A REAL-TIME NOISE SUPPRESSION FILTER
FOR SPEECH ENHANCEMENT AND ROBUST CHANNEL VOCODING

Robert J. McAulay and Marilyn L. Malpass

M.I.T. Lincoln Laboratory
Lexington, Massachusetts

ABSTRACT

A real time implementation of a system for enhancement of speech in additive acoustic noise is described. The technique used is to perform a spectral decomposition of noisy speech via channel vocoder filter bank analysis and to attenuate each spectral component depending on how much the measured speech plus noise power exceeds an estimate of the background noise. A two state model for the speech event (speech absent or speech present) is applied in determining the maximum likelihood estimator of the speech power. This model has resulted in a new class of suppression curves which permits a tradeoff of noise suppression against speech distortion. Experiments utilizing the real time implementation have shown that the noise can be made imperceptible by proper choice of the suppression factor. Integration of the noise suppression filter into the analysis section of a narrowband vocoder is described. This combined system represents an integrated robust vocoder structure where acoustic noise can be suppressed prior to pitch estimation and determination of the modulator gains.

INTRODUCTION

Many of the practical environments in which digital voice terminals are to be deployed, such as the airborne command post or the cockpits of jet fighter aircraft and helicopters, are characterized by a high ambient noise level which reduces the quality and intelligibility of encoded speech [1] and causes listener fatigue. Various noise suppression prefilters have been developed which attenuate each spectral line depending on how much the measured speech plus noise power exceeds an estimate of the background noise power [2-6]. A new class of suppression curves has been derived which incorporates soft-suppression into a maximum likelihood estimate of the speech envelope [7] and permits a one parameter tradeoff of speech distortion versus noise suppression. In this paper a real time implementation of a noise suppression filter which can be used as a prefilter to a narrowband vocoder is described. Application of this suppression filter as a preprocessor for a 2400 bps Linear Predictive Coding (LPC) algorithm is discussed in [7]. It is also shown here that the structure can be incorporated into the analysis section of a 2400 bps channel vocoder so that noise can be suppressed prior to pitch estimation and determination of the modulator gains.

IMPLEMENTATION

Noise suppression prefilters that have been reported on to date have generally been implemented in the frequency domain, exploiting the properties of the FFT for filtering by circular convolution. Since efficient hardware implementations of the channel vocoder are being developed using both digital and analog sampled-data techniques [8,9], it seemed appropriate to attempt a time domain implementation of a noise suppression filter that could exploit these emerging technologies. As in the channel vocoder, a bank of second order Butterworth bandpass filters is used to span the frequency range 120-3270 Hz (the sampling rate was 7575 Hz). These filters are used in the generalized prefilter configuration shown in Fig. 1.

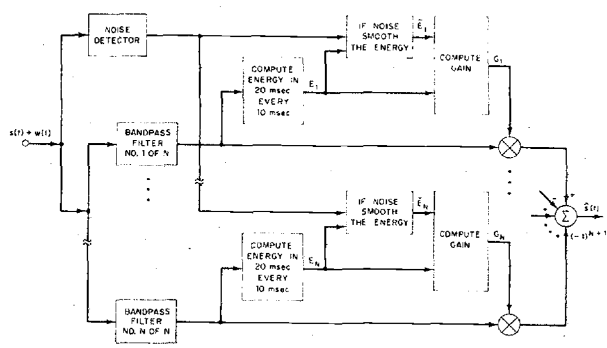

Fig. 1. Block diagram of the noise suppression prefilter.

Whatever rule is used to compute the channel gains it is necessary to compute the channel signal-to-noise ratio (SNR) which requires that measurements must be made to determine the instantaneous signal power and the average noise power at the output of each of the channel filters. Since the speech parameters change very little in 20 ms, some temporal smoothing can be exploited by computing the signal energy from

$$V_n^2 = \sum_{k=1}^{N} y_n^2(k) \qquad (1)$$

CH1559-4/80/0000-0699$00.75 © 1980 IEEE

where $y_n(k)$ represents the signal sample out of the nth channel at time k, where there are N such samples in the 20 ms frame. Determination of the background noise power requires knowledge of whether or not a particular frame contains noise alone. One approach for making this decision has been developed by Roberts [10] who noted that a 4-sec histogram of the frame energies of the input signal was bimodal. By setting a detection threshold between the modes, frames for which speech was absent could be determined with high confidence. For those frames the average noise energy in each channel was estimated by smoothing the measurements in (1) using a 1 sec time constant according to the recursion

$$\lambda_n(m) = \lambda_n(m-1) + \alpha[V_n^2(m) - \lambda_n(m-1)] \quad (2)$$

where $V_n^2(m)$, $\lambda_n(m)$ represent the measured energy and the average noise energy computed for the mth frame.

Using the measurement of $V_n^2(m)$ and the estimated average value of the noise energy $\lambda_n(m-1)$, the nth channel measurement parameter can be estimated from

$$g_n(m) = \frac{V_n^2(m) - \lambda_n(m-1)}{V_n^2(m)} \quad (3)$$

The prefiltering technique known as power subtraction [3-6] applies the square root of this measurement parameter to attenuate the waveforms out of each of the channel filters. This rule was used in the filter structure shown in Fig. 1 and coded for real time operation on the LDVT [11]. When tested using tapes with speech recorded in an environment of acoustically coupled airborne command post noise, it was found that the background noise was not completely suppressed and in fact resulted in reconstructed noise that had an annoying tonal quality. This negative result prompted a theoretical investigation into the noise suppression problem from the point of view of soft decision [7] and led to a class of suppression curves given by

$$G_n(m) = \frac{1}{2}\left(1 + \sqrt{g_n(m)}\right) \frac{\exp(-\xi)\, I_0\left(2\sqrt{\frac{\xi}{1-g_n(m)}}\right)}{1 \;\exp(-\xi)\, I_0\left(2\sqrt{\frac{\xi}{1-g_n(m)}}\right)} \quad (4)$$

where

$$I_0(x) = \frac{1}{2\pi}\int_0^{2\pi} \exp(x \cos\Theta)\, d\Theta \quad (5)$$

is the modified Bessel function of the first kind and ξ is a design parameter. In Fig. 2 several curves for the channel gain are plotted as a function of the measured a posteriori SNR $(V_n^2(m)/\lambda_n(m-1))$ for various values of the parameter ξ. Since the value of ξ controls the amount of suppression applied to each channel, the parameter is referred to as the suppression factor.

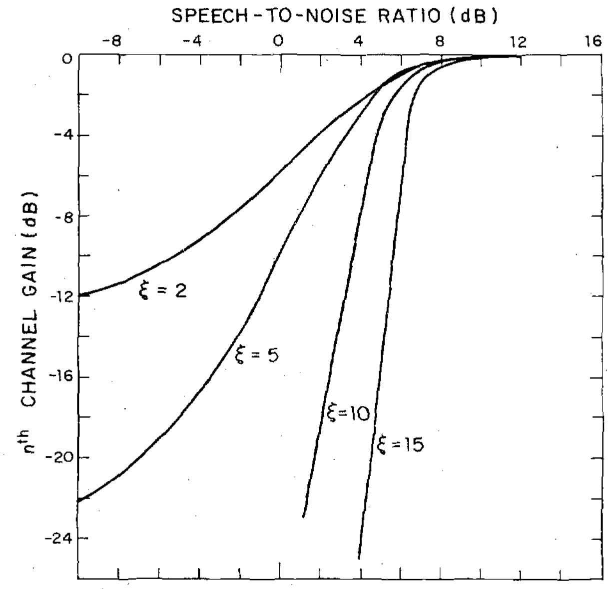

Fig. 2. Soft decision suppression rules.

In the real-time implementation the measurement parameter g(m) is used as a pointer for a table look-up to determine the attenuation prescribed by (4). Fifteen tables corresponding to values of $\xi = 1, 2, 3, \ldots, 15$ have been included in the prefilter with each table consisting of 50 values of the suppression rule computed for equal increments of g(m) from 0 to 1. No attempt was made to optimize the design of these tables. All of the coding was done in assembly language on the LDVT which has the ability to key in a new value of the suppression factor in real time. This meant that the prefilter could easily be adjusted to accommodate a wide class of operational environments which turned out to be a significant capability for effective noise suppression. Since the algorithm was designed to operate in real time, a 10 ms delay had to be incurred between the time the energies were measured and the time the corresponding gains could be computed and applied to the channel waveforms. This was done by computing the energies (block floating point) in 10 ms segments and adding consecutive segments together to produce the desired 20 ms energy measurement. This permitted computation of the gains, G(m), every 10 ms. In order to avoid the introduction of discontinuities in the output waveform a smoothed gain $\bar{G}(m)$ is obtained according to

$$\bar{G}(m) = \bar{G}(m-1) + \beta(m)[G(m) - \bar{G}(m-1)] \quad (6)$$

Since the introduction of smoothing can cause the prefilter to be slow to respond to a leading edge transition which could result in speech distortion, the coefficient in (6) is chosen adaptively according to the rule

$$\beta(m) = \begin{cases} 1 \text{ if } G(m) \geqslant \bar{G}(m-1) \\ \frac{1}{2} \text{ if } G(m) < \bar{G}(m-1). \end{cases} \quad (7)$$

In this way the prefilter responds immediately to an increase in the SNR which should minimize the potential for leading edge distortion. During a trailing edge, in which the gain will be decreasing, the smoothed gain will be used which will tend to maintain the speech signal even though the noise becomes dominant. It is the gain $\bar{G}(m)$ in (7) that is applied to the waveform at the output of each of the channel filters. These waveforms were then added together alternately 180° out of phase to produce the prefilter output waveform $\hat{s}(t)$.

When used to process speech corrupted by airborne command post noise, it was always possible to select a suppression factor which would render the background noise imperceptible, although, for cases in which the SNR was low enough, the cost in doing this was the introduction of various degrees of speech distortion. In these cases, if the suppression factor was subsequently reduced, the speech distortion was reduced at the expense of introducing a perceptible level of background noise.

A CHANNEL VOCODER WITH INTEGRATED PREFILTERING

Since the real time implementation of the prefilter was based on the principles of channel vocoder spectrum analysis, it seemed natural to attempt to integrate the two operations since this would exploit the advantages of efficient channel vocoder implementations. Figure 3 is a block diagram showing the analyzer and synthesizer sections of a standard channel vocoder.

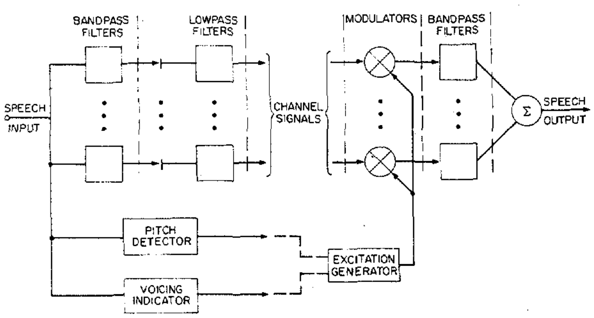

Fig. 3. General scheme of channel vocoder.

The 2400 bps channel vocoder consists of a bank of 19 Butterworth bandpass filters whose bandwidth increases with increasing frequency. The prefilter can be implemented as described in the preceding section simply by tapping off the waveforms at the outputs of each of the filters shown in Fig. 3. Applying the noise suppression gain factors in front of the rectifiers results in a set of modulator coefficients which will cause noise to be suppressed from the synthesized speech at the receiver [12]. Experiments were performed for speech in advanced airborne command post noise and it was found that noise could be suppressed from the synthesized speech. The tradeoffs between the amount of noise suppression and speech distortion are currently being examined.

CONCLUSIONS

A real-time implementation of a noise suppression prefilter based on soft decision has been described which permits a tradeoff of the speech distortion versus noise suppression by appropriate choice of a single parameter. It was then shown that many of the operations performed in the prefilter were intrinsic to the implementation of a standard channel vocoder and hence permitted integration of the noise suppression operations into a real time 2400 bps channel vocoder.

REFERENCES

1. C. Teacher and H. Watkins, "ANDVT Microphone and Audio System Study," Ketron Final Report, 1978.

2. M. R. Weiss, E. Aschkenasy and T. W. Parsons, "Study and Development of the INTEL Technique for Improving Speech Intelligibility," Technical Report No. RADC-TR-75-108, RADC, Griffiss Air Force Base, New York (April 1975).

3. S. F. Boll, "Suppression of Acoustic Noise in Speech Spectral Subtraction," IEEE Trans. Acoust., Speech and Signal Processing ASSP-26, 113 (1978).

4. R. A. Curtis and R. J. Niederjohn, "An Investigation of Several Frequency-domain Processing Methods for Enhancing the Intelligibility of Speech in Wideband Random Noise," Proceedings of International Conference on Acoustics, Speech and Signal Processing, (April 1978), pp. 602-605.

5. R. D. Preuss, "A Frequency Domain Noise Cancelling Preprocessor for Narrowband Speech Communications Systems," Proc. International Conference on Acoustics, Speech and Signal Processing, (April 1979), pp. 212-215.

6. M. Berouti, R. Schwartz and J. Makhoul, "Enhancement of Speech Corrupted by Acoustic Noise," Proc. International Conference on Acoustics, Speech and Signal Processing, (April 1979), pp. 208-211.

7. R. J. McAulay and M. L. Malpass, "Speech Enhancement Using a Soft-Decision Maximum Likelihood Noise Suppression Filter," Speech Communications Papers, Presented at the 97th Meeting of the Acoustical Society of America, Cambridge, Mass., 12-16 June 1979, pp. 399-402.

8. N. A. Kingsbury, "A Digital Channel Vocoder," Proc. of IEE International Specialists Seminar, "Case Histories in Signal Processing," Peebles, Scotland, 17 September 1979.

9. P. E. Blankenship, "An NMOS LSI Channel Vocoder Implementation," Proc. EASCON '78, 25-27 September 1978, pp. 684-692, and SPIE Technical Symposium East '79, April 17-20, 1979, Washington, DC.

10. J. Roberts, "Modification to Piecewise LPC," MITRE Working Paper. WP-21752 (12 May 1978).

11. P. E. Blankenship, "LDVT: High Performance Minicomputer for Real-time Speech Processing," EASCON '77, 26 September 1977.

12. B. Gold, P. E. Blankenship, R. J. McAulay, "New Applications of Channel Vocoders," submitted for publication to IEEE Trans. on Acoustics, Speech and Signal Processing.

This work was sponsored by the Department of the Air Force and the Defense Advanced Research Projects Agency.

The views and conclusions contained in this document are those of the contractor and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the United States Government.