Speech Enhancement in Hearing Aids Using Conjugate Symmetry Property of Short Time Fourier Transform

Dr. (Mrs) S.D. Apte¹, Shridhar²

¹Rajarshi Shahu College of Engineering, Department of Electronics and Communication Engineering, Pune, India.
Email: sdatpe@rediffmail.com
²Basaveshwar Engineering College, Department of Electronics and Communication Engineering, Bagalkot, India.
Email: shridhar.ece@gmail.com

Abstract—Most of the speech enhancement algorithms use the magnitude of STFT while phase is kept unchanged. In this paper the magnitude of STFT of noisy speech is kept unchanged while the phase is modified. Modified complex spectrum of speech is obtained by combining unchanged magnitude spectrum and modified phase spectrum. This modification results into cancellation of low energy components (noise) of complex spectrum more than the high-energy (speech) components leading to reduction in background noise. The models for variation of $\alpha$ with respect to SNR for AWGN, Train noise and Babble noise are developed using non linear regression modeling method. Normal hearing and hearing loss subjective listening tests and spectrogram analysis show that the proposed method gives improved speech quality.

Index Terms — Speech enhancement, hearing aid, STFT, magnitude spectrum, phase spectrum.

I. INTRODUCTION

Speech intelligibility and quality, which are very important for hearing loss people, can be improved by its enhancement [10,11,12,13]. Hearing Aids supported by speech enhancement algorithms really help the hearing loss people to understand the speech in various noisy environments [10,11]. In this direction, lot of research has been carried out [5,2,4]. From literature survey we find that following speech enhancement methods are used.


Let us consider a noisy speech signal

$$X(n)=s(n)+N(n)$$

Assuming the noise to be additive. Where $x (n)$=Noisy speech signal, $N(n)$=noise. Frame wise analysis of the speech signal is carried out [3,6,8,9]. Applying STFT to the noisy speech signal $x (n)$ we get

$$X(n,k)=\sum_{m=-\infty}^{\infty} x(m) w(n-m) e^{-j2\pi km/N}$$

(2)

Where $n$, $k$, $w(n)$ denotes frame duration, discrete frequency index and window function respectively [9,10]. Hamming window of 20ms duration is used. Equation (1) can be written as $X (n, k)=S (n, k)+N (n, k)$ where $X (n, k)$, $S(n, k)$ and $N(n, k)$ are STFT of noisy speech, clean speech and noise respectively [1,3]. In the proposed work, the magnitude of speech STFT is not processed, whereas its phase is modified based on empirical criteria [3]. The unprocessed magnitude is combined with modified phase during synthesis. This results into cancellation of low energy (noise) components more than high energy (speech) components [3]. This leads to enhancement of speech and is comparable with existing algorithms used in hearing aids [14,15]. It is verified by subjective listening tests and spectrogram analysis. The paper is organized as follows. Section II presents details of the proposed work. Section III deals with experimental details. Section IV presents results and discussion and finally section V presents the conclusion and future scope of the work.

II. PROPOSED METHOD

A. Principle

The noisy speech signal $x (n)$ is real valued hence its DFT obeys conjugate symmetry property [3]. The original noisy speech signal $x (n)$ is obtained if IDFT is computed straightaway due to cancellation of imaginary parts of complex conjugate terms of DFT. The degree of cancellation or reinforcement of these imaginary parts of
complex conjugates is controlled by modifying their phase [3].

B. Explanation.

The proposed method includes three stages of processing i.e. 1. Analysis stage. 2. Spectrum modification stage. 3. Synthesis stage [3]. The algorithm of the proposed method is given below.

Step1. Framing speech samples using Hamming Window with frame size of 20 msec and 50% overlap.
Step2. Computation of 1024 point DFT of each frame.
Step3. Magnitude computation of each DFT bin.
Step4. Phase modification empirically.
Step6. Computation of IDFT of each DFT bin.
Step7. Speech synthesis using overlap-add method.

The input noisy speech signal \( x(n) \) is real; hence its DFT \( X(k) \) obeys conjugate symmetry. IDFT of \( X(k) \) results into original noisy speech signal \( x(n) \) due to cancellation of imaginary parts of complex conjugate terms. But the degree of cancellation or reinforcement of complex conjugates can be controlled by modifying their phase [3]. A constant \( \alpha \) given by

\[
\alpha(k) = \lambda; \quad 0 \leq k < N/2
\]

\[
\alpha(k) = -\lambda; \quad N/2 \leq k \leq N-1
\]

Assuming \( N \) to be Even. \( \alpha(k) \) is a real valued frequency dependent function which is antisymmetric about the frequency \( F_s/2 \) rad /sample and \( \lambda \) is a real valued constant. The noisy speech signal STFT \( X(n, k) \) is modified as

\[
X_\alpha(n, k) = X(n, k) + \alpha(k)
\]

The modified phase of \( X_\alpha(n, k) \) is computed and further combined with magnitude of noisy speech signal to get modified complex spectrum.

\[
X_M(n, k) = X(n, k) e^{j\alpha(n, k)}
\]

The IDFT of above complex spectrum results into enhanced real signal, the explanation is as follows. Fourier analysis resolves a signal into a weighted sum of sinusoids. I.e. sum of complex conjugates [5,6]. The magnitude spectrum of DFT of a real valued signal obeys even symmetry whereas phase spectrum obeys odd symmetry. During the process of signal synthesis (IDFT) the conjugates sum together to result in to a real signal due to cancellation of their imaginary parts. The degree of cancellation or summation of these complex conjugates can be controlled by modifying their phase. The above process is elaborated using signal – vector analogy. Considering a pair of complex conjugate numbers \( C_1 = X + jY \) and \( C_1^- = X - jY \) having magnitude

\[
M_1 = \sqrt{X^2 + Y^2}
\]

and phase angles

\[
\phi l = \tan^{-1}(Y/X)
\]

and

\[
\phi l^- = \tan^{-1}(-Y/X)
\]

These complex conjugate numbers are modified as

\[
C_{11} = X + jY + \alpha_1
\]

\[
C_{11}^- = X - jY - \alpha_1
\]

the resulting phase angles are

\[
\phi 11 = \tan^{-1}(Y/X + \alpha_1)
\]

and

\[
\phi 11^- = \tan^{-1}(-Y/X - \alpha_1)
\]

Combining the magnitude in equation (7) with phase in equations (12) and (13) forms the new complex numbers

\[
C_{1p} = \sqrt{X^2 + Y^2} e^{j\tan^{-1}(Y/X + \alpha_1)}
\]

and

\[
C_{1p}^- = \sqrt{X^2 + Y^2} e^{j\tan^{-1}(-Y/X - \alpha_1)}
\]

The resultant of above two complex numbers is given by

\[
C_{R1} = 2\sqrt{X^2 + Y^2} ; \quad \text{if } \alpha_1 < M_1
\]

The resultant obtained in equation (16) is same as the resultant of original complex numbers,

\[
C_1 = X + jY = \sqrt{X^2 + Y^2} e^{j\tan^{-1}(Y/X)}
\]

\[
C_1^- = X - jY = \sqrt{X^2 + Y^2} e^{j\tan^{-1}(-Y/X)}
\]

given by

\[
C = 2\sqrt{X^2 + Y^2}
\]

Therefore from equation (16) and equation (19) it is proved that the phase modification due to \( \alpha_1 \) has very negligible effect on the spectral components having magnitude more (Speech) than noise components. Considering \( \alpha_1 \gg M_1 \) and a pair of complex conjugate numbers

\[
C_2 = P + jQ
\]

\[
C_2^- = P - jQ
\]

Both having the magnitude given by

\[
M_2 = \sqrt{P^2 + Q^2}
\]

and phase angles

\[
\phi 2 = \tan^{-1}(Q/P)
\]

and

\[
\phi 2^- = \tan^{-1}(-Q/P)
\]

respectively. These complex conjugate numbers are modified as
\[ C_{22} = P + j Q + \alpha_1 \] (25)
\[ and \quad C_{22}^* = P - j Q + \alpha_1 \] (26)

The resulting phase angles are
\[ \phi_{22} = \tan^{-1}(Q / P + \alpha) \] (27)
and
\[ \phi_{22}^* = \tan^{-1}(-Q / P - \alpha) \] (28)

Combining the magnitude in equation 22 with phase in equations (27) and (28) results into complex numbers
\[ C_{2P} = \sqrt{P^2 + Q^2} \ e^{j\tan^{-1}(Q/P + \alpha)} \] (29)
and
\[ C_{2P}^* = \sqrt{P^2 + Q^2} \ e^{j\tan^{-1}(-Q/P - \alpha)} \] (30)

The resultant of above two complex conjugate numbers is given by
\[ C_{R2} = \sqrt{P^2 + Q^2} \cos(\theta) \] (31)

Where \[ \theta = \tan^{-1}(Q / P + \alpha) + \tan^{-1}(-Q / P - \alpha) \] (32)

If \[ \alpha \gg \text{M2} \], equation (31) becomes
\[ C_{R2} = \sqrt{2(\alpha^2)} \cos[\theta] \] (33)
\[ C_{R2} \ll \sqrt{P^2 + Q^2} \] (34)

Therefore from equation (22) and equation (34) it is proved that \[ C_{R2} \ll \text{M2} \] [3]. This implies that the phase modification due to \[ \alpha_1 \] has considerable effect on the spectral components having magnitude less (noise) than speech components leading to enhancement of speech, which is more beneficial to hearing loss people through their hearing aids. This is based on the assumption that magnitudes of speech components are more than noise components in given noisy speech signal [1]. The various values of \[ \alpha \text{ for different types of noise are obtained empirically and are given in table III} \] [3]. The models for \[ \alpha \text{ as a function of SNR are obtained using non linear regression modeling techniques. The models are as follows with the conventions } y=\alpha \text{ and } x=\text{SNR (dB).} \]

For AWGN the model equation is
\[ y=1.82*10^{0.6}x^{5.149}*10^{10}(4)x^{4.3}+3.22*10^{0.3}x^{3.151}*10^{10}(2)x^{2.028}+3.496 \] (35)

For Train noise the model equation is
\[ y=4.03*10^{0.2x}x^{5.159}*10^{0.3}x^{4.0+0.021}+8.17*10^{0.2x}x^{2.054}+5.99 \] (36)

For Babble noise the model equation is
\[ y=3.59*10^{0.2x}x^{5.34}+3.159*10^{0.3}x^{4.2}+2.13*10^{0.2}x^{5.74}+10^{0.2}x^{2.137}+9.531 \] (37)

The model error in all the three cases is negligible.

III. EXPERIMENTAL DETAILS

A. Speech database

In the experimental evaluation, the NOIZEUS speech corpus is used [3]. This corpus is composed of non stationary noises at different SNRs. In the evaluation, the train, babble and white Gaussian noise are used at 0dB, 5dB, 10 dB and 15dB SNR levels.

B. Experimentation Procedure

Zero mean and normalized speech samples at a sampling frequency of 8 KHz are obtained. 20ms frame duration and 10 ms overlap are chosen to frame the speech samples using Hamming window. 1024 point DFT of each frame is taken and its magnitude is computed. First 50% of the 1024 point FFT of each frame are modified by adding a constant \[ \alpha \] and remaining 50% are modified by subtracting the same constant \[ \alpha \]. Here \[ \alpha \] is the function of speech SNR and type of noise. Phase of each modified DFT samples are calculated. Modified DFT values are obtained by combining the original DFT magnitudes and modified phase values. Finally discrete time signal is obtained by computing IFFT followed by overlap add method [7]. The experiments were conducted on normal hearing subjects and subjects with hearing loss using white Gaussian noise, train noise and babble noise.

IV. RESULTS AND DISCUSSION

Mean opinion score (Shown in Table I and II) obtained from listening tests on normal hearing subjects and subjects with hearing loss show that the proposed method performs best in the case of white noise as compared to train and babble noise. Results of spectrogram analysis are shown in figures 2,3,4,5,6, and 7. The enhanced speech signal in the presence of white Gaussian noise doesn’t exhibit speech distortion; at the same time background noise has been attenuated. In case of train and babble noise though the background noise is suppressed, a small amount of speech distortion exists because of the factor \[ \alpha \] being constant for all values of frequencies. The subjective speech intelligibility test on Normal hearing and Hearing loss subjects differs by a small amount in case of Babble noise only.

V. CONCLUSION AND FUTURE SCOPE OF WORK

In this paper an improved method of speech enhancement is presented which finds a front-end application in hearing aids used by people suffering from general hearing loss due to age and sensori neural
hearing loss. The magnitude spectrum of noisy speech signal is combined with modified phase spectrum to get modified complex spectrum. During signal synthesis using modified complex spectrum low energy components (noise) cancel out more as compared to high-energy components (speech). Thus achieving enhancement of speech. The work is validated through subjective listening tests on normal hearing subjects and subjects with hearing loss. The mean opinion score of normal hearing subjects and subjects with hearing loss differs by a small value. The reduction in noise is evident from the spectrograms. As a future scope of work more experimental investigation can be done on subjects with hearing loss due to age and sensori neural hearing loss.

### TABLE I.
Opinion Score and Mean Opinion Score of Speech Intelligibility Test on Normal Hearing Subjects

<table>
<thead>
<tr>
<th>TYPE OF NOISE</th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>MEAN OPINION SCORE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNR 0 dB</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>SNR 5 dB</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td>SNR 10 dB</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
</tr>
<tr>
<td>SNR 15 dB</td>
<td>0.94</td>
<td>0.94</td>
<td>0.94</td>
<td>0.94</td>
</tr>
<tr>
<td>SNR 20 dB</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
</tr>
</tbody>
</table>

### TABLE II
Opinion Score and Mean Opinion Score of Speech Intelligibility Test on Subjects with Hearing Loss

<table>
<thead>
<tr>
<th>TYPE OF NOISE</th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>MEAN OPINION SCORE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNR 0 dB</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>SNR 5 dB</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td>SNR 10 dB</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
</tr>
<tr>
<td>SNR 15 dB</td>
<td>0.94</td>
<td>0.94</td>
<td>0.94</td>
<td>0.94</td>
</tr>
<tr>
<td>SNR 20 dB</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
</tr>
</tbody>
</table>

© 2009 ACADEMY PUBLISHER
Figure 1. Spectrogram of noisy speech sample (Train noise) of SNR 0 dB from NOIZEUS database

Figure 2. Spectrogram of enhanced speech sample of figure 1

Figure 3. Spectrogram of noisy speech sample (Babble) of SNR 5 dB from NOIZEUS database.

Figure 4. Spectrogram of enhanced speech sample of figure 3

Figure 5. Spectrogram of noisy speech sample (Gaussian noise) of SNR 12 dB from NOIZEUS database

Figure 6. Spectrogram of enhanced speech sample of figure 5
Table III

Empirical values of $\lambda$ for different Noise and SNRs

<table>
<thead>
<tr>
<th>SNR (dB)</th>
<th>$\lambda$ Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AWGN</td>
</tr>
<tr>
<td>0</td>
<td>3.5</td>
</tr>
<tr>
<td>1</td>
<td>3.2</td>
</tr>
<tr>
<td>2</td>
<td>2.9</td>
</tr>
<tr>
<td>3</td>
<td>2.55</td>
</tr>
<tr>
<td>4</td>
<td>2.3</td>
</tr>
<tr>
<td>5</td>
<td>2.0</td>
</tr>
<tr>
<td>6</td>
<td>1.75</td>
</tr>
<tr>
<td>7</td>
<td>1.55</td>
</tr>
<tr>
<td>8</td>
<td>1.45</td>
</tr>
<tr>
<td>9</td>
<td>1.23</td>
</tr>
<tr>
<td>10</td>
<td>1.05</td>
</tr>
<tr>
<td>11</td>
<td>0.95</td>
</tr>
<tr>
<td>12</td>
<td>0.85</td>
</tr>
<tr>
<td>13</td>
<td>0.75</td>
</tr>
<tr>
<td>14</td>
<td>0.65</td>
</tr>
<tr>
<td>15</td>
<td>0.55</td>
</tr>
</tbody>
</table>

REFERENCES.


