PROCESSING OF NOISY SPEECH USING PARTIAL PHASE

B. Yegnanarayana*, K.V. Madhu Murthy* and Hema A. Murthy*.

ABSTRACT

This paper explores the possibility of processing noisy speech using signal reconstruction algorithms from Fourier Transform (FT) phase and magnitude. Algorithms have been proposed in the literature for signal reconstruction from FT phase alone, or, from FT magnitude with additional information in the form of 1-bit phase or signal values. More recently, algorithms have been proposed for signal reconstruction from partial phase (phase information in selected frequency bands) with compensating number of signal samples. In this paper we examine application of these techniques for processing noisy speech. In particular, we show that by selectively processing high signal-to-noise ratio (SNR) regions we can reduce the effect of background additive noise significantly.

INTRODUCTION

In recent years there is considerable interest in the area of signal reconstruction from Fourier Transform (FT) phase or magnitude (ref 1 to ref 5) in view of its relevance in speech, imaging and optics. In speech enhancement, for example, if speech could be synthesized from its short-time Fourier Transform (STFT) magnitude alone, then the STFT phase spectrum of the noisy speech is not required (ref 1). Algorithms have been proposed for signal reconstruction from FT phase alone, or from FT magnitude with additional information in the form of 1-bit phase or signal values (ref 2, ref 3). More recently, algorithms have also been proposed for signal reconstruction from partial phase together with compensating number of signal values (ref 4, ref 5). It was shown in (ref 5) that it is possible to exactly compensate for the missing phase information by the signal samples. It was also shown that the degradation of the reconstructed signal is graceful as the number of compensating signal samples falls below the required number for exact reconstruction. If the number of compensating signal samples is more than the required number, the number of iterations required for reconstruction of the signal within a given error is reduced. A reasonably good reconstruction can be achieved if the partial phase samples are known in the bands where the magnitude spectrum peaks. These developments in signal reconstruction from partial data suggest methods for processing noisy speech. In particular, we show that by focussing on the high SNR regions of the spectrum, we will be able to achieve a high degree of speech enhancement.

*Department of Computer Science and Engineering, Indian Institute of technology, Madras, India.
There are a variety of contexts in which it is desired to enhance speech. The objective of enhancement in most cases is to improve the overall quality or to increase intelligibility or to reduce listener fatigue. The techniques employed for cleaning noisy signals vary based on the availability of signal and noise information and the type of noise, viz., additive, multiplicative, convolutional, etc. In this paper, we focus our attention on the processing of speech signals corrupted by additive noise. Even in this limited context a number of different approaches have explored (ref 6). Most of these techniques are based on one of the following principles, namely, (i) spectral subtraction, (ii) modelling of degraded speech - AR(auto-regressive) or ARMA(auto-regressive moving average) modelling, (iii) periodicity of voiced speech - adaptive comb filtering, (iv) adaptive noise cancellation. Some of the pre-requisites for these techniques to work are: (i) speech signal should be stationary, (ii) statistical properties of speech and noise should be known apriori (iii) accurate pitch information must be available. It is difficult to obtain this information from the given speech signal due to its quasi-stationary nature and also due to noise. Algorithms for signal reconstruction from partial phase do not require any such apriori information.

PROCESSING OF NOISY SPEECH

The effect of additive noise in speech is to reduce the dynamic range of the short-time spectral envelope. This is the reason why the standard linear prediction (LP) analysis does not work satisfactorily for noisy speech. Moreover, a high order (10 to 14) LP analysis produces spurious peaks in the model spectrum. The reduced dynamic range and the spurious peaks produce considerable distortion in the synthetic speech. It is also difficult to extract excitation information (V/UV decision and pitch) from noisy speech. The method of spectral subtraction will not work satisfactorily either, because some statistically averaged spectrum is subtracted from the spectrum of each short-time segment of speech. The basic issues in processing noisy speech are how to (i) increase the dynamic range of the short-time spectral envelope (ii) emphasize the peaks corresponding to formants. This requires some method to identify the formant peaks for each segment in the presence of additive noise.

We propose an approach to overcome some of the above-mentioned problems. The approach uses both linear prediction analysis as well as signal reconstruction from phase. LP analysis is used to identify the high SNR regions as well as to increase the dynamic range by emphasizing those regions in relation to low SNR regions. It is likely that in most cases only one peak may be reliably detected, as other formant peaks may be submerged in the noise spectrum. This arises because the SNR is different in different frequency bands, and usually the higher formants may be at least 10 dB below the first formant peak. By emphasizing the major peak in relation to the rest of the spectrum, the dynamic range is improved, thus reducing the perceptual effect of the noise.
At the same time, the higher formants submerged in noise will still be present which the ear may perceive to add to the intelligibility of speech. The proposed approach of noisy speech processing does not need extraction of any excitation information. We use the phase function and the fine structure of the magnitude spectrum of the noisy speech to preserve the excitation information.

THE PROPOSED ALGORITHM

1. Take a segment (10-20 msec) of noisy speech.
2. Perform a pth order (p=3 to 5) LP analysis on the segment.
3. Decide from the LP spectrum whether the segment is voiced or unvoiced. Unvoiced segments are retained as it is in the processed speech.
4. Compute the magnitude spectrum of the LP residual error.
5. Identify the high SNR regions from the LP spectrum.
6. Reconstruct the bandpass signal from the high SNR regions of the spectrum.
7. Perform a low order (3-5) LP analysis on the resulting signal. The LP spectrum will have much higher dynamic range than the LP spectrum of the noisy speech.
8. Multiply the magnitude spectrum of the LP residual obtained in step 4 with this new LP spectral envelope.
9. Using the resulting magnitude spectrum and the phase spectrum of the noisy speech, obtain the processed signal using a straightforward inverse FT.
10. Perform a few iterations on the signal using finite support constraint in the signal domain and the phase spectrum of noisy speech.

Figs. 1 and 2 illustrate the original clean signal and the noisy signal and their respective spectra. Figs. 3 and 4 show the reconstructed signal and its spectrum after 1 and 10 iterations. It can be seen from Fig. 3 that there is considerable reduction in the additive noise in the reconstructed signal. However, the details due to higher formants are missing. This is restored to some extent by the phase as shown in Fig. 4. The spectral dynamic range of the reconstructed signal is still large compared to that of the noisy speech. This is how noise reduction has been achieved without a significant loss in intelligibility.

Performance of this algorithm will be demonstrated by playing processed speech obtained for several cases of noisy speech.

REFERENCES

5. B. Yegnanarayana, S.T. Fathima and Hema A. Murthy, ICASSP-87 (Dallas, Texas, 1987).

---x---

Signal | log Mag. spectrum | phase spectrum
Fig.1a | Fig.1b | Fig.1c
Fig.2a | Fig.2b | Fig.2c
Fig.3a | Fig.3b | Fig.3c
Fig.4a | Fig.4b | Fig.4c

Time in msec | Frequency in kHz | Frequency in kHz
0 | 0 | 0
256 | 5 | 5

Fig. 1 Original signal and its log mag. and phase spectra.
Fig. 2 Noisy signal and its log mag. and phase spectra.
Fig. 3 Reconstructed signal (after 1 iteration) and its log mag. and phase spectra.
Fig. 4 Reconstructed signal (after 10 iterations) and its log mag. and phase spectra.