AN EFFECTIVE QUALITY SPEECH ENHANCEMENT USING PARTICLE FILTERS

Ms.A.Helen Jenifer Archana , G.Prem Kumar,D.Raja

Abstract-- Speech Enhancement aims to improve speech quality. The objective of enhancement is improvement in intelligibility and overall perceptual quality of degraded speech signal using audio signal processing techniques. The central methods for enhancing speech are the removal of background noise and echo suppression. In this paper we use particle filters (PFs) for noise reduction on essentially linear speech production models, and the mounting evidence that the introduction of nonlinearities can lead to a refined speech model. This paper presents a study of PF solutions to the problem of speech enhancement in the context of nonlinear, neural-type speech models. It also decreases the computational cost by allowing us to use less particles. The simulation is done using MATLAB and the performance of PF for various speech signals are computed.

Index Terms—Nonlinear, particle filters (PFs), speech denoising, speech enhancement.

1.INTRODUCTION

Enhancement means the improvement in the value or quality of something. When applied to speech, this simply means the improvement in intelligibility or quality of a degraded speech signal by using signal processing tools. By speech enhancement, it refers not only to noise reduction but also to de-reverberation and separation of independent signals. Since this field is fundamental for research in the applications of digital signal processing it is also of great interest to the industry which is always looking for new solutions that are both effective and practical. This is a very difficult problem for two reasons. First, the nature and characteristics of the noise signals can change dramatically in time and between applications. It is also difficult to find algorithms that really work in different practical environments.

Ms.A.Helen Jenifer Archana is with Electronics and Communication Engineering Department of MEPCO SCHLENK Engineering College, Sivakasi as Assistant Professor. G.Prem Kumar and D.Raja are the ECE UG students of MEPCO SCHLENK engineering college sivakasi.

(premg1991@gmail.com, rajatsi.92@gmail.com)

Second, the performance measure can also be defined differently for each application. Two criteria often used to measure the performance are quality and intelligibility. It is very hard to satisfy both at the same time.

Speech is the most natural form of human communication. The advancement in technology has revolutionized the way we communicate. Speech has also become an important means of human-machine communication. Reliability and affordability of mobile voice communication devices has made “anywhere and anytime” communication a reality. The freedom and flexibility provided by mobile communication has also introduced new challenges. One of the most prominent is the suppression of background acoustic noise. Often mobile voice communication devices are used in an environment with a high level of ambient noise leading to degraded speech quality.

The perception of speech signal is usually measured in terms of its quality and intelligibility. The quality is a subjective measure that indicates the pleasantness/naturalness of the perceived speech. Intelligibility is an objective measure which predicts the percentage of words that can be correctly identified by listeners. In noisy environment degraded quality and poor intelligibility of the perceived speech make voice communication difficult and fatiguing. Speech sounds such as consonants, fricatives and stops are often masked by noise resulting in the loss of intelligibility. Therefore suppression of background acoustic noise is a relevant problem. Apart from perceptual enhancement noise
reduction is also crucial to obtain good performance of the speech coding algorithms that make mobile communication feasible.

Robustness to environmental noise has remained a limiting factor in the widespread deployment of speech enabled services such as automatic speech recognition systems for man/machine interaction and biometric applications like speaker identification/verification systems. These technologies show impressive performance in relatively noise-free environment. However in practice the speech signals are often degraded by acoustic background noise and the performance of these systems degrades rapidly in practical noisy conditions. One way to solve this problem is to use a speech enhancement (SE) algorithm as a pre-processing step for reducing the noise content and then use the enhanced signal as input to the speech processing system. Another major application of speech enhancement is in hearing aid design which needs an effective noise reduction front-end along with speech processing module to suit the listener's disability.

Speech enhancement is an area of speech processing where the goal is to improve the intelligibility or pleasantness of a speech signal. The most common approach in speech enhancement is noise removal where by the estimation of noise characteristics, can cancel noise components and retain only the clean speech signal. The basic problem with this approach is that if we remove those parts of the signal that resemble noise, we are also bounded to remove those parts of the speech signal that resemble noise. In other words, speech enhancement procedures often inadvertently also corrupt the speech signal when attempting to remove noise. Algorithms must therefore compromise between effectiveness of noise removal and level of distortion in the speech signal.

Current speech processing algorithms can roughly be divided into three domains, spectral subtraction, and sub-space analysis and filtering algorithms. Spectral subtraction algorithms operate in the spectral domain by removing, from each spectral band, that amount of energy which corresponds to the noise contribution. While spectral subtraction is effective in estimating the spectral magnitude of the speech signal, the phase of the original signal is not retained, which produces a clearly audible distortion known as “ringing”.

Fig. 1: Neural Network based speech restoration from noise

Sub-space analysis operates in the autocorrelation domain, where the speech and noise components can be assumed to be orthogonal, whereby their contributions can be readily separated. Unfortunately, finding the orthogonal components is computationally expensive. Moreover, the orthogonality assumption is difficult to motivate. Finally, filtering algorithms are time-domain methods that attempt to either remove the noise component (Wiener filtering) or estimate the noise and speech components by a filtering approach (Particle filtering, Kalman filtering).

I. PROPOSED METHOD

The Particle Filtering is one of the successful applications to speech enhancement. The idea is to represent the posterior density by a set of random particles with associated weights. PF computes the estimates based on these samples and weights. A discrete set of samples or particles represents the object state and evolves over time driven by means of “Survival of the fittest”. Non linear motion modes can be used to predict object states.

A. Neural Network Based PF

Fig.1 represents a typical situation where the speech signal is recovered from noise signal. The noise is acoustically added to the speech and the goal is to suppress the noise using speech enhancement method resulting in the output signal with the higher SNR. A high SNR estimate indicates that the signal content is less corrupted by noise. A low SNR estimate indicates the opposite.

B. Speech Model
Global nonlinear speech production model is shown in Fig. 2. The architecture of this class of network consists of input layers, output layers and intermediary layers called hidden layers. The computational units of the hidden layers are known as hidden neurons. The hidden layer does the intermediate computation before directing the input to output layer. The input layer neurons are linked to the hidden layer neurons. The weights on these links are referred to as input-hidden layer weights. The hidden layer neurons and output layer neurons are linked and the corresponding weights are referred to as output-hidden layer weights.

![Figure 2: Nonlinear speech production Model](image)

The feed forward model was chosen for several reasons. First, it can potentially achieve a higher prediction gain than traditional linear predictive models of the same and for a reference on linear prediction. Next, it lends itself well to sequential estimation as formulated by state-space equations. Finally, it constitutes a natural extension of the classical linear prediction model. The model detailed here supports variable input sizes and multiple neurons, and biases can be included.

For this generic model, different variations are considered.

- **Single-neuron versus multiple-neurons models.** In the single-neuron model, \( P = 1 \) and \( c_k[0] = c_k \) is considered to be known and equal to 1. In the multiple-neurons model, obviously \( P > 1 \).

- **Biased versus unbiased models.** In the unbiased model, \( b_k \) is set to 0.

The clean signal at discrete instant \( k \) is denoted by \( y_k \). The model can be described with a single equation,

\[
y_k = \sum_{i=1}^{P} w_{ik}y_{k-i} - b_k + \sigma_{gk}g_k
\]

Where,

- \( P \) is the number of neurons in the model.
- \( w_{ik} \) is a \( P \) by \( M \) matrix representing the internal coefficients of the network.
- \( c_k \) is a length vector denoting the output coefficients of the network.
- \( b_k \) is a length vector denoting the bias inputs at each of the neurons in the network.
- \( g_k \) represents a zero mean unit variance white Gaussian noise.
- \( \sigma_{gk} \) is a time-varying positive number.
- \( \sigma_{gk}g_k \) is the excitation noise of the speech model.
- \( f(.) \) is the activation function of the network.

An activation function given in Fig. 2 performs the mathematical operation on the signal output. The most common activation functions are

- Linear function
- Threshold function
- Piecewise linear function
- Sigmoidal function
- Tangent function

The activation functions are chosen depending upon the type of problem to be solved by the network. Activation function is also called as saturating linear function and can have either a binary or bipolar range for the saturation limits of the outputs.

Here in all the cases the function \( f(.) \) is chosen to be

\[
f(x) = \tanh(x)
\]

### C. Noise Model
White noise is considered as the noise model. It is otherwise called as the white-Gaussian noise. The measurement $z_k$ is simply formed by the addition of the clean speech $y_k$ and a white Gaussian noise. It is given by

$$z_k = y_k + \sigma_{v_k}$$

Where,

- $v_k$ is zero mean, unit variance white Gaussian noise.
- $\sigma_{v_k}$ is the standard deviation of the resulting additive noise $\sigma_{x,k}v_k$.

The quantity $\sigma_{v,k}$ can be specified as known or must be determined if it is known, then its value is directly given to the algorithm. If it is unknown, it must be determined using an assumption on its evolution model in the same way we treat the unknown speech parameters.

II. PF ALGORITHM

The particle filter is a sequential Monte Carlo algorithm, i.e., a sampling method for approximating a distribution that makes use of its temporal structure. A “particle representation” of distributions is used. The posterior density is represented by the set of random particles with associated weights. PF focus adaptively on probable regions of state-space. The algorithm depends on the choice of importance density. The particle filtering algorithm consists of two steps as shown in Fig. 3.

1) Sequential importance sampling step.
2) Selection step.

In the sampling step, for each particle at time $t$, we sample from transition priors and then evaluate and normalize the importance weights. In the selection step, multiply or discard the particles with respect to high or low importance weights to obtain $M$ particles.

The most important property of the particle filter is its ability to handle complex, multi-modal (non-Gaussian) posterior distributions. However, it has difficulties when $x_k$ is high-dimensional. Essentially, the number of particles $N$ required to adequately approximate the distribution grows exponentially with the dimensionality of the state space.

IV. RESULTS AND DISCUSSION

The proposed particle filtering algorithm for speech enhancement is simulated using MATLAB and the results were obtained. Fig. 4. shows the response of the original speech signal. White Gaussian noise is acoustically added to the original signal and the neural network based PF is used as an enhancement algorithm to recover the speech signal and the response of the enhanced signal is shown in Fig. 6. PF achieves high signal
to noise ratio with minimum mean square error. The values of PSNR and MSE are calculated as shown in Fig. 7. Three different speech signals are taken and their performance are compared and tabulated shown in Table I.

![Fig. 4 Frequency Response of Original Speech Signal](image)

![Fig. 5 Frequency Response of Noisy Signal](image)

![Fig. 6 Frequency Response of Enhanced Speech Signal](image)

![Fig. 7 Values of MSE and PSNR](image)

**TABLE I**
Comparison of MSE and PSNR for three different speech signals

<table>
<thead>
<tr>
<th>Input Signal</th>
<th>MSE</th>
<th>PSNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>In1.wav</td>
<td>0.0011</td>
<td>77.7687</td>
</tr>
<tr>
<td>In2.wav</td>
<td>0.0530</td>
<td>60.7262</td>
</tr>
<tr>
<td>In3.wav</td>
<td>0.0052</td>
<td>70.9921</td>
</tr>
</tbody>
</table>

V. CONCLUSION

A particle filtering algorithm for speech enhancement has been proposed in this paper. There is no strong limitation on the source signal and additive noise. Simulation results have illustrated that the estimators based on Gaussian noise model provide better performance with high signal to noise ratio.

VI. ACKNOWLEDGMENT

Apart from the efforts of me, the success of any work depends largely on the encouragement and guidelines of many others. I take this opportunity to express my gratitude to the people who have been instrumental in the
successful completion of this work. I would like to extend my sincere thanks to all of them. I owe a sincere prayer to the Lord Almighty for his kind blessings and giving me full support to do this work, without which this would have not been possible. I wish to take this opportunity to express our gratitude to all, who helped me directly or indirectly to complete this paper.

VII. REFERENCES


