

Research Article

Comparative Study between the Discrete-Frequency Kalman Filtering and the Discrete-Time Kalman Filtering with Application in Noise Reduction in Speech Signals

Leandro Aureliano da Silva ¹, Gilberto Arantes Carrijo,¹ Eduardo Silva Vasconcelos,¹ Roberto Duarte Campos,² Cleiton Silvano Goulart ² and Rodrigo Pinto Lemos³

¹Department of Electrical Engineering, Universidade Federal de Uberlândia, Av. João Naves de Ávila, 2160 Bloco 3N, Campus Santa Mônica, Uberlândia, MG, Brazil

²Department of Electrical Engineering, Faculdade de Talentos Humanos, R. Manoel Gonçalves de Rezende, 230 São Cristóvão, Uberaba, MG, Brazil

³Department of Electrical Engineering, Universidade Federal de Goiás, Av. Esperança, s/n. Campus Universitário, Goiânia, GO, Brazil

Correspondence should be addressed to Leandro Aureliano da Silva; leandro.aureliano@yahoo.com.br

Received 26 January 2018; Accepted 15 April 2018; Published 3 June 2018

Academic Editor: Shunyi Zhao

Copyright © 2018 Leandro Aureliano da Silva et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article aims to carry out a comparative study between discrete-time and discrete-frequency Kalman filters. In order to assess the performance of both methods for speech reconstruction, we measured the output segmental signal-to-noise ratio and the Itakura-Saito distance provided by each algorithm over 25 different voice signals. The results show that although the two algorithms performed very similarly regarding noise reduction, the discrete-time Kalman filter produced smaller spectral distortion on the estimated signals when compared with the discrete-frequency Kalman filter.

1. Introduction

Even with the advent of the Internet, voice transmission is still one of the most important ways of communication. The quality and intelligibility of speech signals play a key role in the ease and precision during information exchange. Practically in almost all voice transmission applications, the quality can be affected by factors such as ambient noise, losses due to digital link encoding, and interference from other conversations or even from other signal sources [1].

In order to overcome their harmful effects, digital speech processing techniques can be employed to reduce or even eliminate them. In recent years, some techniques and methods such as spectral subtraction, Kalman filtering, psychoacoustics, and wavelet transforms gained more prominence, especially in noise reduction, so that many research efforts have been made for improving them.

In [2, 3], the authors enhance speech quality by removing the musical noise introduced by spectral subtraction. In [1],

the authors combined spectral subtraction and wavelets on a prefiltering approach for noise reduction in speech signals and used the result as an initial guess for a Kalman filter. When compared to Kalman filtering using only wavelets or spectral subtraction alone to produce the initial guess, their method showed the least spectral distortion and a similar segmental output signal-to-noise ratio.

Since wavelet-based denoising is highly dependent on thresholding the approximation and detail coefficients, recent research in this area focuses on new thresholds [4, 5].

Shao and Chang [6] concatenated the Kalman filter to a bank of wavelet filters with a perceptual weighting filter. They used a technique of masking the psychoacoustic model to derive the weighting filter. According to the authors, that work brought two contributions. The first one was the wavelet-based auditory model with a perceptual wavelet filter bank that maps the frequency response of the human auditory system through subband decomposition. The second was the Kalman filter using a voice state space model in

the wavelet domain, whose computational cost was reduced when compared to the discrete-time Kalman filter. They were able to reduce the noise in different environments with low signal degradation.

Dhivya and Justin [7] proposed a noise reduction based technique that applies spectral subtraction to the wavelet approximation coefficients and soft thresholding to the detail coefficients. They used five wavelet filters and compared them according to their output signal-to-noise ratios. Besides the output SNR, they also considered the correlation coefficient and the perceptual evolution of speech quality (PESQ) criteria.

However, although these algorithms show significant advances in noise removal, most of them do not evaluate spectral distortion nor do they attempt to minimize it. So, since the method in [1] provided low spectral distortion, this article proposes a comparative study between discrete-time and discrete-frequency Kalman filters simply using the noisy signal as initial estimate. According to Fujimoto and Ariki [8], the main difference between the two approaches is that the operation of the Kalman filter is more computationally efficient in the frequency domain than in the time domain.

On the other hand, transforming the set of Kalman filter equations to/from the frequency domain produces a significant distortion in the estimated signal. Then, we used prefiltering based on spectral subtraction to reduce this distortion. In order to assess the performance of the proposed algorithms, we measured both the segmental signal-to-noise ratio of the outputs and the Itakura-Saito distance.

This article is structured as follows: Sections 2 and 3 describe the discrete-time and discrete-frequency Kalman filtering algorithms, respectively. Section 4 brings the experimental results and finally, in Section 5, the conclusions are presented.

2. Discrete-Time Kalman Filtering (DTKF)

In the 1960s, Rudolf Emil Kalman published the paper "A New Approach to Linear Filtering and Prediction Problems", describing a recursive solution to the discrete-time linear filtering problem [1]. Since then, due to the major advances of digital computing, Kalman filtering has become a very important technique in several areas such as navigation, monitoring processes, economics, and signal reconstruction from noisy samples.

In this article, the Kalman filtering development follows the heuristics described by Vaseghi [9]. Thus, the speech signal is modeled as an autoregressive process of order P , $AR(P)$, according to

$$x(n) = \sum_{k=1}^P a_p(k) x(n-k) + w(n) \quad (1)$$

where $a_p(k)$ are the linear prediction coefficients of order P , $w(n)$ is the prediction error associated with the excitation of the source-filter model of speech production, and $x(n)$ is the n th sample of the speech signal.

It can be observed that, in the acquisition process of audio and speech signals, most of the signals are captured in the

presence of some type of additive noise. Consequently, we can model the noisy signal as shown in

$$y(n) = x(n) + v(n) \quad (2)$$

where $y(n)$ is the noisy speech signal and $v(n)$ is a white Gaussian additive noise.

From (1) and (2), we can set up a state space model described by (3) and (4), respectively [9]:

$$\mathbf{x}(n) = \mathbf{A}(n-1) \mathbf{x}(n-1) + \mathbf{w}(n) \quad (3)$$

$$\mathbf{y}(n) = \mathbf{H}(n) \mathbf{x}(n) + \mathbf{v}(n) \quad (4)$$

where $\mathbf{x}(n)$ is the $P \times 1$ state vector at time n ; $\mathbf{A}(n-1)$ is the state transition matrix with dimensions $P \times P$ that relates current time n with past time $(n-1)$; $\mathbf{w}(n)$ is the $P \times 1$ input vector of the state equation and it is modeled as a white noise; $\mathbf{y}(n)$ is the $M \times 1$ observation vector; $\mathbf{H}(n)$ is the channel distortion matrix of dimensions $M \times P$; and $\mathbf{v}(n)$ is an $M \times 1$ additive white noise vector [9].

According to Vaseghi [9], $\mathbf{w}(n)$ and $\mathbf{v}(n)$ are assumed to be independent white noise processes so that

$$E[\mathbf{v}(n) \mathbf{v}^T(k)] = \mathbf{R}(n) = \begin{cases} r(n) \mathbf{I}_{M \times M} & k = n \\ 0 & k \neq n \end{cases} \quad (5)$$

$$E[\mathbf{w}(n) \mathbf{w}^T(k)] = \mathbf{Q}(n) = \begin{cases} q(n) \mathbf{I}_{P \times P} & k = n \\ 0 & k \neq n \end{cases} \quad (6)$$

where $\mathbf{R}(n)$ and $\mathbf{Q}(n)$ are diagonal covariance matrices, respectively, related to the additive noise and the prediction error.

The Kalman filtering estimates a process by using a kind of feedback control: first, the filter estimates the state of the process at a given time, then the feedback is obtained in the form of a new measurement.

Brown and Hwang [10] and Vaseghi [9] divided the Kalman filtering equations into two groups. The first ones are the time-update equations (prediction) and the second are the measurement-update equations (correction). Equation (7) describes the time-update:

$$\hat{\mathbf{x}}\left(\frac{n}{n-1}\right) = \mathbf{A}(n-1) \hat{\mathbf{x}}\left(\frac{n-1}{n-1}\right) \quad (7)$$

while measurement-update equations are shown in (8) and (9), respectively.

$$\mathbf{K}(n) = \mathbf{P}\left(\frac{n}{n-1}\right) \mathbf{H}^T(n) \times \left[\mathbf{H}(n) \mathbf{P}\left(\frac{n}{n-1}\right) \mathbf{H}^T(n) + \mathbf{R}(n) \right]^{-1} \quad (8)$$

$$\hat{\mathbf{x}}\left(\frac{n}{n}\right) = \hat{\mathbf{x}}\left(\frac{n}{n-1}\right) + \mathbf{K}(n) \left[\mathbf{y}(n) - \mathbf{H}(n) \hat{\mathbf{x}}\left(\frac{n}{n-1}\right) \right] \quad (9)$$

$$\mathbf{P}\left(\frac{n}{n}\right) = [\mathbf{I} - \mathbf{K}(n) \mathbf{H}(n)] \mathbf{P}\left(\frac{n}{n-1}\right) \quad (10)$$

where $\mathbf{P}(n/n)$ is the error covariance matrix at time n ; $\mathbf{K}(n)$ is the Kalman gain matrix, responsible for minimizing $\mathbf{P}(n/n)$; and $\hat{\mathbf{x}}(n/n)$ is the state estimate at time n , according to the previous observations of $\mathbf{y}(n)$.

3. Discrete-Frequency Kalman Filtering (DFKF)

Fujimoto and Ariki [8] introduced the discrete-frequency Kalman filtering (DFKF) in 2000 to provide more computationally efficient algorithm. This is accomplished by transforming the Kalman filter equations to be iterated in the frequency domain and then inverse transforming the estimated spectrum back to the time domain to find the estimated signal. In order to do so, they divide the frequency domain into multiple frames in such a way that the l th frame $X(k, l)$ is the complex spectrum of the noiseless signal $x(n, l)$ and $y(n, l)$ is the white Gaussian noise. Thus, the noise-corrupted signal $y(n, l)$ is given by the following equation [8]:

$$y(n, l) = x(n, l) + v(n, l) \quad (11)$$

Since $x(n, l)$ can be replaced by the Inverse Discrete Fourier Transform (IDFT) of $X(k, l)$, we have

$$y(n, l) = \sum_{k=0}^{N-1} X(k, l) \exp\left(j2\pi \frac{kn}{N}\right) + v(n, l) \quad (12)$$

In matrix notation, (12) can be represented as shown in

$$\mathbf{y}(n, l) = \underbrace{\begin{pmatrix} 1 \\ \exp\left(j2\pi \frac{n}{N}\right) \\ \vdots \\ \exp\left(j2\pi \frac{(N-1)n}{N}\right) \end{pmatrix}^T}_{\mathbf{F}_n} \underbrace{\begin{pmatrix} X(0, l) \\ X(1, l) \\ \vdots \\ X(N-1, l) \end{pmatrix}}_{\mathbf{X}_l} + \mathbf{v}(n, l) \quad (13)$$

that can be simply written as

$$\mathbf{y}(n, l) = \mathbf{F}_n \mathbf{X}_l + \mathbf{v}(n, l) \quad (14)$$

where n represents time within l th frame, N is the number of samples in the frame, and \mathbf{F}_n is the $N \times 1$ vector containing the basis of the Discrete Fourier Transform (DFT). \mathbf{X}_l is the complex spectrum vector of the l th frame. Since time n has no meaning for \mathbf{X}_l , there is no state transition matrix in the Kalman equations for the frequency domain, so that the computational effort of the DFKF is significantly reduced.

Analogous to the DTKF, the DFKF can be represented by the following equations:

$$\mathbf{K}_{(n,l)} = \mathbf{P}_{(n-1,l)} \mathbf{F}_n^H \left[\mathbf{F}_n \mathbf{P}_{(n-1,l)} \mathbf{F}_n^H + \mathbf{R}_{(n,l)} \right]^{-1} \quad (15)$$

$$\hat{\mathbf{X}}_{(n,l)} = \hat{\mathbf{X}}_{(n-1,l)} + \mathbf{K}_{(n,l)} \left[\mathbf{y}(n, l) - \mathbf{F}_n \hat{\mathbf{X}}_{(n-1,l)} \right] \quad (16)$$

$$\mathbf{P}_{(n,l)} = \mathbf{P}_{(n-1,l)} - \left[\mathbf{K}_{(n,l)} \mathbf{F}_n \mathbf{P}_{(n-1,l)} \right] \quad (17)$$

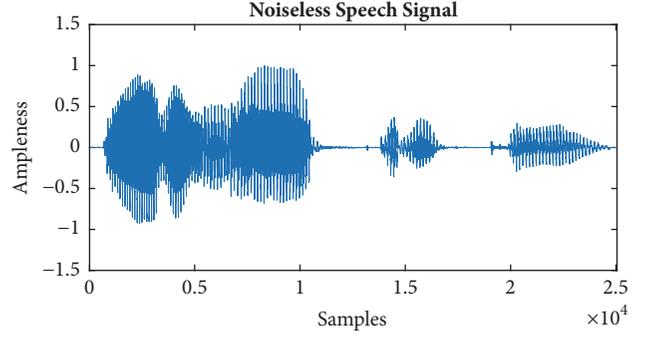


FIGURE 1: Noiseless signal used for comparison with the estimated signal.

where $(\cdot)^H$ means the complex conjugate transpose of a matrix.

In order to obtain the estimated signal of the Kalman filter in the time domain, we must apply the Inverse Discrete Fourier Transform (IDFT) on (16).

4. Results

In order to compare the performances of the studied techniques, we used 25 different recorded speech signals sampled at 22050 Hz and coded with 16 bits per sample. Each signal was windowed by a Hamming window of size 512 with 50% overlap. All tests were performed using Matlab R2013B on a Core i7 processor computer with 8 GB RAM.

The quality of the estimated speech signal in the output of each filter was evaluated using the segmental signal-to-noise ratio (SNRseg). We have chosen the SNRseg because it can be calculated over short segments of the speech signal, in order to balance the weights assigned to each segment of higher or lower signal strength. SNRseg is given by [11]

$$\text{SNRseg} = \frac{10}{M} \sum_{j=0}^{M-1} \log_{10} \left[\frac{\sum_{n=mj-N+1}^{mj} x^2(n)}{\sum_{n=mj-N+1}^{mj} [x(n) - \hat{x}(n)]^2} \right] \quad (18)$$

where mj are the limits of each one of the M frames of length N . To carry out the tests, the signals were contaminated by additive white noise and the input segmental signal-to-noise ratio (SNRI) was adjusted to 3 dB.

As reported by Rabiner and Schafer [12], a suitable way to measure spectrum variations is the Itakura-Saito distance. Such measure can be calculated as

$$d(\mathbf{b}, \mathbf{a}) = \log \left[\frac{\mathbf{bRb}^T}{\mathbf{aRa}^T} \right] \quad (19)$$

where \mathbf{a} and \mathbf{b} are the linear prediction coefficients (LPC) vectors of the original and estimated signals, respectively, and \mathbf{R} is the autocorrelation matrix of the original signal. The closer to each other the spectra of the original and estimated signals, the smaller $d(\mathbf{b}, \mathbf{a})$. Thus, an Itakura-Saito distance equal to zero indicates that the spectra are the equal [12].

The DTKF algorithm was employed in the first test, which used the utterance elétrica (electrical in Portuguese). The results are shown in Figures 1, 2, and 3, respectively.

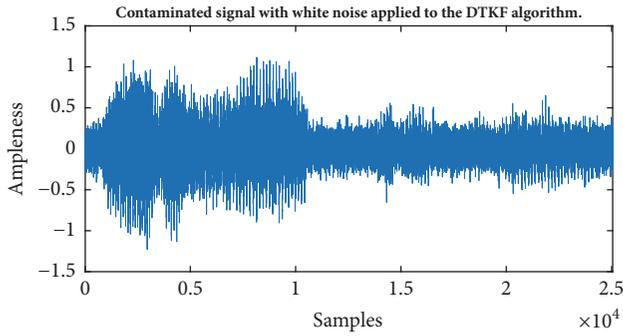


FIGURE 2: Contaminated signal with white noise applied to the DTKF algorithm.

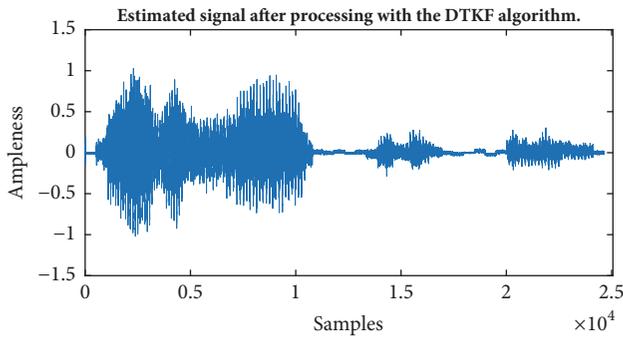


FIGURE 3: Estimated signal after processing with the DTKF algorithm.

Figures 2 and 3 evidence the noise reduction, especially during the silence parts of the signal. The SNRO in this case was 10 dB and the Itakura-Saito distance was 0.3250.

The second test preserved the same parameters of the first test except for the use of DFKF. The results are shown in Figures 4 and 5, respectively. The SNRO was 8 dB and the comparison of Figures 4 and 5 shows a considerable reduction in the noise. However, the Itakura-Saito distance was 0.3782, which indicates a larger distortion in the filtering.

Therefore, the DTKF algorithm produced smaller spectral distortion than the DFKF but provided a larger SNRO.

The results of the tests for the 25 words are presented in Figures 6 and 7. Figure 6 shows that the SNRO in targeted tests was almost always the same for DTKF and DFKF, with an average of 9 dB.

Figure 7 shows that the DTKF algorithm produced smaller signal distortion for all tests. Thus, we can affirm that the DTKF is more suitable than the DFKF for speech processing.

Tests were also performed after prefiltering the noisy signals. The prefiltering was based on spectral subtraction like in [1]. All results showed that the DTKF produced smaller spectral distortion than DFKF. The spectral distortions for the 25 words are shown in Figure 8 for an SNRI of 3 dB.

The comparison of Figures 7 and 8 indicates that prefiltering allowed only a tiny improvement in the reduction of spectral distortion provided by the DTKF algorithm.

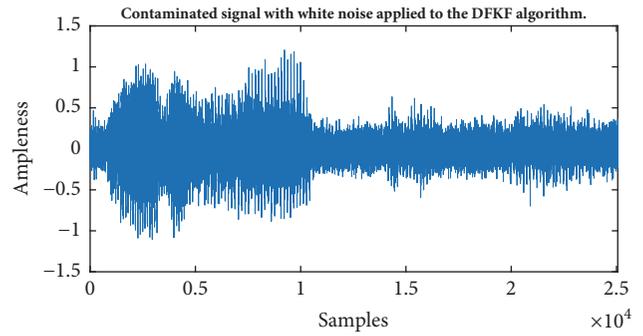


FIGURE 4: Contaminated signal with white noise applied to the algorithm DFKF.

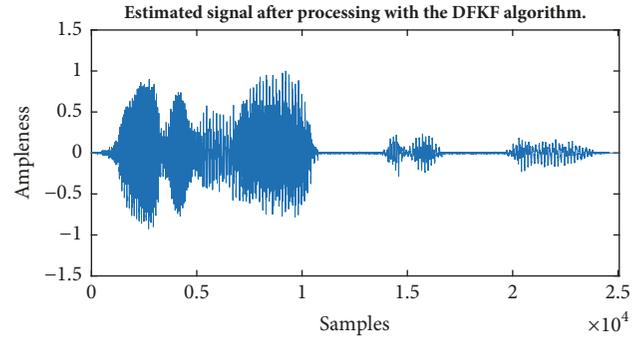


FIGURE 5: Estimated signal after processing with the DFKF algorithm.

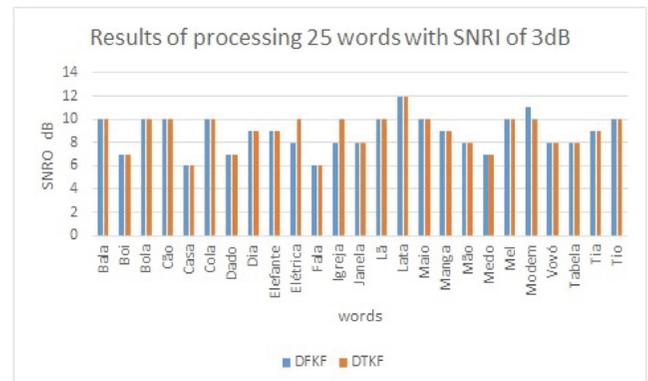


FIGURE 6: Comparison for segmental signal-to-noise ratio output (SNRO) with 25 words contaminated by white noise with signal-to-noise ratio input (SNRI) of 3 dB.

5. Conclusions

This paper presented a comparative study between discrete-time and discrete-frequency Kalman filtering algorithms. Tests were carried out with 25 different words using Itakura-Saito distance to measure the spectral distortion and the segmental signal-to-noise ratio to evaluate the noise reduction.

Although the two algorithms performed very similarly regarding noise reduction, discrete-time Kalman filtering produced smaller spectral distortion on the estimated signals for all targeted tests. This shows that discrete-time Kalman

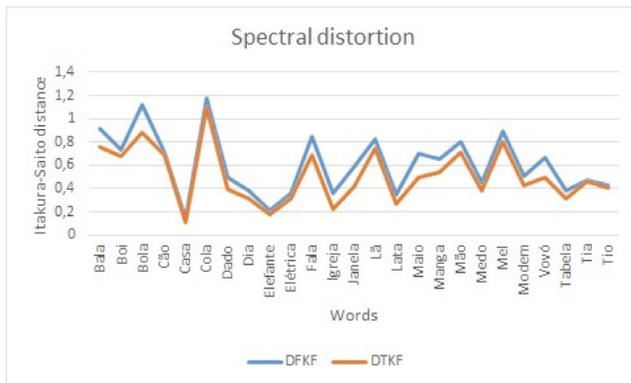


FIGURE 7: Comparison for spectral distortion for 25 words contaminated by white noise with signal-to-noise ratio input (SNRI) of 3 dB.

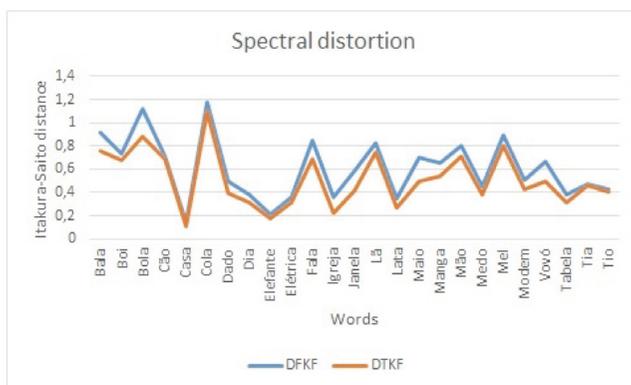


FIGURE 8: Comparison for spectral distortion for 25 words contaminated by white noise with signal-to-noise ratio input (SNRI) of 3 dB, using spectral subtraction with prefiltering of the contaminated signal.

filtering is more suitable than discrete-frequency Kalman filtering for the reconstruction of speech signals corrupted by additive white noise.

Data Availability

The voice data (.wav files) used to support the results of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] L. A. da Silva and M. B. Joaquim, "Redução de Ruído em Sinais de Voz Usando, Filtros de Kalman de Tempo e Freqüência Discretos, Combinados com Subtração Espectral de Potência e/ou Wavelets," in *Proceedings of the XXV Simpósio Brasileiro de Telecomunicações-SBrT*, pp. 3–6, 2007.
- [2] C. Lu, K. Tseng, and C. Chen, "Reduction of Musical Residual Noise Using Hybrid Median Filter," in *Proceedings of the 2012 Spring Congress on Engineering and Technology (S-CET)*, pp. 1–4, Xi'an, China, May 2012.
- [3] R. Miyazaki, H. Saruwatari, T. Inoue, Y. Takahashi, K. Shikano, and K. Kondo, "Musical-noise-free speech enhancement based on optimized iterative spectral subtraction," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 20, no. 7, pp. 2080–2094, 2012.
- [4] R. Li, C. Bao, B. Xia, and M. Jia, "Speech enhancement using the combination of adaptive wavelet threshold and spectral subtraction based on wavelet packet decomposition," in *Proceedings of the 2012 11th International Conference on Signal Processing, ICSP 2012*, pp. 481–484, chn, October 2012.
- [5] R. Aggarwal, J. Karan Singh, V. Kumar Gupta, S. Rathore, M. Tiwari, and A. Khare, "Noise Reduction of Speech Signal using Wavelet Transform with Modified Universal Threshold," *International Journal of Computer Applications*, vol. 20, no. 5, pp. 14–19, 2011.
- [6] Y. Shao and C.-H. Chang, "A Kalman filter based on wavelet filter-bank and psychoacoustic modeling for speech enhancement," in *Proceedings of the ISCAS 2006: 2006 IEEE International Symposium on Circuits and Systems*, pp. 121–124, grc, May 2006.
- [7] R. Dhivya and J. Justin, "A novel speech enhancement technique," *International Journal of Research in Engineering and Technology*, vol. 3, no. 19, pp. 98–102, 2014.
- [8] M. Fujimoto and Y. Ariki, "Noisy speech recognition using noise reduction method based on Kalman filter," in *Proceedings of the 25th IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2000*, vol. 3, pp. 1727–1730, June 2000.
- [9] S. V. Vaseghi, *Advanced Digital Signal Processing and Noise Reduction*, John Wiley & Sons, 2008.
- [10] R. G. Brown and P. Y. C. Hwang, *Introduction to random signals and applied Kalman filtering: with MATLAB exercises and solutions*, Wiley, New York, NY, USA, 1997.
- [11] J. R. Deller, J. G. Proakis, and J. H. Hansen, *Discrete Time Processing of Speech Signals*, Prentice Hall PTR, 1993.
- [12] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, Prentice Hall, 1978.



Hindawi

Submit your manuscripts at
www.hindawi.com

