Speech Enhancement in Wireless Communication System Using Hybrid Spectral-Kalman Filter

Olawole E. Toyin[1], Akande D. Oluwole[2], and Adeyemo Z. Kayode[2]
[1] Computer Engineering, University of Ilorin, Ilorin, Nigeria
[2] Electronic and Electrical Engineering, Ladoke Akintola University of Technology, Ogbomosho, Nigeria
Email: olawole.et@unilorin.edu.ng; {doakande; zkadeyemo}@lautech.edu.ng

***Abstract*—Speech enhancement is of paramount importance in improving the intelligibility and audibility of speech signals in an uncontrolled noisy environment over wireless communication systems. However, the existing techniques used to achieve this such as spectral subtraction are characterized by remnant noise or musical note, while Kalman filter which can remove the remnant noise is associated with instability and non-linear characteristics. Therefore, this paper proposes speech enhancement model in noisy environment using hybrid spectral-Kalman filter over wireless communication system. The proposed model was carried out using acquired noisy speech signals by the Samsung Android mobile phone. The phone call was made in an uncontrolled noisy environment. The signal received was processed by the hybrid spectral-Kalman filter which was achieved by using the output speech signal from the spectral subtraction in time domain as the input of the Kalman filter. The proposed hybrid spectral-Kalman filtering technique was evaluated using Signal-to-Noise Ratio (SNR), Mean Square Error (MSE), Perceptual Evaluation of Speech Quality (PESQ) and Short-Time Objective Intelligibility (STOI). Comparison was then performed with the existing techniques and validation was performed using corrupted speech signals obtained from the NOIZEUS corpus data set. The result revealed that the proposed technique performed better than the existing techniques in a noisy environment.**

***Index Terms*—Kalman filter, Mean Square Error (MSE), spectral-Kalman, spectral subtraction, speech enhancement, Signal-to-Noise Ratio (SNR)**

I. INTRODUCTION

Wireless communication plays an importance role in the daily activities of human endeavors by providing seamless connectivity to numerous network subscribers across the globe at any point in time [1]. The unrelenting growth in the application of wireless devices most especially for audio or speech signal transmission have led to tremendous improvement in wireless communication systems for a reliable and effective communication [2]. Speech quality which is one of the crucial concerns for audio signal in multimedia services over wireless communication system is an essential feature and of great interest.

Speech or voice communication can be performed via mobile phone through traditional calls or digital voice call over various social medium platforms via the internet in convening sensitive and personal information [3]. The quality and intelligibility of a good or acceptable voice communication propagated over wireless medium in the presence of additive Gaussian noise (stationary) and background disturbance (non-stationary) must be prompt, audible and clearly deciphered [4], [5]. However, noise is an interference which can totally destructs the quality, intelligibility and audibility of transmitted voice or speech signals to achieve a reliable and better voice communication and is present everywhere [2], [3], [6].

Noise can be classified into natural and non-natural [7]. Natural noise is generated from natural phenomenon which include: thunder, wind, rain and cosmic bodies while non-natural (man-made) noises are generated as a result of the impact of human activities on the environment in form of sound with different energies and spectra. Such sounds can be generated from train, factory, airplane, machines, market places and babble. To address the characteristics of transmitted speech signals in the presence of background noise and additive Gaussian noise, speech enhancement plays a vital role and an interesting area of research to improve the quality and intelligibility of noisy speech [8]-[10].

Some of the areas where speech application becomes imperative includes: speech recognition, hands free devices, hearing aids etc. [3], [11], [12] and requires great attention. Some of the existing filtering techniques to mitigate the effect of noise from noisy voice communication include; Minimum filter, Maximum filter, Weiner filter [13], spectral subtraction [14] and Kalman filter [15], [16]. Spectral subtraction and Kalman filtering are the majorly used techniques in this regard. Spectral subtraction filter has been reported to minimize the noise by subtracting noisy speech with noisy estimate [17]. The noise situated between the pauses of the noisy speech is detected by applying Voice Audacity Detector (VAD) [18]-[20].

In addition, the spectral obtained from the subtractive rules of spectral subtraction always result in negative values due to errors in estimating the noise spectrum of the noisy speech signals. This subsequently results in the presence of remnant noise also known as musical noise at the output of the speech signal [21]. Kalman filtering which is another frequently used in different applications [18] has proven to solve the problem of noisy or corrupted speech signal by balancing the quantity and quality of speech. This is done by taking the ratio of

Manuscript received December 27, 2021; revised March 21, 2022; accepted April 22, 2022.

Corresponding author: Akande D. Oluwole (email: doakande@lautech.edu.ng).

speech to noise trend for the filtering algorithm and estimating the Mean Square Error (MSE) [22]. However, Kalman filter exhibits instability and nonlinear characteristics as a result of the error observed at the prediction update stage [23] and causes speech distortion. Nonetheless, different variance of Kalman filter has surfaced in the last decade such as iterative Kalman filter, to improve the output Signal-to-Noise Ratio (SNR) of the filter [24]. There have been numerous existing works on speech enhancement, few of which are reviewed as follows:

In [23] the author presented a speech enhancement technique using a fast adaptive Kalman filtering algorithm to eliminate the noise and other distortions present in a contaminated speech which is present in the conventional Kalman filtering algorithm. The proposed filtering algorithm was made to update the first value of state vector constantly in order to eliminate the matrix operations and also reduces the time complexity of the algorithm. The proposed technique outperformed the conventional and perceptual Kalman filtering algorithm in terms of reduced running time.

Authors in [25] used a modified spectral subtraction technique to reduce the noise present in a speech signal. The work explored the effects of interfering sounds on speech signal by considering white Gaussian noise and colored noise in a single channel system. The modification to the conventional spectral subtraction technique was the removal of the VAD while assuming *a priori* information knowledge of the noise. The result obtained showed a better improvement in SNR for white noise; however, the SNR deteriorates significantly as a result of the musical note effect in the colored noise.

In [26] a multi-band spectral subtraction with cross spectral component reduction was used to enhance the speech signal. In the work, the modified multi-band spectral subtraction and the cross-correlation techniques were employed. The modified multi-band spectral subtraction was used to compute the over subtraction factor which is dependent on the SNR value in a frame-by-frame manner. The second method was implemented to compute the correlation between the noisy speech and the speech signal. The result of the work was evaluated using subjective listening and spectrogram analysis. However, these evaluation metrics are insufficient to determine the quality of the enhanced speech signal.

In [7], the removal of pink noise generated from electronic devices which corrupt speech signal was proposed using Kalman filtering. Kalman filter was used to eliminate pink noise from the corrupted speech signal acquired from a SpEAR database (SpEAR is a database of in-ear speech signals recorded in varying conditions of the audio-phonation loop). The speech signal was corrupted with pink noise of SNR of 5 dB, after which Kalman filter was employed to remove the noise present in the speech signal. The result showed that the pink noise was significantly reduced, however, with more iterative processes. Also, the approach was not applied in real-time situation where SNR is very high. Based on these limitations observed in the existing works, this paper proposes a hybrid spectral-Kalman filtering technique to enhance the transmitted speech signal in the presence of uncontrolled noisy environment over the wireless communication system.

In view of the drawbacks of the reviewed works, this paper proposes a combination of the spectral subtraction and Kalman filtering techniques to enhance the quality of propagated speech signal in the presence of both stationary and non-stationary noises over wireless communication system. The contributions of this paper are as follows;

1) The noisy speech signal used in this paper comprises a combination of different uncontrolled noises generated from different sources (non-stationary) and transmitted over the wireless channel (stationary) unlike those utilized in the literature,

2) To reduce the uncontrolled noise in the transmitted speech signal, a hybrid spectral-Kalman filter was developed to remove the remnant noises (musical notes) exhibited at the output of spectral subtraction and ensure a stable predictive update,

3) Lastly, mathematical expression for the proposed hybrid spectral-Kalman filter was derived and the results were evaluated using SNR and MSE. In addition, to test for the quality and the intelligibility [22] of the proposed technique, Perceptual Evaluation of Speech Quality (PESQ) and Short-Time Objective Intelligibility (STOI) were employed and the result was validated using NOIZEUS corpus data set.

The remaining of this paper is organized as follows: Existing speech enhancement techniques are presented in Section II while Section III presents the review on the existing related work. Section IV proposes hybrid spectral-Kalman filter technique. The simulation results are drawn in Section V and Section VI concludes the paper.

II. Existing Speech Enhancement Techniques

Noise reduction algorithms play an important role in enhancing the quality of an acceptable speech signal over wireless transmission medium. The primary objectives of speech enhancement techniques are to improve the intelligibility and the overall perceptual quality of the degraded speech signal using audio signal processing techniques. A real-time noise reduction technique provides a vital attribute in mobile devices to offer better quality speech signal at the receiving end [24].

Take for instances, experiencing a situation where an audio call is made in a noisy environment such as crowded market, train/bus station or densely populated area characterized with uncontrolled background noise. This noise can degrade the audibility and quality of the ongoing call at the receiver end. Therefore, speech enhancement becomes imperative and the algorithms to be employed can be categorized into three fundamental classes namely; filtering techniques, spectral restoration, and model-based methods [25].

The spectral subtraction, Weiner filtering, Kalman filtering and Signal Subspace Approaches (SSA) are categorized as filtering techniques. For the spectral restoration, the Minimum Mean-Square-Error Short-Time Spectral Amplitude Estimator (MMSE-STSA) is used while the model-based approaches are Hidden Markov Model (HMM) and stochastic model [25]. Therefore, in this paper, filtering techniques are considered which include the combination of the spectral subtraction and Kalman filtering techniques.

A. Spectral Subtraction Technique

This technique was first proposed by Bolls and Berouti in the year 1979, to enhance the broadband noise present in corrupt speech and is rated one of the popular speech enhancement technique [3]. The method involves the subtraction of an estimate of the noise power spectrum from the speech power spectrum and setting the negative differences to zero. Subsequently, the new power spectrum is recombined with the original phase, and then reconstruction of the time waveform is performed. Presented in Fig. 1 is the diagram of spectral subtraction noise filtering technique. The figure comprises the noisy speech signal pre-processing and framing, VAD and Discrete Fourier Transform (DFT) stages.

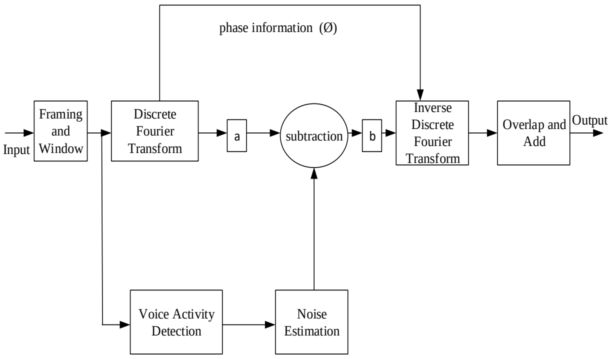


Fig. 1. Spectral subtraction. a: magnitude information of the pre-processed signal, b: magnitude information of the spectral subtractive process.

The noisy speech signal which is time-varying is first sent into the pre-processing stage to undergo framing and windowing. The noisy speech $y(n)$ contaminated by noise is basically expressed as:

$$y(n)= s(n)+v(n) \tag{1}$$

where $s(n)$ is the clean speech signal and $v(n)$ is the additive noise signal. Framing occurs when the speech signal is divided into samples, frames with overlapping short-time of 1/2 or 3/4. This is to preserve the sampled values in the last two blocks and provide a base for the assumed stationary speech signal in a time series of isolated stationary fragment for ease of extracting process. More so, windowing process is performed to minimize the spectral distortion and discontinuities in speech signals. Thereafter, the pre-processed signal is passed to the VAD for binary classification into either 1 or 0. If the output is 1, the frame contains speech samples and if 0, the frame contains the noise estimate [27], [20].

The work of the DFT is to convert each frame to its frequency domain form in order to obtain the spectral magnitudes and phase. The noise estimate is then subtracted from the original stream of the noisy speech spectral magnitudes [27]. The output of the spectral subtraction process is then combined with the phase obtained from the DFT to generate an enhanced speech spectral magnitude. This is then passed through Inverse Discrete Fourier Transform (IDFT) to obtain the enhanced speech signal in time domain. The expression for the spectral subtraction $|\hat{S}(f)|^b$ is given in [27] as:

$$\left|\hat{S}(f)\right|^b = \left|Y(f)\right|^b - \alpha\left|\bar{V}(f)\right|^b \tag{2}$$

where $|Y(f)|^b$ is the enhanced clean speech spectra, $|\bar{V}(f)|^b$ is the time average noise spectra, α is a control factor that depends on the kind of subtraction which is 1 for full subtraction and greater than 1 for over-subtraction, and b=1 is for magnitude spectral subtraction and b=2 for power magnitude spectral subtraction.

B. Kalman Filtering Technique

Kalman filtering technique is an algorithm that is widely used in different fields in estimating and updating. This involves using a number of observed measurements over a period of time, especially measurements that contain recorded noise and other irregularities. This helps to produce an estimate of unknown variables which tend to be more precise and accurate than those based on single measurement. This is achieved by estimating a joint probability distribution over the variables for each particular time frame.

Fig. 2 shows the Kalman filtering technique in two-step process of the prediction and the observation steps [28]. In the prediction step, the filter produces an estimate of the current state variables along with their uncertainties. Once the outcome of the next measurement is observed, the estimates are updated using weighted average by assigning more weight to estimate with higher certainty.

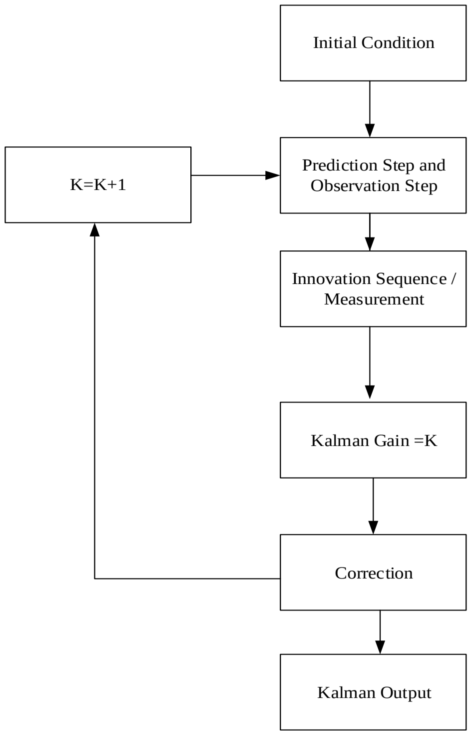


Fig. 2. Kalman filtering process [28].

This filter presents an optimal way of extracting a signal from noise by exploiting the state space model form. These state space models allow speech to be further described by an Autoregressive (AR) process [28]. Kalman filtering technique is a state equation that models the dynamics of the signal generation process and an observation equation of statistical noise and distorted observation [29].

For a signal $\mathbf{x}(m)$ and a noisy observation $\mathbf{y}(m)$, the equations describing the state processes model and the observation model are given by [22] as:

$$\mathbf{x}(m) = \mathbf{Ax}(m-1) + \mathbf{Bu}(m) + \mathbf{e}(m) \tag{3}$$

$$\mathbf{y}(m) = \mathbf{Hx}(m) + \mathbf{n}(m) \tag{4}$$

where $\mathbf{x}(m)$ is the p-dimensional signal vector or the state parameter vector at time m, $\mathbf{A}$ is a $p \times p$ dimensional state matrix that relates the state of the process at time m–1 and m and is expressed as:

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ a_p & a_{p-1} & a_{p-2} & \cdots & a_2 & a_1 \end{bmatrix} \tag{5}$$

$\mathbf{B}$ is a p-dimensional input vector, $\mathbf{u}(m)$ is the p-dimensional control input, $\mathbf{e}(m)$ is the p-dimensional uncorrelated input excitation, $\mathbf{y}(m)$ is the M-dimensional observation vector. Nevertheless, in applications such as channel equalization and speech enhancement, the external control unit $\mathbf{B}$ is not required when Kalman filter is applied. This makes (3) reduced to:

$$\mathbf{x}(m) = \mathbf{Ax}(m-1) + \mathbf{e}(m) \tag{6}$$

Equation (6) is the error covariance update for the M-dimensional observation vector whose diagonal is the MSE. The sum of these MSE yields the trace of the matrix. Hence, the mean square error can be reduced by minimizing equation (6). Also, by taking the differential of the observation vector $\mathbf{y}(m)$ with respect to $\mathbf{H}$, Equation (4) becomes:

$$\mathbf{y}'(m) = \mathbf{H}^T\mathbf{x}(m) + \mathbf{n}(m) \tag{7}$$

where $\mathbf{y}'(m)$ is the noisy speech signal, $\mathbf{H}^T\mathbf{x}(m)$ is the clean speech signal and $\mathbf{n}(m)$ is assumed to be the noise.

III. Proposed Hybrid Spectral-Kalman Filter

The proposed hybrid spectral-Kalman filter model shown in Fig. 3 combines two different models namely; the spectral subtraction and the Kalman filtering in order to achieve enhanced system.

The motivation behind the proposed technique for audio signal transmission in wireless communication is to provide a robust and efficient algorithm that minimizes the problems associated with spectral subtraction and Kalman filtering techniques. This will help in achieving an accurate, precise and preserve the quality and intelligibility of audio or speech signal.

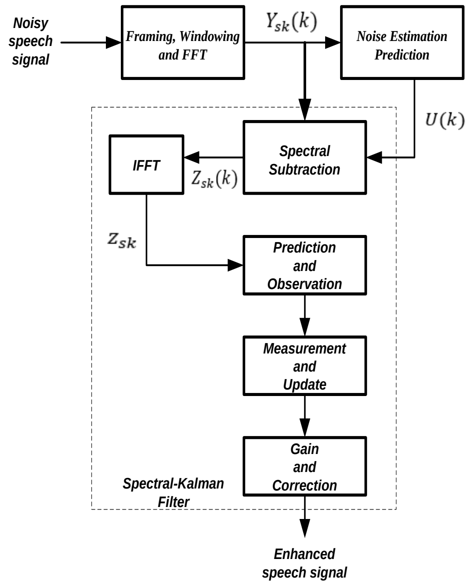


Fig. 3. Block diagram of the proposed hybrid spectral-Kalman filter.

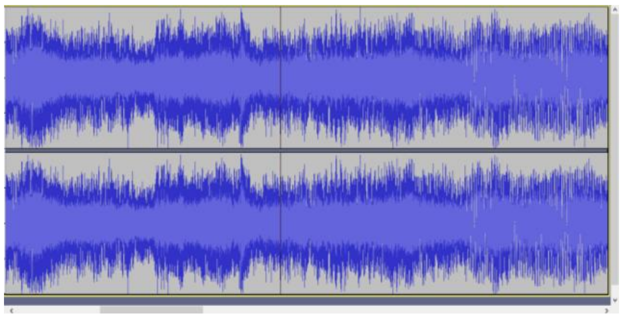


Fig. 4. Interface of the recorded noisy speech signal conversion.

A. Noisy Speech Signal Acquisition

The proposed model was initiated with the acquisition of noisy speech signal through an Android mobile phone with specification Samsung A10S with storage capacity of 32 GB, RAM of 2 GB with a battery life of 4000 mAh. A mobile phone call was made in an uncontrolled noisy environment (in a local market by the road side) to a receiver for a period of 60 s. The first 16 s of the recorded noisy speech signal is the length of the sampled signal used for simulation. At the receiver side, the received signal was transferred to a computer system (Lenovo Intel core i3, 2.1 GHz speed and 4 GB RAM) in order to convert into a wav.file (digital form) with the aid of an Analogue to Digital Converter (ADC) at a sampling rate of 8 kHz and 16 bits per sample. The interface of the received noisy speech signal conversion is as shown in Fig. 4.

B. Noisy Speech Signal Pre-processing

In this section, the sample of the noisy speech signal over the wireless communication system is subjected to signal conditioning processes. This framing process was performed on the sampled signal to evaluate the characteristics of the noisy speech signals in bits. This helps to obtain the actual characteristics of the noise in the silent regions of the speech signal. The average number of frames, n, of the noisy speech can be obtained from (8) according to [30] as

$$\begin{aligned} T &= xn - y(1-n) \\ 16 &= 20\times10^{-3}n - 10\times10^{-3}(1-n) \\ n &= 534 \text{ frames} \end{aligned} \quad (8)$$

where T is the speech signal length, x=20 ms is the frame width, y=10 ms is the overlap or shift width. The framing pattern which is used to obtain the characteristics of the noisy speech signal in bits at the silent regions of the sampled data is illustrated in Fig. 5. The sampled recorded speech signal after undergoing the framing process is shown in Fig. 5 (a) while Fig. 5 (b) shows the overlapping of the frames in order to minimize the non-stationarity features present in each frame.

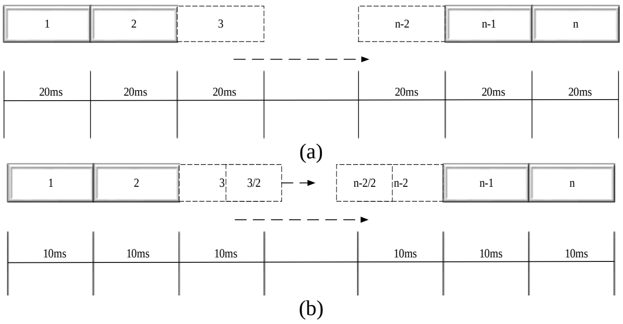


Fig. 5. Framing pattern: (a) framing and (b) overlapping frames.

After the successful framing, the sampled signal is windowed to minimize spectral distortion and discontinuities in speech signals. In this paper, the Hamming window is utilized being a raised cosine windowing technique. This is to ensure that the destructive ripples are removed to obtain the actual signal frequency spectrum. The expression for the Hamming window $w_{\text{sk}}(n)$ is given as [30]:

$$w_{\text{sk}}(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N}\right), \text{ for } 0 \le n \le N \quad (9)$$

where N is the window length as obtained in (8).

To minimize the spectral distortion of the framed sampled signal, each frame is multiplied by its Hamming window. The output of the Hamming windowed signal $\mathbf{y}_{\text{sk}}(n)$ with is a p-dimensional vector is expressed as [30]:

$$\mathbf{y}_{\text{sk}}(n) = \mathbf{x}_{\text{sk}}(n) w_{\text{sk}}(n) \quad (10)$$

where $\mathbf{x}_{\text{sk}}(n)$ is a p-dimensional framed input signal vector. The noisy speech signal is then passed to the Fast Fourier Transform (FFT) for conversion to the frequency domain and the transformed signal is expressed in [31] as:

$$\mathbf{Y}_{\text{sk}}(k) = \sum_{n=0}^{N-1} \mathbf{y}_{\text{sk}}(n) \exp\left(\frac{-j2\pi n}{N}\right) \quad (11)$$

This helps to obtain the initial magnitude measurement for noise estimation prediction by using VAD for the classification of the silent and speech regions.

C. Hybrid Spectral-Kalman Filter

The proposed hybrid spectral-Kalman filter is initiated by first performing the spectral subtraction process. This is done by taking the difference between the magnitude of the noisy speech signal and that of the noise signal. The output of this subtractive stage $\mathbf{Z}_{\text{sk}}(k)$ is given as:

$$\mathbf{Z}_{\text{sk}}(k) = \mathbf{Y}_{\text{sk}}(k) - \mathbf{U}(k) \quad (12)$$

where $\mathbf{Y}_{\text{sk}}(k)$ is the noisy speech signal and $\mathbf{U}(k)$ is the noise component. The output of the spectral subtraction is then converted back to the time domain using IFFT and combined with the phase component after the FFT operation which is expressed in [31] as:

$$\mathbf{z}_{\text{sk}}(n) = \frac{1}{N}\sum_{n=0}^{N-1} \mathbf{Z}_{\text{sk}}(k) \exp\left(\frac{j2\pi n}{N}\right) \quad (13)$$

The output of the spectral subtraction in time domain is taken to be the observation signal which is passed through the Kalman filter. Multiplying the prior signal $\mathbf{z}'_{\text{sk}}$ of (13) with an observation signal obtained from (7), the prediction signal $\hat{\mathbf{z}}_{\text{sk}}$ is expressed as:

$$\hat{\mathbf{z}}_{\text{sk}} = \mathbf{z}'_{\text{sk}}\left(\mathbf{F}\mathbf{p}_{\text{sk}} + \mathbf{v}_{\text{sk}}\right) \quad (14)$$

where $\mathbf{F}$ is the $1\times p$ noiseless signal vector with connection between the state vector and the measurement vector which is assumed to be stationary, $\mathbf{p}_{\text{sk}}$ is the $p\times1$ state vector and $\mathbf{v}_{\text{sk}}$ is the $p\times1$ associated measurement error vector. The update estimate of the predicted signal is obtained as:

$$\hat{\mathbf{z}}_{\text{sk}} = \hat{\mathbf{z}}_{\text{sk}} + \boldsymbol{\tau}_{\text{sk}}\left(\mathbf{z}_{\text{sk}} - \mathbf{F}\hat{\mathbf{z}}_{\text{sk}}\right) \quad (15)$$

where $\boldsymbol{\tau}_{\text{sk}}$ is the hybrid spectral-Kalman filer gain vector with $p\times p$ dimension which is obtained from the MSE. The error obtained which is the relationship between the observed signal $\mathbf{z}_{\text{ks}}$ and the predicted signal $\hat{\mathbf{z}}_{\text{sk}}$ is termed the MSE and is given as:

$$\mathbf{e}_{\text{sk}} = \mathrm{E}\left[\mathbf{e}_{\text{sk}}^2\right] \quad (16)$$

where $\mathbf{e}_{\text{sk}} = \mathbf{z}_{\text{sk}} - \hat{\mathbf{z}}_{\text{sk}}$. Equation (16) is further expressed as the error covariance matrix $\mathbf{P}_{\text{sk}}$ is obtained as:

$$\mathbf{P}_{\text{sk}} = \mathrm{E}\left[\mathbf{e}_{\text{sk}}\mathbf{e}_{\text{sk}}^T\right] = \mathrm{E}\left[\left(\mathbf{z}_{\text{sk}} - \hat{\mathbf{z}}_{\text{sk}}\right)\left(\mathbf{z}_{\text{sk}} - \hat{\mathbf{z}}_{\text{sk}}\right)^T\right] \quad (17)$$

Simplifying (17) with the substitution of equations (14) and (15) yields:

$$\mathbf{P}_{\text{sk}} = \left(\mathbf{I} - \mathbf{F}\mathbf{z}_{\text{sk}}\boldsymbol{\tau}_{\text{sk}}\right)\hat{\mathbf{P}}_{\text{sk}}\left(\mathbf{I} - \mathbf{F}\mathbf{z}_{\text{sk}}\boldsymbol{\tau}_{\text{sk}}\right)^T + \mathbf{z}_{\text{sk}}\boldsymbol{\tau}_{\text{sk}}\mathbf{R}\boldsymbol{\tau}_{\text{sk}}^T \quad (18)$$

where $\mathbf{I}$ is an identity matrix. $\hat{\mathbf{P}}_{\text{sk}} = \mathrm{E}\left[\mathbf{e}_{\text{sk}-1}\mathbf{e}_{\text{sk}-1}^T\right]$ is the prior estimate of the error covariance matrix and $\mathbf{R} = \mathrm{E}\left[\mathbf{v}_{\text{sk}}\mathbf{v}_{\text{sk}}^T\right]$ is the covariance of the noise model over time also assumed to be stationary. The diagonal of the covariance matrix obtained from (18) comprises the MSE and is then expressed as:

$$\mathbf{P}_{\text{sk}} = \begin{bmatrix} \mathbf{e}_{\text{sk}-1}\mathbf{e}_{\text{sk}-1}^T & \mathbf{e}_{\text{sk}}\mathbf{e}_{\text{sk}-1}^T & \mathbf{e}_{\text{sk}+1}\mathbf{e}_{\text{sk}-1}^T \\ \mathbf{e}_{\text{sk}-1}\mathbf{e}_{\text{sk}}^T & \mathbf{e}_{\text{sk}}\mathbf{e}_{\text{sk}}^T & \mathbf{e}_{\text{sk}+1}\mathbf{e}_{\text{sk}}^T \\ \mathbf{e}_{\text{sk}-1}\mathbf{e}_{\text{sk}+1}^T & \mathbf{e}_{\text{sk}}\mathbf{e}_{\text{sk}+1}^T & \mathbf{e}_{\text{sk}+1}\mathbf{e}_{\text{sk}+1}^T \end{bmatrix} \quad (19)$$

The MSE can be computed using the sum of the diagonal of the matrix in (19) which is the trace of the covariance matrix. By taking the differential of the trace of $\mathbf{P}_{\text{sk}}$ with respect to $\boldsymbol{\tau}_{\text{sk}}$, the condition for the minimum value of $\boldsymbol{\tau}_{\text{sk}}$ is obtained as:

$$\boldsymbol{\tau}_{\text{sk}} = \hat{\mathbf{P}}_{\text{sk}} \mathbf{F}^T \mathbf{z}_{\text{sk}}^T \left(\mathbf{z}_{\text{sk}} \mathbf{F} \hat{\mathbf{P}}_{\text{sk}} \mathbf{F}^T \mathbf{z}_{\text{sk}}^T + \mathbf{R} \right)^{-1} \tag{20}$$

The update of the error covariance matrix necessary to minimize the error with optimal gain is obtained by substituting (20) into (18) and after simplifying gives:

$$\mathbf{P}_{\text{sk}} = \hat{\mathbf{P}}_{\text{sk}} - \boldsymbol{\tau}_{\text{sk}} \mathbf{F} \hat{\mathbf{P}}_{\text{sk}} \tag{21}$$

and the update of the error covariance matrix by extension to a later time is expressed as:

$$\mathbf{P}_{\text{sk}+1} = \mathrm{E}\left[\mathbf{e}_{\text{sk}+1} \mathbf{e}_{\text{sk}+1}^T \right] \tag{22}$$

Therefore, the enhanced speech sample using spectral-Kalman filtering technique is given as:

$$\hat{s} = \mathbf{F} \hat{\mathbf{z}}_{\text{sk}} \tag{23}$$

The flowchart of the proposed hybrid spectral-Kalman filtering technique is as shown in Fig. 6.

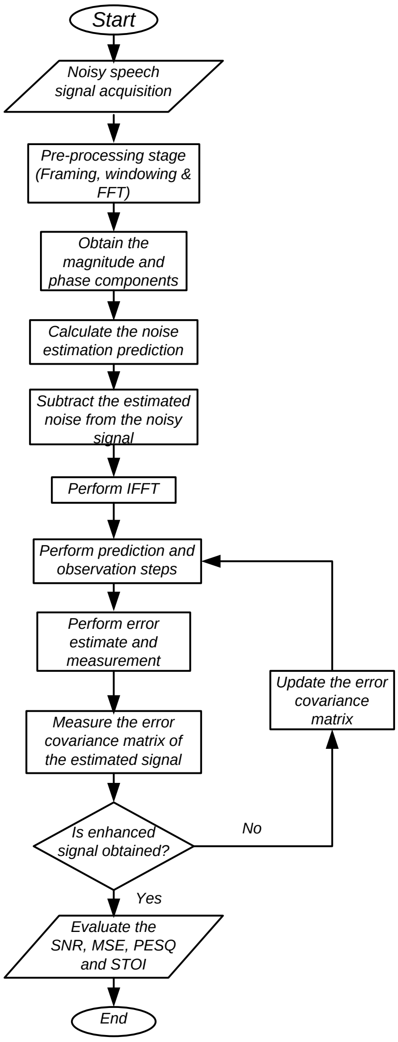


Fig. 6. Flowchart of the proposed hybrid spectral-Kalman filter.

IV. Simulation Results

The performance evaluation of the proposed hybrid spectral-Kalman filtering technique was performed through simulation in MATLAB R2021a in terms of waveform of amplitude voltage at fixed SNR with their corresponding spectrogram, MSE and with two objective evaluation tests namely: PESQ and STOI. The results obtained were then compared with Kalman filtering, and spectral subtraction techniques. In addition, corrupted noisy speech signal from NOIZEUS corpus data set [8] which comprises the mixture of clean speech corrupted with different selected noises such as babble, car and street were used. The simulation parameters used in this paper are presented in Table I.

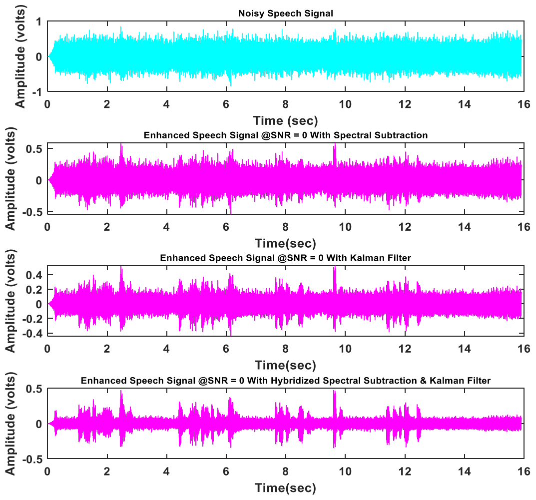
TABLE I: Simulation Parameters

Parameter	Specification
Channel type	Single channel
Noise	Stationary and non-stationary
Number of frames	534
Recording time	16 s
Sampling rate	8 kHz
Bit length	16 bits

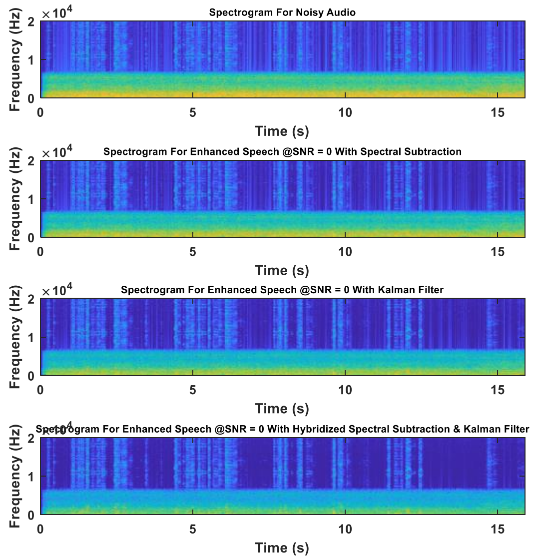
The amplitude voltage of the noisy speech signal varies from –1 V to 1 V for a period of 16 s as seen in Fig. 7 to Fig. 9 in time domain, when the noisy speech signal was simulated with conventional Kalman filter, spectral subtraction and the proposed hybrid spectral-Kalman techniques at different SNRs. This is used to determine the effectiveness of the proposed hybrid spectral-Kalman technique against the existing technques.

Fig. 7 (a) and Fig. 7 (b) present the waveform and spectrogram obtained when the noisy speech signal was passed through the Kalman filtering, spectral subtraction and hybrid spectral-Kalman filtering at a SNR of 0 dB in time domain. Fig. 7 (a) represents the waveform obtained at SNR of 0 dB, the obtained result revealed that the noisy speech wave peak that was initialy –1 V to 1 V which has been reduced to –0.5 V to 0.5 V for Kalman.filtering, spectral subtraction and the proposed hybrid spectral-Kalman filtering techniques. From the result obtained, it was shown that even if there was no substantial improvement in the speech perception quality at SNR of 0 dB, for the overall waveform, still the hybrid spectral-Kalman filtering technique had the best noise reduction waveform compared to the other existing speech enhancement techniques. This is due to a reduction in residual noise in the noisy speech signal.

Also Fig. 7 (b) shows the spectrogram of the waveforms of the spectral subtraction, Kalman filter and the hybrid spectral-Kalman filter. The spectrogram shows that the hybrid technique experienced an improvement in cleaning the noisy speech signal at the silent region compared to other filters. Therefore, the overall result obtained shows that the hybridized spectral-Kalman filtering technique gave a better result compared to other existing techniques.



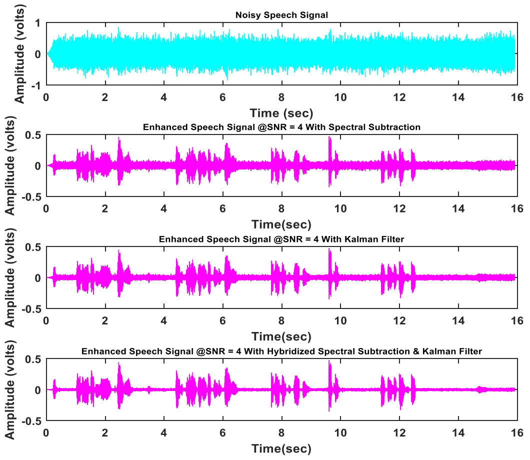
(a)



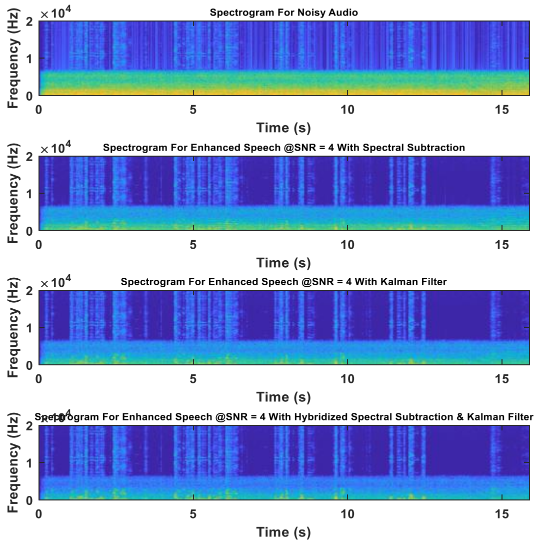
(b)

Fig. 7. Enhanced speech signal at SNR of 0 dB: (a) waveform and (b) spectrogram.

Similarly, Fig. 8 (a) and Fig. 8 (b) present the waveform and spectrogram obtained when the noisy speech signal was passed through the hybrid spectral-Kalman, Kalman filtering and spectral subtraction techniques at SNR of 4 dB in the time domain. The figures show the result obtained at SNR of 4 dB which initiates the first significant auditory improvement in the noisy speech signal. At 4 dB, Fig. 8 (a) shows that the amplitude voltage reduced significantly for all the filtering techniques with the noise component at the silent region of the enhanced signal of the hybrid spectral-Kalman technique reduced more than that of the existing techniques. From the spectrogram presented in Fig. 8 (b), the result revealed that the hybrid filtering technique outperforms the existing techniques when simulating at SNR of 4 dB. This is because the residual noise generated as a result of the instability in the prediction update of the Kalman filter and the musical note effect due to subtractive errors in spectral subtraction techniques have been significantly minimized which resulted in the improvement of the speech signal perception quality.



(a)

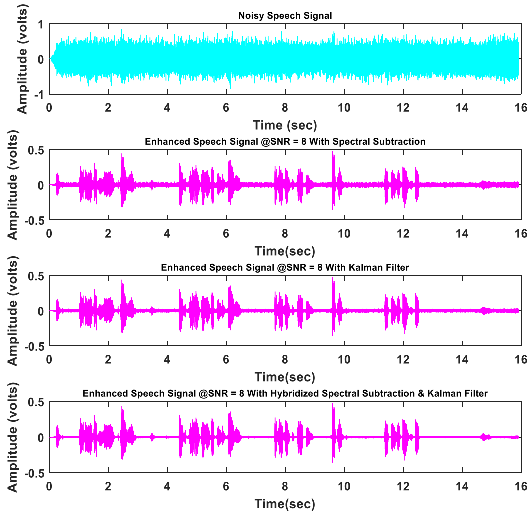


(b)

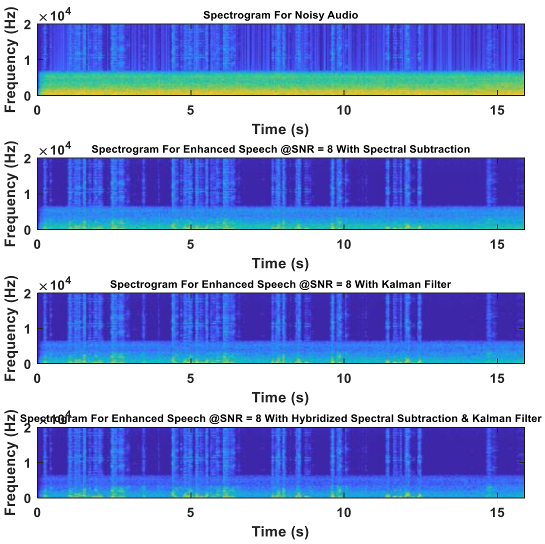
Fig. 8. Enhanced speech signal at SNR of 4 dB: (a) waveform and (b) spectrogram.

Fig. 9 (a) and Fig. 9 (b) present the results obtained when the noisy speech signal was simulated at SNR of 8 dB using the hybrid spectral-Kalman, Kalman filtering and spectral subtraction techniques. In Fig. 9 (a), a better acoustic enhancement reveals that a larger fraction of noise in the recorded speech signal has been removed by the hybrid spectral-Kalman filter. This makes the enhanced speech signal to become clearer with minimal existence of musical noise as compared to other exiting techniques. Fig. 9 (b) shows an improvement in the spectrum of the noisy speech signal as the spectrogram plot obtained revealed that a huge amount of noise has been eradicated in the proposed filtering technique.

Fig. 10 presents the plot of MSE against SNR for the speech enhancement techniques. The result reveals that the MSE decreases with increase in SNR values. In the result, the hybrid spectral-Kalman filter technique outperforms both the spectral subtraction and the Kalman filtering techniques. As observed from the plot, at SNR of 0 dB, the MSE value obtained for hybrid spectral-Kalman filter, spectral subtraction and Kalman filter were 0.000798, 0.00718 and 0.00319, respectively while at SNR of 12 dB, the corresponding MSE value were 0.00000554, 0.000050 and 0.00002220, respectively.



(a)



(b)

Fig. 9. Enhanced speech signal at SNR of 8 dB: (a) waveform and (b) spectrogram.

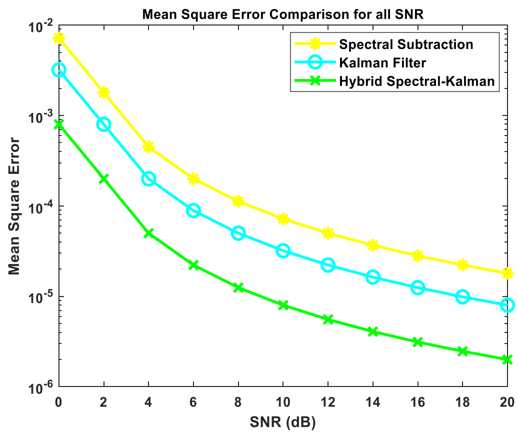


Fig. 10. Mean square error against SNR.

This is an indication that the hybrid spectral-Kalman filter was able to significantly suppress the uncontrolled noises and other computational errors exhibited by the existing techniques. This improves the quality, audibility and intelligibility of speech signal over wireless communication channel. Table II shows the MSE values obtained for the proposed hybrid spectral-Kalman filtering and the conventional techniques with varying SNR values.

TABLE II: MSE FOR THE FILTERING TECHNIQUES AT DIFFERENT SNR VALUES

SNR (dB)	Spectral subtraction	Kalman filter	Hybrid spectral-Kalman
0	0.007180	0.00319000	0.00079800
2	0.001796	0.00078900	0.00020000
4	0.000449	0.00020000	0.00004990
6	0.000200	0.00008870	0.00002220
8	0.000112	0.00004990	0.00001250
10	0.000718	0.00003190	0.00000798
12	0.000050	0.00002220	0.00000554
14	0.000037	0.00001630	0.00000407
16	0.000028	0.00001250	0.00000312
18	0.000022	0.00000985	0.00000246
20	0.000018	0.00000798	0.00000209

Fig. 11 and Fig. 12 present the objective measures of the proposed speech enhancement technique in terms of quality and intelligibility, respectively. Fig. 11 depicts the plot of PESQ against SNR. It was observed from the result that the quality of the enhanced speech signals for all the techniques increases with increase in SNR. In the result, the hybrid spectral-Kalman filter technique outperforms both the spectral subtraction and the Kalman filtering techniques. This result further shows the superiority of the proposed hybrid spectral-Kalman filter over the existing techniques in terms of a better quality of the enhanced speech.

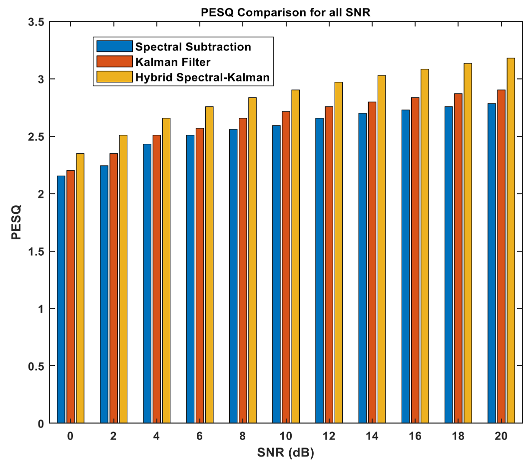


Fig. 11. PESQ against SNR.

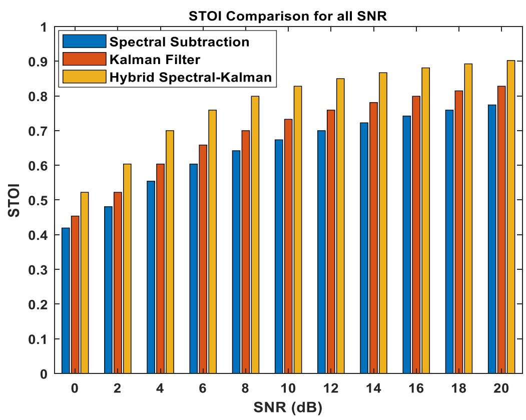


Fig. 12. STOI against SNR.

Fig. 12 depicts the plot of STOI against SNR. It was observed from the result that the intelligibility of the enhanced speech signals for all the techniques increases with increase in SNR. The hybrid spectral-Kalman filter technique shows better performance in intelligibility over both the spectral subtraction and the Kalman filtering techniques. This is an indication that the intelligence of the enhanced speech is well preserved by applying the hybrid spectral-Kalman filter technique.

To further validate the performance of the proposed hybrid spectral-Kalman filter, the noisy speech signal corrupted with different noises such as babble, car and street were obtained from NOIZEUS corpus. The output SNR of the enhanced noisy speech signals corrupted with these noises were then tested using the PESQ and STOI at SNR of 5 dB to 20 dB with a step size of 5. The results obtained are as depicted in Fig. 13 and Fig. 14, respectively. The results obtained showed that the quality and the intelligibility indicators increase with increase in the SNR.

In Fig. 13 the PESQ result demonstrated that the hybrid spectral-Kalman filtering technique enhances the quality of the enhanced speech signal as compared to other techniques with a value of 2.5840 against 2.3300 and 2.1288 for Kalman filter and spectral subtraction techniques at 10 dB using babble noise. For car and street noises, the spectral-Kalman filtering technique recorded 2.7004 and 2.5849 for PESQ as against Kalman filter and spectral subtraction with 2.3661 and 2.2723, and 2.1524 and 2.2274 at 10 dB, respectively.

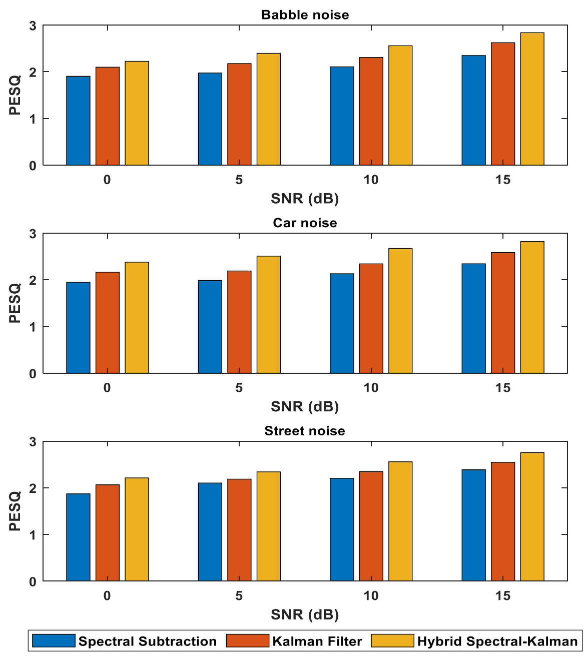


Fig. 13. PESQ against SNR for NOIZEUS corpus data set.

Fig. 14 shows that the results obtained gave STOI values of 0.7875, 0.7221 and 0.6707 for babble noise at 10 dB using the hybrid spectral-Kalman, Kalman filter and spectral subtraction techniques, respectively. Also, for car noise at the same decibel, 0.7793, 0.7065 and 0.6650 were recorded for the hybrid spectral-Kalman, Kalman filter and spectral subtraction techniques, respectively. While 0.7893, 0.7338 and 0.6948 were obtained for the hybrid spectral-Kalman, Kalman filter and spectral subtraction techniques for the street noise. The result demonstrated the level of intelligibility of the speech signal which is achieved by minimizing the remnant noises present after enhancing the signal.

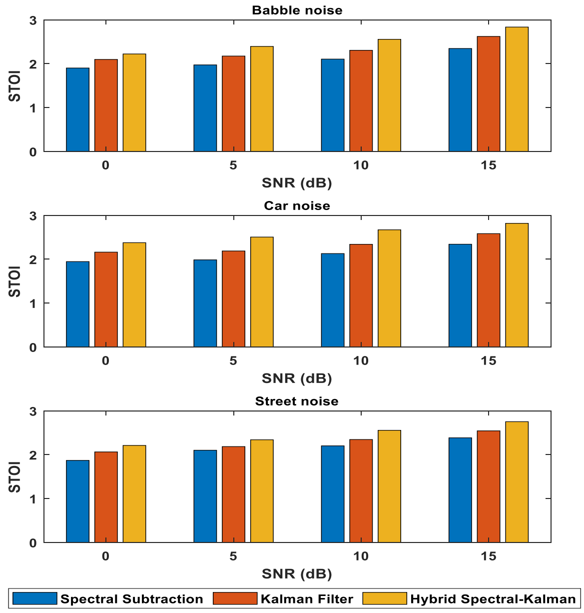


Fig. 14. STOI against SNR for NOIZEUS corpus data set.

V. Conclusion

In this paper, a hybrid spectral-Kalman filter for speech enhancement over a wireless communication system has been presented. The speech enhancement technique has been proposed using the existing spectral subtraction and Kalman filtering techniques. Speech signal transmitted in the presence of uncontrolled noise was acquired and then converted into wave format. The converted noisy signal was pre-processed and then passed through the proposed hybrid spectral-Kalman filter to remove both the stationary and non-stationary noises. Mathematical expression for the proposed speech enhancement technique was derived and evaluated in terms of SNR, MSE and objective evaluation indicators such as PESQ and STOI. The result obtained shows that the proposed hybrid spectral-Kalman filter outperforms both the existing spectral subtraction and Kalman filtering techniques. This demonstrates that the proposed technique provides a better speech enhancement over the conventional techniques in speech signal processing.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Esther T. Olawole worked on the methodology, simulation and prepared the first draft. Damilare O. Akande co-supervised, proofread, edited and assisted in the methodology while Zachaeus K. Adeyemo supervised, proofread and added technical and professional inputs.

References

[1] Z. K. Adeyemo, D. O. Akande, F. K. Ojo, and H. O. Raji, "Comparative evaluation of fading channel model selection for mobile wireless transmission system," *Int. Journal of Wireless & Mobile Networks*, vol. 4, no. 6, pp. 127-138, 2012.

[2] T. S. Rappaport, *Wireless Communications: Principles and Practice*, 2nd Edition, Prentice Hall, 2002.

[3] H. F. Pardede, K. Ramli, Y. Suryanto, N. Hayati, and A. Presekal, "Speech enhancement for secure communication using coupled spectral subtraction and Wiener filter," *Electronics*, vol. 8, no. 8, 2019.

[4] M. T. Sadiq, N. Sabbir, and W. J. Kulesza, "Spectral subtraction for speech enhancement in modulation domain," *Int. Journal of Computer Science Issues*, vol. 10, no. 4, pp. 282-285, 2013.

[5] N. Das, S. Chakraborty, J. Chaki, N. Padhy, and N. Dey, "Fundamentals, present and future perspectives of speech enhancement," *Int. Journal of Speech Techno.*, 2020.

[6] Y. Adediran, O. B. Okedere, and H. Lasisi, "Interference management techniques in cellular networks," *Cogent Engineering*, vol. 4, no. 1, pp. 1-10, 2017.

[7] M. A. Ahmed and M. K. Hota, "Removal of pink noise from corrupted speech signal using Kalman filter," *Int. Journal of Innovative Techno. and Exploring Engineering*, vol. 9, no. 4, pp. 2034-2038, 2020.

[8] P. C. Loizou, *Speech Enhancement: Theory and Practice*, 2nd ed., Inc.: Boca Raton, FL, USA: CRC Press, 2013.

[9] B. Shirley and L. Ward, "Intelligibility versus comprehension understanding quality of accessible next generation audio broadcast," *Universal Access in the Information Society.* vol. 20, pp. 691–699, July 2020.

[10] Z. Goh, K. C. Tan and B. T. G. Tan, "Kalman filtering speech enhancement method based on a voiced-unvoiced speech model," *IEEE Trans. on Speech and Audio Processing*, vol. 7, no. 5, pp. 510-524, 1999.

[11] E. Vincent, S. Watanabe, A. A. Nugraha, J. Barker, and R. Marxer, "An analysis of environment, microphone and data simulation mismatches in robust speech recognition," *Computer Speech and Language*, vol. 46, pp. 535–557, Nov. 2017.

[12] H. Levitt, "Noise reduction in hearing aids: A review," *Journal of Rehabilitation and Research Development*, vol. 38, no. 1, pp. 111–122, Feb. 2001.

[13] R. McAulay and M. Malpass, "Speech enhancement using a soft-decision noise suppression filter," *IEEE Trans. on Acoustic Speech Signal Processing*, vol. 28, no. 2, pp. 137–145, Apr. 1980.

[14] S. Boll, "A spectral subtraction algorithm for suppression of acoustic noise in speech," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, 1979, vol. 4, pp. 200–203.

[15] R. Ishaq, B. G. Zapirain, M. Shahid, and B. Lovstrom, "Subband modulator Kalman filtering for single channel speech enhancement," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, 2013, pp. 7442–7446.

[16] K. K. Paliwal and A. Basu, "A speech enhancement method based on Kalman filtering," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, 1987, vol. 12, pp. 177–180.

[17] S. Abd El-Moneim, M. I. Dessouky, F. E. Abd El-Samie, M. A. Nassar, and M. A. ElNaby, "Hybrid speech enhancement with empirical mode decomposition and spectral subtraction for efficient speaker identification," *Int. Journal of Speech Techno.*, vol. 18, no. 4, pp. 555-564, 2015.

[18] C. Boubakir and B. Daoued, "On the use of Kalman filter for enhancing speech corrupted by coloured noise," *WSEAS Trans. on Signal Processing*, vol. 4, no. 12, pp. 657-666, 2008.

[19] N. S Kim and J. H. Chang "Spectral enhancement based on global soft decision," *IEEE Signal Processing Letters*, vol. 7, no. 5, pp. 108-110, May 2000.

[20] J. Sohn, N. S. Kim and W. Sung, "A statistical model-based voice activity detector," *IEEE Signal Process. Letters*, vol. 6, no. 1, pp. 1–3, Jan. 1999.

[21] N. Upadhyay and A. Karmakar, "Spectral subtractive-type algorithms for enhancement of noisy speech: An integrative review," *Int. Journal Image, Graphics and Signal Processing*, vol. 5, no. 11, pp. 13–22, 2013.

[22] S. K. Roy and K. K. Paliwal, "Robustness and sensitivity tuning of the Kalman filter for speech enhancement," *Signals*, vol. 2021, no. 2, pp. 434–455, 2021.

[23] N. Banale, S. Sudhansu, and M. Jagde, "Noise free speech enhancement based on fast adaptive Kalman filtering algorithm," *Int. Journal of Computer Applications*, vol. 93, no. 9, pp. 47-51, 2014.

[24] Y. Chen, "Speech enhancement of mobile devices based on the integration of a dual microphone array and a background noise elimination algorithm," *Sensors*, vol. 18, no. 1467, pp. 1-14, 2018.

[25] S. M. Chowdhury, H. Mahmudul, and A. Khalil, "Noise reduction from speech signals using modified spectral subtraction technique," *European Journal of Engineering Research and Science*, vol. 3, no. 7, pp. 78-80, 2018.

[26] M. S. Karabashetti and S. K. Shridhar, "Speech enhancement using multiband spectral subtraction with cross spectral component reduction," *Int. Journal of Engineering Research and Techno.*, vol. 7, no. 10, pp. 1-4, 2019.

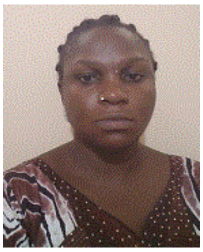
[27] S. V. Vaseghi, *Advanced Digital Signal Processing and Noise Reduction*, 2nd ed, John Wiley & Sons, 2000.

[28] G. M. Tamilselvin and P. Aarthy, "Online tuning of fuzzy logic controller using Kalman algorithm for critical system," *Journal of Applied Research and Techno.*, vol. 15, no. 5, pp. 492-503, 2017.

[29] J. Peksa, "Prediction framework with Kalman filter algorithm," *Information*, vol. 11, no. 358, pp. 2-8, April 2020.

[30] O. K. Hamid, "Frame blocking and windowing in speech signal," *Journal of Information, Communication and Intelligence System*, vol. 4, no. 5, pp. 87-94, 2018.

[31] A. V. Oppenheim, R. W. Schaffer, and J. R. Buck, *Discrete Time Signal Processing*, Prentice Hall, 2nd ed, Upper Saddle River, NJ: Prentice Hall, 1999.



Esther T. Olawole received her B.Eng. and M.Eng. degrees in Electrical and Electronics Engineering in 2010 and 2016, respectively, from University of Ilorin, Ilorin, Kwara State, Nigeria. She is a registered member of Council for the Regulation of Engineering in Nigeria (COREN) and also a Member of the Nigerian Society of Engineers (NSE). She is currently pursuing her Ph.D. degree in speech enhancement via noise reduction. Her research interest is on signal processing in wireless communication.

Damilare O. Akande obtained his B.Tech. and M.Eng. degrees in Electronic and Electrical Engineering from Ladoke Akintola University of Technology (LAUTECH), Ogbomoso and Federal University of Technology, Akure, Nigeria in 2008 and 2012, respectively. He obtained his Ph.D. degree from Universiti Sains Malaysia, Malaysia in 2019. He is a member of Nigeria Society of Engineers and a registered member of COREN. His research interest is on signal processing, cooperative communications and MAC layer design.

Zachaeus K. Adeyemo obtained the B.Eng. and M.Eng. degrees in Electrical Engineering from University of Ilorin, Ilorin, Nigeria and his Ph.D. degree in Electronic and Electrical Engineering in 2009 from Ladoke Akintola University of Technology (LAUTECH), Ogbomoso, Nigeria. Prof. Adeyemo is a professor of Communication Engineering at the Department of Electronic and Electrical Engineering, LAUTECH, Ogbomoso. He is a member of the IEEE and registered engineer, Council for the Regulation of Engineering in Nigeria (COREN). His research interest is on wireless communications.