34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Evaluation of a New Approach for Speech Enhancement Algorithms in Hearing Aids

Vahid Montazeri, Soudeh A. Khoubrouy, Issa M.S. Panahi, Senior *Member, IEEE*

Abstract— **Several studies on hearing impaired people who use hearing aid reveal that speech enhancement algorithms implemented in hearing aids improve listening comfort. However, these algorithms do not improve speech intelligibility too much and in many cases they decrease the speech intelligibility, both in hearing-impaired and in normally hearing people. In fact, current approaches for development of the speech enhancement algorithms (e.g. minimum mean square error (MMSE)) are not optimal for intelligibility improvement. Some recent studies investigated the effect of different distortions on the enhanced speech and realized that by controlling the amplification distortion, the intelligibility improves dramatically. In this paper, we examined, subjectively and objectively, the effects of amplification distortion on the speech enhanced by two algorithms in three background noises at different SNR levels.**

I. Introduction

Due to hearing aid limitations in terms of power and size, single microphone speech enhancement algorithms seem to be the best candidate among all categories of the speech enhancement methods. Although, these algorithms improve ease of listening for the users, they do not improve speech intelligibility adequately and in some cases they degrade intelligibility [1]. As a result, the designers have to use two or more microphones for speech enhancement in hearing aids which cause more power consumption. Some recent studies focused on the intelligibility problem and discussed several factors that cause lack of intelligibility in current single microphone enhancement algorithms. One of the major factors is that these algorithms pay no attention to the difference between attenuation distortion and amplification distortion [1] and assume, incorrectly, that they have the same effect on the speech. For example, most of current enhancement algorithms work based on the minimization of mean square error (MSE) between the enhanced and clean signal without differentiating between the negative and positive errors which are equivalent with attenuation distortion and amplification distortion respectively. We will investigate the effect of these two different distortions in this paper. Here we use a parameter (defined later) which has a high correlation with both speech intelligibility and speech quality [12]. Different values of this parameter demonstrate the amount of attenuation and amplification distortions in the enhanced speech. Loizou and Kim in [1] showed that this parameter is equivalent to the Articulation Index (AI). We name this parameter as Signal to Residual Spectrum Ratio [1] or SR. Assuming $\mathrm{X}(k), \mathrm{D}(k)$ and $\hat{\mathrm{X}}(k)$ are clean magnitude spectrum, noise magnitude spectrum, and estimated magnitude spectrum at frequency bin k respectively, SR is defined as:

$$\mathrm{SR}(k) = \frac{\mathrm{SNR}(k)}{(\sqrt{\mathrm{SNR}(k)} - \sqrt{\mathrm{SNR_{ENH}}(k)})^2} \quad (1)$$

where $SNR(k) \triangleq X^2(k)/\mathrm{D}^2(k)$ is the true instantaneous SNR at frequency bin k and $SNR_{ENH}(k) \triangleq \hat{X}^2(k)/\mathrm{D}^2(k)$ is the enhanced SNR. Fig. 4 plots the SR as a function of $\mathrm{SNR_{ENH}}$ for $\mathrm{SNR} = 0\mathrm{dB}$. Based on this figure, we can distinguish two different regions with positive and negative SR values: Region 1 in which the SR is positive and region 2 in which the SR is negative. It can be shown that region 1 implies $\hat{X}(k) \leq 2X(k)$ which is attenuation distortion and amplification distortion up to 6.02 dB and region 2 implies $\hat{X}(k) > 2X(k)$ which is amplification distortion greater than 6.02 dB [1]. From Fig. 4 it can be understood that for a maximum value of SR and consequently a maximum speech intelligibility and quality, $\hat{X}(k)$ should be in the region 1 for all values of k or equivalently:

$$E\left\{\hat{X}(k) | \hat{X}(k) > 2X(k)\right\} = 0 \quad (2)$$

Using this constraint, we examined two speech enhancement algorithms: Logmmse which is one of the most well-known enhancement algorithms [7] and Spectral Subtraction which is one of the traditional speech enhancement algorithms [6]. The performance of each algorithm is examined by three different background noises: train noise, restaurant noise, and destroyer engine noise with SNR levels of $+5$, 0, and -5 dB. As will be seen, removing the data in the region 2 improves the performance of the both algorithms dramatically.

II. Evaluation Procedure and Results

In our experiments, the noisy signal with the sampling frequency of $8KHz$ is divided into 20 ms frames (with 50% overlap between frames) and then processed with one of the enhancement algorithms. After processing, the estimated magnitude spectrum $\hat{X}(k)$ is then compared to the clean magnitude spectrum $X(k)$ (we assumed knowledge of the clean magnitude spectrum). Next, the modified magnitude spectrum is computed based on the following equation:

$$X_R(k) = \begin{cases} \hat{X}(k), & \text{if } \hat{\mathrm{X}}(k) \leq 2X(k) \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

Vahid Montazeri is with the University of Texas at Dallas, Richardson, TX, 75080, USA. Email: vahid.montazeri@utdallas.edu

Soudeh A. Khoubrouy is with the University of Texas at Dallas, Richardson, TX, 75080, USA. Email: sa.khoubrouy@utdallas.edu

Issa M.S. Panahi is with the University of Texas at Dallas, Richardson, TX, 75080, USA. Email: issa.panahi@udallas.edu

Table 1 shows the average percentage of the frequency bins of the estimated signal ($\hat{X}(k)$) falling in region 1 and 2, the mean square error for the frequency bins in the region (MSE1), and also the mean square error for the frequency bins in the region 2 (MSE2). Moreover, Fig. 1 to Fig. 3 show the spectrograms of the clean signal (X), noisy signal at $SNR = -5\ dB$, estimated signal ($\hat{X}$), and the modified signal(X_R) for three different background noises.

For subjective evaluation, five normal hearing volunteers participated in the listening experiment. They were asked to write the sentence they hear in each of these conditions: two algorithms, for each algorithm three background noises and for each background noise three SNR levels. Each participant listened to the sentences by a stereo headphone at appropriate volume level. 10 sentences were chosen from IEEE database [13]. Table 2 shows the results of the experiments in terms of the percentage of the words identified correctly.

A comparison between the spectrogram of $\hat{X}$ and X_R in the figures reveals that the modified speech is much nearer to the clean speech especially in consonants onsets, offsets,

TABLE 1

Algorithm	Noise Type	SNR	Region 1	Region 2	MSE1	MSE2
Logmmse	Train	5 dB	48.05%	51.95%	0.0207	0.0065
		0 dB	44.32%	55.68%	0.0299	0.0125
		-5dB	37.77%	62.23%	0.0340	0.0186
	Restaurant	5 dB	56.79%	43.21%	0.0171	0.0210
		0 dB	63.06%	36.94%	0.0495	0.0222
		-5dB	39.93%	60.07%	0.0572	0.0321
	Destroyer Engine	5 dB	57.93%	42.07%	0.0297	0.0034
		0 dB	46.55%	53.45%	0.0533	0.0126
		-5dB	43.14%	56.86%	0.0601	0.0184
Spectral Subtraction	Train	5 dB	69.38%	30.62%	0.0042	0.0048
		0 dB	60.64%	39.36%	0.0144	0.0141
		-5dB	57.60%	42.40%	0.0172	0.0223
	Restaurant	5 dB	76.88%	23.12%	0.0043	0.0092
		0 dB	73.79%	26.21%	0.0187	0.0249
		-5dB	70.77%	29.23%	0.0277	0.0362
	Destroyer Engine	5 dB	70.30%	29.70%	0.0366	0.0009
		0 dB	59.11%	40.89%	0.0439	0.0037
		-5 dB	55.99%	44.01%	0.0468	0.0056

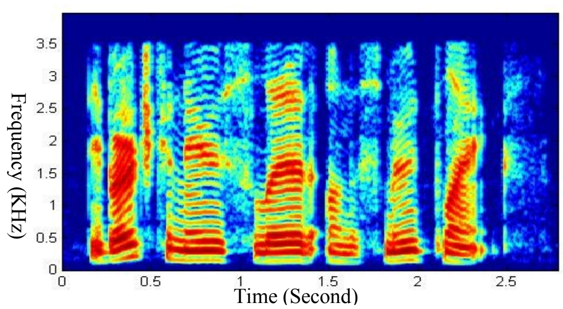

Figure 1.a

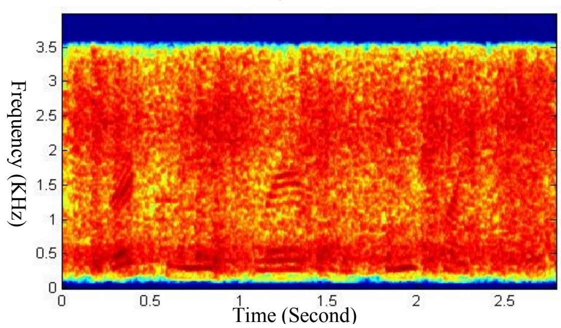

Figure 1.b

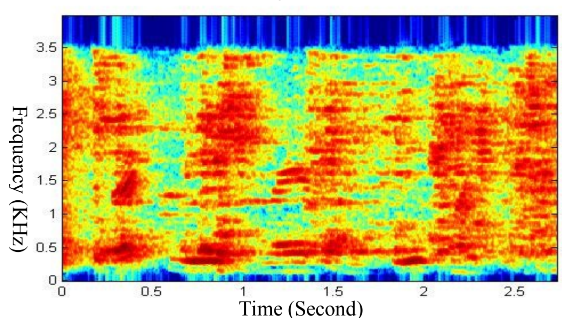

Figure 1.c

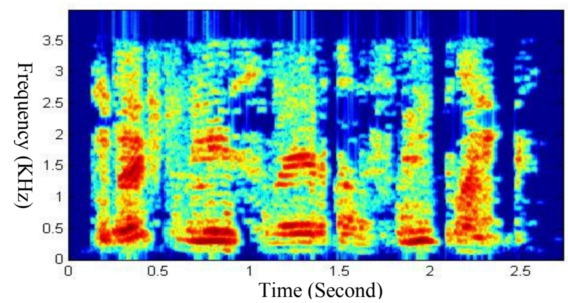

Figure 1.d

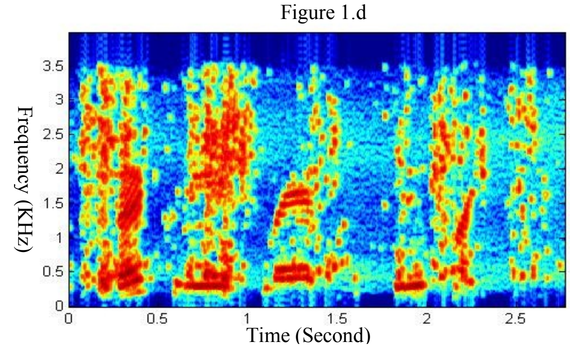

Figure 1.e

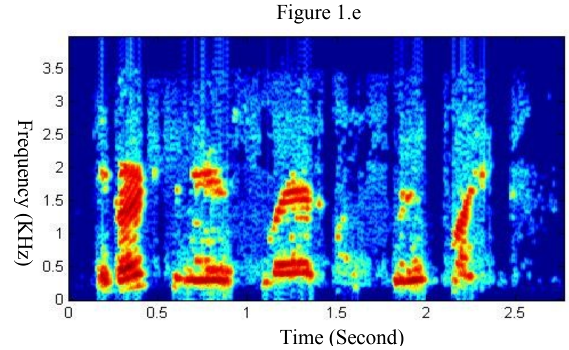

Figure 1.f

Figure 1. The spectrogram for -5dB speech corrupted by train noise, Figure (1.a) shows the clean speech, Figure (1.b) shows the noisy speech, Figure (1.c) shows the speech enhanced by Logmmse method ($\hat{X}$), Figure (1.d) shows the Logmmse modified speech (*X_R*), Figure (1.e) shows the speech enhanced by spectral subtraction method ($\hat{X}$), Figure (1.f) shows the Spectral Subtraction modified speech (*X_R*).

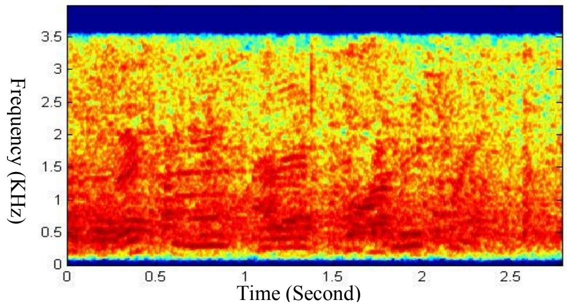

Figure 2.a

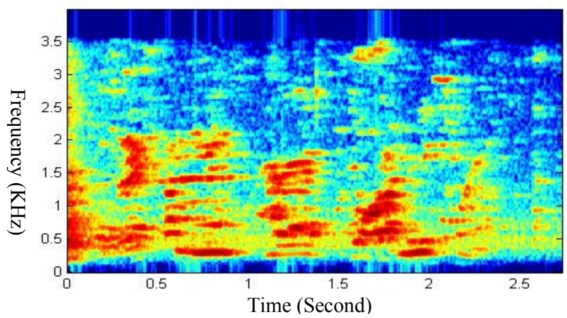

Figure 2.b

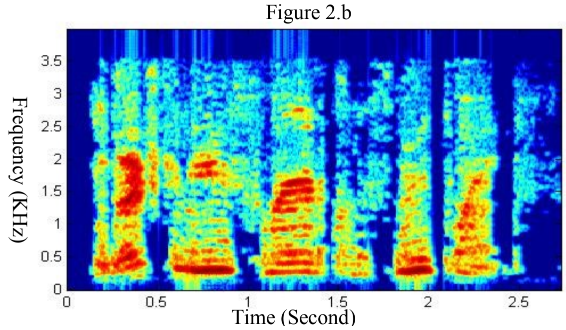

Figure 2.c

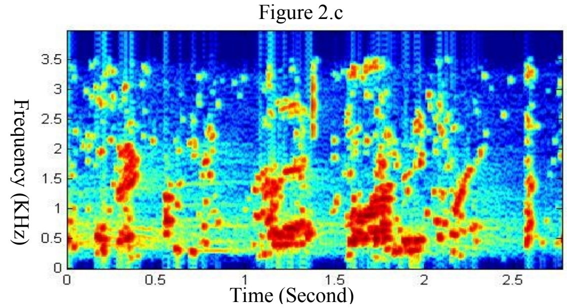

Figure 2.d

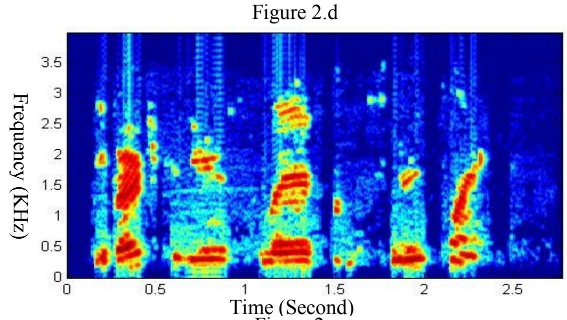

Figure 2.e

Figure 2. The spectrogram for -5dB speech corrupted by restaurant noise, Figure (2.a) shows the noisy speech, Figure (2.b) shows the speech enhanced by Logmmse method ($\hat{X}$), Figure (2.c) shows the Logmmse modified speech (X_R), Figure (2.d) shows the speech enhanced by spectral subtraction method ($\hat{X}$), Figure (2.e) shows Spectral Subtraction modified speech (X_R). (The clean signal is the same as in Fig. 1)

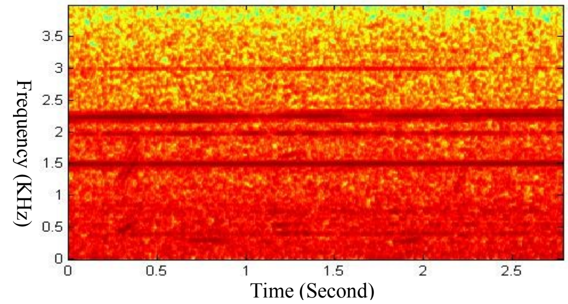

Figure 3.a

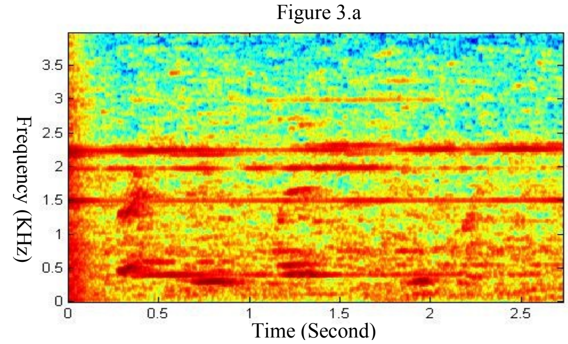

Figure 3.b

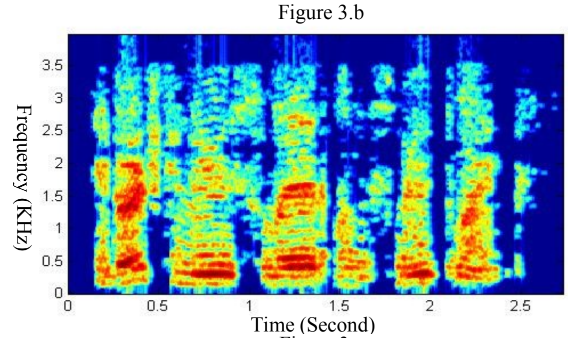

Figure 3.c

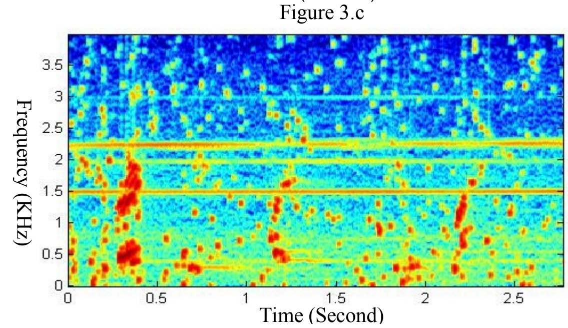

Figure 3.d

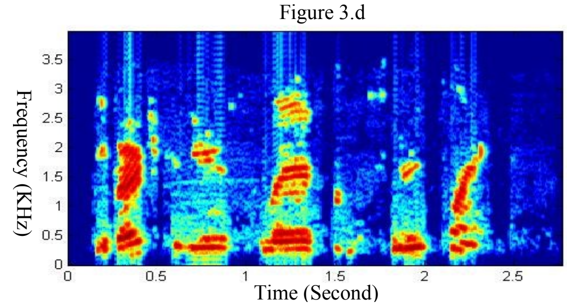

Figure 3.e

Figure 3. The spectrogram for -5dB speech corrupted by destroyer engine noise, Figure (3.a) shows the noisy speech, Figure (3.b) shows the speech enhanced by Logmmse method ($\hat{X}$), Figure (3.c) shows the Logmmse modified speech (X_R), Figure (3.d) shows the speech enhanced by spectral subtraction method ($\hat{X}$), Figure (3.e) shows Spectral Subtraction modified speech (X_R). (The clean signal is the same as in

and also in non-speech periods. Although the knowledge of the clean magnitude spectrum is important in obtaining the clean X_R, the main reason is removing data fall in the region 2. This proves that the negative and positive difference between clean and estimated signal do not have equal effect on the speech. In general, some types of distortions (errors) need to be treated differently. These highly detrimental distortions are greater than 6.02 dB and occur in the region 2 (see Fig. 4). Although the mean square error for the data in the region 2 may be very low, the data in this region should to be removed. One example may give a better understanding: based on table 1, using Spectral Subtraction algorithm, the mean square error for the destroyer engine noise in 5dB is 0.0366 in region 1 and 0.0009 in region 2. However, when we zeroed all of the data in region 2 based on (3) (this means losing information) and played the modified speech for the participants, the intelligibility increased from 20.14% to 88.83% based on the information in table 2. (Also see Fig. (3.d) and Fig. (3.e)) Similar observations exist in other noise types and SNRs for both algorithms.

III. Conclusion

An important factor for the improvement of intelligibility in speech enhancement algorithms was experimented objectively and subjectively with two algorithms in three background noises at different SNR levels. The results verify the fact that in order to improve the intelligibility in the speech enhancement algorithms, the amplification distortion (or positive errors in MSE based algorithms) should to be controlled. This important factor may be used in order to obtain more efficient speech enhancement algorithms with higher performance in terms of quality and intelligibility.

TABLE 2

Algorithm	Noise Type	SNR Level	Without Constraint (3)	With constraint (3)
Logmmse	**Train**	**5 dB**	**70.36%**	**97.29%**
		0 dB	**50.39%**	**92.54%**
		-5dB	**12.50%**	**90.37%**
	Restaurant	**5 dB**	**61.99%**	**98.65%**
		0 dB	**52.97%**	**93.57%**
		-5 dB	**14.41%**	**91.26%**
	Destroyer Engine	**5 dB**	**67.81%**	**95.49%**
		0 dB	**33.12%**	**91.25%**
		-5 dB	**10.03%**	**89.11%**
Spectral Subtraction	**Train**	**5 dB**	**61.45%**	**96.34%**
		0 dB	**37.25%**	**89.31%**
		-5 dB	**10.45%**	**86.43%**
	Restaurant	**5 dB**	**31.87%**	**91.25%**
		0 dB	**7.19%**	**84.19%**
		-5 dB	**5.61%**	**81.12%**
	Destroyer Engine	**5 dB**	**20.14%**	**88.83%**
		0 dB	**6.65%**	**85.36%**
		-5 dB	**3.24%**	**77.29%**

Acknowledgment

The authors appreciate the following people Mr. / Mrs. Abel Foster, Joey Britt, Mohamed Good, Niku Roknabadi, and Dillon Bird, who participated voluntarily in the hearing test on March, 11, 2012.

References

[1] P.C. Loizou, G. Kim, "Reasons why current speech-enhancement algorithms do not improve speech intelligibility and suggested solutions" *IEEE Trans. Audio, Speech and Language Processing,* Vol. 10, no. 1, pp. 47-56, Jan. 2011.

[2] K. Y. Lee, S. Jung, "Time-domain approach using multiple Kalman filters and EM algorithm to speech enhancement with nonstationary noise", *IEEE Trans. Speech and Audio Processing.* Vol. 8, no. 3, pp. 282-291, May. 2000.

[3] Y. Hu, P.C. Loizou, "A comparative intelligibility study of single-microphone noise reduction algorithms", *Conf. Rec. 2007 ICASSP Acoustics, Speech and Signal Processing*, pp. 561-564.

[4] J. Hao, T.W. Lee, T.J. Sejnowski, "Speech Enhancement Using Gaussian Scale Mixture Models", *IEEE Trans, Audio, Speech, and Language Processing,* Vol. 18, no. 6, pp. 1127-1136, Aug. 2010.

[5] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction", *IEEE Trans. Acoustic, Speech, and Signal Processing,* Vol. ASSP-27, pp. 113-120, Apr. 1979.

[6] B. L. Sim, Y.C. Tong, J. S. Chang, C. T. Tan, "A parametric formulation of the generalized spectral subtraction method", *IEEE Trans. Speech and Audio Processing,* Vol. 6, no. 4, pp. 328-337, July 1998.

[7] Y. Ehpraim, D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator", *IEEE Trans. Acoustics, Speech, and Signal Processing,* Vol. ASSP-33, no. 2, pp. 443-445, Apr. 1985.

[8] P.C. Loizou, "Speech enhancement Theory and practice", New York, CRC Press, Jun, 2007.

[9] J. G. Proakis, D.G. Manolakis, "Digital Signal Processing", *4th ed.* Englewood, New Jersey, Prentice-Hall, 2007.

[10] H. Stark, J. W. Woods, "Probability and Random Processes with application to signal processing," *3rd ed.* New Jersey, Prentice-Hall, 2002.

[11] Y. Hu, P.C. Loizou, "Subjective comparison of speech enhancement algorithms", *Conf. Rec. 2006 ICASSP Conf. Acoustics, Speech and Signal Processing,* pp. 153-156.

[12] J. Tribolet, P. Noll, B. McDermott, and R.E. Crochier, "A study of complexity and quality of speech waveform coders", in *Proc. IEEE int. Conf. Acoustic, Speech, and Signal Processing,* pp.586-590, 1978.

[13] "IEEE recommended practice for speech quality measurement," *IEEE Trans., Audio Electroacoustic,* Vol. AU-17, no. 3, pp. 225-246, Sep 1969.

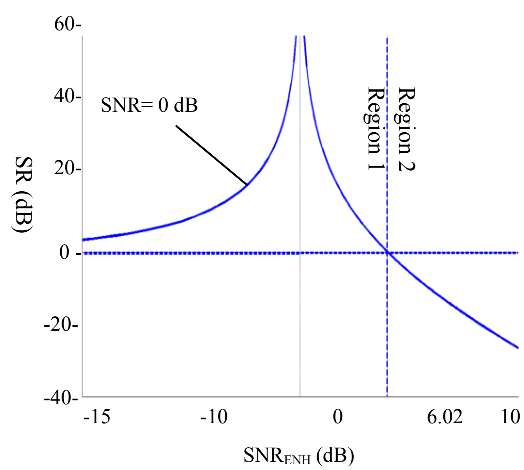

Figure 4