New Speech Enhancement Method based on Wavelet Transform and Tracking of Non Stationary Noise Algorithm

Riadh AJGOU(1,2), Salim SBAA(1), Said GHENDIR(1,2), Ali CHEMSA(1,2) and A. TALEB-AHMED(3)

Abstract—In this work, we have developed an efficient approach for enhancing speech by combining tracking of non stationary noise algorithm and Continues Wavelet Transform (CWT). Tracking of non stationary noise method that is based on data-driven recursive noise power estimation was proposed by Jan S. Erkelens and Richard Heusdens. The Continues Wavelet decomposition of speech signal uses adaptive level with Harr mother wavelet. In this paper, our novel method was evaluated in presence of different kind of noise using the NOIZEUS noisy speech corpus developed in Hu and Loizou laboratory that is suitable for evaluation of speech enhancement algorithms. The noisy database contains 30 IEEE sentences (produced by three male and three female speakers) corrupted by eight different real-world noises at different SNRs. The noise was taken from the AURORA database and includes suburban train noise, babble, car, exhibition hall, restaurant, street, airport and train-station noise. For evaluating the performance of speech enhancement methods we have used Perceptual Evaluation of Speech Quality scores (PESQ, ITU-T P.862). Simulation results demonstrate that the proposed approach offers an improved performance of speech enhancement in comparison with state-of-the-art methods in terms of PESQ measure.

Keywords— Speech enhancement, Tracking of non stationary noise method, Wavelet Transform; PESQ.

I. INTRODUCTION

The paper addressed the problem of suppressing the background noise in noisy speech. Speech signal can be corrupted by noise in various situations, such as trains, cars, airport, babble, factory, street...etc. The problem of enhancing speech degraded by the noise is largely open to research, although many significant techniques have been introduced over the past decade because there are many areas where it is necessary to enhance the quality of speech that has been degraded by background noise. Some of these areas include automobile interiors for hands free cellular, aircraft cockpits, voice communications using mobile telephone, automatic speech recognition (ASR) and speech coders[1]. speech enhancement has become more important than ever before. A speech enhancement system helps in increasing the quality of noisy speech[2].

We propose a novel approach to improve the performance of speech enhancement systems by combining tracking of non stationary noise algorithm[3] and De-noising Speech Signals by Wavelet Transform[4].

The problem of de-noising consists of removing noise from corrupted signal without altering it. Thus, we have evaluated our approach by evaluating speech quality. Reconstructed speech quality is measured with Perceptual Evaluation of Speech Quality (PESQ) score[5]. The PESQ measure was not generally intended to assess speech enhancement algorithms. However, it has been used in the past years in several speech enhancement algorithms. It converts the disturbance parameters in speech to a MOS-like listening quality score in a very wide range of conditions that may include codec distortions, errors, filtering, and variable signal delay. The higher score means better perceptual speech quality[6].

The simulation results show that the proposed speech enhancement method provide better speech quality compared to the traditional state-of-the-art methods using PESQ evaluation method.

In this paper various methods for speech enhancement methods have been introduced.

II. STATE-OF-THE ART OF SPEECH ENHANCEMENT ALGORITHMS

In this section we introduce seven of the most famous speech enhancement methods.

A. Tracking of Non-stationary Noise Based on Data-Driven Recursive Noise Power Estimation

We have to describe this method that was proposed by Jan S. Erkelens and Richard Heusdens[3]. The authors considers estimation of the noise spectral variance from speech signals contaminated by highly non-stationary noise sources. The method can accurately track fast changes in noise power level (up to about 10 dB/s). The enhancement algorithm is based on the minimum mean-square error (MMSE) [7]-[8] estimation in
the DFT (Discrete Fourier Transform) domain of speech spectral amplitudes. MMSE estimation of the noise power is to update the noise spectrum estimates with a reduced risk of speech leakage [3]. The MMSE estimates are obtained with the standard method of multiplying the noisy powers by a spectral gain function. This removes most of the speech contribution from the noisy spectrum, allowing for fast and accurate tracking of changing noise levels.

1) Prior SNR Estimator \( \hat{\xi}_{SE} \) for Speech Enhancement
For speech estimation, “decision-directed” estimator was used[3]:

\[
\hat{\xi}_{SE} (k, m) = \max \left[ \alpha_{SE} \frac{\hat{A}^2(k,m-1)}{\hat{\lambda}_D(k,m)} + (1-\alpha_{SE}) \left[ \frac{R^2(k,m)}{\hat{\lambda}_D(k,m)} - 1 \right], \xi_{\text{min}} \right]
\]

Where: \( k \) is frequency index in signal; \( m \) is frame index.
\( \hat{\lambda}_D \) is the noise variance.
\( \hat{A}^2 \) is the speech power estimate.
\( R^2 \) is the noisy power.
\( \alpha_{SE} \) is speech enhancement factor between 0 and 1.
\( \xi_{\text{min}} \) is a small value larger than 0 in [dB].

Where the latest available estimate of the noise variance \( \hat{\lambda}_D(k,m) \) was used[3]:

\[
\hat{\lambda}_D(k,m) = \alpha_s(k,m) \hat{\lambda}_D(k,m-1) + (1-\alpha_s(k,m)) \hat{D}^2(k,m)
\]

Where: \( \hat{D}^2(k,m) \) is the noise power.
\( \alpha_s(k,m) \) is the smoothing parameter (equation 8).

Note that the speech power \( \hat{A}^2 \) estimate is used in the first term instead of the square of the amplitude estimate \( \hat{A}^2 \) (the standard definition). The standard “decision-directed” estimator is the most commonly used estimator of prior SNR[3]:

\[
\hat{\xi}(k,m) = \max \left[ \alpha \frac{\hat{A}^2(k,m-1)}{\hat{\lambda}_D(k,m)} + (1-\alpha) \left[ \frac{R^2(k,m)}{\hat{\lambda}_D(k,m)} - 1 \right], \xi_{\text{min}} \right]
\]

2) Amplitude Gain Functions: The gain functions for \( \hat{A} \) and \( \hat{A}^2 \) are based on a generalized-Gamma speech amplitude prior [3]. The generalized-Gamma prior is given by:

\[
f_{\hat{A}}(a) = \frac{\gamma^\beta}{\Gamma(\nu)} a^{\nu-1} \exp(-\beta a^\nu), \beta > 0, \gamma > 0, \nu > 0, a \geq 0
\]

where \( \Gamma(.) \) is the gamma function, and \( \beta \) depends on \( \gamma \), \( \nu \) and \( \hat{\lambda}_s \). The random variable \( \hat{A} \) represents the DFT magnitude. The MMSE gain functions for \( \gamma = 1, \nu = 1 \) and for which the expressions can be found in [8]. For these parameter values, we have:

\[
\beta = \sqrt{\frac{\gamma}{\sqrt[4]{\hat{\lambda}_s}}}
\]

(\( \hat{\lambda}_s \) is the speech spectral variance that is the expectation of the speech power \( \hat{A} \)).

3) Noise Tracking
The steps taken in the noise tracking algorithm:
First, the prior SNR parameter \( \hat{\xi}_{NT} (k,m) \) and the posterior SNR \( \hat{\xi}(k,m) \) are estimated, using the latest available noise variance estimate \( \hat{\lambda}_D(k,m) \)[3]:

\[
\hat{\xi}_{NT} (k,m) = \max \left[ \alpha_{NT} \frac{R^2(k,m-1)}{\hat{\lambda}_D(k,m)} + (1-\alpha_{NT}) \left[ \frac{R^2(k,m)}{\hat{\lambda}_D(k,m)} - 1 \right], \xi_{\text{min}} \right]
\]

\( \alpha_{NT} \): is a factor of noise tracking between 0 and 1.

Next, the speech presence probability estimate \( \hat{p} \) is updated using[3]:

\[
\hat{\zeta}(k,m) = \sum_{i=-w}^{w} b(i) \hat{\zeta}(k-i,m), \text{with } \sum_{i=-w}^{w} b(i) = 1
\]

A rectangular window with \( w = 1 \) is used for \( b(i) \). Then a hard decision about speech presence is made:
Recent Advances on Electroscience and Computers

if $\zeta(k,m) > T(k,m)$

$$I(k,m) = \begin{cases} 1 & \text{speech present} \\ 0 & \text{speech absent} \end{cases}$$ (7)

else

$$I(k,m) = 0$$

end

Where: $T(k,m)$ is a threshold.

Otherwise, the speech presence probability determines the smoothing parameter, that is estimated by [3]:

$$\alpha_s(k,m) = \alpha_d + (1 - \alpha_d) \hat{P}(k,m)$$ (8)

The speech presence probability estimate is updated with a first order recursion [3]:

$$\hat{P}(k,m) = \alpha_p \hat{P}(k,m-1) + (1 - \alpha_p) I(k,m)$$ (9)

where $\alpha_p$ lies between 0 and 1. This estimate is used in equation (8) to find the smoothing parameter in (2).

The noise variance estimate is now updated using equation (2), where $\hat{D}^2$ is computed with a gain function found in [3].

Finally, for the speech spectral amplitude estimation, we compute prior SNR $\hat{\zeta}_E(k,m)$ from equation (1) and recompute posterior SNR $\zeta(k,m)$ from equation (10) using the new noise variance estimate [3]:

$$\zeta(k,m) = \frac{R^2(k,m)}{\hat{\lambda}_D(k,m)}$$ (10)

4) Safety Net

The method will react quite slowly to sudden, large jumps in the noise level. For these cases, the safety net ensures that the algorithms continue to work properly:

The idea is to push the noise variance estimate into the right direction when we detect that its value is much too low.

As a reference value, we use the minima $P_{\min}(k,m)$ of the smoothed values $\overline{P}(k,m)$ of the noisy power $R^2(k,m)$ in a short window of length $w_{\min}$, where $\overline{P}(k,m)$ is given by [3]:

$$\overline{P}(k,m) = \eta \overline{P}(k,m-1) + (1 - \eta) R^2(k,m)$$ (11)

where $\eta$ is a small smoothing parameter. After updating with $\hat{\lambda}_D$ (equation (2)), we check whether it fulfills the following condition [3]:

$$B P_{\min}(k,m) < \hat{\lambda}_D(k,m)$$ (12)

where $B > 1$ is a correction factor. In case of a large increase in noise level that the algorithm cannot follow $B P_{\min}(k,m)$, will become larger than $\hat{\lambda}_D(k,m)$ after a time of the order of the window length. If that happens, we reset the $\hat{\lambda}_D(k,m)$ values that violated (14) to max $\left[ B P_{\min}(k,m), \hat{D}^2(k,m) \right]$, and the corresponding $\hat{P}(k,m)$ to 0. The factor $B$ is taken larger than 1, but much smaller than the bias correction that would apply if the window $w_{\min}$ would contain only noise. This ensures that the safety net will not unintentionally come into action when some speech energy leaks into $P_{\min}$. We use very little smoothing of $R^2$ values (small $\eta$) to compute the minimum $P_{\min}$, because that allows us to keep the window $w_{\min}$ short. We have observed that the value of $B$ and the window length are not very critical for good performance, but a window length of at least 0.5 s is required [3].

B. Speech Enhancement Based on a Priori Signal to Noise Estimation

This method was proposed by P. Scalart and J. Vieira Filho (1996) [10]. Because The a Priori SNR estimation leads to the best subjective results. According to this conclusions, an approach was developed [10].

C. Geometric Approach (GA)

This recent method was proposed by Yang Lu, Philipos C. Loizou (2008) [11] that is A geometric approach to spectral subtraction Abstract. Yang Lu, Philipos C. Loizou presented a Geometric Algorithm (GA) to spectral subtraction based on geometric principles [11]. Unlike the conventional power spectral subtraction algorithm which assumes that the cross terms involving the phase difference between the signal and noise are zero, the algorithm makes no such assumptions. This was supported by error analysis that indicated that while it is safe to ignore the cross terms when the spectral SNR is either extremely high or extremely low, it is not safe to do so when the spectral SNR falls near 0 dB [11]. A method for incorporating the cross terms involving phase differences between the noisy (and clean) signals and noise was proposed [11]. Analysis of the suppression curves of the GA algorithm indicated that it possesses similar properties as the traditional MMSE algorithm (Ephraim and Malah, 1984) [11]. Objective evaluation of the GA algorithm showed that it performed significantly better than the traditional spectral subtraction algorithm in all conditions.

D. Harmonic Regeneration Noise Reduction (HRNR)

This method was proposed by Cyril Plapous, Claude Marro, and Pascal Scalart (2006) [12]. This approach addressed the
problem of single microphone speech enhancement in noisy environments. The well-known decision-directed (DD) approach drastically limits the level of musical noise but the estimated a priori SNR is biased since it depends on the speech spectrum estimation in the previous frame[12]. Therefore, the gain function matches the previous frame rather than the current one which degrades the noise reduction performance[12]. The consequence of this bias is an annoying reverberation effect. The authors proposed a method called Two-Step Noise Reduction (TSNR) technique which solves this problem while maintaining the benefits of the decision-directed approach. The estimation of the a priori SNR is refined by a second step to remove the bias of the DD approach, thus removing the reverberation effect. However, classic short-time noise reduction techniques, including TSNR, introduce harmonic distortion in enhanced speech because of the unreliability of estimators for small signal-to-noise ratios. This is mainly due to the difficult task of noise PSD estimation in single microphone schemes. To overcome this problem, a method called Harmonic Regeneration Noise Reduction (HRNR) was proposed. A non-linearity is used to regenerate the degraded harmonics of the distorted signal in an efficient way. These methods are analyzed and objective and formal subjective test results between HRNR and TSNR techniques are provided. A significant improvement is brought by HRNR compared to TSNR thanks to the preservation of harmonics[12].

**E. Phase Spectrum Compensation (PSC)**

This work was proposed by Anthony P. Stark, Kamil K (2008) [13]. In this paper a novel approach for speech enhancement has been presented, where the noisy magnitude spectrum is recombined with a phase spectrum compensated for additive noise distortion to produce a modified complex spectrum. Noise estimates are incorporated into the phase spectrum compensation procedure. During synthesis the low energy components of the modified complex spectrum cancel out more than the high energy components, thus reducing background noise [13].

**F. Speech enhancement using a priori SNR estimator**

This method was proposed by I. Cohen (2004) [14], where it based on a priori SNR estimator, minimum mean-square error (MMSE). The author proposed a non causal estimator for the a priori signal-to-noise ratio (SNR), and a corresponding non causal speech enhancement algorithm. In contrast to the decision directed estimator of Ephraim and Malah [15], the non causal estimator is capable of discriminating between speech onsets and noise irregularities [14]. Onsets of speech are better preserved, while a further reduction of musical noise is achieved. Experimental results show that the non causal estimator yields a higher improvement in the segmental SNR, lower log-spectral distortion, and better Perceptual Evaluation of Speech Quality scores (PESQ, ITU-T P.862) [14].

**G. Unbiased MMSE-Based Noise Power Estimation with Low Complexity and Low Tracking Delay.**

This method was proposed by T. Gerkmann and C. Richard (2012) [16]. It has been proposed to estimate the noise power spectral density by means of minimum mean-square error (MMSE) optimal estimation[16]. Otherwise, the resulting estimator can be interpreted as a voice activity detector (VAD)-based noise power estimator, where the noise power is updated only when speech absence is signaled, compensated with a required bias compensation[16]. The bias compensation is unnecessary when we replace the VAD by a soft speech presence probability (SPP) with fixed priors [16]. Choosing fixed priors also has the benefit of decoupling the noise power estimator from subsequent steps in a speech enhancement framework, such as the estimation of the speech power and the estimation of the clean speech[16]. In addition, the proposed SPP approach maintains the quick noise tracking performance of the bias compensated MMSE-based approach while exhibiting less overestimation of the spectral noise power and an even lower computational complexity[16].

**III. PROPOSED APPROACH**

Our approach to enhance speech is based on two speech enhancement methods. First method is Continues Wavelet Transform (CWT). Second method is Tracking of Non-stationary Noise Based on Data-Driven Recursive Noise Power Estimation[3] that is developed by Jan S. Erkelens and Richard Heusdens[3].

The performance of the proposed speech enhancement is evaluated in presence of different kind of noise using the NOIZEUS noisy speech corpus developed in Hu and Loizou laboratory[5] that is suitable for evaluation of speech enhancement algorithms.

**A. Continues Wavelet Transform (CWT)method :**

The motivation to use wavelet to achieve better noise reduction performance [4]. In our work, the Continues Wavelet decomposition of speech signal $S(t)$ uses adaptive level with Harr mother wavelet. Decomposition process produces 'N' vectors of wavelet coefficients according to adaptive threshold.

Wavelet transform is based on the idea of filtering a signal $S(t)$ with a dilated and translated versions of a prototype function $\Psi_{a,\tau}(t)$. This function is called the mother wavelet and it has to satisfy certain requirements [8]. The Continuous Wavelet Transform(CWT) for $S(t)$, is defined as [17]:

$$CWT(S, a, \tau) = \int_{-\infty}^{+\infty} S(t) \times \Psi_{a,\tau}(t) dt$$

Where:

$$\Psi_{a,\tau}(t) = \frac{1}{\sqrt{a}} \Psi\left(\frac{t-\tau}{a}\right), a \in \mathbb{R}^*$$

where: $a$ is the scale parameter and $\tau \in \mathbb{R}$ is the translation parameter. In addition to its simple interpretation, the CWT
satisfies some other useful properties such as linearity and conservation of energy [17].

1) level decomposition with Adaptive threshold
The number of level (scales) decomposition to be considered is according to the formula:

\[ p = 3 \times \left( \frac{\log (n)}{\log (2)} \right) \]  

(15)

Where: \( n \) is samples number. (We keep the integer number of \( p \)). In the analysis of speech signals, we calculate wavelet coefficients corresponding for each scale \( a = 1, \ldots, p \).

1) Which scale to be considered?
After wavelet coefficient calculation for each scale (equation 11), we assume it's sufficient to consider wavelet coefficients corresponding to a maximum energy of scale \( a \):

\[ E(a) = \sum_{i=1}^{n} |c_i|^2 \]  

(16)

\( E(a) \) : Energy corresponding to scale \( a \).
\( n \) : samples number of speech signal.
\( c \) : wavelet coefficient.

So, we adopt wavelet coefficients that concentrate more signal energy, it provides better reconstruction quality and introduce less distortion into processed speech. The speech signal to be reconstructed using these wavelet coefficients and passed through Jan S. Erkelens and Richard Heusdens algorithm (Tracking of Non-stationary Noise Based on Data-Driven Recursive Noise Power Estimation). Fig.1 shows wavelet coefficients energy of a speech signal taken from TIMIT database[18] as function of scale \( a \) where the wavelet coefficients to be used that are with a scale of \( a = 18 \).

B. De-noising procedure using CWT and Tracking of Non-Stationary Noise Based on Data-Driven Recursive Noise Power Estimation algorithm

The general de-noising procedure involves six steps. The procedure follows the steps described below:

1. Choose a wavelet (Haar wavelet).
2. Compute of level (scales) decomposition to be considered according to formula (15).
3. Calculate wavelet coefficient for each scale (formula 11).
4. Compute wavelet coefficient energy related to each scale (formula 16).
5. We consider wavelet coefficients matching to a maximum energy.
6. Passed wavelet coefficient through Tracking of Non-stationary Noise Based on Data-Driven Recursive Noise Power Estimation algorithm.

IV. SPEECH QUALITY ASSESSMENT

The perceptual speech quality was objectively measured using Perceptual Evaluation of Speech Quality method (PESQ)[19]-[5]. The PESQ method evaluates the quality of the speech signal by comparing the reference signal with the degraded signal. The PESQ algorithm models the human perception of the speech signal and thus enables the prediction of speech quality comparable to the subjective assessment as it would be performed by the human audience[18]. In this work we have adopted Loizou's PESQ implementation [5].

V. RESULTS AND DISCUSSION

A. Experimental Setup

To evaluate the proposed method, we have used speech signals taken from TIMIT database[18] in presence of white Gaussian Noise and NOIZEUS noisy speech corpus developed in Hu and Loizou laboratory[5]. The NOIZEUS corpus is suitable for evaluation of speech enhancement algorithms. The noisy database contains 30 IEEE sentences (produced by three male and three female speakers) corrupted by eight different real-world noises at different SNRs. The noise was taken from the AURORA[5] database and includes suburban train noise, babble, car, exhibition hall, restaurant, street, airport and train-station noise.

Parameter Settings: For the wavelet we have chosen Harr wavelet. For Tracking of Non-Stationary Noise algorithm (as the authors in [3] used), the following parameter settings are used in the experiments: \( \alpha_q \) in (8) is set to 0.85, \( \alpha_p = 0.1 \) in (9), and \( T(k,m) = 4 \) in (7) independent of time and frequency. We have used \( w = 1 \) and \( b(i) = 1 / (2^{w} + 1) \) in (6). The same value 0.98 is used for the smoothing parameters \( \alpha_{NT} \) in (5) and \( \alpha_{SE} \) in (1), and \( \xi_{\text{min}} \) is set to -19 dB. We use \( \eta = 0.1 \) in (13), \( B = 1.5 \) in (14), and the length of \( w_{\text{min}} \) spans 0.8 s.
B. Performance evaluation using TIMIT database

The proposed approach is objectively evaluated against several popular speech enhancement methods under noise conditions. We compare the proposed approach (CWT + Tracking of Non-Stationary Noise algorithm) with seven methods of state of the art (Tracking of Non-stationary Noise Based on Data Driven Recursive Noise Power Estimation, Speech Enhancement Based on a Priori Signal to Noise Estimation (P. Scalart 1996), Geometric Approach (GA), Harmonic Regeneration Noise Reduction (HRNR), Phase Spectrum Compensation (PSC), Speech enhancement using a priori SNR estimator (I. Cohen 2004) and Unbiased MMSE-Based Noise Power Estimation with Low Complexity and Low Tracking Delay).

Fig. 2 illustrates PESQ for various noise levels, obtained using proposed method, and various seven methods of state of the art. Where, white noise has been added to speech signal at several SNRs, from -5 to 30 dB in steps of 5 dB.

From fig. 2, it can be seen that the proposed method scores higher than the other methods in terms of the PESQ measure in presence of additive White Gaussian noise. Also, PSC and GA have the same PESQ scores (Curves are superposed).

C. Performance evaluation using NOIZEUS corpus

Otherwise the proposed approach was also evaluated in presence of several kinds of noise that are :Babble, Airport, Cart, Street, Restaurant. Where, The noise level varies between SNRs of -5 dB and 30 dB in steps of 5dB. The proposed approach was objectively evaluated against famous speech enhancement methods. (seven methods of state of the art).

The fig. 3 represents PESQ measure for the proposed approach and the seven state-of-the-art methods in presence of Babble noise. From this figure we can conclude the efficiency of our approach but HRNR method is the worst under 10 dB. PSC and GA methods have nearly the same PESQ scores. Otherwise, Unbiased MMSE-Based Noise Power and Tracking noise methods have almost the same PESQ.

The fig. 4 represents PESQ measure for the proposed approach and the seven state-of-the-art methods in presence of Airport noise. From this figure we can conclude the efficiency of our approach. PSC and GA methods have always nearly the same PESQ.

The fig. 5 represents PESQ measure for the proposed approach and the seven state-of-the-art methods in presence of Car noise. From this figure we can conclude the efficiency of our approach. PSC and GA methods have always nearly the same PESQ, but the P. Scalart(1996) method was the worst.

The fig. 6 represents PESQ measure for the proposed approach and the seven state-of-the-art methods in presence of Street noise. From this figure we can conclude the efficiency of our approach. PSC and GA methods have always nearly the same PESQ, but Tracking method was the second one after our approach in term of PESQ.

The fig. 7 represents PESQ measure for the proposed approach and the seven state-of-the-art methods in presence of Restaurant noise. From this figure we can conclude the efficiency of our approach. PSC and GA methods have always nearly the same PESQ. Unbiased MMSE-Based Noise Power and Tracking noise methods have nearly the same PESQ in this type of noise.

D. Performance evaluation against Run Time

As a third test we compare our approach with seven state of the art methods in terms of runtime. Table. 1 shows simulation results in terms of runtime, where we can observe that Geometric approach has less run time than the other algorithms, but our approach has more run time than four methods (not the best). In our simulation we have used a Laptop that is Intel (R) core (TM) i5-3210M CPU @ 2.5GHZ 2.50GHZ.

<table>
<thead>
<tr>
<th>Speech enhancement methods</th>
<th>Elapsed time [sec]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed method</td>
<td>0.4194</td>
</tr>
<tr>
<td>Noise tracking method</td>
<td>0.1972</td>
</tr>
<tr>
<td>Geometric Approach (GA)</td>
<td>0.0944</td>
</tr>
<tr>
<td>Phase Spectrum Compensation (PSC)</td>
<td>0.0981</td>
</tr>
<tr>
<td>Speech Enhancement Based on a Priori Signal to Noise Estimation</td>
<td>0.7808</td>
</tr>
<tr>
<td>(P. Scalart 1996)</td>
<td></td>
</tr>
<tr>
<td>Unbiased MMSE-Based Noise Power Estimation with Low Complexity</td>
<td>0.2058</td>
</tr>
<tr>
<td>and Low Tracking Delay (2012)</td>
<td></td>
</tr>
<tr>
<td>Harmonic Regeneration Noise Reduction (HRNR)</td>
<td>0.6113</td>
</tr>
<tr>
<td>Speech enhancement using a priori SNR estimator (I. Cohen 2004)</td>
<td>1.3023</td>
</tr>
</tbody>
</table>

Fig. 2. PESQ measure for the proposed approach and seven state-of-the-art methods in presence of additive White Gaussian noise. The noise level varies between SNRs of -5 dB and 30 dB in steps of 5 dB.
In this paper, we have provided a novel speech enhancement method which is robust to several kinds of noise (white Gaussian noise, car, babble, street, airport, and restaurant) and in comparison with seven state-of-the-art methods. For evaluating the performance of speech enhancement methods, we have used Perceptual Evaluation of Speech Quality scores (PESQ, ITU-T P.862).

Our approach is based on two methods, the first is Tracking of Non-stationary Noise Based on Data Driven Recursive Noise Power Estimation and the second is Continuous wavelet Coefficients where the wavelet coefficients to be used are from which the Energy is maximum. Otherwise, we have evaluated our approach in terms of runtime where it has more runtime than four methods thus, not the best in view of runtime. The usefulness of the proposed algorithm for some applications needs to be verified.

VI. CONCLUSION

In this paper, we have provided a novel speech enhancement method which is robust to several kinds of noise (white Gaussian noise, car, babble, street, airport, and restaurant) and in comparison with seven state-of-the-art methods. For evaluating the performance of speech enhancement methods, we have used Perceptual Evaluation of Speech Quality scores (PESQ, ITU-T P.862).

Our approach is based on two methods, the first is Tracking of Non-stationary Noise Based on Data Driven Recursive Noise Power Estimation and the second is Continuous wavelet Coefficients where the wavelet coefficients to be used are from which the Energy is maximum. Otherwise, we have evaluated our approach in terms of runtime where it has more runtime than four methods thus, not the best in view of runtime.

The usefulness of the proposed algorithm for some applications needs to be verified.
REFERENCES


