A Novel Speech Enhancement Algorithm Using Imaginary Part of DFT

Dr. (Smt). S.D. Apte1, Shridhar2

1Rajarshi Shahu College of Engineering/E & CE Dept, Pune, India.
Email: sdapte@rediffmail.com
91-9975185707
2Basavashvar Engineering College/E & CE Dept, Bagalkot, India
Email: shridhar.ece@gmail.com
91-7259704913

Abstract-- In this paper a simple and computationally efficient speech enhancement algorithm is presented. Most of the speech enhancement algorithms modify magnitude of STFT while keeping the phase as it is. In contrast, in this paper the magnitude of STFT is kept as it is while the phase is modified. This modification results into cancellation of low energy components (noise) more, than the high-energy (speech) components in the reconstructed time domain signal. Objective and subjective measures and spectrogram analysis of processed speech show that the proposed method yields improved speech quality. In the proposed method speech enhancement is achieved through modification of phase angle of DFTs of speech through their imaginary parts. This is the driving force for the algorithm to be computationally efficient.

Keywords--Speech enhancement, Hearing aid, DFT, Magnitude spectrum, Phase spectrum, Imaginary part,

I. INTRODUCTION

Speech intelligibility and its perceptual quality are very important for its understanding, especially in noisy conditions [1],[2]. The quality of speech can be increased by its enhancement [3],[4],[6],[7]. In this direction, lot of research has been carried out [5-11]. Following are the main speech enhancement methods reported in the literature: (1) Spectral Subtraction. (2) Modified Spectral Subtraction. (3) Minimum Mean Square Error. (4) Formants intensification. (5) Signal Subspace Approach. (6) Wiener filtering. (7) Modified Spectral Subtraction with masking property. (8) Kalman filtering. (9) Warped DFT. (10) Human auditory properties. (11) Principal Component Analysis. (12) Hidden Markov Model. (13) Neural Networks. (14) Multi-band Speech decomposition etc.

Assuming the noise to be additive, a noisy speech signal \( x(n) \) composed of clean speech \( s(n) \) and noise \( N(n) \) is given by

\[
x(n) = s(n) + N(n)
\]  

(1.1)

Taking DFT of equation (1.1) gives \( X(k) = S(k) + N(k) \). Some of the speech enhancement methods process the magnitude spectrum of noisy speech, whereas phase spectrum is kept as it is [5],[8],[9]. In the proposed work, the magnitude of noisy speech STFT is kept as it is, whereas phase is processed [6]. The unprocessed magnitude is combined with processed phase to get modified spectrum. This results into cancellation of low energy (noise) components more than high energy (speech) components in the reconstructed signal [6]. This leads to enhancement of speech and is comparable with existing algorithms.

II. PROPOSED METHOD

The basic assumption of the proposed method is energy of speech components is more than that of noise components. The noisy speech signal \( x(n) \) is real; hence its DFT obeys conjugate symmetry. The degree of cancellation or reinforcement of imaginary parts of DFT is controlled by modifying their phase through their imaginary parts [6],[7]. Assuming \( N \) to be even, an imaginary constant \( C(k) \) given by

\[
C(k) = j\beta ; 0 \leq k < N/2 \quad \text{and} \quad C(k) = -j\beta ; N/2 \leq k \leq N-1
\]  

(2.1)

is used for phase modification. It is anti-symmetric about \( F_s/2 \) frequency. Here it is assumed to be a constant. Using this constant the noisy speech signal of equation (1.1) is modified as \( X'(k) = X(k) + C(k) \). The modified phase of \( X'(k) \) is computed and further combined with magnitude of original noisy speech signal to get modified complex spectrum

\[
S(k) = |X(k)| e^{j\phi(k)} ; 0 \leq k \leq (N-1)
\]  

(2.2)

Since conjugate symmetry is disturbed due to phase modification, IDFT of modified spectrum results into complex signal [7-11]. Only real parts are retained for further processing. This approach results in reduced computational complexity. The above principle can be proved using signal – vector analogy [6]. Considering a pair of complex conjugate numbers \( C_1 = X+jY \) and \( C_1^* = X-jY \) both having same magnitude \( M \), and resultant \( R \) given by

\[
M = \sqrt{X^2 + Y^2} \quad \text{and} \quad R = \sqrt{2(X^2 + Y^2)}
\]  

(2.3)

phase angles \( \phi = \tan^{-1}(Y/X) \) and \( \phi^* = \tan^{-1}(-Y/X) \) (2.4)

The modified complex numbers are given by

\[
C_1 = X+jY+\beta \quad \text{and} \quad C_1^* = X-jY-\beta.
\]  

(2.5)

The modified phase angles are given by

\[
\phi_{11} = \tan^{-1}\left(\frac{Y+\beta}{X}\right) \quad \text{and} \quad \phi_{22} = \tan^{-1}\left(\frac{-Y-\beta}{X}\right)
\]  

(2.6)

The modified complex numbers are

\[
C = \sqrt{X^2 + Y^2} e^{j\tan^{-1}\left(\frac{Y+\beta}{x}\right)} \quad \text{and} \quad C^* = \sqrt{X^2 + Y^2} e^{j\tan^{-1}\left(-\frac{Y+\beta}{x}\right)}
\]  

(2.7)

The resultant of above two complex numbers is given by

\[
R = \sqrt{2(X^2 + Y^2) + 2(X^2 + Y^2) \cos\left(\tan^{-1}\frac{Y+\beta}{x} - \tan^{-1}\frac{-Y-\beta}{x}\right)}
\]  

(2.8)

The proposed method is computationally efficient and is comparable with existing algorithms.
Case I. Resultant R, when $\beta < \sqrt{x^2 + y^2}$, from equation 2.7

$$ R = \sqrt{2(x^2 + y^2) + 2(x^2 + y^2) \cos \left(\tan^{-1}\frac{\beta}{x} - \tan^{-1}\frac{\gamma}{x}\right)} \tag{2.8} $$

$$ \therefore R = 2\sqrt{x^2 + y^2} \tag{2.9} $$

This is equal to the resultant of original complex conjugate numbers, as given in the equation (2.3). The implication of above result is spectral components having magnitudes more (speech components) than magnitude of $\beta$, the spectral components (speech components) remain unaltered [6, 7].

Case II. Resultant R, When $\beta \gg \sqrt{x^2 + y^2}$

Equation 2.7 reduces to

$$ R = \sqrt{2(x^2 + y^2) + 2(x^2 + y^2) \cos \left(\tan^{-1}\frac{\beta}{x} - \tan^{-1}\frac{\gamma}{x}\right)} \tag{2.10} $$

Let

$$ \theta = \tan^{-1}\frac{\beta}{x} - \tan^{-1}\frac{\gamma}{x} \Rightarrow \theta = \tan^{-1}\frac{x^2 - \beta \gamma}{x^2 + \beta \gamma} \tag{2.11} $$

Considering the right angled triangle PQR shown in Fig.1 and from equation (2.11) we have

\[ \text{Fig 1. Triangle PQR} \]

In triangle PQR,

$$ \cos \theta = \frac{x^2 + \beta \gamma}{x^2 + (\beta^2 + \gamma^2) \sqrt{\beta \gamma}} $$

\therefore Equation 2.10 reduces to

$$ R = \sqrt{2(x^2 + y^2) + 2(x^2 + y^2) \cos \theta} \tag{2.12} $$

where $\cos \theta << 1$. Hence the Resultant R reduces to

$$ R \ll 2\sqrt{x^2 + y^2} \tag{2.13} $$

Resultant R is very small compared to the resultant of original complex conjugate numbers. The implication of above result is spectral components having magnitudes much smaller (noise components) than magnitude of $\beta$, the spectral component gets suppressed more [6-8].

This leads to enhancement of signal to noise ratio. Empirically determined values of $\beta$ as a function of input speech SNR for white Gaussian, train and babble noise for which satisfactory speech enhancement was obtained are tabulated in Table. 1.

III. Relation Between $\alpha$ and $\beta$.

Enhancement of speech signal can also be achieved by the same principle of phase modification using real parts of DFTs [6, 7]. This is similar to the process illustrated in section 2. Let $\alpha$ be the factor used to modify the real parts of complex conjugates similar to the factor $j\beta$. The relation between $\alpha$ and $\beta$ is important with respect to computational complexity of the algorithm, which is derived in the following section.

A. Lemma

For any given pair of complex conjugate numbers $X + jY$ and $X - jY$,

$$ \tan^{-1}\frac{Y}{X + \alpha} + \tan^{-1}\frac{Y}{X - \alpha} = \tan^{-1}\frac{|Y + \beta|}{X} + \tan^{-1}\frac{|Y - \beta|}{X} $$

Then $\beta << \alpha$. Where $\alpha$ and $\beta$ are real.

Proof:

Let

$$ \theta = \tan^{-1}\frac{Y}{X + \alpha} + \tan^{-1}\frac{Y}{X - \alpha} \Rightarrow \tan \theta = \frac{2XY}{x^2 - \alpha^2 - y^2} \tag{3.1} $$

$$ \therefore \alpha = \sqrt{x^2 - y^2 - 2\frac{XY}{K}} \tag{3.2} $$

On the same lines

$$ \tan \theta = \frac{2XY + \beta}{x^2 - y^2 - \beta^2 - 2\beta \gamma} \tag{3.3} $$

$$ \therefore \beta = \frac{-(x + k\alpha)(x^2 + 1 + k^2)}{k} \tag{3.4} $$

From equations (3.1) and (3.2) it can be verified that $\beta << \alpha$ for any given values of $X$, $Y$ and $\theta$.

IV. Experimental Details

In the experimental evaluation the NOIZEUS speech corpus is used. Enhanced speech is evaluated by employing perceptual estimation of speech quality (PESQ) [6, 7]. Mean PESQ scores over a subset of NOIZEUS speech data base are calculated and tabulated in Table.2. 10 normal hearing subjects of age group 20-25 years and 10 subjects with moderate hearing loss of age group 55-60 years participated in listening tests. The improvements in Mean Opinion Score (MOS) in both the cases are plotted in Figure.3 and 4 respectively. Spectrogram analysis is also carried out and is shown in figure.5. The values of $\beta$ given in Table.1 were determined empirically so as to obtain maximum PESQ and maximum mean opinion scores.

| TABLE I - Empirically Determined Values of $\beta$ as a Function of Input Speech SNR for White Gaussian, Train and Babble Noise. |
|-----------------|-----------------|-----------------|-----------------|
| SNR (dB)        | AWGN            | TRAIN           | BABBLE          |
| 0.0             | 0.0070          | 0.70            | 0.490           |
| 3.0             | 0.0090          | 1.00            | 0.500           |
| 10.0            | 0.0200          | 1.80            | 0.70            |
| 15.0            | 0.0400          | 2.00            | 0.90            |
TABLE II. MEAN PESQ SCORES FOR WHITE NOISE CASE FOR THE SPECTRAL SUBTRACTION (SS), SPECTRAL CONTRAST ENHANCEMENT (SC), MINIMUM MEAN SQUARE (MM) AND PROPOSED (PR) METHOD.

<table>
<thead>
<tr>
<th>INPUT SPEECH SNR(dB)</th>
<th>NOISY SPEECH</th>
<th>CLEAN SPEECH</th>
<th>METHODS USED</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>SS SC MM PR</td>
</tr>
<tr>
<td>0</td>
<td>1.61</td>
<td>4.6</td>
<td>1.74 1.62 1.86 1.92</td>
</tr>
<tr>
<td>5</td>
<td>1.86</td>
<td>4.6</td>
<td>2.19 2.02 2.23 2.34</td>
</tr>
<tr>
<td>10</td>
<td>2.17</td>
<td>4.6</td>
<td>2.57 2.42 2.54 2.80</td>
</tr>
<tr>
<td>15</td>
<td>2.60</td>
<td>4.6</td>
<td>3.12 2.57 3.00 3.10</td>
</tr>
</tbody>
</table>

Fig 3. MOS improvement for AWGN (Top), Train Noise (Middle) and Babble Noise (Bottom) in case of listening tests on hearing impaired subjects

V. RESULTS.

Improvement of mean opinion scores in case of normal hearing subjects and subjects with moderate hearing loss shown in Fig.3 indicate that the proposed method performs best in case of additive white Gaussian noise as compared to train and babble noise. The results of spectrogram analysis shown in Fig. 4 indicate that the enhanced signal in case of white Gaussian noise does not exhibit speech distortion, while background noise has been attenuated. In case of train and babble noise though the noise is suppressed, a small amount of signal distortion is also introduced. The mean PESQ scores are comparable with conventional enhancement methods. The added advantage of the proposed method is its reduced computational load as compared to phase modification through real parts of DFTs.

CONCLUSIONS AND FUTURE SCOPE OF WORK.

In this paper a computationally efficient speech enhancement algorithm is presented. This proposed work can find application in hearing aids as a front-end algorithm to suppress the background noise. As a future scope of work more experimental investigations can be done on subjects with hearing loss, sensorineural problems to understand human auditory system still better.

REFERENCES