed signal is passed through a noise estimation predictor to estimate the noise components. Similarly, the pre-processed signal passes through the MD-KF prediction and observation stage to model the modulated clean speech and additive noise using the state space equation. The output of this stage is further passed through the measurement and update stage to obtain the a priori state estimate covariance matrix of the noisy speech signal.

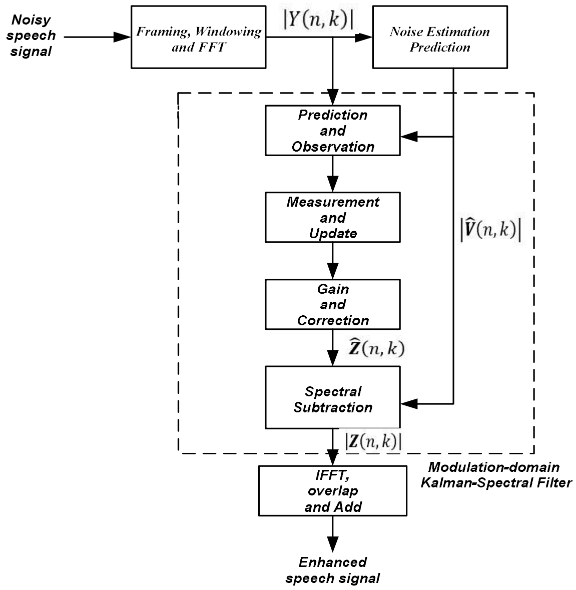


Figure 1: Proposed coupled MD-KSF at the receiver.

III. MATERIALS AND METHODS

The proposed coupled MD-KSF which comprises the Kalman filter and Spectral subtraction to improve the quality and intelligibility of the deteriorated noisy speech signal is presented.

A. Pre-processing Stage

Consider a noisy speech signal y(n) which is a mixture of clean speech s(n) and statistically independent additive noise v(n) which is given as:

$$y(n) = s(n) + v(n) \qquad (1)$$

With the background noise modeled as zero mean and uncorrelated to the clean speech. Due to the nature of a clean speech which is characterized as nonstationary and time varying, the degraded noisy speech signal is subjected to processing in a frame-by-frame manner. This is to evaluate the properties of the noisy speech signal. Windowing operation is then performed on the framed degraded noisy speech signal. Hamming window is utilized to ensure that the actual frequency spectrum is maintained. The Hamming window function $w(n)$ is expressed as (Hamid, 2018):

$$w(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N}\right),\ for\ 0 \le n \le N \qquad (2)$$

where N is the window length. The received noisy speech signal is multiplied with the Hamming window in equation (2) and transformed using the short-time Fourier transform (STFT) domain. This is expressed according to Oppenheim *et al.* (1999) in (3) as:

$$Y(n,k) = \sum_{n=0}^{N-1} y(n)\, w(n) e^{\left(\frac{-j2\pi n}{N}\right)} \qquad (3)$$

The acoustic magnitude of the modulating noisy speech signal $'|Y(n,k)|'$ after the pre-processed stage is given as:

$$|Y(n,k)| = |X(n,k)| + |V(n,k)| \qquad (4)$$

where $'|X(n,k)|'$ is the acoustic magnitude of the clean modulating speech signal and $'|V(n,k)|'$ is acoustic magnitude of the modulating white Gaussian noise signal for each frame. The noise estimate is then determined through the help of Voice Audacity Detector (VAD) and the average of the modulating noisy speech signal frames of the speech pauses over the silent region, and is given by Upadhyay and Karmakar (2015) as shown in (5):

$$\left|\hat{V}(n,k)\right| = \left(\frac{1}{N}\sum_{n=1}^{N-1}|Y(n,k)|^2\right)^{1/2} \qquad (5)$$

The output of the noise estimation predictor of the modulating noisy speech signal $\left|\hat{Y}(n,k)\right|$ is obtained by subtracting (5) from the (4) to give:

$$\left|\hat{Y}(n,k)\right| = |Y(n,k)| - \left|\hat{V}(n,k)\right| \qquad (6)$$

Since the exact spectral noise estimate of the noisy speech signal is difficult to achieve, remnant noise and speech distortion which reduce the quality and intelligibility of the output of the modulating signal makes the signal inaccurate and therefore, requires to minimize the noise effects.

B. Modulation-Domain Kalman Filtering Stage

In this stage, the Modulation-Domain Kalman Filtering (MD-KF) technique is employed to enhance the pre-processed spectral of the noisy speech signal. The p^{th} order linear predictor is used to model both the modulating clean speech signal and the additive noise. The state space equation is then expressed by Roy and Paliwal (2021) and indicated in (7) and (8) as:

$$\widehat{\boldsymbol{X}}(n,k) = \boldsymbol{A}(k)\widehat{\boldsymbol{X}}(n-1,k) + \boldsymbol{d}\widehat{\boldsymbol{W}}(n,k) \quad (7)$$

$$\left|\widehat{\boldsymbol{Y}}(n,k)\right| = \boldsymbol{c}^T\widehat{\boldsymbol{X}}(n,k) + \left|\widehat{\boldsymbol{V}}(n,k)\right| \quad (8)$$

where $\boldsymbol{A}(k)$ is a $p \times p$ dimensional state transition matrix expressed as:

$$\boldsymbol{A}(k) = \begin{bmatrix} -a_{1,k} & -a_{2,k} & -a_{3,k} & \cdots & -a_{p-1,k} & -a_{p,k} \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \quad (9)$$

$\widehat{\boldsymbol{X}}(n-1,k) = \left[\left|\widehat{\boldsymbol{X}}(n,k)\right|, \left|\widehat{\boldsymbol{X}}(n-1,k)\right|, \ldots, \left|\widehat{\boldsymbol{X}}(n-p+1,k)\right|\right]^{\mathrm{T}}$ is a clean speech modulation vector, $n-1$ denotes the prior state of the noisy speech signal frame, $\boldsymbol{d} = [1,0,\ldots,0]^{\mathrm{T}}$is the measurement vector of the excitation noise $\widehat{\boldsymbol{W}}(n,k)$, $\boldsymbol{c} = [1,0,\ldots,0]^{\mathrm{T}}$ is the measurement observation. Then the estimation of an unbiased and the linear Minimum Mean Square Error (MMSE) is recursively performed by Kalman filter according to (So *et al*., 2010). The a priori state estimate covariance matrix of the noisy speech signal is given as:

$$\boldsymbol{\Lambda} = [\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T] \quad 10)$$

The equation in (10) can be further expressed as:

$$\Lambda(n|n-1,k) = A(k)\Lambda(n-1|n-1,k)A(k)^{\mathrm{T}} + d\sigma_W^2 d^{\mathrm{T}} \quad (11)$$

where $\boldsymbol{\sigma}_{\mathrm{W}}^2$ is the zero-mean variance of the additive White noise $\widehat{\mathbf{W}}(\mathrm{n},\mathrm{k})$. By taking the differential of (11), the optimal Kalman filter gain $'\boldsymbol{\tau}(n,k)'$ in the modulation-domain is obtained as:

$$\boldsymbol{\tau}(n,k) = \boldsymbol{\Lambda}(n|n-1,k)\boldsymbol{c}\left[\boldsymbol{\sigma}_{V(n,k)}^2 + \boldsymbol{c}^T\boldsymbol{c}\boldsymbol{\Lambda}(n|n-1,k)\right]^{-1} \quad (12)$$

where $\boldsymbol{\sigma}_V^2$ is the zero-mean variance of the additive White noise $V(n,k)$. Then a priori estimate of the state vector $'\widehat{\boldsymbol{X}}(n|n-1,k)'$ which is dependent on the a posteriori estimate $'\widehat{\boldsymbol{X}}(n-1|n-1,k)'$ is obtained as:

$$\widehat{\boldsymbol{X}}(n|n-1,k) = \boldsymbol{A}(k)\widehat{\boldsymbol{X}}(n-1|n-1,k) \quad (13)$$

and the update of the a posteriori estimate covariance matrix of the noisy speech signal is given as:

$$\boldsymbol{\Lambda}(n|n,k) = [\boldsymbol{I} - \boldsymbol{\tau}(n,k)\boldsymbol{c}^{\mathrm{T}}]\boldsymbol{\Lambda}(n|n-1,k) \quad (14)$$

where I is an identity matrix of $p \times p$ dimensional matrix, while the update of the estimated speech signal is obtained as:

$$\widehat{\boldsymbol{X}}(n|n,k) = \widehat{\boldsymbol{X}}(n|n-1,k) + \boldsymbol{\tau}(n,k)\left[\widehat{\boldsymbol{Y}}(n,k) - \boldsymbol{c}^T\widehat{\boldsymbol{X}}(n|n-1,k)\right] \quad (15)$$

The estimated speech signal after the modulation-domain Kalman filter is obtained as:

$$\widehat{\boldsymbol{Z}}(n,k) = \boldsymbol{c}^T\widehat{\boldsymbol{X}}(n|n,k) \quad (16)$$

C. Modulation-Domain Spectral Subtractive Stage

The output of the MD-KF is then passed into the modulation-domain Spectral subtraction stage for further enhancement of the speech signal. The estimated noise signal $'\widehat{V}(n,k)'$ is subtracted from the (16) to obtain (Upadhyay and Karmakar, 2015) as:

$$|\boldsymbol{Z}(n,k)| = \left|\widehat{\boldsymbol{Z}}(n,k)\right|^{\beta} - \alpha\left|\widehat{\boldsymbol{V}}(n,k)\right|^{\beta} \quad (17)$$

where $\alpha \geq 1$ is the controlling factor. In this paper, α is taken to be 1 which indicates full subtraction and $\beta \in \{1,2\}$ also, β is taken to be 1 for the magnitude spectral subtraction. Then performing the inverse FFT (IFFT) operation on the enhanced modulating speech signal in (17), by overlapping and adding back the phase component, the enhanced clean speech signal $'s(n)'$ is obtained as:

$$s(n) = \sum_{n=0}^{N-1}(|\boldsymbol{Z}(n,k)|)\, e^{\left(\frac{j2\pi n}{N}\right)} \quad (18)$$

D. Simulation

In this section, the subjective evaluation of the proposed coupled MD-KSF technique is performed through simulation in MATLAB R2021a using a personal computer with the following configurations: Inter (R) Core i3, HP620 processor, 2.13 GHz and 4 GB RAM. The metrics used in the evaluation of the proposed coupled MD-KSF technique are as defined. Spectral waveform is the plot of the magnitude of the enhanced speech signal against frequency. As defined by Amit and Vinay (2019), the MSE is the sum of the diagonal of the $p \times p$ dimensional state space transition matrix of the coupled MD-KSF is taken to be the trace of that matrix and LSD is the distance between the power spectral of the noisy speech signal and that of the clean speech signal (Enqvist and Karlsson, 2008). In addition, objective evaluation tests: STOI is used to measure the intelligibility of the enhanced speech signal obtained from the noisy speech signal in a short-time temporal envelope. PESQ is a measure of the quality of the enhanced speech signal. The closer the PESQ score to 4.5 the better the quality of the speech signal. These tests are done on selected noisy speech signal obtained from the NOIZEUS corpus data set (Loizou, 2013). Comparison of the proposed coupled MD-KSF technique with existing Kalman filter and Spectral subtraction techniques is made. The data set contains clean speech corrupted with mixture of different selected noises such as babble, car and street. The noisy speech is then transmitted over wireless channel while Figure 2 shows the flowchart of the proposed coupled MD-KSF speech enhancement technique. The simulation parameters are contained in Table I.

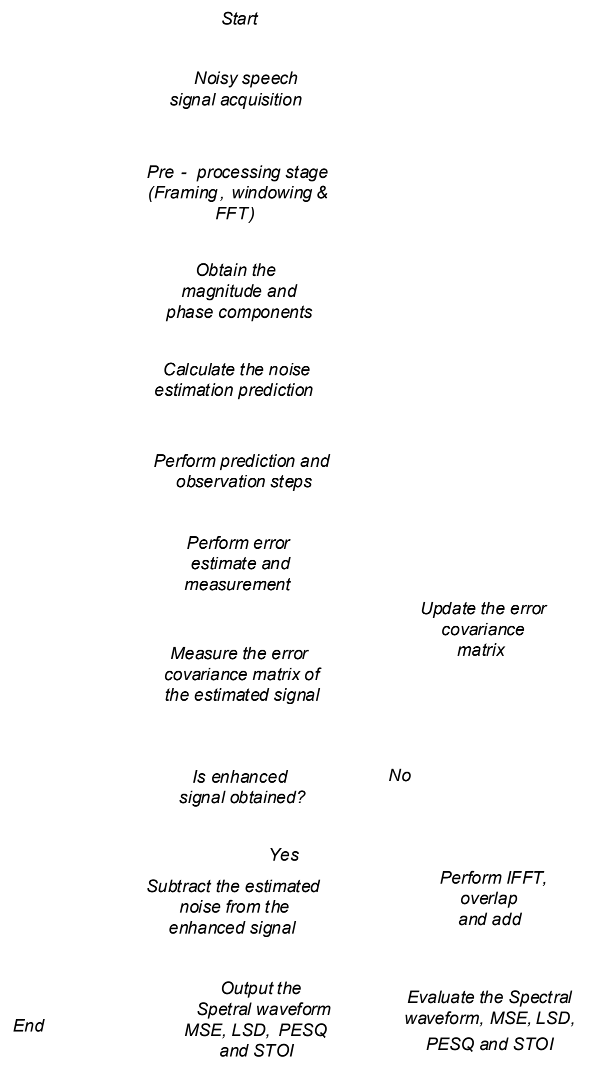


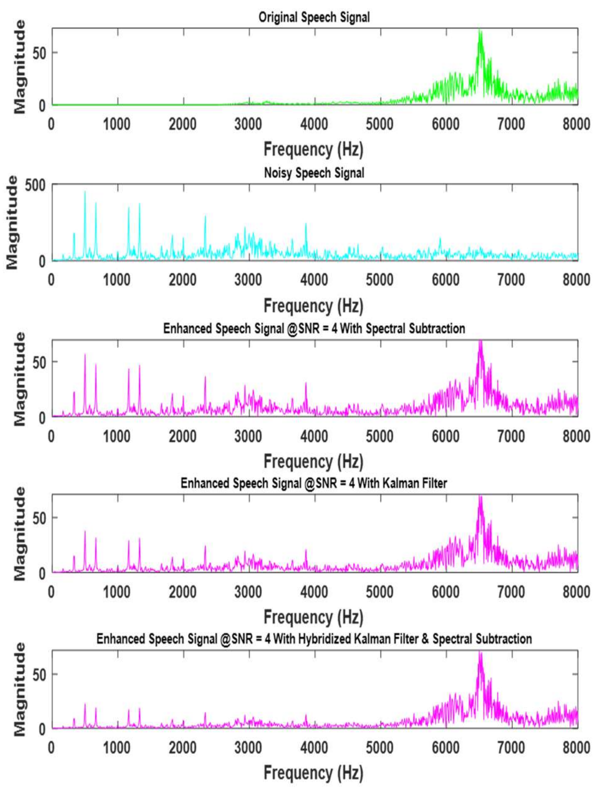
Figure 2. Proposed MD-KSF speech enhancement flowchart.

Table 1: Simulation parameters

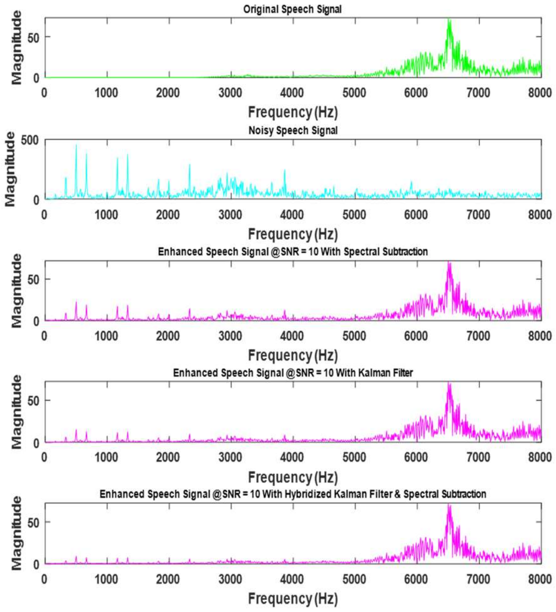
Parameter	Specification
Channel type	Single channel
Noise type	Stationary and non-stationary
Framing duration	20 ms
Number of frames	534
Sampling rate	8 kHz
Bit length	16

IV. RESULTS AND DISCUSSION

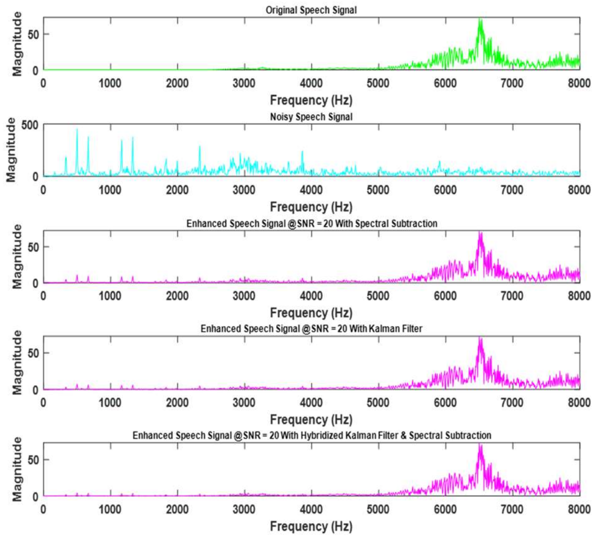
The proposed coupled MD-KSF is evaluated using spectral waveforms in the frequency domain at 5, 10 and 15 dB as shown in Figure 3. From the figures, it was observed that as the spectral power increases, the magnitude of the enhanced clean speech signal improves with little significant effect of remnant residual noise in speech. The enhanced clean speech obtained for the proposed coupled MD-KSF shows noticeable improvement in the speech obtained as against Kalman filter and Spectral subtraction filtering method. The coupled MD-KSF removes the wide-band residual noise components considerably in the enhanced speech signal and providing better resolution in the speech spectral peaks and a very low residual noise floor in the enhanced speech. The spectrum of the coupled MD-KSF, Kalman filter, and spectral subtraction noise filtering technique at 15 dB shows an improvement in the spectrum of the enhanced speech signal as compared to the 5 and 10 dB. This is due to the large amount of remnant noise has been suppressed in the coupled MD-KSF technique resulting in significant reduction in musical note effect, thereby improving the quality and intelligibility of the enhanced speech signal..



(a)



(b)



(c)

Figure 3: Waveform of noisy speech signal for MD-KSF and existing techniques at (a) 5 dB, (b) 10 dB and (c) 15 dB.

In Figure 4, the plot of MSE for the proposed coupled MD-KSF, Kalman filter and Spectral subtraction techniques are evaluated with varying values of SNR. The result shows that as the SNR increases, the MSE decreases significantly. Comparing the result obtained at different values of SNR shows that at 5 dB, the values of MSE for MD-KSF, Kalman filer and spectral subtraction techniques were 2.6E-05, 7.22E-05 and 1.625E-04, respectively, while the corresponding values at SNR of 10 dB were 5.76E-06, 1.60E-05 and 3.60E-05 were obtained. In addition, the MSE values of 2.59E-06, 7.205E-06 and 1.625E-05 for MD-KSF, Kalman filer and Spectral subtraction techniques, respectively at SNR of 15 dB. The proposed coupled MD-KSF experiences a significant reduction in MSE value at all SNR compared to existing filtering techniques due to tremendous suppression of the remnant noise. This improves the psychoacoustic and physiological characteristics of the enhanced speech signal.

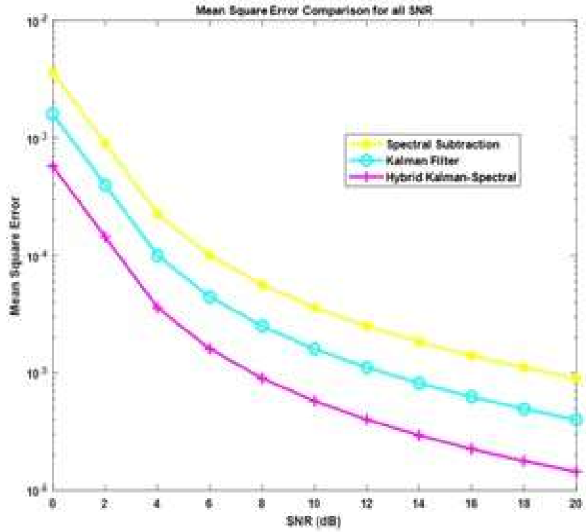


Figure 4: MSE against SNR for the proposed coupled MD-HKSF and the existing techniques.

Figure 5 shows the plot of LSD against SNR for the coupled MD-KSF technique. The result obtained depicts that the LSD value decreases as the SNR increases. Worthy of note is the result obtained for the proposed coupled MD-KSF technique which outperforms both the Kalman filter and the Spectral subtraction techniques. This shows a significant improvement in the psychoacoustic and physiological characteristic features observed in the proposed coupled MD-KSF compared to the existing techniques measured at the receiver for all values of SNR. As revealed from the plot, at a low SNR of 5 dB, the LSD value obtained for the proposed coupled MD-KSF gave a reduction of 18.05% over Kalman filter and 33.34% over the spectral subtraction technique. Also, at SNR of 10 dB, the LSD value obtained for the proposed coupled MD-KSF recorded a reduction of 19.00% and 35.89% over Kalman filter and Spectral subtraction techniques, respectively. With the SNR increased to 15 dB, a reduction of 19.62% and 37.19% were recorded for the proposed coupled MD-KSF over Kalman filter and Spectral subtraction techniques, respectively. This result suggests that the proposed coupled MD-KSF significantly subdue the stationary and non-stationary noises, and the computational errors exhibited by the existing techniques. This therefore improves the quality, audibility and intelligibility of enhanced speech signal over a wireless noisy environment at the receiver.

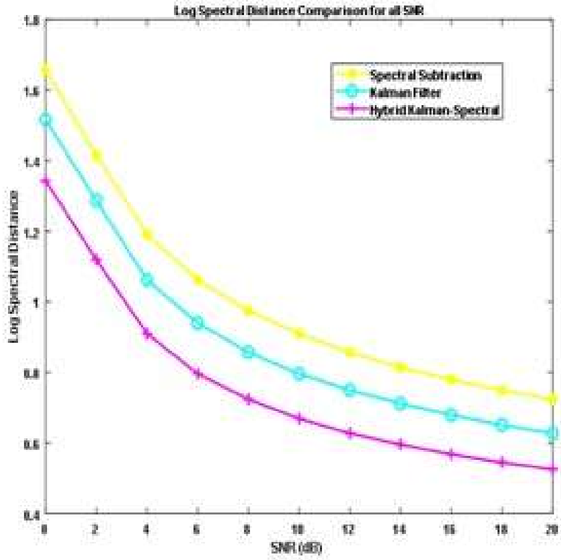


Figure 5: LSD against SNR for the proposed coupled MD-KSF and the existing techniques.

Figure 6 presents the plot of STOI obtained when the noisy speech signal from babble, car, and street noises were mixed with clean speech signal transmitted over wireless communication channel and received with the proposed coupled MD-KSF at the receiver. It was observed that the quality of the enhanced speech signals for all the techniques increases with an increase in SNR. The closer the STOI score to 1, the more intelligible the enhanced speech signal. For instance, as observed from the result obtained, the proposed coupled MD-KSF outperforms both the Kalman filter and Spectral subtraction techniques by a reduction of 33.09% and 39.83% for babble noise; 35.40% and 41.22% for car noise and 32.57% and 37.29% for street noise, respectively at 10 dB. This result further shows the superiority of the proposed coupled KSF over the existing techniques with better quality of enhanced speech signal over wireless channel.

Figure 7 presents the plot of PESQ against SNR. It is observed from the result that the quality of the enhanced speech signals for all the techniques increases with increase in SNR. The coupled MD-KSF outperforms both the Kalman filter and Spectral subtraction techniques. Furthermore, the result shows the supremacy of the proposed coupled MD-KSF over the existing techniques in terms of a better quality of enhanced speech over wireless channel. The result obtained for the coupled MD-KSF, Kalman filter and Spectral subtraction enhanced the quality of the noisy speech signal at 10 dB for babble noise were 3.5309, 2.3625 and 2.1247, for car noise were 3.6523, 2.3594 and 2.1470, and for street noise were 3.5174, 2.3717 and 2.2057, respectively. The result further confirms the preservation of the psychoacoustic and physiological characteristics of the enhanced speech signal as compared to the existing techniques.

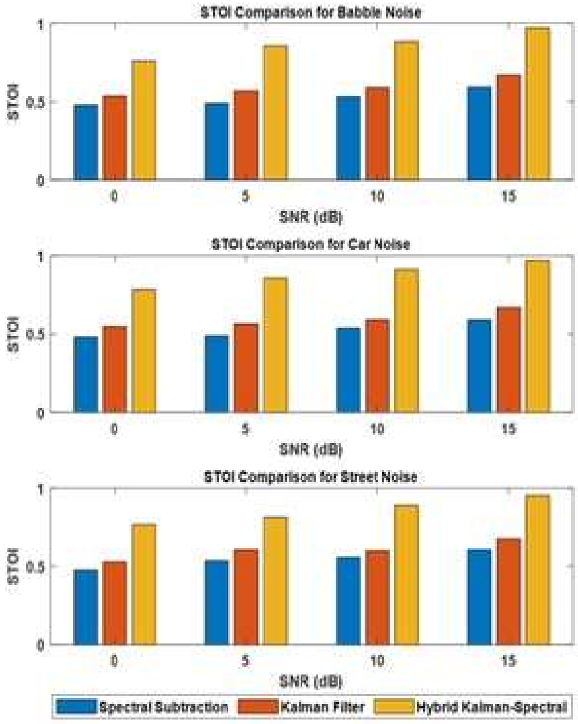


Figure 6: STOI plots of the proposed coupled MD-KSF and the existing techniques.

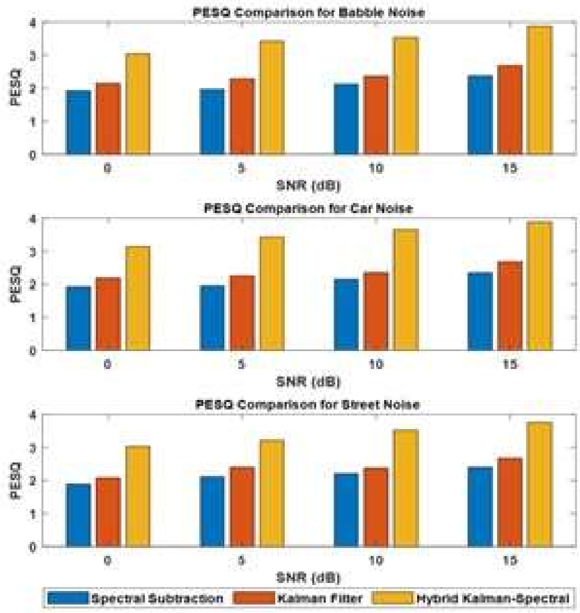


Figure 7: PESQ plots of the proposed coupled MD-KSF and the existing techniques.

V. CONCLUSION

In this paper, the effect of coupled MD-KSF for noise suppression in speech analysis and processing over wireless single channel communication system has been presented. The speech enhancement technique has been proposed by coupling existing Kalman filter and spectral subtraction techniques. Speech signal captured in the presence of uncontrolled noise was transformed into wave format and pre-processed. The pre-processed noisy speech signal was passed through the proposed coupled MD-KSF to eliminate the stationary alongside the non-stationary noises. Mathematical expression for the proposed speech enhancement technique was derived and evaluated in terms of SNR, MSE, SER and LSD while PESQ and STOI objective tests were performed. The result obtained shows that the proposed coupled MD-KSF technique outperforms the existing techniques. This indicates that using the proposed technique in noise suppression provides better preservation of the psychoacoustic and physiological characteristic features and musical note reduction.

AUTHOR CONTRIBUTIONS

Olawole E. T.: Conceptualization, Methodology, Software and Writing – original draft. **Akande D. O.:** Co-supervised, Methodology, Writing – review & editing. **Adeyemo Z. K.:** Supervisor, Writing – review & editing. **Ojo F. K:** Methodology & Validation. **Ojo S. I.:** Software & Validation.

REFERENCES

Adeyemo, Z. K.; Akande, D. O.; Ojo, F. K., and Raji, H. O. (2012). Comparative evaluation of fading channel model selection for mobile wireless transmission system. International Journal of Wireless and Mobile Networks, 4 (6): 127-138.

Alam, M. J. D. (2010). Perceptual improvement of wiener filtering employing a post-filter. Digital Signal Processing. 21 (1): 54-65

Amit, K. S. and Vinay, K. S. (2019). Performance of evaluation of different window function for STDFT based on prediction techniques taking paired numeric mapping scheme. International conference on signal processing and integrated networks. 739-743.

Banale, N.; S. Sudhansu and M. Jagde (2014). Noise free speech enhancement based on fast adaptive Kalman filtering algorithm. International Journal of Computer Applications, 93 (9): 47-51.

Bingyin, X. and C. Bao (2014). Wiener filtering based speech enhancement with weighted denoising auto-encoder and noise classification. Speech Communication, 60: 13-29.

Chen, R.; C. Chan and H. C. So, (2012). Model-based speech enhancement with improved spectral envelope estimation via dynamics tracking. IEEE Transaction on Audio, Speech, and Language Processing. 20 (4): 1324-1336.

Chowdhury, S. M.; H. Mahmudul, and A. Khalil (2018). Noise reduction from speech signals using modified spectral subtraction technique. European Journal of Engineering Research and Science. 3 (7): 78-80.

Enqvist, P., and Karlsson, J. (2008). Minimal distance and covariance interpolation. IEEE conference on decision and control. *137-142.*

Ephraim, Y. and D. Malah (1984). Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. IEEE Transaction on Acoustic Speech Signal Processing. 32: 1109-1121.

Esch, T. and Vary, P. (2008). Speech enhancement using a modified Kalman filter based on complex linear prediction and super gaussian priors. In Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing. pp. 4877-4880. Las Vegas.

Foth, M.; R. Schroeter, and J. Ti (2013). Opportunities of public transport experience enhancements with mobile services and urban screens. International Journal of Ambient Computing and Intelligence. 5 (1): 1-18.

Goh, Z.; K. C., Tan and B. T. G. Tan (1999). Kalman filtering speech enhancement method based on a voiced-unvoiced speech model. IEEE Transaction on Speech and Audio Processing, 7 (5): 510-524.

Hamid, O. K. (2018). Frame blocking and windowing in speech signal. Journal of Information, Communication and Intelligence System, 4 (5): 87-94.

Leonardro, A.; A. Culberto, S. V.; Edwardo and D. C. Roberto (2018). Comparative study between the discrete frequency Kalman filtering and discrete time Kalman filtering with application in noise reduction in speech signals. Journal of Electrical and Computer Engineering. 1-5.

Loizou, P. C. (2013). Speech enhancement: theory and practice. 2nd edition. CRC Press, Inc.: Boca Raton, FL, USA.

Olawole E. T.; D. O., Akande and Z. K. Adeyemo (2022). Speech enhancement in wireless communication system using hybrid spectral-Kalman filter. International Journal of Electrical and Electronic Engineering and Telecommunications. 11 (5): 363-372.

Oppenheim, A. V.; R. W., Schaffer and J. R., Buck (1999). Discrete time signal processing. Prentice Hall. 2nd edition. *Prentice Hall,* Upper Saddle River, New Jersey.

Paliwal, K. K. and Basu A. (1987). A speech enhancement method based on Kalman filtering. In Proceeding of IEEE International Conference on Acoustics. Speech, and Signal Processing. 12: 177-180.

Paliwal, K.; B. Schwerin and K. Wójcicki (2012). Speech enhancement using a minimum mean-square error short-time spectral modulation magnitude estimator. Speech Communication. 54: 282-305.

Pardede, H.; K. Ramli, Y.; Suryanto, N. Hayati and A. Presekal (2019). Speech enhancement for secure communication using coupled spectral subtraction and Wiener filter. Electronics. 8 (897): 1-19.

Rao, G. M. and U. S. Kumar (2016). Speech enhancement using iterative Kalman filter with time and frequency mask in different noisy environment. International Journal of Signal Processing, Image Processing and Pattern Recognition. 9(9): 317-324.

Roy, S. K. and K. K. Paliwal (2021). Robustness and sensitivity tuning of the Kalman filter for speech enhancement. Signals, 2021 (2) 434–455.

Santos, E.; M. Khosravy; M. A. Lima, A. S. Cerqueira, C. A. Duque, and A. Yona, (2019). High accuracy power quality evaluation under a colored noisy condition by filter bank ESPRIT. Electronics. 8 (1259): 1 - 28.

Sen, S.; A. Dutta and N. Dey (2019). Audio indexing. in audio processing and speech recognition. Springer Briefs in Applied Sciences and Technology (pp. 1–11). Singapore: Springer.

So, S. and K. K. Paliwal (2010). Suppressing the influence of additive noise on the Kalman gain for low residual noise speech enhancement. Speech Communication. 53: 355–378.

So, S.; K. K. Wojcicki and K. K. Paliwal (2010). Single-channel speech enhancement using Kalman filtering in the modulation domain. In Proc. Conf. Intl. Speech Communication Association, (September) pp. 993-996. Japan.

So, S. and K. K. Paliwal (2011). Modulation-domain Kalman filtering for single-channel speech enhancement. Speech Communication, 53 (6): 818-829.

Sohn, J.; N. S. Kim and W. Sung (1999). A statistical model-based voice activity detector. IEEE Signal Process. Letters. 6 (1): 1-3.

Upadhyay, N. and Karmakar, A. (2015). Speech enhancement using spectral subtraction-type algorithms: a comparison and simulation study. Eleventh International Multi-Conference on Information Processing-2015 (IMCIP-2015), 54: 574-584.

Upadhyay, N. and Jaiswal, R. K. (2015). Single channel speech enhancement: using wiener filtering with recursive noise estimation. 7th International conference on Intelligent Human Computer Interaction. IHCI 2015, 84: 22-30.

Vaseghi, S. V. (2000). Linear prediction models, Advanced digital signal processing and noise reduction, 2nd Edition, John Wiley & Sons, UK.

Venkateswarlu, S. C.; K. S. Prasad and A. S. Reddy (2011). Improve speech enhancement using Weiner filtering. Global Journal of Computer Science and Technology 11(7): 30-38.

Wang, Y. and M. Brookes (2018). Model-based speech enhancement in the modulation domain. IEEE Trans. on Audio, Speech and Language Processing. 26(3): 580-594.

Yuhong, L.; Z. Dongmei and J. Zhanjun (2013). Improved spectral subtraction speech enhancement algorithm, Advanced Material Research. pp. 760-762.