Int J Speech Technol (2015) 18:157–166
DOI 10.1007/s10772-014-9255-3

A wavelet based method for removal of highly non-stationary noises from single-channel hindi speech patterns of low input SNR

Sachin Singh · Manoj Tripathy · R. S. Anand

Received: 6 May 2013 / Accepted: 30 September 2014 / Published online: 14 October 2014

Abstract This paper presents a binary mask thresholding function in Doubachies10 wavelet transform for enhancement of highly non-stationary noise mixed single-channel Hindi speech patterns of low (negative) SNR. In the wavelet transform, a five level of decomposition is used and detailed coefficients of all five levels are given to binary mask thresholding function for removing noise and enhancing the speech patterns. The robustness of the proposed method is compared with the wildly popular methods such as log-mmse, test-psc, Wiener, IdBM, and spectral-subtraction on the basis of performance measure parameters viz SNR, PSNR, PESQ, and Cepstrum distance. The algorithms were implemented in MATLAB 7.1.

Keywords Speech enhancement · Hindi speech patterns · SNR · PESQ · Cepstrum distance · Wavelet transform

1 Introduction

Speech is an effective way of communicating ideas from one person to another. When speech signal propagates through a highly non-stationary noisy medium then it may be distorted to a severely degraded level. Daily life noise patterns like pop music, exhibition hall, multi-talker babble, and restaurant are some of the examples of highly non-stationary noises that create maximum distortion in speech patterns. The distorted speech may become meaningless and sharp deterioration is evident in the performance of speech communication system Zhang (2010). Hence for removal/suppression of this highly-non-stationary noise from speech patterns an effective speech enhancement system is required. From the listener's point of view, the other purpose of speech enhancement is to improve the speech intelligibility and clarity in speech patterns for better understanding of speech signals.

There are various speech enhancement methods available in the literature (Singh et al. 2014; Boll 1979; McAulay and Malpass 1980; Ephraim 1992; Dendrinos et al. 1991; Ephraim and Trees 1995; Jensen and Hansen 1995; Yi and Loizou 2004; Bahoura and Rouat 2006; Johnson et al. 2007; Hongyan et al. 2008; Jie and Heping 2012; Farah and Celia 2012; Singh et al. 2014; Gabor 1946; Singh et al. 2013; Goupillaud et al. 1984). In general, the speech enhancement methods may be classified into four groups, i.e., spectral-subtractive algorithms Singh et al. (2014), statistical-model based algorithms McAulay and Malpass (1980); Ephraim (1992), subspace algorithms Ephraim (1992); Dendrinos et al. (1991); Ephraim and Trees (1995), and wavelet transform (WT) (Yi and Loizou 2004; Bahoura and Rouat 2006; Johnson et al. 2007; Hongyan et al. 2008; Jie and Heping 2012; Farah and Celia 2012). Wavelet transform analysis gives information in frequency and time both; it is a time-frequency analysis which is highly suitable for analysis of mixed highly non-stationary noise with speech signal in comparison to traditional Fourier transform analysis (Singh et al. 2014; Gabor 1946). Wavelet transform provides a multi-resolution analysis of highly non-stationary speech signals by using long windows for low frequencies and short windows for high frequencies Goupillaud et al. (1984). Hence, wavelet is much effective in all noisy environments for enhancement of speech patterns. Basically, soft and hard thresholds are the most commonly used function in wavelets for speech

S. Singh (✉) · M. Tripathy · R. S. Anand
Department of Electrical Engineering, Indian Institute of Technology Roorkee, Roorkee 247667, India
e-mail: oxygen_sachin@rediff.com; sachinsingh.iitr@gmail.com

M. Tripathy
e-mail: manojfee@iitr.ac.in

R. S. Anand
e-mail: anandfee@iitr.ac.in

enhancement. The residual noise remains in enhanced speech due to discontinuity of hard threshold function while the soft threshold is continuous but it gives an unavoidable error at time of speech reconstruction.

In order to improve quality and intelligibility of speech patterns, in this paper, a binary mask threshold function based db10 wavelet transform method is proposed for enhancement of highly non-stationary noises mixed speech patterns of low (negative) SNR. The binary mask threshold value is consider as −5 dB. The input pattern of noisy speech is decomposed in five levels. After thresholding, reconstruction of speech patterns is performed and results show the effective improvement in performance parameters of enhanced speech patterns.

The remaining paper is organized as follows: Sect. 2 presents five levels of wavelet decomposition and binary mask thresholding function, followed by the adopted procedure for proposed method in Sect. 3. Simulation conditions, parameters used and performance analysis of different methods are given in Sect. 4 for enhancement in highly non-stationary noise of low (negative) SNR mixed single-channel Hindi speech patterns. Finally, conclusions are drawn in Sect. 5.

2 Background

2.1 Observation model

Generally noisy speech signal in time domain, recorded by single microphone is given as:

$$y(n) = x(n) + v(n) \tag{1}$$

where $y(n)$, $x(n)$ and $v(n)$ denotes the noisy speech, clean speech and highly non-stationary additive background noise, respectively. To obtain clean speech patterns from the noisy speech patterns a specific gain i.e. wiener gain is applied to each spectral component which is represented as:

$$x(k,t) = g(k,t) * y(k,t) \tag{2}$$

where $g(k,t)$ denotes the gain function, and $x(k,t)$ and $y(k,t)$ denotes the estimate of clean Hindi speech and noisy speech spectrum respectively. t is time frame and k is frequency bin.

2.2 Wavelet transform

Wavelet is a mathematical function which is used to divide a given function into different scale components. The wavelet transform can be divided into two main categories. First is continuous wavelet transform (CWT) and second is discrete wavelet transform (DWT). The substantial redundant information is generated from CWT, since it is given by continuously scaling and translating the mother wavelet. But the mother wavelet can also be scaled and translated using specific subset of scale and translation values or representation grid. Hence DWT is more efficient. It has a high frequency resolution in low bands and low frequency resolution in high bands. It is very helpful for speech signal processing. The DWT $W(m,n)$of signal $f_{(t)}$ with respect to a wavelet $\phi_{(t)}$is given as:

$$W(m,n) = 2^{-(m/2)} \sum_{t=0}^{T-1} f_{(t)} \phi_{((t-n.2^m)/2^m)} \tag{3}$$

where T is the length of the signal $f_{(t)}$. m and n are the integer variables in the functions of scaling and translation parameters. t is the discrete time index.

After many experimentation with different types of mother wavelets such as Daubechies (order 1–40), Symlets (order 2–8), Coiflets (order 1–5), and BiorSplines (order 1.1–6.8) db10 is found suitable for my application due to the similarity of order and shape of noisy speech patterns. Since the performance of the DWT type wavelets depends on shape and order of different mother wavelets. In each step of wavelet transform, a particular scaling function is applied to input data. If input data has N number of samples, the scaling function will calculate N/2 smoothed values. In the ordered wavelet transform the smoothed values will be stored in the lower half of the N element input vector.

In wavelet analysis, noisy speech pattern is split into two types of coefficients, namely approximated and detailed coefficients. In this paper five levels decomposition of wavelet coefficients is used since needful information of speech pattern is remains in low frequency band and highest level decomposition gives maximum resolution in lower frequency band of input speech pattern. The high-pass filter gives detail coefficients and low-pass filter gives approximation coefficients. In five levels of decomposition we get five detailed coefficients D1, D2, D3, D4, D5 and fifth level approximate coefficients for analysis of highly non-stationary noisy speech pattern. The levels of decomposition are given in Fig. 1.

3 Proposed method

3.1 Binary mask threshold function

In Eq. 2, a Wiener gain function is used for computation of clean speech spectrum since this gain function is very effective in terms of speech quality and intelligibility Scalart and Filho (1996). This wiener gain function is depends on

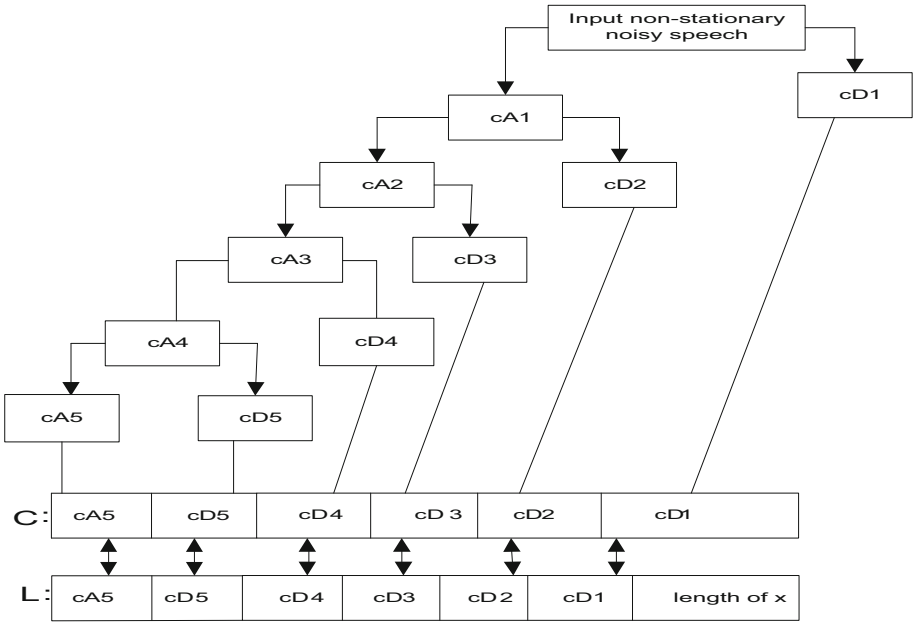

Fig. 1 Block diagram of five level wavelet decomposition

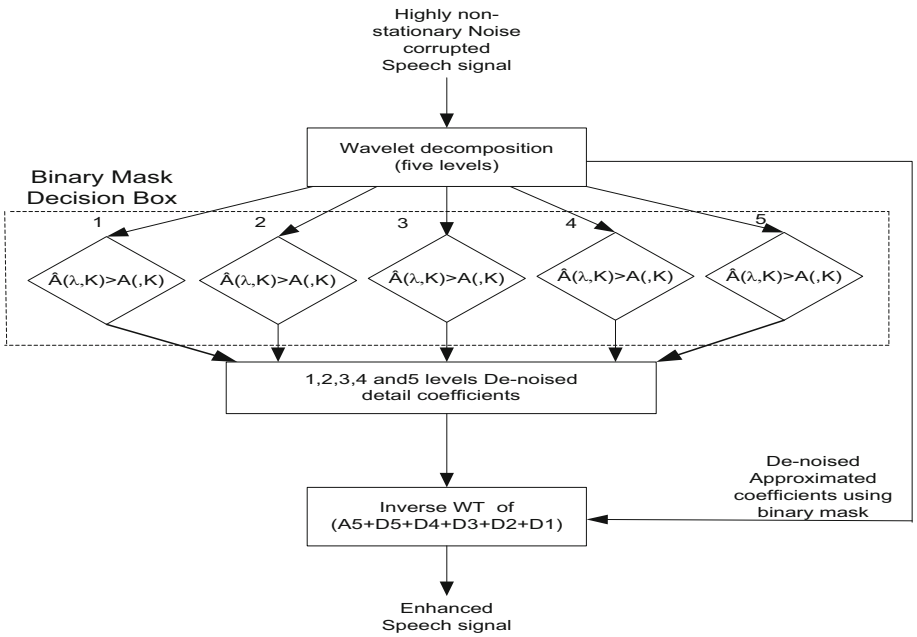

Fig. 2 Flow chart of the proposed method for enhancement in single-channel speech patterns

priori SNR and this gain calculation is based on the following equation Rangachari and Loizou (2006)

$$g(k,t) = \sqrt{\frac{priori\ SNR}{1 + priori\ SNR}} \tag{4}$$

Now, the overall estimate of the noise spectral magnitude by using noisy wiener gain function is given as:

$$\hat{A}(k,t) = g(k,t).y(k,t) \tag{5}$$

where $g(k,t)$ is a noisy Wiener gain function. On the basis of this estimated noise spectrum $\hat{A}(k,t)$ a binary mask is constructed. If estimated noise magnitude spectrum is greater than true noise magnitude spectrum then it is a condition of noise underestimation distortion. This over/under estimation is compared for each time-frequency bin (k,t) (Fig.2). In comparison of time-frequency bin procedure, the time-frequency bins satisfying the constraint were recovered, while time-frequency bins violating the constraints were zeroed out. The threshold value for binary mask threshold function is considered as -5 dB for maximum improvement.

Table 1 Output SNR measures values obtained for noisy and enhanced speech pattern

Noise type	Enhancement techniques	SNR(dB)				
	Noisy	−5	−10	−15	−20	−25
Babble	Log-mmse	3.7108	−0.3911	−2.5796	−2.9525	−2.0269
	Test PSC	1.6703	0.3835	0.0480	−0.0016	−0.0037
	Wiener	6.3249	3.0403	.1152	−1.2639	−1.0463
	1 dBM	7.6608	6.5900	3.6249	2.2790	−0.9071
	Spectral Sub	6.7048	3.6259	−0.2930	−0.9697	−0.6855
	Proposed	6.8364	5.4287	4.1940	2.7960	1.0903
	Noisy	−5	−10	−15	−20	−25
Pop music	Log-mmse	2.7118	−2.0980	−6.5386	−6.3986	−4.6131
	Test PSC	2.6905	0.5407	−0.0502	−0.0578	−0.0228
	Wiener	6.2297	3 0346	0.2514	−1.6418	−1.9570
	IdBM	5.1158	3.2037	3.3963	1.7873	−1.9981
	Spectral sub	6.6818	3.6425	0.7956	−1.6078	1.7354
	Proposed	5.0613	5.3938	3.9275	2.7797	1.1919
	Nisey	−5	−10	−15	−20	−25
Restaurant	Log-mmse	4.6247	1.6238	−1.8196	−4.6092	−4.5864
	Test PSC	1.5466	0.4657	0.1198	0.0343	0.0119
	Wiener	6.9373	4.1357	1.3837	−1.4293	−2.5928
	IdBM	6.9165	3.4483	205021	2.1622	−0. 5092
	Spectral sub	7.3886	4.4730	1.6435	−1.1716	−2.1098
	Proposed	804151	6.2230	4.4927	3.0149	1.0635
	Noisy	−5	−10	−15	−20	−25
Exhibition	Log-mmse	5.3217	2.6023	0.3139	−1.6629	−4.1718
	Test PSC	2.3494	0.6718	0.1239	−0.0023	−0.0143
	Wiener	7.8869	5.3969	2.9062	0.8218	−1.2383
	IdBM	7.3703	5.7937	6.2206	2.9411	1.9398
	Spectral sub	8.5415	6.0165	3.6562	1.4411	−0.9323
	Proposed	9.5343	7.0761	4.0173	3.9555	2.1062

Using this concept, the modified speech magnitude spectrum is recovered as Hu and Loizou (2007):

$$x_{enhanced} = \begin{cases} x(k,t), & \text{if } \widehat{A}(k,t) > A(k,t) \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

where $\hat{A}(k,t)$, $A(k,t)$is estimated and a true noise magnitude spectrum for time-frequency bin (k,t). An inverse wavelet transform is applied to compute enhanced speech spectrum and finally overlap-and-add technique is used to synthesize the noise-suppressed speech signal.

The input signal is obtained by amalgamation of single-channel Hindi speech pattern and additive back ground noises such as pop music, exhibition, restaurant, and babble etc. The db10 wavelet transforms is used for decomposition as db10 is suitable and gives better results for this application. In next step these three level detailed coefficients are recovered for same number of samples as in input speech. Now these detailed coefficients D1, D2, D3, D4, D5 are given to binary mask threshold function for noise suppression. The denoised coefficients are obtained by applying binary mask design on the detailed coefficients (D1 to D5) and approximated coefficients A5 of input noisy speech pattern. The Inverse Wavelet Transform is applied on the sum of detailed and approximated coefficients, to get enhanced speech pattern. The capability of proposed method is measured in terms of performance measure parameters to show the improvement in quality and intelligibility of Hindi speech pattern.

4 Simulations and results discussion

4.1 Simulation conditions

In this experiment, the clean speech patterns of Hindi language have been taken from IIIT-H (International Institute of

Table 2 Cepstrum Distance measures values obtained for noisy and enhanced speech pattern

Noise type	Enhancement techniques	Cepstrum distance				
	Noisey	−5	−10	−15	−20	−25
Babble	Log-mmse	4.5921	4.9528	5.4306	6.1142	6.7387
	Test PSC	8.2647	8.6265	8.8290	8.8998	8.9039
	Wiener	4.3933	5.0254	5.7910	6.5810	7.2071
	IdBM	5.4830	6.2020	7.1333	7.7539	8.6447
	Spectral sub	5.6965	6.3803	7.0233	7.6207	7.9614
	Proposed	4.4359	4.8991	5.2019	5.4335	5.4520
	Noisy	−5	−10	−15	−20	−25
Pop music	Log-mmse	5.2935	5.7240	6.2458	6.7787	7.3119
	Test PSC	7.5246	7.8952	8.0725	8.1274	8.1438
	wiener	4.8802	5.6672	6.3720	6.9722	7.5200
	IdBM	6.4523	7.4001	8.0207	8.4666	8.8016
	Spectral sub	7.0658	7.6249	8.2241	8.5619	8.6890
	Proposed	5.0162	5.5138	5.9538	5.9862	5.9638
	Noisy	−5	−10	−15	−20	−25
Restaurant	Log-mmse	4.8125	4.7132	4.8997	5.1571	5.6619
	Test PSC	8.1935	8.5898	8.8087	8.89100	8.8966
	Wiener	3.9227	4.3357	4 9270	5.4738	5.8093
	IdBM	5.5893	6.4502	7.4589	8.0319	8.5047
	Spectral sub	5.0776	5.8999	6.3150	6.7133	7.2347
	Proposed	4.4982	4.9467	5.2900	5.3374	5.3237
	Noisy	−5	−10	−15	−20	−25
Exhibition	Log-mmse	5.7104	6.2065	6.6398	7.2652	7.8753
	Test PSC	8.0783	8.4198	8.5976	8.6747	8.7115
	Wiener	5.5191	6.2476	6.8742	7.5502	8.1564
	IdBM	6.6778	7.3998	7.8321	8.1800	8.34351
	Spectral sub	7.7872	8.3901	8.9271	9.2187	9.3881
	Proposed	5.2649	6.0708	6.2453	6.3502	6.2115

Information Technology Hyderabad) Indic speech database Prahallad et al. (2012), which is spoken by a female speaker. This database is consists of 1000 speech patterns. The clean speech patterns of Hindi language have been added with four different types of highly non-stationary noise source patterns at different levels of signal-to-noise ratio (SNR) ranging from −5 to −25 dB (in 5 dB steps). These highly non-stationary noise sources (babble, exhibition, restaurant, and pop music) are taken from AURORA database Pearce and Hirsch (2000). The sampling rate of noise and speech pattern is 16 kHz.

4.2 Performance parameters

The performance of the methods is compared on the basis of subjective and objective measurement. The output SNR, perceptive evaluation of speech quality (PESQ), Peak-SNR, and Cepstrum distance measure are taken for evaluation of enhanced speech signal.

SNR is a ratio of RMS amplitude value of signal A_{signal} and noise A_{noise}. It is an objective parameter measure of speech quality. It is given in dB as:

$$SNR_{dB} = 10\log_{10}\left[\left(\frac{A_{signal}}{A_{noise}}\right)^2\right] \tag{7}$$

Peak-SNR is a ratio between maximum possible power of a clean speech signal and the power of the corrupting noise signal. It is calculated as:

$$PSNR = 10\log_{10}\left(\frac{MAX^2_{signal}}{\sqrt{MSE}}\right) \tag{8}$$

where MSE is mean square error difference between clean and enhanced Hindi speech signal.

Table 3 PESQ measures values obtained for noisy and enhanced speech pattern

Noise type	Enhancement techniques	PESQ				
	Noisy	−5	−10	−15	−20	−25
Babble	Log-mmse	2.1676	1.8356	1.5905	1.3440	1.0751
	Test PSC	0.9854	0.3448	0.6554	0.3286	0.3126
	wiener	2.3672	1.9566	1.6550	1.2701	0.8180
	IdBm	2.5137	2.1104	1.7599	1.3934	1.0462
	Spectral sub	2.1904	1.7972	1.4384	0.9406	0.6061
	Proposed	2.3538	2.1889	1.9834	1.8244	1.6987
	Noisy	−5	−10	−15	−20	−25
Pop Music	Log-mmse	2.0154	1.6723	1.5449	1.4590	1.4132
	Test PSC	0.9996	0.6905	0.4257	0.4623	0.4630
	Wiener	2.4156	2.1180	1.7778	1.5468	1.2612
	IdBM	2.3403	1.9484	1.7709	1.5985	1.2230
	Spectral sub	2.0594	1.8465	1.4637	0.9515	0.3823
	Proposed	3.0659	2.8641	2.6112	2.3759	2.1906
	Noisy	−5	−10	−15	−20	−25
Restaurant	Log-mmse	2.1506	1.9256	1.7537	1.6989	1.5302
	Test PSC	0.9895	0.7448	1.0114	0.9777	0.8514
	Wiener	2.4108	2.1613	1.8726	1.5775	1.2469
	IdBM	2.6388	2.2600	1.8659	1.5832	1.2471
	Spectral sub	2.2379	1.8774	1.4880	1.2344	.8345
	Proposed	2.5046	2.2852	2.1121	1.9329	1.6065
	Noisy	−5	−10	−15	−20	−25
Exhibition	Log-mmse	2.3856	2.1236	1.8751	1.6452	1.5514
	Test PSC	0.9279	0.6747	0.8491	1.0593	0.8192
	Wiener	2.6743	2.4085	2.1916	1.9271	1.7671
	IdBM	2.5058	2.1417	1.8827	1.6182	1.4338
	Spectral sub	2.2129	1.9737	1.7409	1.4215	1.1908
	Proposed	2.7390	2.4769	2.2463	2.0383	1.7918

PESQ is an algorithm that analyzes the enhanced Hindi speech signal sample-by-sample after a temporal alignment of enhanced and clean Hindi speech signal. It is standardized as ITU-T recommendation P.862 (02/01) for objective voiced speech quality testing. Mapping function of PESQ is given as PESQ (2003):

$$PESQ = 0.999 + \left(\frac{4.999 - 0.999}{1 + e^{-1.4945*x+4.6607}}\right) \tag{9}$$

where x is enhanced speech Hindi speech signal. PESQ algorithms basically gives mean opinion score (MOS) in the range from 1 (bad) to 4.5 (excellent).

Cepstrum distance measure is distortion measure between input and output speech signal that is classified into frequency domain. Cepstrum distance measure corresponds to the best parameter for subjective measures among the several spectral envelope calculating methods based on the LPC methods Kitawaki and Nagabuchi (1988). It is calculated as:

$$CD = 10/\log_{10}\sqrt{2\sum_{i=1}^{P}\{y(k,t) - x(k,t)\}^2} \tag{10}$$

where CD is a measure for Cepstral distance and $y(k,t)$ and $x(k,t)$ are the input and output speech signal respectively. P is maximum number of coefficients.

4.3 Results and discussion

The performance of proposed method is compared with most commonly used methods such as log-mmse Ephraim and Malah (1985), test-psc Stark (2008), wiener Scalart and Filho (1996), IdBM Wojcicki and Loizou (2012), and spectral-subtraction Boll (1979). Four performance measure parameters viz PESQ, Output SNR, PSNR and Cepstrum distance are taken for comparative analysis. There are four types of highly non-stationary noise sources such as babble,

Table 4 PSNR measures values obtained for noisy and enhanced speech pattern

Noise Type	Enhancement Techniques	PSNR				
	Noisy	−5	−10	−15	−20	−25
Babble	Log-mmse	84.3128	81.5772	79.8320	73.3725	76.3187
	Test PSC	77.5070	72.7041	67.8543	62.9349	57.9677
	Wiener	85.0759	82.4410	79.4442	76.9782	73.9669
	IdBM	85.1320	82.9113	81.4650	80.4815	81.0391
	Spectral sub	85.3994	82.1010	78.8265	75.9751	72.3979
	Proposed	85.4745	83.7882	82.4789	81.9098	81.3309
	Noisy	−5	−10	−15	−20	−25
Pop Music	Log-mmse	84.1593	81.8020	80.1762	79.1982	78.2038
	Test PSC	80.5551	75.9462	71.2744	66.4910	61.5367
	Wiener	85.7978	83.0839	80.5494	78.4820	76.4702
	IdBM	82.7060	79.9949	81.9213	81.1163	80.6633
	Spectral sub	85.5715	82.6149	80.0435	77.8019	75.1476
	Proposed	83.1171	83.8704	82.1754	81.4347	80.7233
	Noisy	−5	−10	−15	−20	−25
Restaurant	Log-mmse	84.6711	82.5734	80.7331	79.4503	78.9621
	Test PSC	76.2237	71.3355	66.4141	61.4539	56.4787
	Wiener	86.0480	83.3609	81.1628	79.4979	78.4699
	IdBM	84.9111	80.0883	78.8253	80.5975	79.5827
	Spectral sub	86.0760	83.0936	80.8586	73.6507	76.7566
	Proposed	87.3027	84.7457	82.7047	81.5159	80.2267
	Noisy	−5	−10	−15	−20	−25
Exhibition	Log-mmse	85.7696	83.8372	82.2304	81.0059	79.7005
	Test PSC	79.3010	74.5761	69.7618	64.8544	59.8960
	Wiener	87.3602	85.1690	83.0141	81.4138	79.9837
	IdBM	85.6916	84.0035	85.2219	81.3822	80.0837
	Spectral sub	87.6333	85.2319	83.1721	81.4621	79.9725
	Proposed	88.7803	86.1140	82.3175	83.0691	80.7640

pop music, restaurant, and exhibition, which are taken at different levels of input SNR varying from −5 to −25 dB. The performance parameters are given in Tables 1, 2, 3 and 4, which report the objective measures obtained for noisy and enhanced single-channel Hindi speech pattern.

The output SNRs values are given in Table 1. The proposed method give highest output SNR values at all levels of input SNR (i.e. varied from −25 to −5 dB) except at −10 dB. The proposed method gives second highest value of output SNR 5.4287 at −10 dB whereas IdBM method gives maximum output SNR of 6.5900 dB for Hindi language. For the pop music and exhibition noise sources case, proposed method shows maximum improvement in output SNR except at −5 and −15 dB, respectively. The maximum output SNR values are shown for the restaurant noise case by proposed method in comparison to other given methods. The overall performance of proposed method is very good in terms of output SNR.

Table 2 demonstrates the values of Cepstrum Distance measure parameter values at various levels of input SNRs in presence of given four highly non-stationary noise sources (i.e. babble, pop music, restaurant and exhibition). Cepstrum distance values must be minimum for the maximum improvement in noisy speech spectrum. The proposed method gives highest improvement for all noise sources except restaurant noise. For the restaurant noise case proposed method shows second highest improvement for the noisy speech signal. Hence restaurant noise case is the exceptional case for Hindi language in terms of overall performance of the proposed method.

The PESQ values are given in Table 3. PESQ value must be high for maximum improvement which is given by proposed method for all noise levels and sources. Perceptive evaluation of speech is much similar to speech intelligibility evaluation hence it can be said that PESQ value must be high for maximum speech intelligibility. Proposed method gives maximum speech intelligibility improvement.

The performance of the proposed method is also measured in terms of PSNR parameter and output values are given in

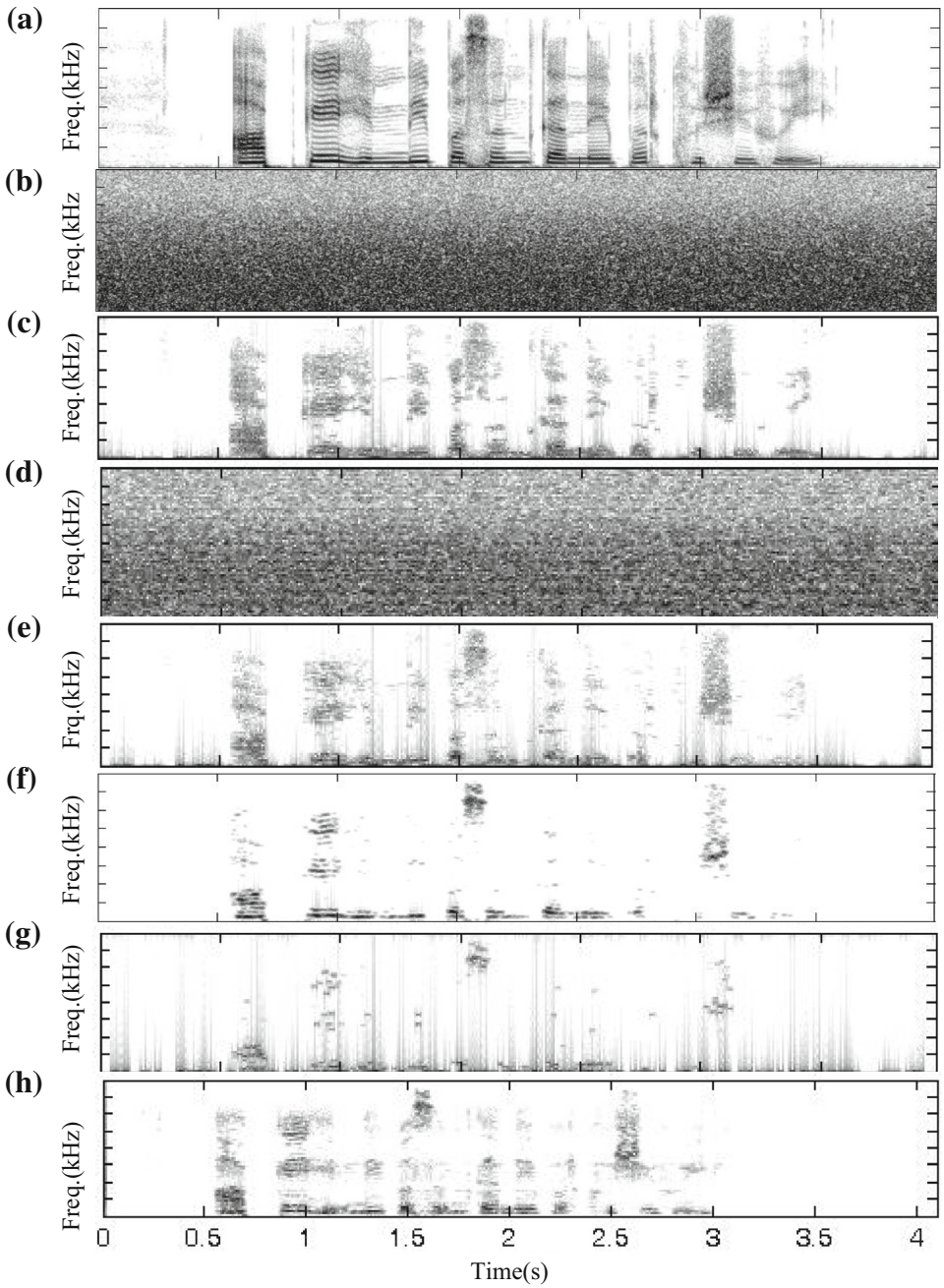

Fig. 3 Spectrogram of Hindi utterance, *apke hindi pasand karne par khusi hui* by a female speaker is given as: **a** clean speech **b** noisy speech (pop music noise at −25 dB SNR), **c** log-mmse, **d** test-psc, **e** Wiener, **f** idbm, **g** spectral subtraction, **h** proposed method

Table 4. The maximum value of PSNR is shown by proposed method in the babble and restaurant noise cases whereas it is not consistent for the pop music and exhibition noise environment at −5, −15 dB, respectively but in rest of the noise levels it gives highest improvement. These two noise levels are be considered as exceptional case since overall performance of the proposed method is very high.

PESQ and PSNR parameters values are given in Tables 3 and 4, respectively. Spectrogram of the single-channel Hindi speech pattern *"apke hindi pasand karne par khusi hui"* by a female speaker is shown in Fig. 3. The single-channel Hindi speech patterns are processed at various levels of input SNRs by all mentioned method for measuring the performance in quality and intelligibility improvement.

The proposed method gives maximum quality and intelligibility than those obtained by log-mmse, test-psc, Wiener, IdBM spectral-subtraction methods. One measure difference between spectrograms of proposed method and aforementioned methods is that proposed method does not give residual noise and impulses in output spectrograms. It is clear that the proposed method's output plot gives clear pattern but Wiener, spectral subtraction and log-mmse give output

pattern with some impulses that creates distortion in speech spectrum. The remaining two methods i.e. test- psc and IdBM do not improve in noisy speech to that level where someone can listen clearly. The proposed method reduces highly non-stationary noise efficiently and improves in quality as well as intelligibility while other methods introduce some distortion like impulses in processed speech. however, the proposed method is not consistent in some cases of noisy Hindi speech signal but overall performance is very good in comparison to other enhancement methods hence these points are consider as exceptional case for Hindi language analysis.

5 Conclusion

In this paper, a binary mask thresholding function based db10 mother wavelet transform is proposed and compared with other commonly used methods and for measuring the effectiveness in enhancement of single-channel Hindi speech patterns of low (negative) SNR range from −5 to −25 dB (in 5 dB gap). The binary mask thresholding function is considered as decision making function for reconstruction of enhanced speech patterns from noisy Hindi speech patterns.

Simulation results given by various methods show that the proposed method gives consistently better results in terms of quality and intelligibility measure parameters i.e. output SNR, PESQ, PSNR and lower value in Cepstrum distance measure at different levels (i.e. −5, −10, −15, 20 and −25 dB) of input SNR. Proposed method is not consistent in some cases of noise levels for Hindi language database. Although as discussed earlier, few results are inconsistent, but in terms of overall perspective the proposed method perform better than all other methods. The spectrograms and listening quality also shows the proposed method gives highest improvement in quality and intelligibility of the reconstructed speech signal.

References

Bahoura, M., & Rouat, J. (2006). Wavelet speech enhancement based on time-scale adaptation. *Speech Communication*, *48*(12), 1620–1637.

Boll, S. F. (1979). Suppression of acoustic noise in speech using spectral subtraction. *IEEE Transactions on Acoustics Speech and Signal Processing*, *27*(2), 113–120.

Dendrinos, M., Bakamidis, S., & Carayannis, G. (1991). Speech enhancement from noise: A regenerative approach. *Speech Communication*, *10*(1), 45–57.

Ephraim, Y., & Malah, D. (1985). Speech enhancement using a minimum mean square error log-spectral amplitude estimator. *IEEE Transactions on Audio, Speech and Language Processing*, *33*, 443–445.

Ephraim, Y. (1992). Statistical-model-based speech enhancement systems. *Proceedings of the IEEE*, *80*, 1526–1555.

Ephraim, Y., & Van Trees, H. L. (1995). A signal subspace approach for speech enhancement. *IEEE Transactions on Acoustics Speech and Signal Processing*, *3*(4), 251–266.

Gabor, D. (1946). Theory of communication. *The Journal of Electrical Engineering*, *93*, 429–457.

Goupillaud, P., Grossmann, A., & Morlet, J. (1984). Cycle-octave and related transforms in seismic analysis. *Journal of Applied Geophysics*, *23*(1), 85–102.

Hu, Y., & Loizou, P. C. (2007). A comparative intelligibility study of single-microphone noise reduction algorithms. *Journal Acoustic Socity of America*, *122*, 1777–1786.

Jensen, S. H., & Hansen, P. C. (1995). Reduction of broad-band noise in speech by truncated QSVD. *IEEE Transactions on Acoustics Speech and Signal Processing*, *3*(6), 439–448.

Johnson, M. T., Yuan, X., & Ren, Y. (2007). Speech signal enhancement through adaptive wavelet thresholding. *Speech Communication*, *2*(49), 123–133.

Kitawaki, N., & Nagabuchi, H. (1988). Objective quality evaluation for low bit-rate speech coding systems. *IEEE Journal on Selected Areas in Communications*, *6*, 262–273.

Li, J., & Liu, H. (2012). New wavelet packet transform algorithm based on critical bandwidth. *Computer Engineering and Applications*, *14*(48), 5–7.

McAulay, R., & Malpass, M. (1980). Speech enhancement using a soft-decision noise suppression filter. *IEEE Transactions on Acoustics Speech and Signal Processing*, *28*(2), 137–145.

Pearce, D., & Hirsch, H. G. (2000). The aurora experimental framework for the performance evaluation of speech recognition system under noisy conditions. *International conference on spoken language processing*, Beijing, 16–20 Oct 2000.

Perceptual evaluation of speech quality (PESQ) An objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs. ITU-T Recommendation P.862.1 (2003).

Prahallad, K., Elluru, N. K., Keri, V., Rajendran, S., Black, A. W. (2012). The IIIT-H indic speech databases. In *Proceedings of Interspeech*, Portland, Oregon, USA (2012). http://speech.iiit.ac.in/index.php/research-svl/69.html.

Rangachari, S., & Loizou, P. C. (2006). A noise-estimation algorithm for highly non-stationary environments. *Speech Communication*, *48*, 220–231.

Sanam, T. F. (2012). Enhancement of noisy speech based on a custom thresholding function with a statistically determined threshold. *The International Journal of Speech Technology*, *15*(4), 463–475.

Scalart, P., & Filho, J. (1996). Speech enhancement based on a priori signal to noise estimation. *In Proceedings of IEEE International conference on acoust speech, signal processing* (pp. 629–632).

Singh, S., Tripathy, M., & Anand, R. S. (2013). Noise removal in single channel Hindi speech patterns by using binary mask thresholding function in various mother wavelets. *IEEE International Conference on Signal Processing, Computing and Control (ISPCC)*, Shimla, India, 26–28 Sept 2013.

Singh, S.,Tripathy, M., & Anand, R. S. (2014). Wavelet packet based multiple noises suppression in single channel speech using binary mask threshold. *IEEE international conference on signal propagation and computer technology (ICSPCT)*, Ajmer, India, 12–13 July 2014.

Singh, S., Tripathy, M., & Anand, R. S. (2014). "Subjective and objective analysis of speech enhancement algorithms for single channel speech patterns of Indian and english languages", Taylor & Francis. *IETE Technical Review*, *31*(1), 34–46.

Stark, A. P., et al. (2008). Noise driven short-time phase spectrum compensation procedure for speech enhancement. *In Proceedings of Interspeech, Brisbane, Australia.*

Tao, H., & Qin, H. (2008). Chengbo Research of signal denoising method based on an improved wavelet thresholding. *Piezoelectronics & Acoustooptics*, *1*(30), 93–95.

Wojcicki, K., & Loizou, P. C. (2012). Channel selection in the modulation domain for improved speech intelligibility in noise. *The Journal of the Acoustical Society of America*, *131*(4), 2904–2913.

Yi, H., & Loizou, P. C. (2004). Speech enhancement based on wavelet thresholding the multitaper Spectrum. *IEEE Signal Processing Letters*, *12*, 59–67.

Zhang, X. (2010). *Digital : Speech signal processing and MATLAB simulation*. Beijing: Publishing House of Electronics Industry.