# EMD AND HURST BASED MODE SELECTION FOR SPEECH ENHANCEMENT

Anuja P Soman

A.Yogeshwaran M.E.

***Abstract***-This paper presents a speech sweetening technique for signals corrupted by non stationary acoustic noises. The pro-exhibit approach applies the empirical mode decomposition (EMD) to the rip-roaring speech signal and obtains a set of intrinsic mode functions (IMF). the most contribution of the projected procedure is that the adoption of the Hurst exponent within the selection of IMFs to reconstruct the speech. This EMD and Hurst-based (EMDH) approach is evaluated in speech sweetening experiments considering environmental acoustic noises with totally different indices of non stationary. The results show that the EMDH improves the segmental S/N Associate in Nursing and overall quality composite live, encompassing the sensory activity analysis of speech quality (PESQ). Moreover, the short-time objective understandability (STOI) measure reinforces the superior performance of EMDH. Finally, the EMDH is additionally examined in an exceedingly Speaker Identification task in rip-roaring conditions. The projected technique ends up in the highest Speaker Identification rates when compared to the baseline speech sweetening algorithms and conjointly to a multi condition coaching procedure.

***Index Terms*** — Empirical Mode Decomposition, Hurst Exponent, Index of non stationary, Speaker Identification, Speech Enhancement.

## I.INTRODUCTION

The suppression of acoustic distortion in rip-roaring speech signals is still an important research topic. The main issue of the speech sweetening techniques is bothered with the correct estimation of the noise statistics, significantly, in real non stationary environments. The classical estimators square measure supported voice activity detectors (VAD). The ability spectrum of the noise parts is then computed as a smoothened adaptation of its past values obtained throughout the speech pauses. These procedures show cheap accuracy for stationary background noises however they can not exactly estimate time-varying spectra. The difficult in following non stationary noises becomes additional. Manuscript received Gregorian calendar month thirty, 2013; revised January twenty nine, 2014; accepted March ten, 2014. Date of publication March 19, 2014; date of current version March thirty one, 2014. This work was supported partly by the National

Council for Scientific and Technological Development (CNPQ) underneath the 304254/2012-6 analysis grant. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Roberto Togneri. L. Zão is with the Graduate Program in Defense Engineering, Military Institute of Engineering (IME), First State Janeiro|Rio|city|metropolis|urban center} de Janeiro 22290-270, Brazil (e-mail: zao@ime.eb.br). R. Coelhois with the Electrical Engineering Department, Military Institute of Engineering (IME), First State Janeiro|Rio|city|metropolis|urban center} de Janeiro 22290-270, Brazil (e-mail: coelho@ime.eb. br). P. Flandrin is with the department of physics (UMR 5672 CNRS), Ecole Normale Supérieure Diamond State city, 69634 Lyon, France (e-mail: St. Patrick.Flan-drin@enslyon.fr). Color versions of 1 or additional of the figures during this paper square



ISSN: 2278 – 909X 

measure on the market on-line at http://ieeexplore.ieee.org. Digital Object Identifier ten.1109/TASLP.2014.2312541 evident for long speech segments and low signal/noise ratio (SNR). The minimum statistics (MS) [1] and also the improved minima controlled algorithmic averaging (IMCRA) [2] algorithms were planned to subsume these things. Thus, the estimation of the noise power spectrum is applied to every timeframe even throughout speech activity. However, these approaches are inaccurate in chase extremely non stationary noises [3]. Recent contributions, like the unbiased minimum mean-square error (UMMSE) [4] formula, are planned to estimate the power spectrum of non stationary noises with shorter delays. within the literature, time-frequency (TF) analysis, e.g. wavelets, have conjointly been adopted for speech sweetening. In such proposal [5] [6] the wavelet decomposition is applied to the noisy speech signal, and a choice criteria identifies the smallest amount corrupted parts before the reconstruction of the improved version of the speech signal. completely different from the facility spectrum based ways, the TF-based ones don't need specific estimation of the noise statistics. within the past few years, alternative TF speech sweetening solutions [7]–[9], supported the empirical mode decomposition (EMD) [10], have been introduced in the literature. The EMD is a nonlinear time-domain adaptive method for decomposing signals into a series of periodical intrinsic mode functions (IMF) and a residual. As opposition the rippling decomposition, the EMD doesn't need a group of basis functions to properly analyze the target signal. In fact, the IMFs obtained with the EMD rely solely on the target knowledge. Moreover, the EMD isn't restricted to stationary signals. In [7], the EMD-based detrending (EMD-DT) technique was planned to separate any reasonably target signal from a corrupting slowly-varying trend. The EMD-based filtering (EMDF) was given in [9] as a post-enhancement approach to get rid of residual low-frequency noise from antecedently increased speech signals. Though the EMDF showed promising objective quality results for speech corrupted with stationary noises, lower improvement was obtained with the non stationary Babble noise (refer to [9]). Speech sweetening techniques are usually evaluated in terms of the improvement in the speech quality. The segmental signal/noise ratio (SegSNR) and its frequency-domain version (the frequency-weighted SegSNR - fwSegSNR [11]) are samples of the normally used speech objective quality measures. The spectral subtraction (SS) [12], the minimum mean-square error short-time spectral amplitude (MMSE-STSA) [13] and also the optimally-modified log-spectral amplitude (OMLSA) [14] estimators are samples of approaches that bring home the bacon attention grabbing objective quality improvement. However, a comparative study [15] of those noise-reduction algorithms showed that they're ineffectual of accelerating the intelligibility. This case becomes more challenging in non stationary clanging eventualities attributable to the wrong noise statistics chase [16]. This paper introduces a unique EMD-based speech enhancement technique during which the noise elements of every United Nations agency square measure identified and designated by its Hurst exponent [17] statistics. The ensuing or least corrupted IMFs square measure wont to reconstruct the enhanced version of the speech signal. Within the planned EMDH technique, the IMFs choice and the speech reconstruction square measure performed on a frame-by-frame basis. The EMDH is investigated considering both quality and intelligibility objective measures. It is shown that the proposed approach achieves speech intelligibility gain even in extremely non stationary clanging conditions. The EMDH technique is additionally evaluated as a post-enhancement approach to the OMLSA and the Wiener filtering algorithm[18] with the UMMSE noise


ISSN: 2278 – 909X 

calculator [4]. The EMDH analysis experiments square measure conducted with speech signals corrupted with four real acoustic noises considering five completely different values of SNR. The experiments additionally embody the computation of the index of non stationary (INS) [19] of the acoustic noises. 5 baseline algorithms, namely SS, OMLSA, UMMSE, EMDF and EMD-DT, and 4 objective measures square measure adopted for the speech sweetening experiments. In terms of speech quality, the EMDH achieves the very best SegSNR and composite live results for the extremely non stationary noises (e.g., Babble). Moreover, it out performs the baseline EMDF and EMD-DT techniques for all the noise sources. The fwSegSNR and also the short-time objective comprehensibility (STOI) [20] measures square measure wont to measure the comprehensibility gain of the planned and baseline ways. The simplest fwSegSNR improvement is obtained for the EMDH as a post-enhancement approach to the UMMSE. relating to STOI, the EMDH outperforms the five baseline techniques. The speech enhancement with the EMDH is also examined in Speaker Identification (SI) experiments conducted in changing environments. The accuracy results show that the employment of speech utterances processed with the EMDH well improves the general SI performance compared to the clanging signals while not use of the EMDH. Moreover, the adoption of EMDH ends up in the simplest SI results when put next to the opposite speech sweetening techniques and additionally the employment of a multicondition coaching procedure [21]. This paper is organized as follows. Section II introduces the EMDH formula, as well as the fundamental ideas of the EMD and also the definition of the Hurst exponent. Descriptions of the baseline speech sweetening techniques square measure presented in Section III. the target measures wont to measure the EMDH performance in terms of speech quality and intelligibility square measure briefly delineate in Section IV. The speech enhancement experiments square measure elaborated in Section V. Then, the results obtained with the EMDH and also the baseline approaches square measure given and mentioned. In Section VI, the fundamental ideas regarding the speaker identification task are introduced. The SI accuracy results obtained with the speech enhancement techniques are given in Section VI. Finally, Section VII concludes this work.

## II.EMDHSPEECHSWEETENING TECHNIQUE

The first step of the planned EMDH speech sweetening technique is to decompose the clanging speech signal into a collection of IMFs using the EMD method Then, the Hurst exponent is computed on a frame-by-frame basis from each of the resulting IMFs to work out that of them square measure chiefly composed by noise. Finally, Associate in Nursing increased version of the speech signal is reconstructed victimization the remaining IMFs. Within the literature, the moving ridge decomposition has been wide used for time-frequency analysis. During this work, the EMD is adopted due to 2 main blessings over the wavelets-based approach. Firstly, the rippling decomposition is predicated on a group of pre-defined basis functions, that doesn't essentially fits well to any or all forms of signals.

Moreover, since it uses linear time-invariant filters, the rippling decomposition isn't adaptable to native or temporary variations within the input. On the opposite hand, the EMD analyzes the speech signal in a wholly adaptative means, and it's fully supported the native properties of the input signal. It makes the EMD suitable for non stationary signal analysis and additionally assures the completeness of the signal reconstruction exploitation the IMFs.


ISSN: 2278 – 909X 

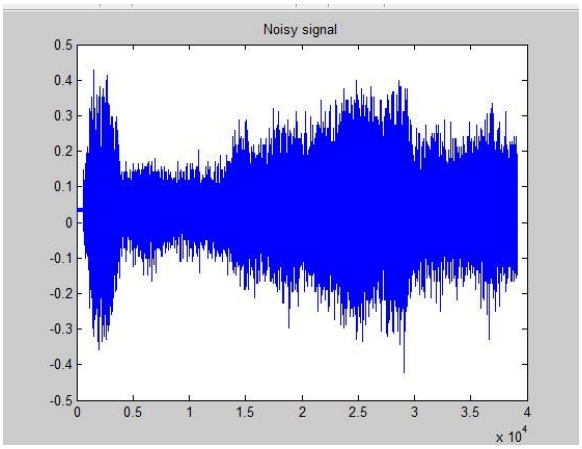


Fig 1.Noise Signal

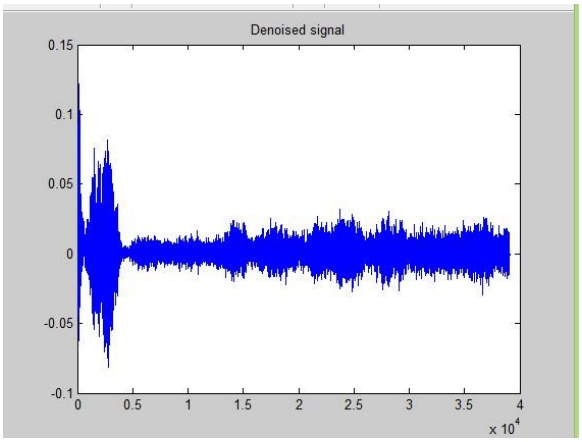


Fig 2.Denoising Signal

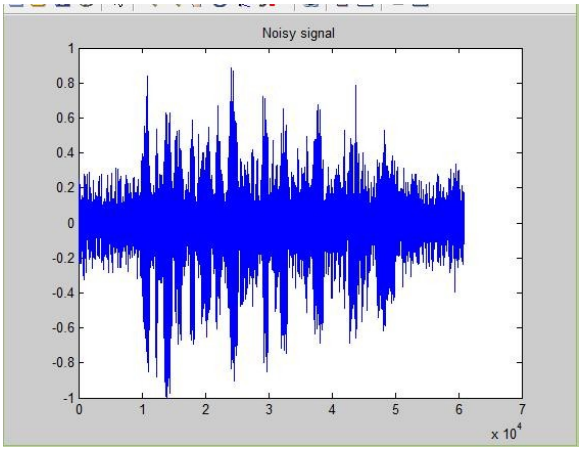


Fig 3.Noisy Speech Signal

## A. Empirical Mode Decomposition

The general plan of the EMD is to research a sign between 2 consecutive extrema (minima or maxima), and defines an area high-frequency half, additionally referred to as detail , and an area trend , such that the first UN agency is then composed of the native details, obtained from all the consecutive extrema of . The high versus, low-frequency separation procedure is iteratively continual over the residual, resulting in a brand new UN agency and a brand new residual. Fig. one illustrates the first five IMFs obtained from mouldering a sample speech section of five hundred ms collected from the TIMIT [22] info. The rule projected in [10] for mouldering the input are often summarized within the following steps: 1) Determine all extrema (local minima and maxima) of ; 2) Obtain the upper and lower envelopes by interpolating the native maxima and minima, respectively; 3) Calculate the native trend because the average between the higher and lower envelopes ; 4) Calculate the detail part as ; 5) Retell on the residual native trend . The IMFs should have zero mean and everyone their native maxima and minima should be positive and negative, respectively2. If the detail component, obtained in step4,doesnot follow these properties, steps one to four area unit continual with in situ of .This method, called sifting, is repeated until the new can be considered as Associate in Nursing UN agency. For following UN agency, the sifting method is applied on the residual . From the EMD rule, it are often noticed that the whole variety of extrema is reduced from one UN agency to following. Then every mode are often taken as a zero-mean amplitude and frequency modulated (AM-FM) signal. Note from Fig. one that the first UN agency consists of quicker oscillations than the second, which in its turn has faster fluctuations than the third, and soon. It means that , at each time interval, the EMD applies a high-frequency versus low-frequency separation between IMFs. Thus the


ISSN: 2278 – 909X

first modes must present the high-frequency content of the signal. Moreover, as can also be noted from Fig.1,the cut off frequency between consecutive IMFs is time-varying and signal dependent. Since the EMD rule will solely be applied if there area unit a minimum of 2 extrema within the last computed residual , any input are often rotten during a finite variety of IMFs. If the UN agency is denoted as and a complete of IMFs area unit extracted from where is the last residual obtained from the EMD algorithm. In [23], it absolutely was shown that, once applied to a complete mathematician noises (fGn), the EMD behaves sort of a 2 filter-bank with overlapping band-pass filters. During this analysis, the first UN agency is taken because the output of a high-pass filter with a non-negligible content in its lower half band .For the remaining modes, every UN agency is roughly composed of the higher half-band a part of the last residual that results from the previous iteration.

**B. EMDH**

Hurst-based International Monetary Fund choice The EMD formula states that, if a speech signal is decomposed, its reconstruction victimisation solely a set of the first IMFs, would result in the removal, at every timeframe, of the low-frequency elements . In [9], the authors showed that the energy content of a clean speech signal is usually targeted within the first four IMFs. Thus, they ended that any worth of in (2) is enough for a decent speech signal reconstruction. The continual line in Fig. 2(a) indicates the variance calculable from the samples of every IMF obtained from an different speech auditory communication collected from the TIMIT info, i.e., , wherever is that the total range of speech samples. it's noticeable that, in agreement with Fig. 1, there's a rise within the IMF energy (variance) from the first to the second IMF. Moreover, Fig. 2(a) conjointly shows that the modes with the very best indices ( ) gift lower energy values than the first ones. The broken line in Fig. 2(a) represents the variance values obtained with the speech section corrupted by a true works noise, extracted from NOISEX-92 info [24], with SNR of zero sound unit. Note the unexpected variance increase from IMFs five to nine, that is because of the low-frequency elements of the corrupting noise. The main issue of the proposed EMDH technique is the adoption of the Hurst exponent [17] to make a decision that IMFs ought to be chosen for the speech signal reconstruction. Let the speech signal be portrayed by a model , with the normalized autocorrelation coefficient operate (ACF, ) defined by where as that the mean of and is that the pause. The ACF of a half Gaussian noise is given by [25].Where is the Hurst exponent of .The value is defined by the ACF decaying rate whose asymptotic behavior .The Hurst exponent expresses the time-dependence or scaling degree of and is said to its spectral characteristics. inside the full vary , the facility spectral density may be shown to be proportional to once [25]. For , is constant over the full frequency spectrum (e.g., white noise), wherever as low frequencies area unit distinguished within the case where , and above all once or pink noise). as a result of such characteristics, the Hurst exponent was projected in [26] to compose a speech feature vector and with success applied to speaker recognition. During this work, the wavelet-based computer [27] was adopted to get the values of the IMFs on a frame-by-frame basis. The rippling-based Hurst computer may be represented in 3 main steps as follows: 1) rippling decomposition: the distinct wavelet rework (DWT) is applied to in turn decompose the input sequence of samples into approximation and detail coefficients3, wherever is that the decomposition scale (and is that the coefficient index of every scale. 2) Variance estimation: for every scale , the variance is evaluated from the detail coefficients, where is the number of available coefficients for every scale . In [27], it's shown


ISSN: 2278 – 909X 

that , wherever could be a constant. 3) Hurst computation: a weighted regression is employed to get the slope of the plot of versus . The Hurst exponent is calculable as the average values of different IMFs calculable from a TIMIT clean speech signal and also the same corrupted by the mill noise (Fig. 2(a)). The EMD is firstly used to decompose the speech signals. Then, the wavelet-based Hurst computer is applied to every IMF. The Hurst exponent is calculable from non-overlapping frames of 512 samples, that corresponds to thirty two ms with rate of sixteen kilohertz, victimization the Daubechies filters [28] with twelve coefficients and also the 3-12 scales. It may be seen that the first IMFs (e.g., 1-3), adore the high frequency parts, have . Moreover, for the very best IMF indices (e.g., 7-9) the values area unit near the unity, wherever the noise parts area unit typically targeted [29], [30]. This truth also can be discovered within the speech signal corrupted with the Factory noise, wherever the low-frequency energy content is focused on the IMFs . It shows that the exponent estimation permits the identification criteria to pick out the IMF low-frequency noise parts.

### C. EMDH Speech Signal Reconstruction

The EMDH algorithmic program starts with the decomposition of the input screaky speech into modes. Windowed IMFs (w-IMF) area unit then obtained by cacophonous every mode into non-overlapping short-time frames ,WIMFs elsewhere, where is that the frame index and is that the fixed time-duration of the frames. During a consecutive step, the rippling decomposition is applied to any or all the windowed IMFs, w-IMF , so as to estimate and store their Hurst exponent. Thus, a vector of Hurst values with components are obtained for each frame index .The next step is to determine for each frame , the index of the last .The subscript is employed to discriminate the detail and trend parts of EMD, from the detail and approximation coefficients of the rippling decomposition .Windowed IMF whose worth of is below a given threshold. If represents the improved speech signal, then every of its frames is reconstructed as wIMF and is finally given .In the projected EMDH, the IMF choice is completely supported the Hurst exponent calculable from short-time segments. This frame-by-frame analysis avoids that fast changes within the power spectrum of non stationary noises have an effect on the IMF choice of the whole speech signal. To avoid discontinuities, the subsequent procedure is applied within the signal reconstruction. Suppose that the speech frame is reconstructed with a smaller range of w-IMFs than ensuing. Thus, there's a minimum of one index specified w-IMF is enclosed within the reconstruction of frame , however w-IMF isn't in frame . Then, the samples of the half-right a part of} w-IMF area unit increased by the samples of the half-right part of the Hanning window whose size equals the frame duration. Therefore, the worth of the last sample of w-IMF turns to zero and also the continuity of the reconstructed speech signal is preserved. The analogous procedure is adopted once any IMF is employed within the reconstruction of frame and not of frame .

## III. SPEECH IMPROVEMENT BASELINE TECHNIQUES

This Section briefly describes the five baseline speech enhancement techniques adopted during this work. The SS, OMLSA and UMMSE apply the short-time Fourier rework (STFT) to firstly acquire an estimate of the noise power spectrum. Following, the identified noise parts area unit deducted or compansated-from the STFT of the screaky signal to boost the speech quality.

Spectral Subtraction be a speech auditory communication corrupted by AN additive noise . Thus, it may be written , wherever represents the clean speech signal. By


ISSN: 2278 – 909X 

applying the STFT to the on top of relation, it may be written where and are the frequency bin and the time frame indices, severally. The first step of SS [12], [31] is to estimate the noise power spectrum victimization the classical VAD-based approach. Then, the clean speech power spectrum is calculable as [31].In (10), the spectral floor parameter and also the time-varying over subtraction issue area unit set as in [31]. The spectrum of the improved signal is then calculable victimization the part of the screaky speech signal, and also the increased speech signal is( or pink noise). Attributable to such characteristics, the Hurst exponent was planned in [26] to compose a speech feature vector and with success applied to speaker recognition. During this work, the wavelet-based figurer [27] was adopted to get the values of the IMFs on a frame-by-frame basis. The moving ridge-based Hurst figurer is represented in 3 main steps as follows: 1) moving ridge decomposition: the separate wavelet remodel (DWT) is applied to in turn decompose the input sequence of samples into approximation and detail coefficients3, wherever is that the decomposition scale (and is that the coefficient index of every scale. 2) Variance estimation: for every scale , the variance is evaluated from the detail coefficients, where is the number of available coefficients for every scale . In [27], it's shown that , wherever may be a constant. 3) Hurst computation: a weighted regression toward the mean is employed to get the slope of the plot of versus . The Hurst exponent is calculable as the average values of different IMFs calculable from a TIMIT clean speech signal and also the same corrupted by the plant noise . The EMD is firstly used to decompose the speech signals. Then, the wavelet-based Hurst figure is applied to every International Monetary Fund. The Hurst exponent is calculable from non-overlapping frames of 512 samples, that corresponds to thirty two ms with rate of sixteen rate, exploitation the Daubechies filters [28] with twelve coefficients and also the 3-12 scales. It is seen that the first IMFs (e.g., 1-3), equivalent to the high frequency elements. Moreover, for the very best International Monetary Fund indices (e.g., 7-9) the values are near the unity, wherever the noise elements are typically focused [29], [30]. This reality also can be determined within the speech signal corrupted with the Factory noise, wherever the low-frequency energy content is focused on the IMFs . It shows that the exponent estimation allows the identification criteria to pick out the International Monetary Fund low-frequency noise elements.

### C. EMDH Speech Signal Reconstruction

The EMDH algorithmic rule starts with the decomposition of the input clattery speech into modes in keeping with (1). Windowed IMFs (w-IMF) are then obtained by cacophonic every mode into non-overlapping short-time frames, w-IMFIMF elsewhere, is that the frame index and is that the fixed time-duration of the frames. in an exceedingly consecutive step, the moving ridge decomposition is applied to all or any the windowed IMFs, w-IMF , so as to estimate and store their Hurst exponent. Thus, a vector of Hurst values, , with components are obtained for each frame index. The next step is to determine, for each frame, the index of the last finally reconstructed by overlapping and adding its inverse Fourier rework. The second baseline technique adopted during this work applies the IMCRA [2] to get AN estimate of the noise power spectrum. Then, the OMLSA [14] is employed to reconstruct the enhanced version of the clean speech. The IMCRA noise estimator consists of 2 iterations. Firstly, a VAD is defined supported the minimum noisy speech power spectrum values obtained from a collection of past frames. in an exceedingly second stage, this VAD is employed to work out the speech presence likelihood for every frequency bin and every timeframe. The noise


ISSN: 2278 – 909X 

power spectrum estimation is recursively given by where could be a time-varying smoothing parameter that depends on . When the noise spectrum estimation, the OMLSA technique reconstructs the improved speech signal by minimizing the mean-square error of the log-spectral amplitude. The gain function that ends up in the spectral amplitude of the optimally reconstructed speech is defined where could be a perform of the a priori SNR , and also the minimum worth is defined by a subjective criteria. All the parameters employed in the OMLSA and IMCRA implementation, as well as the bias compensation issue for the noise estimation, are an equivalent as those adopted . Within the third speech improvement baseline procedure, the un- biased minimum mean-square error (UMMSE) noise power estimation [4] is adopted to trace the noise spectrum. during this proposal, the authors combined speech presence uncertainty to the calculator originally projected in [32], and located that the estimation of the noise power spectrum may be updated on every occasion frame via the algorithmic smoothing could be a smoothing issue and also the noise periodogram estimate depends on the speech presence and absence possibilities and on the noise power spectrum calculable from the last frame. The most issue of adopting the UMMSE is that, not like IMCRA, it doesn't need a minimum search among a given range of past frames. It ends up in shorter delays within the noise estimation. Besides, UMMSE doesnot require a bias compensation issue. Following the procedure in [4], the UMMSE noise calculator is followed by the speech improvement algorithmic program projected in [18]. The Wiener filtering gain relies on the estimation of the a priori SNR, that is obtained with the decision-directed approach projected in [13].

## VII. CONCLUSION

This paper has introduced a brand new speech sweetening technique supported EMD and on a Hurst-based IMF selection criteria. The Hurst exponent statistics is adopted to spot and select those IMFs that are most tormented by the noise components. The speech signal is finally reconstructed considering the least corrupted IMFs. many experiments were conducted using four different acoustic noises, three of them non stationary. The EMDH performance was compared to five baseline speech sweetening algorithms, and it had been conjointly evaluated as a post- sweetening approach. The planned technique improved four objective measures that are extremely correlated with speech quality and intelligibility. Moreover, the EMDH outperformed the baseline EMDF and EMD-DT approaches for many of the condition. The superior performance of the planned speech sweetening technique was conjointly verified within the speaker identification experiments conducted in yelling environments.

## REFERENCES

[1] R. Martin, ―Noise power spectral density estimation based on optimal smoothing and minimum statistics,‖ *IEEE Trans. Speech Audio Process., vol. 9, no. 5, pp. 504–512, Jul. 2001.*

[2] I. Cohen, ―Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging,‖ *IEEE Trans. Speech Audio Process., vol. 11, no. 5, pp. 466–475, Sep. 2003.*

[3] K. Manohar and P. Rao, ―Speech enhancement in non stationary noise environments using noise properties,‖ *Speech Commun., vol. 48, pp.* 96–109, Jan. 2006.

[4] T. Gerkmann and R. Hendriks, ―Unbiased MMSE- based noise powerestimation with low complexity and low tracking delay,‖ *IEEE Trans. Audio, Speech, Lang. Process., vol. 20, no. 4, pp. 1383–1393, May*2012.



ISSN: 2278 – 909X 

[5] D. Donoho and I. Johnstone, ―Threshold selection for wavelet shrinkage of noisy data,‖ in *Proc. 16th Annu. Int. Conf. IEEE Eng.Med. Biol. Soc. (EMBC'94), Nov. 1994, vol. 1, pp. A24–A25.*



ISSN: 2278 – 909X 