

# Adaptive Speech Enhancement Using Frequency-Specific SNR Estimates

Carlos Avendano <sup>†</sup>, Hynek Hermansky <sup>†</sup>, Marvin Vis <sup>‡</sup> and Aruna Bayya <sup>\*</sup>

<sup>†</sup> *Oregon Graduate Institute of Science and Technology, Portland, OR*

<sup>‡</sup> *U S West Advanced Technologies, Boulder, CO*

<sup>\*</sup> *Rockwell International, Newport Beach, CA*

## **Abstract**—

We describe an adaptive speech enhancement technique based on selecting a set of pre-computed FIR filters to process the compressed short-time power spectral trajectories of noisy speech. The responses of the pre-computed filters depend only on the signal to noise ratios (SNRs) and does not depend on the center frequency of the sub-bands. This allows for a compact design in which the estimate of the SNR at the particular frequency channel is used as the filter selection criterium for that sub-band.

\*

## I. INTRODUCTION

As one of the most recent and profitable applications in the telecommunications industry, mobile telephony has reached a stage in which it is widely available to the public. The quality of the services is of great concern for companies to remain competitive in the market.

Mobile telephone calls frequently originate from noisy environments, and enhancing the quality of the received speech is of interest. One of the problems that needs to be considered is that, in general, background noise has different characteristics from one call to the next. A successful noise suppression system needs to use some strategy to deal with this factor.

In this work we describe an adaptive noise suppression technique intended for applications in services such as voice mail where the noisy speech recording is available for non-real time processing. With some modifications, the system is in principle also suitable for real-time processing.

## II. BACKGROUND

The system described in this study is based on RASTA-like processing of the temporal trajectories of the short-time spectrum of speech. The previous work [1] proposed a speech enhancement technique in which compressed power spectral trajectories of corrupted speech were processed by a filter bank with finite impulse response (FIR)

filters designed on parallel recordings of clean and noisy data. Thus, the filter bank was noise-specific and the algorithm was most efficient on disturbances similar to those present in the training data.

Two important aspects of the filters were observed:

1. The magnitude frequency response of filters corresponding to frequency regions of high speech energy showed suppression of low ( $< 2$  Hz) and high ( $> 8$  Hz) modulation frequencies <sup>†</sup>, while enhancing modulations around 5 Hz. Filters at regions of low spectral energy were low-pass or flat.
2. The dc gain of the filters was high at high signal to noise ratio (SNR) sub-bands and low at low SNR sub-bands, thus following the Wiener principle of optimal noise suppression.

The observations above suggested that the filter characteristics may depend on the energy of the speech signal relative to the noise level at each sub-band, thus a filter bank could be designed based on these local SNRs (frequency-specific SNRs).

## III. PRELIMINARY STUDIES

The first question that we formulated based on our observations was whether the filter responses depend only on the local SNR or if they also depend on the center frequency of the sub-band for which they are designed.

To answer this question we constructed a database by corrupting a sample of clean speech (approximately 180 s in length, taken from the TIMIT database) with additive white Gaussian noise (AWGN) at different overall SNRs (20, 15, 10, 5, 3, 0, -3, -5, -7, -10, -15 dB). From this training data we designed a set of RASTA-like filter banks (one for each overall SNR condition) following the

---

<sup>†</sup>We use the term “modulation frequency” to describe the frequency content of the time trajectories of the sub-band magnitude outputs of the short-time Fourier transform, where we have used 8 kHz sampling, 256 samples per window, and 75% window overlap

---

\*Appeared in Proc. IVTTA '96, New Jersey 1996.

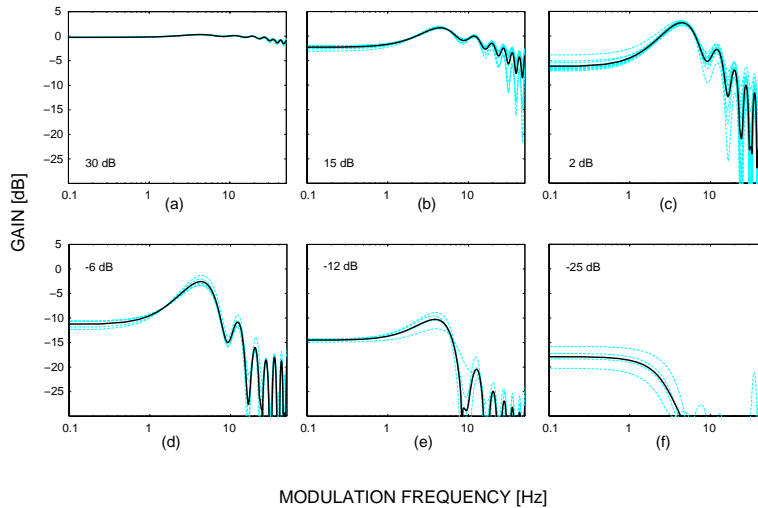


Fig. 1: Filter frequency responses (dotted lines) and mean response (solid lines) for several frequency-specific SNRs

procedure described in [1]. Thus, the exact frequency-specific SNR for the data used to design each filter in the filter banks was known (This frequency-specific SNR was computed as the ratio of the total power of the time trajectories of the magnitude short-time Fourier transform (STFT) of speech and noise signal at the given frequency band).

#### A. SNR-dependent RASTA-like Filters

Fig. 1 shows the filter characteristics for different sub-band SNRs. Each plot in the figure shows the magnitude frequency responses of filters derived at a given SNR for several <sup>‡</sup> sub-bands (dotted lines), together with the mean response (solid line) of the filters.

As the frequency-specific SNR decreases, the magnitude frequency response of the filters changes from

- a flat response (i.e. no filtering, see Fig. 1(a)), through
- a strong band-pass response enhancing modulation frequencies around 5Hz (i.e. speech enhancement, see Fig. 1(c) and Fig. 1(d)), to
- a low gain, low cut-off frequency low-pass response (i.e. suppression of the given component, Fig. 1(f)).

Notice that the attenuation of the dc component increases with decreasing frequency-specific SNR.

The results obtained in this preliminary study confirm the idea that the RASTA-like filters are strongly

<sup>‡</sup>We computed filters for a given frequency-specific SNR only at some representative sub-bands covering the frequency range of interest.

dependent on the SNR of the sub-band and relatively independent of the sub-band center frequency.

#### IV. SYSTEM DESIGN

The observations described above allow us to design a speech enhancement system which adapts to a specific noise condition. This extension makes the system applicable in realistic situations with noises and speech of unknown variance and coloration.

The new system configuration can be seen in Fig. 2. To assemble the appropriate filter bank for a particular noisy speech recording we compute the frequency-specific SNR for each magnitude STFT time trajectory over the whole recording and select a filter from a basis set of a few pre-computed basic filter shapes. After all filters for all sub-bands are selected we proceed to filter the compressed magnitude STFT trajectories, expand and re-synthesize using an overlap-add technique.

#### A. SNR Estimation

In practical situations we do not know the frequency-specific SNRs so an estimation procedure is required. We are primarily interested in the internal consistency of the estimate (rather than in the accuracy of the actual SNR estimate) as a measure of its usefulness for selecting a set of filters.

For this purpose we apply a noise estimation procedure proposed by Hirsch [2], in which the noise power at each magnitude STFT trajectory is estimated by computing a histogram of its amplitudes. The peak of the smoothed histogram is chosen as the noise amplitude estimate. Since we do not know the power of the clean speech signal, we use the power of the available noisy

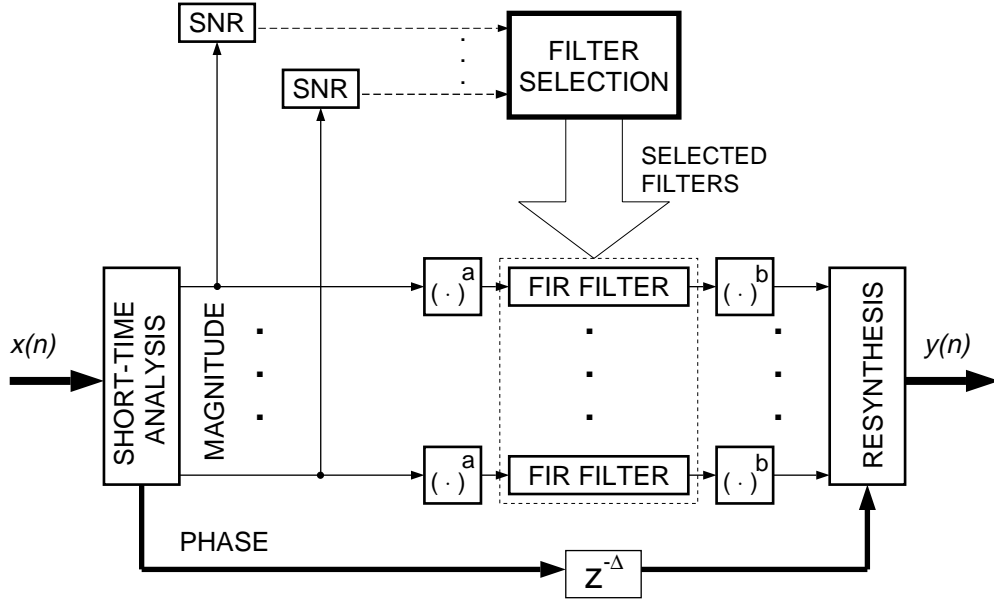


Fig. 2: Block diagram of the adaptive system.  $x(n)$  is the input corrupted speech,  $y(n)$  is the estimate of the clean speech ( $a = 2/3$  and  $b = 1/a$ )

signal, thus obtaining an estimate of the noisy signal to noise ratio. For our purpose, the performance of this estimator was found to be reasonable.

## B. Filter Design

### B.1. Pre-computation of the filters

To design the set of basic filters we used the same clean and noisy data as reported in the preliminary study above. For this, we assumed that the additive noise sources of interest have Gaussian distributions. The coloration of the noise is irrelevant given that, individually, the sub-band noise components from a colored Gaussian noise signal behave in the same way as if they were derived from a white source.

To derive a set of SNR-specific filters we averaged the magnitude frequency responses of filters computed at a given SNR and designed a non-causal linear phase FIR filter from the averaged response. We excluded filters with center frequencies below 100 Hz from the average because their responses were found to deviate slightly from the average (mainly in the dc gain factor). The linear phase assumption is justified from the observation that all the filters computed for the preliminary study above are approximately linear phase. A total of 25 filters, each corresponding to a frequency-specific SNR in 1 dB steps, was found to perform reasonably well.

### B.2. Construction of the Filter Table

In order to calibrate the SNR estimator which is used during processing (i.e. to find a mapping between the estimated and actual frequency-specific SNRs) the SNRs corresponding to each filter were estimated using the histogram technique. The filters were stored in a table along with their corresponding frequency-specific SNRs.

### C. Operation of the System

During the operation of the speech enhancement system on data with unknown noise, the SNR is estimated for each sub-band and a proper filter bank is built by selecting the appropriate filters from the table.

## V. RESULTS

To evaluate the performance of the system under different conditions we conducted the following set of tests:

### A. Known noise

To test the system apart from the SNR estimator, we artificially corrupted the clean speech (with colored Gaussian noise) and applied the processing with prior knowledge of the frequency-specific SNR. The result indicated a strong suppression of background noise while preserving the speech signal with very minor distortions. The residual noise has a very different character than the original disturbance. While the noise is not musical as in spectral subtraction, it presents periodic level fluctuations. These fluctuations are related to the enhancement

of certain modulation frequencies imposed by the filters in the medium SNR range (see Fig. 1). The modulation frequencies of the residual noise around 5 Hz are also enhanced and can be heard as the periodic disturbance.

### B. Unknown noise

Applying the algorithm based on the frequency-specific SNR estimates, we found very similar results, however, the noise level was underestimated and the suppression was slightly milder. Tuning the estimated to real SNR map, or biasing the SNR estimator itself might be helpful, but a better and more robust solution to the SNR estimation problem needs to be found if we want to take full advantage of the adaptive structure.

For a wide range of noise types and levels present in real cellular telephone calls we found a noticeable suppression of the perceived noise. In several informal preference tests we found that some subjects were disturbed by the residual noise, however there was an agreement about the reduction of background noise and preservation of the speech signal.

## VI. APPLICATIONS

Although originally designed for the off-line applications in enhancement of noisy voice-mail recordings, the new technique is not constrained to non-real time processing. We did not yet extensively experiment with the real-time processing, but the frequency-specific SNR estimation procedure can be done in real time if a first estimate is computed during the first few seconds of a conversation and updated over the length of the sample. As such, this adaptive update has the ability to adapt to time-varying conditions.

## VII. CONCLUSIONS

A few observations of the properties of temporal processing of the short-time spectrum of speech lead to the design of a new adaptive speech enhancement technique. Our informal tests indicate that the algorithm generalizes quite well over different types and levels of noises. While the system was originally designed to process voice-mail recordings it can be modified to operate on real time situations.

## ACKNOWLEDGMENTS

The work was in part supported by grants to OGI from US WEST Advanced Technologies (9069-111), from DoD under MDA-904-94-C-6169, and from NSF/ARPA under IRI-9314959.

## REFERENCES

[1] Hynek Hermansky, Eric A. Wan and Carlos Avendano. "Speech enhancement based on temporal pro-

cessing," *Proc. IEEE ICASSP-95*, pp. 405-408, Detroit, MI, 1995

[2] H. Gunter Hirsch, "Estimation of noise spectrum and its application to SNR estimation and speech enhancement," Tech. Report TR-93-012, International Computer Science Institute Berkeley, CA, 1993