FREQUENCY DOMAIN NOISE SUPPRESSION APPROACHES IN MOBILE TELEPHONE SYSTEMS

Jin Yang*

SHARP Microelectronics Technology, Inc., Camas, WA 98607

ABSTRACT

Frequency domain noise suppression systems with a single microphone are studied for adverse mobile noise environments. A new noise suppression algorithm based on a modified maximum likelihood estimate is developed. The algorithm takes into account not only minimum voice distortion but also subjective criteria for the noise naturalness. A speech detector based on the temporal signal-to-noise ratio information of every subband of a bandpass filter bank is proposed. The speech detector is proven to be very effective and robust in a rapidly changing noisy background. The proposed noise suppression method in the mobile telephone systems demonstrates much better subjective and objective test results than the existing methods.

1. INTRODUCTION

Background acoustic noise represents one of the major impairments in mobile voice communications, especially in hands-free mobile telephone systems. Since the designs of voice coders and voice recognition units for digital mobile telephones assume high SNR values (>12dB), low SNR will deteriorate the performance dramatically.

Frequency domain noise suppression is carried out by first decomposing the corrupted speech signal into different frequency subbands. The noise power of each subband is then estimated during non-voiced periods. Noise suppression is achieved through the use of the suppression factor corresponding to the temporal signal power over estimated noise power ratio of each subband. Figure 1 shows the noise suppression system configuration using a bandpass filter bank with each band orthogonal to another.

A comprehensive overview of various noise suppression techniques can be found in J.S. Lim and A.V. Oppenheim [1]. Existing methods include power estimation [1], maximum likelihood estimate [1], Wiener estimation [1], magnitude estimation [1][2], soft-decision method [3], and minimum mean-squared error estimate [4]. However those methods typically either do not provide sufficient noise reduction ratio in pure noise section or do not take into account the subjective criteria. Thus in situations where the SNR varies rapidly, noise flutter and significant speech distortion may occur.

The performance of the noise suppression system depends on the accuracy and robustness of the speech detector. Existing speech detector designs utilize various combinations of energy estimation [3], zero-crossing rate [5], correlation function, LPC coefficient and signal power change ratio [5]. However, in very noisy environments, speech detectors based on these design approaches may suffer serious performance degradation.

In this paper, a noise suppression system using a modified maximum likelihood estimate is developed in Section 2. Section 3 presents a novel speech detector. Section 4 shows the results of the proposed noise suppression system with applications to mobile telephone systems.

2. NOISE SUPPRESSION APPROACHES

Frequency domain noise suppression assumes that noise characteristics change slower than that of voice. Therefore, it is possible to estimate the noise spectrum during non-voiced periods and reduce the noise content of sampled signals.

We assume that the corrupted speech signal $\tilde{x}(i)$ has spectrum $x(i,w)$ in the following form,

$$x(i,w) = A(i,w)e^{j\phi(i,w)} + n(i,w), \qquad (1)$$

where i represents the i-th moment; w is the frequency; $A(i,w)$, $\phi(i,w)$ represent the amplitude and phase of the voice, respectively;and $n(i,w)$ denotes the noise spectrum.

The perception of speech is insensitive to phase $\phi(i,w)$ [2], therefore the problem of extricating speech signal spectrum from corrupted signal spectrum $x(i,w)$ can be simplified to estimating $A(i,w)$, denoted as $\hat{A}(i,w)$. Moreover, since in noise suppression system, a detector had been used to determine whether the given signal consists of noise only or speech plus noise, a binary hypothesis model is appropriate. Thus, we have,

$$\begin{aligned} H_0&: \; x(i,w) = n(i,w) \\ H_1&: \; x(i,w) = A(i,w)e^{j\phi(i,w)} + n(i,w). \end{aligned} \qquad (2)$$

* This work was carried out when the author was with NovAtel Comm. Ltd., Calgary, Canada T2K 7V8.

0-7803-0946-4/93 $3.00 © 1993 IEEE

Authorized licensed use limited to: National Tsing Hua Univ.. Downloaded on June 13,2020 at 07:30:56 UTC from IEEE Xplore. Restrictions apply.

The minimum mean-squared spectral error estimate $\hat{A}(i,w)$ is the conditional mean

$$\hat{A}(i,w) = E\{A(i,w)|X(i,w)\}, \tag{3}$$

where, $X(i,w) = |x(i,w)|$, i.e.,

$$\hat{A}(i,w) = E\{A(i,w)|X(i,w),H_1\}p(H_1|X(i,w)) + E\{A(i,w)|X(i,w),H_0\}p(H_0|X(i,w)). \tag{4}$$

When voice is absent, the maximum noise suppression factor $G_{\min}$ determined by the subjective criteria for the noise naturalness shall be applied as,

$$E\{A(i,w)|X(i,w),H_0\} = G_{\min} X(i,w). \tag{5}$$

When voice is present, since the maximum likelihood estimator is asymptotically efficient for large SNR, it suffices to replace $E[A(i,w)|X(i,w),H_1]$ with

$$E\{A|X,H_1\} = \frac{1}{2}\left[X + \sqrt{X^2 - \sigma}\right], \tag{6}$$

where $\sigma = E\{|n(i,w)|^2\}$ is the estimate of noise power .

Using Bayes rule for $p(H_1|X(i,w))$ [3], we have

$$P(H_1|X(i,w)) = \frac{e^{-\eta} I_0[2\sqrt{\eta \frac{X(i,w)^2}{\sigma(i,w)}}]}{1 + e^{-\eta} I_0[2\sqrt{\eta \frac{X(i,w)^2}{\sigma(i,w)}}]}, \tag{7}$$

where $\eta = A^2 / \sigma$ is the a priori signal - to - noise ratio.

Thus the modified maximum likelihood estimate is

$$\hat{A}(i,w) = \frac{1}{2}\left[X + \sqrt{X^2 - \sigma}\right]P(H_1|X) + G_{\min} X[1 - p(H_1|X)]. \tag{8}$$

Appending the phase of the input signal to the estimated envelope, we have the noise-reduced signal spectrum $y(i,w)$ as,

$$y(i,w) = \hat{A}(i,w)\frac{x(i,w)}{|x(i,w)|}, \tag{9}$$

i.e., $y(i,w) = G(i,w)x(i,w)$, (10)

where

$$G(i,w) = \left[\frac{1}{2} + \frac{1}{2}\sqrt{\frac{X(i,w)^2 - \sigma(i,w)}{X(i,w)^2}}\right]P(H_1|X(i,w)) + G_{\min}(1 - p[H_1|X(i,w)]). \tag{11}$$

Then, two steps are required for noise suppression:

1. Estimate the noise power spectrum $\sigma(i,w)$ during non-voiced periods,

$$\sigma(i,w) = E\{|n(i,w)|^2\}. \tag{12}$$

2. Reduce noise using suppression factor $G(i,w)$. The real time output signal $\tilde{y}(i)$ after noise reduction with an orthogonal bandpass filter bank is,

$$\tilde{y}(i) = \sum_{k=1}^{K} G(i,w_k)x(i,w_k), \tag{13}$$

where K is the number of subbands.

Table 1 shows various existing noise suppression methods which can be represented by different noise suppression functions $G(i,w)$. Note that $G(i,w)$ is a function of X^2/σ, i.e., temporal SNR of the input signal. The relations between noise suppression factors and input signal temporal SNR are depicted in Figure 2. Ideally, when input signal SNR is high, it is most likely that speech is present and the maximum likelihood estimate should be used. When original input SNR is low, it most likely corresponds to noise alone and the maximum suppression factor should be applied in order to get a quiet natural noise background. As can be seen from the curves of Figure 2 and Equation (11), the proposed modified maximum likelihood method approaches the maximum likelihood estimate in the high SNR section and approaches the maximum noise suppression factor in the low SNR section.

Table 1. Various Existing Algorithms and Related Noise Suppression Functions

Algorithm	Formula G(i,w)
Spectral subtraction	$\sqrt{\frac{X(i,w)^2 - \rho\,\sigma(i,w)}{X(i,w)^2}}$ $\rho=1$ *for power subtraction method*
Magnitude subtraction	$1 - \sqrt{\frac{\sigma(i,w)}{X(i,w)^2}}$
Maximum likelihood	$\frac{1}{2}\left[1 + \sqrt{\frac{X(i,w)^2 - \sigma(i,w)}{X(i,w)^2}}\right]$
Wiener	$\frac{X(i,w)^2 - \sigma(i,w)}{X(i,w)^2}$
Soft decision scheme	$\frac{1}{2}\left[1 + \sqrt{\frac{X(i,w)^2 - \sigma(i,w)}{X(i,w)^2}}\right]P(H_1\|X)$ *$P(H_1/X)$ as defined in Equation (7)*

3. SPEECH DETECTOR

The performance of the noise suppression system is based upon the accuracy of the background noise estimate $\sigma(i,w)$. The noise power estimate is updated during the time when only background noise is present. Therefore, an effective and robust speech detector plays an important role in noise suppression system.

Traditionally, the fixed power energy threshold method and various modified technologies are effective for distinguishing the vowels and high SNR signal from

Authorized licensed use limited to: National Tsing Hua Univ.. Downloaded on June 13,2020 at 07:30:56 UTC from IEEE Xplore. Restrictions apply.

noise background. However these approaches are not sufficient in detecting the unvoiced or low SNR signal, especially in a non-stationary and rapidly changing noise background, such as mobile communications situations.

Mobile noise is typically a pink noise. Unvoiced sound usually has high frequency components, and will be covered by the strong low frequency noise in mobile situation. Noise suppression method based on bandpass filter bank provides a unique way to distinguish the speech pause by evaluating the signal-to-noise factors in different subbands. Since the high frequency component of noise is also relatively small, the strong high frequency unvoiced sound can be easily detected.

We proposed a voice detection criterion which is based on a comparison of the mean of the SNR factors in divided subbands with a predetermined SNR factor threshold, denoted as SNR_f_threshold, i.e.,

$$\frac{1}{K}\sum_{k=1}^{K} SNR\_f_k \begin{cases} \geq SNR\_f\_threshold \\ \quad \textit{Detect as Voice Section} \\ < SNR\_f\_threshold \\ \quad \textit{Detect as Noise Section} \end{cases} , \quad (14)$$

where SNR_f_k is the SNR factor of the k-th subband as

$$SNR\_f_k = \max\left\{ \frac{X(i,w_k)^2 - \sigma(i,w_k)}{X(i,w_k)^2}, 0 \right\}. \quad (15)$$

The continuously updated noise power estimate σ is used, because the noise power is changing with various conditions, such as by-pass car, head-on traffic, etc.

This speech detector is proven to be very effective in the adverse noise environment compared with the adaptive frame energy method, where the unvoiced or nasal sounds have been clipped frequently. Figure 3 shows the speech detector outputs for the case of the vehicle moving on the highway at a speed of 100km/h with windows open.

4. IMPLEMENTATION AND CONCLUSIONS

The schematic diagram for the proposed noise suppression system is shown in Figure 1. An orthogonal filter bank is used to achieve simple and distortionless reconstruction.

The overlapping of voice and noise frequency bands causes difficulty in reducing noise without distorting voice, therefore both objective and subjective criteria are used for measuring the performance of the noise reduction algorithms. The segmental noise power reduction ratio is used to measure the noise reduction level during the pure noise section, which is defined as,

$$NR_{seg} = \frac{1}{L}\sum_{l=1}^{L} 10 \log_{10}\left(\frac{N_{pre\_power}}{N_{out\_power}}\right), \quad (16)$$

where L is the number of divided noise section segments; N_{pre_power} and N_{out_power} are noise power before noise reduction and after noise reduction, respectively.

The segmental SNR improvement is defined as the SNR difference between the SNR after noise reduction and the original input signal SNR, i.e.,

$$SNR_{seg} = \frac{1}{M}\sum_{m=1}^{M} 10\left[\log_{10}\left(\frac{S_{out\_power}}{N_{out\_power}}\right) - \log_{10}\left(\frac{S_{pre\_power}}{N_{pre\_power}}\right)\right], \quad (17)$$

where M is the number of divided voice section segments; S_{pre_power} and S_{out_power} are the signal power before noise reduction and after noise reduction, respectively.

The proposed noise suppression system for a mobile environment demonstrates much better subjective and objective test results than the existing methods. A real-time test shows that 10dB noise reduction, and 9dB SNR improvements have been achieved even in the most adverse mobile situation. The original signals as well as the corresponding noise suppressed signals are shown in Figure 4. Informal subjective tests show excellent intelligibility and voice quality.

REFERENCES

[1] J. S. Lim, and A. V. Oppenheim, "Enhancement and Bandwidth Compression of Noise Speech," *Proc. IEEE*, vol. 67, pp. 1586-1604, Dec. 1979.

[2] N. D. Degan, and C. Prati, "Acoustic Noise Analysis and Speech Enhancement Technique for Mobile Radio Applications," *Signal Processing 15*, pp.43-56, 1988.

[3] R. J. McAulay, and M. L. Malpass, "Speech Enhancement Using a Soft-Decision Noise Suppression Filter," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 28, pp. 137-145, April 1980.

[4] Y. Ephraim, and D. Malah, "Speech Enhancement Using a Minimum Mean Square Error Short-Time Spectral Amplitude Estimator," *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. 32, pp.1109-1120, 1984.

[5] W. Chan, and D. D. Faconer, "Speech Detection for a Voice/Data Mobile Radio Terminal," *Proc. of IEEE Conf. on Comm.*, pp.F8.7.1-F8.7.5, June 1983.

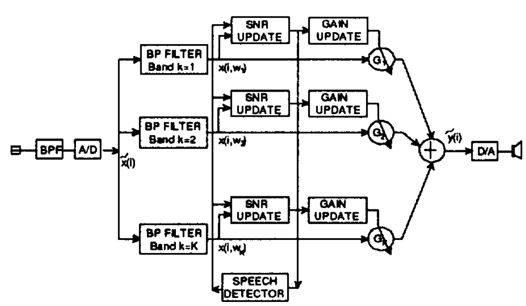
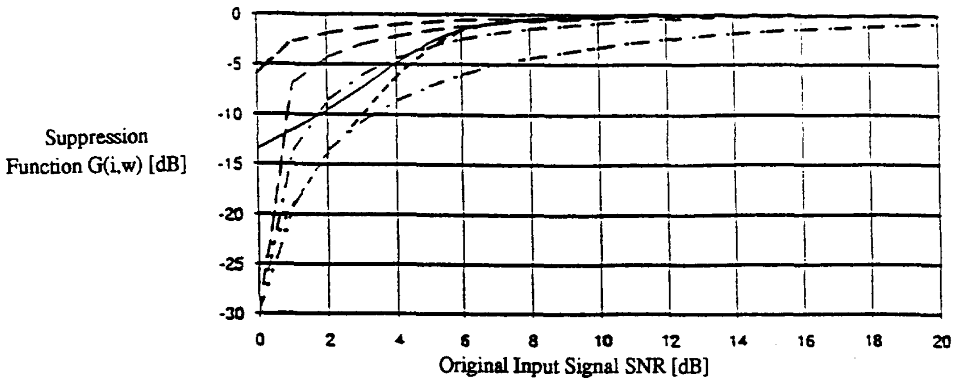


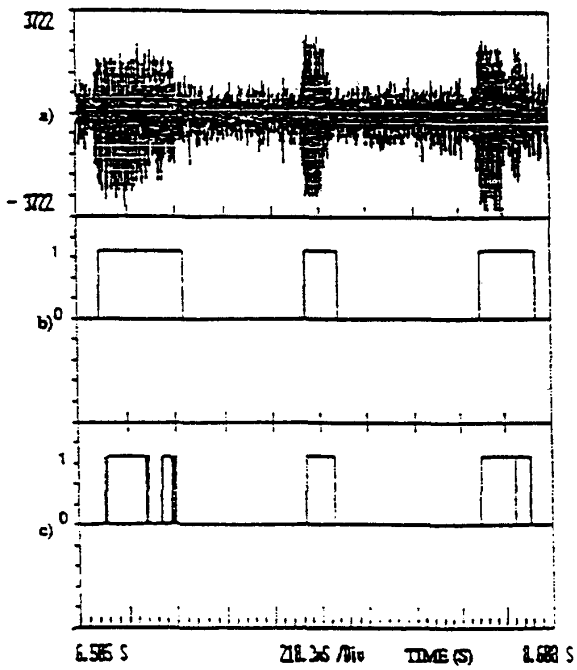
FIGURE 1. Noise Suppression System Configuration

Authorized licensed use limited to: National Tsing Hua Univ.. Downloaded on June 13,2020 at 07:30:56 UTC from IEEE Xplore. Restrictions apply.



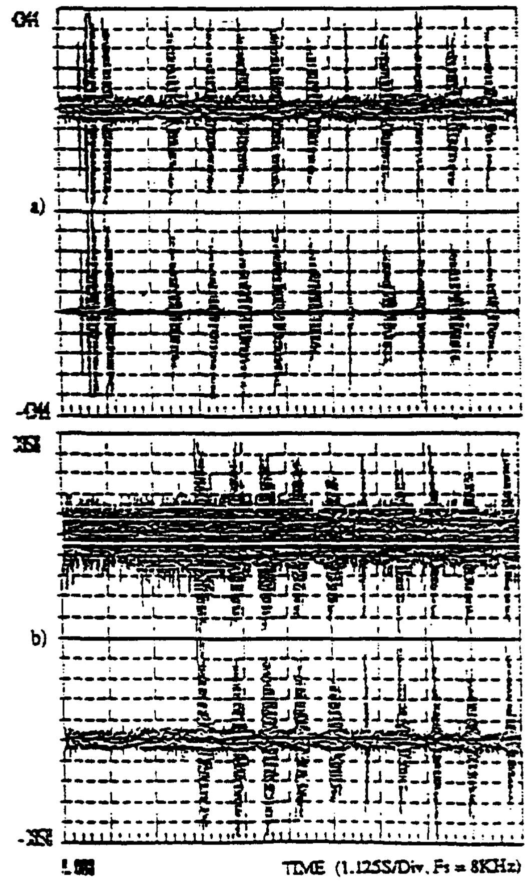
— — Maximum Likelihood — — Power Sub. - - - - Soft Decision: $\eta = 5$
- - — Magnitude Sub. - - — Wiener ——— Mod_ML: $\eta = 5$, $G_{min}=-15$

FIGURE 2. Suppression Function G(i,w) vs Temporal SNR of Various Noise Suppression Algorithms



a) Input Signal: Picked Up by Microphone at Driver's Side Visor, Driving on highway with windows open
b) Speech Detector Output Using the Proposed Method
c) Speech Detector Output Using Adaptive Frame Energy

FIGURE 3. Speech Detector Outputs



a) Driving on Highway with Windows Closed
b) Driving on Highway with Windows Open
Up, Noise Corrupted Signal; Down: Noise Reduced Signal

FIGURE 4. Signals of The Noise Suppression System

Authorized licensed use limited to: National Tsing Hua Univ.. Downloaded on June 13,2020 at 07:30:56 UTC from IEEE Xplore. Restrictions apply.