SPEECH ENHANCEMENT EMPLOYING LOUDNESS SUBTRACTION AND OVER-SUBTRACTION

Wei Zhang and Tyseer Aboulnasr
School of Information Technology and Engineering, University of Ottawa, K1N 6N5, Ottawa, ON, Canada
Email:zhangw,aboulnas@site.uottawa.ca

1. INTRODUCTION

In classical speech enhancement algorithms, the focus is on the removal of additive noise. The simplest type of speech enhancement is based on spectral subtraction. The estimate of the average noise spectrum is subtracted from the noisy speech spectrum to give the enhanced speech spectrum. This model often leads to large residual noise and musical artifacts. It has been discussed extensively and summarized with a generalized form in [1].

While reducing the power of the noise in the signal results in SNR improvement, it is known that the correlation between the subjective quality of the speech and SNR is very weak [2]. Better measures to evaluate the subjective quality of speech are designed considering the human hearing system. For example, loudness can better simulate the subjective quality of the speech and was used in the ITU-T P.862 recommendation [3]. A loudness subtraction method has been presented in [4] and we will derive a more statistically precise model based on the Laplacian speech model [5].

For the spectral subtraction approach, the remaining noise after the enhancement is still high when the SNR is low. The spectral over-subtraction method has been proposed to provide further improvement [6]. It implements a SNR dependent subtraction factor which applies a higher subtraction factor in the lower SNR frames and vice versa. We will extend the over-subtraction approach to the proposed approach in loudness domain.

This paper is organized as follows. The speech enhancement approaches based on the loudness subtraction and the proposed over-subtraction model are discussed in Section 2 and 3, respectively. In Section 4, simulation results are given. Conclusions are given in Section 5.

2. LOUDNESS SUBTRACTION MODEL

In the PESQ measure, the relative quality of a speech signal depends on the difference in the loudness domain between the signal and the reference (clean) speech signal. This suggests implementing a speech enhancement in the loudness domain. For simplicity, we will use the power law of loudness: $N' = C \cdot I^{\alpha}$, there N' is the specific loudness in a frequency band, I is the physical intensity. C and α are constants. In this paper, we assume α = 0.27.

A generalized loudness subtraction algorithm is proposed as follows:

$$\hat{X}^2 = \left(\left(Y^2\right)^{\alpha} - a\left(Y^2\right)^{\alpha}\right)^{\frac{1}{\alpha}}, \tag{1}$$

where a is defined as the subtraction factor. To maintain the loudness of the reconstructed speech at the same levels as the original clean speech, we can show that

$$a = E\left\{Y^{2\alpha} - X^{2\alpha}\right\} / E\left\{N^{2\alpha}\right\} \tag{2}$$

Define the Noisy-Signal-to-Noise Ratio (NSNR) as: *NSNR* $= E\left\{Y^2\right\}/E\left\{N^2\right\}$.The parameter a is determined in terms of the measurable signals as follows. Two independent sources are used to simulate the speech and noise in the frequency domain. The clean speech X is assumed to have Laplacian distribution, while the noise N is Gaussian [5]. Both sources are zero mean and with variances determined by the specified NSNR. The resulting selection of a is shown in Figure 1 (the solid line) as a function of the NSNR.

3. PROPOSED LOUDNESS OVER-SUBTRACTION MODEL

The approach in the previous section subtracts a portion of the average loudness of the noise from the noisy speech signal. Given that the noise is random, the actual loudness of the noise will not be the same for all the samples. There will always be fluctuations around average which will lead to large noise residues in the enhanced signal.

In this paper, we focus on removing an over-estimate of the noise from the noisy signal in the loudness domain. This can be done through increasing the value of a. As a increases, more noise will be deducted but the enhanced speech will likely be more distorted. If a is chosen carefully, the improvement of the SNR will compensate for the larger distortion of the speech.

Using the general nature of the spectral over-subtraction factor in [6], we adjusted a to optimize the performance in the algorithm in loudness domain. This over-subtraction scaling factor is chosen to be 5 below 0dB NSNR, 1 beyond 20dB NSNR and changing linearly between 0 and 20 dB. This scaling factor will be multiplied by a selected in the last section using (2) before being used in (1). The resulting subtraction factor is depicted in Figure 1 with the dotted line.

4. SIMULATIONS

In this section, the subtraction and over-subtraction approaches are compared in both loudness and spectral domains. White stationary noise is added to the clean speech signal with different SNRs from 0 dB to 20dB. The FFT of the noisy speech is obtained on a frame-by-frame basis. The phase of the FFT is maintained separately for reconstruction. The loudness of the noise is estimated with the use of a noise memory [5]. The spectrum of enhanced speech components is estimated using (1). The subtraction factor of subtraction and over-subtraction in loudness domain is shown in Figure 1. The approach in [4] is a different loudness subtraction algorithm with a = 1. The noisy speech signal is enhanced with the spectral subtraction and over-subtraction, loudness subtraction (both the algorithm in [4] and the proposed algorithm) and over-subtraction.

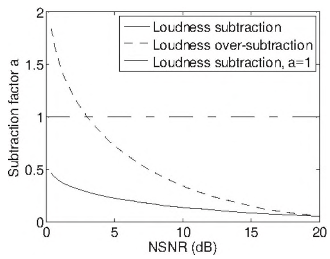

Figure 1: The subtraction factors for loudness domain approaches

The enhanced speech is then reconstructed with the original phase and inverse FFT. The PESQ, a popular objective measure of speech subjective quality, is shown in Table 1 for enhanced signals. The results confirm that using the over-subtraction of noise in an adaptive way in the loudness domain leads to the highest PESQ scores. Similar tests have been done with the segmental SNR and Log-Likelihood Ratio (LLR) measures showing that the loudness over-subtraction continues to provide an improvement. For the LLR measures, the loudness over-subtraction has slightly higher distortion than the loudness subtraction. Both methods in the loudness domain have lower distortion than the corresponding methods in spectral domain.

5. CONCLUSIONS

In this paper, we presented speech enhancement approaches based on loudness subtraction and over-subtraction. The subtraction factor is selected adaptively for each frame based on the NSNR of the speech signal. These approaches in the loudness domain result in improved Segmental SNR, improved PESQ scores and less distortion compared to the corresponding algorithms in the spectral domain.

REFERENCES

[1] B. L. Sim, Y. C. Tong, J. S. Chan and T. T. Chin, "A parametric formulation of the generalized spectral subtraction method," *IEEE Transactions on Speech and Audio Processing,* Vol. 6, No. 4, pp. 328-337, July, 1998.

[2] S. Quackenbush, T. Barnwell, and M. Clements, *Objective Measures of Speech Quality,* Englewood Cliffs, NJ, Prentice Hall, 1988.

[3] A.W. Rix, J. G. Beerends, M. P. Hollier and A. P. Hekstra, " Perceptual evaluation of speech quality (PESQ)- a new method for speech quality assessment of telephone networks and codecs," *Proceedings of ICASSP*, Vol. 2, pp. 749-752, 7-11 May, 2001.

[4] T. Petersen and S. Boll, "Acoustic noise suppression in the context of a perceptual model," *Proceedings of ICASSP*, Vol. 6, pp. 1086-1088, April, 1981.

[5] S. Gazor and W. Zhang, "Speech probability distribution," *IEEE Signal Processing Letters*, Vol. 10, No. 7, pp. 204-207, July 2003.

[6] M. Berouti, R. Schwartz and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," *Proceedings of ICASSP,* Vol. 4, pp. 208-211 April, 1979.

Table 1: Comparison of PESQ scores for spectral subtraction, spectral over-subtraction, loudness subtraction, loudness over-subtraction and the approach in [4].

Speech Type	Input SNR (dB)	PESQ score: Input Signal	Spectral Subtraction	Approach In [4]	Loudness Subtraction	Spectral over-Subtraction[6]	Loudness Over-Subtraction
Female	0	1.122	1.299	1.573	1.420	1.581	1.812
	10	1.819	2.334	2.076	2.209	2.345	2.418
	20	2.637	2.764	2.823	2.924	2.920	2.976
Male	0	1.390	1.461	1.575	1.478	1.573	1.696
	10	1.839	2.205	2.005	2.050	2.222	2.298
	20	2.488	2.581	2.657	2.701	2.742	2.762