REVIEWER 2

This paper presents a speech enhancement method for secure communication by combining Wiener filter and spectral subtraction for noise estimation. I think this paper requires major revisions. My comments are given below:

1. (Line 131 in page 4) The statement “… clean signal y(t)” is not correct. It should be “x(t)”.

We are sorry for our mistake and thank you for pointing it out. We have corrected it.

2. (Page 4) The expression of eq. (6) is not adequate. X(m,k) should be the estimated spectrum of clean speech. A hat for the symbol X(m,k) is required.

Thank you very much for your correction. We have modified it accordingly

\[ | \hat{X}(m,k) |^2 = H(m - 1,k) | Y(m,k) |^2. \]  \hspace{1cm} (6)

3. (Page 5) The marked area should be presented by a dotted line in Fig. 1.

Thank you very much for the suggestions. We have revised it.

4. (Line 1 after Fig. 1 of Page 5) The statement “where m is the frequency index and k is the frame index.” is not correct. It should be revised as “where m and k denote the frame and frequency indices, respectively.”.

Thank you very much for pointing it out. We have modified it accordingly and move that explanation to Section II as it is first mentioned there.

where m and k denote the frame and frequency indices respectively.

5. (Page 5) The expression of eq. (8) is wrong.

We are very sorry for the mistakes. We have corrected it.

\[ H(m,k) = \left( \frac{| \hat{D}(m - 1,k) |^2}{| \hat{X}(m - 1,k) |^2 + | \hat{D}(m - 1,k) |^2} \right)^{0.5}. \]  \hspace{1cm} (8)
6. (Page 6) The expression of eq. (11) is not correct. The value of q for the second condition is missed.

\[ q = \begin{cases} 
1 & \text{if } \Psi_n(m,k) \geq 20 \text{ dB} \\
-0.038\Psi_n(m,k) + 1.8 & \text{if } -5 \text{ dB} \leq \Psi(m,k) < 20 \text{ dB} \\
1.88 & \text{if } \Psi(m,k) < -5 \text{ dB}. 
\end{cases} \]  

7. (Page 6) The authors utilized posteriori SNR given in eqs. (13) and (14) to estimate the priori SNR psi given in eq. (12). I think this method is not good. Please give an utterance as an example to present the performance by providing true priori SNR and estimated one.

We admit that the method for SNR estimation very simple and may not very good. But we chose it since it is very simple and computationally fast. We could not provide the actual SNR estimation as the distortion is actually cause by encryption process. But given the knowledge of actual clean speech, we extract actual noise (similar to how we get the results in Fig. 5) and calculate the SNR between the clean speech and the noise. We compare that with SNR estimation given in Eq. (12)-(14). We add the results in our manuscripts and modify it.

The computation for calculating the posterior SNR is simple and fairly fast. We evaluate the performance of these formulations by comparing their estimation with the actual SNR. The actual SNR estimation can be calculated since we have access to the clean speech. The sample result is illustrated in Fig. 2. As can be see, the SNR estimator may not perform very well. This is as expected as the distortions is highly non-stationary.

![Figure 2. Comparisons of actual SNR with estimated SNR](image)
8. (Page 10) It is not available to distinguish which compared method is better by observing Fig. 5. Please try to provide another example by using an utterance which contain speech-pause regions.

We are sorry for unclear figures. We modify the figure by only use two examples of encoding schemes: PCM and IMA-ADPCM to make the figures larger. We also use colored line. We add the spectrum of the clean speech to identify when is the speech and non-speech period. We revise the figure as follows:

![Figure 6](image)

**Figure 6.** Comparisons of the resulting noise estimation spectra of the proposed method method and several noise estimators: Martin, Hirsch, and IMCRA. The figures show the magnitude spectra (in dB) of a frequency bin (k=10) when we use $\epsilon = 0.8$.

9. (Line 1 after eq.(13) in page 6) The symbol $\Phi_n$ is a typo.

We are sorry for the mistake. We have corrected it.
The notation $\phi(m,k)$ is the noisy signal-to-noise ratio, i.e.:

10. (Line 2 after eq. (9) in page 5) The expression of mid$Y(m,k)$ is unclear. I think it is a typo.

*We are sorry for the typo. We have corrected it.*

11. The presentation of the caption for each sub-figure in Figs. 6-8 is not correct. Please prepare them according the authors’ guide of the journal.

*Thank you very much for the suggestion. We have put the caption for sub-figure on the top of the figure.*

12. (Line 58 in page 2) The abbreviation SS is re-defined. It has been defined at line 54.

*We are sorry for the duplication. We remove the definition in line 58*

*However, SS requires good noise estimation or otherwise some artifacts*

13. Sections 2 and 3 should be combined into one section.

*Thank you very much for the suggestion. We combine Section 2 and 3 and rename it itu Speech Enhancement Methods.*

14. Sections 5 and 6 also can be combined into one section.

*Thank you very much for the suggestion. However, as the results and discussion sections are quite large, we chose to keep both sections in separate section to make it easier for readers to follow the flow of the discussions.*

15. (Line 118 in page 4) The sentence from “Usually, …” should be changed to next paragraph.

*Thank you for the suggestion. We have made it into a new paragraph.*
16. (Page 7) The font size in Fig. 2 should be smaller.

Thank you very much for the suggestion. We have revised the figure.

17. The sub-grid lines in Tables 1-3 should be removed.

Thank you for your suggestion. We have removed the sub-grid lines in our tables.

18. The presentation of caption for Table 2 is too redundant. Some statements can be moved to text. Please revise it.

Thank you for the suggestion. The caption is now as the following:

”The average LSD (in dB) of the actual noise and estimated noise of our method and other noise estimators for various communication channels.”

19. The English usage is not satisfied. It is strongly recommended to be proofread by a native English speaker to improve the written quality.

A. (Line 94 in page 3) The word during is repeated.

Thank you very much for pointing it out. We have revised it.

B. (Line 3 of page 4)”… with limited window’s length.” can be revised as “…” with limited window length.”.

Thank you very much for pointing it out. We have revised it.

C. (Lines 125 to 126 in page 4) The descriptions “Usually, they are determined heuristically based on the estimate of the signal-to-noise ratio (SNR). Their examples are [19,21,22].” should be revised as “They are determined heuristically based on the estimate of the signal-to-noise ratio (SNR) [19,21,22].”

Thank you very much for the suggestion. We have rewritten the sentences accordingly.
D. (The last line of page 4)”…fourier transform…” should be revised as ”…Fourier transform…”.

*Thank you for pointing out our mistake. We have corrected it.*

E. (Line 1 after Fig. 1 of Page 5)”… is fed into the our noise estimator…” should be revised as “”… is fed into the proposed noise estimator…”.

*Thank you for pointing out our mistake. We have corrected it.*

F. (Line 206 in page 9)The statement “Figure 5 compares the noise estimates of our method with other noise estimators:” should be revised as “Figure 5 shows the comparisons of the noise estimates for our method with other noise estimators:”.

*Thank you very much for the suggestions. We have modified it accordingly.*