Int J Speech Technol (2015) 18:609–617
DOI 10.1007/s10772-015-9305-5

Binary mask based method for enhancement of mixed noise speech of low SNR input

Sachin Singh[1] · Manoj Tripathy[1] · R. S. Anand[1]

Received: 23 November 2014 / Accepted: 30 August 2015 / Published online: 14 September 2015

Abstract This paper presents a noise reduction method based on binary mask thresholding function for enhancement in single channel speech patterns of mixed highly non-stationary noises with low (negative) input SNR. For this purpose, a mixed highly non-stationary noisy speech database is generated by using noise and clean speech database of AURORA and INDIC speech, respectively. Results are compared with widely used methods such as Daubechies13, Daubechies40, Symlet13, Coiflet5, Wiener, Spectral Subtraction, and log-MMSE for performance evaluation in terms of SNR, PESQ, and Cepstrum distance parameters. In comparison to other methods the proposed single-channel speech enhancement method shows satisfactory results and obtained significant improvement in speech quality and intelligibility.

Keywords Speech enhancement · SNR · PESQ · Cepstrum analysis · SII · Mother wavelets

1 Introduction

To communicate among humans, we need a fundamental mode that transfers ideas from person to person that is speech. If speech signal is transferred in a noisy medium then speech signal may be distorted. This kind of noise may be daily life noise patterns like vehicle, fan, machine gun, tank, factory, fighter plane etc. that create distortion in speech signal. The distorted speech may become meaningless. Hence, for enhancement of these noisy speech signals we need effective speech enhancement methods. There are various speech enhancement methods available in the literature (Lim and Oppenheim 1979; Loizou 2007; Weiss et al. 1974; Boll 1979; Wiener 1949; Hansen and Clements 1991; Ephraim and Malah 1984, 1985; Hazrati and Loizou 2012; Paliwal et al. 2011, 2012; Wojcicki and Loizou 2012). Some of these techniques are spectral subtraction, minimum mean square error (MMSE) based techniques, modulation channel based speech enhancement techniques, wiener filtering methods and wavelet transform based etc. The spectral subtractive algorithms were initially proposed by Weiss et al. (1974) in the correlation domain and later by Boll (1979) in the Fourier transform domain. After filtering, this spectral subtractive method generates isolated peaks (i.e. musical noise). The optimal filter that minimizes the estimation error is called the Wiener filter. Wiener filtering algorithms exploit the fact that noise is additive and one can obtain an estimate of the clean signal spectrum simply by subtracting the noise spectrum from the noisy speech spectrum (Wiener 1949). The main drawback of the iterative Wiener filtering approach was that as additional iterations were performed, the speech formants shifted in location and decreased in formant bandwidth (Hansen and Clements 1991). The wiener filter is the optimal complex spectrum estimator, not the optimal magnitude spectrum estimator. Ephraim and Malah proposed a MMSE estimator which is optimal magnitude spectrum estimator (Ephraim and Malah 1984). Unlike the Wiener filtering, the MMSE estimator does not require a linear model between the observed data and the estimator.

✉ Sachin Singh
oxygen_sachin@rediff.com; sachinsingh.iitr@gmail.com

Manoj Tripathy
manojfee@iitr.ernet.in

R. S. Anand
anandfee@iitr.ernet.in

[1] Department of Electrical Engineering, Indian Institute of Technology Roorkee, Roorkee, Uttarakhand 247 667, India

But it assumes probability distributions of speech and noisy DFT coefficients. The DFT coefficients are statistically independent and hence uncorrelated. One drawback of this estimator is that it is mathematically tractable and it is not the most subjectively meaningful one. To overcome this problem a log-MMSE is derived by Ephraim and Malah (1985). Furthermore, some more efficient techniques are available in literature based on ideal binary mask (IdBM) (Hazrati and Loizou 2012). But modulation channel selection based method is more efficient for both quality and intelligibility improvement (Paliwal et al. 2011, 2012; Wojcicki and Loizou 2012).

Many researchers have been worked on speech enhancement using wavelet transform based methods. There are many algorithms given on various thresholding concepts for speech enhancement (Donoho 1995; Aggarwalet et al. 2011; Sanam and Shahnaz 2012; Tabibian et al. 2009; Bahoura and Rouat 2001; Zhao et al. 2011; Sheikhzadeh and Abutalebi 2001; Yi and Loizou 2004; Wang and Zhang 2005; Shao and Chang 2007). But quality and intelligibility of speech depends on the criterion adopted for masking threshold. A more efficient concept for threshold selection is adaptive thresholding (Johnson et al. 2007; Sumithra 2009; Sanam and Shahnaz 2012; Yu et al. 2007; Zhou 2010; Ghanbari and Reza 2006). A novel data adaptive thresholding approach to single channel speech enhancement is given by Hamid et al. (2013). In this paper complex signal were used in place of mixed speech signal. This complex signal was a combination of fractional Gaussian noise and noisy speech. A wavelet packet based binary mask method is given for mixed noise suppression (Singh et al. 2014, 2015). In this paper more than one noise and clean signal are mixed to generate noisy speech data for performance evaluation.

Over the past four decades, various single channel speech enhancement methods have been proposed for reduction/removal of one noise at one time but not analyzed for mixture of noises at same time. In this paper, a comparative study and implementation of speech enhancement techniques are presented for single channel Hindi speech patterns with mixed highly non-stationary noises. The mixed noises, we have taken as exhibition + pop music, restaurant + train, pop music + train + babble and pop music + babble + car. These four noise groups are used for quality and intelligibility evaluation of Hindi speech patterns. The well known and popular techniques like spectral-subtraction, wiener filtering, MMSE, p-MMSE, log-MMSE, ideal channel selection, modulation channel based method and wavelet transform based methods are implemented and their subjective and objective performances are analyzed to find out the optimal technique for Hindi speech enhancement in particular environmental condition (where more than one noise is present).

The paper is organized as follows; Sect. 2 presents the background of single channel speech enhancement techniques for noise reduction and binary mask function.

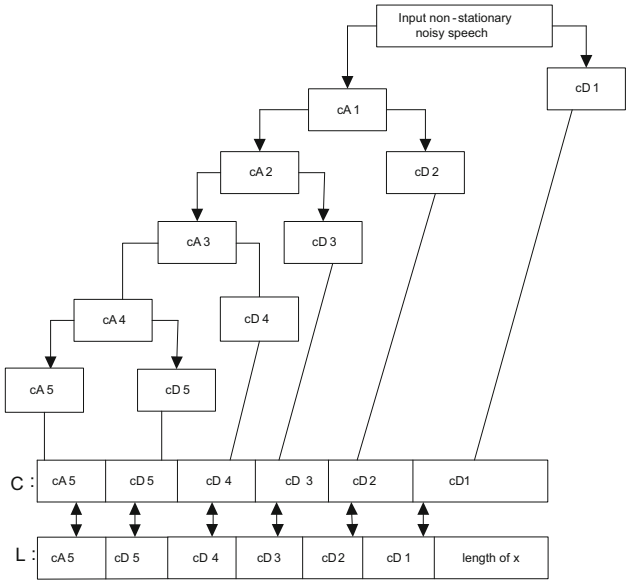

Fig. 1 Block diagram of wavelet decomposition up to five levels

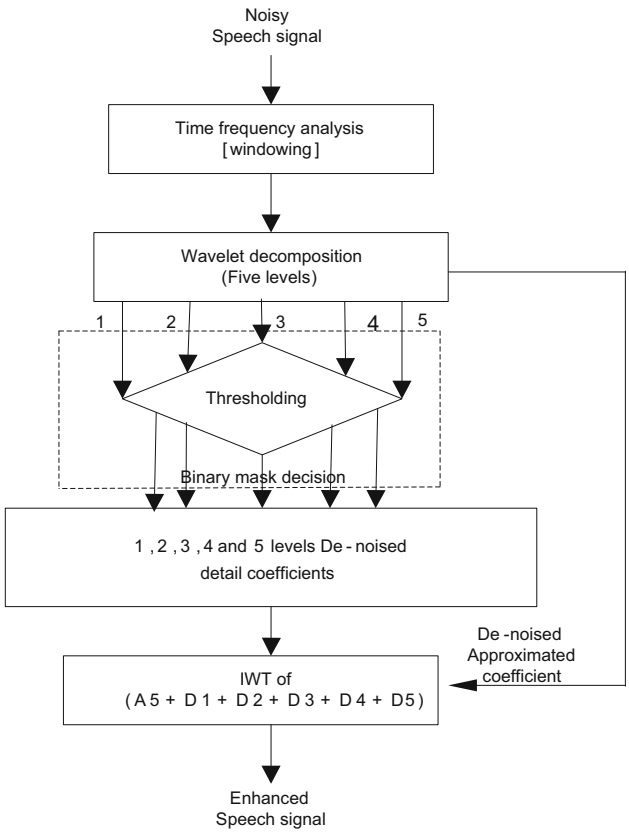

Fig. 2 Flow chart of the proposed method for enhancement of single-channel Hindi speech patterns

Simulation conditions are given in Sect. 3. Section 4 shows the results and discussions. Finally, the conclusion is summarized in Sect. 5.

2 Noisy speech enhancement

A mixed highly non-stationary noisy single-channel Hindi speech signal can be modeled as the sum of clean speech and more than one additive background noises.

$$y(n) = x(n) + n_1(n) + \cdots n_n(n) \quad (1)$$

where, $y(n)$, $x(n)$ and $n_n(n)$ denote the noisy speech, clean speech and various additive background highly non-stationary noises, respectively.

2.1 Simulated algorithm

The motivation for the simulation of various speech enhancement algorithms is to enhance the noisy single-channel noisy speech patterns from mixed highly non-stationary signals. Eight commonly used speech enhancement algorithms are evaluated for enhancement in mixed noisy single-channel Hindi speech pattern. Wiener, Spectral Subtraction, log-MMSE and wavelet transform based method (Daubachies10, Daubachies40, Symlet18, Coiflet5, BiorSpline6.8) are implemented for comparative analysis. Wiener filtering method is an iterative method that is based on minimization of mean square error of the noisy speech (Wiener 1949). Spectral subtraction is a widely used

Table 1 Performance parameter output SNR values for various mixed highly non-stationary noises

Noises	Enhancement techniques	SNR (dB)				
	Noisy input SNR	−25	−20	−15	−10	−5
Mixture of pop music and exhibition	Daubachies10	1.778	4.0101	4.1546	7.1555	9.5067
	Daubachies40	1.7712	3.5343	5.1675	6.6275	9.9024
	Symlet18	2.2479	3.4205	4.8325	6.6382	9.5289
	Coiflet5	2.6428	3.3247	5.8187	7.1766	9.7381
	BiorSpline6.8	2.5391	4.1788	3.0136	6.992	5.8842
	Wiener	−1.2498	0.8167	2.9098	5.3999	7.8911
	Spectral Sub.	−0.8658	1.4284	3.6513	6.0195	8.5354
	Log-MMSE	−4.1829	−1.6592	0.3106	2.6061	5.3269
Mixture of restaurant and train	Daubachies10	0.8505	2.9039	3.1321	5.3812	7.6243
	Daubachies40	0.4257	2.0691	3.9177	5.3893	5.5261
	Symlet18	0.8571	1.9933	4.1862	3.625	7.7983
	Coiflet5	0.853	2.4197	3.0056	3.6767	8.0155
	BiorSpline6.8	1.4657	2.5661	2.5368	5.5749	4.8582
	Wiener	−2.523	−1.4984	0.419	3.5375	6.6196
	Spectral Sub.	−2.5127	−1.4094	1.3005	4.16	7.1591
	Log-MMSE	−5.1815	−5.6372	−3.8247	−0.5493	3.617
Mixture of pop music, babble and train	Daubachies10	0.796	1.9896	3.9811	5.7142	8.0452
	Daubachies40	0.426	2.723	4.406	2.6845	5.9224
	Symlet18	0.3057	2.1709	4.0897	4.6034	3.711
	Coiflet5	0.5541	2.7207	4.2634	5.9325	7.8983
	BiorSpline6.8	1.0895	2.6212	4.1885	5.8425	8.0562
	Wiener	−1.2192	−1.6315	−0.2318	2.6956	6.1566
	Spectral Sub.	−0.8287	−1.1184	0.1767	3.2851	6.5537
	Log-MMSE	−2.7389	−3.4605	−3.0626	−0.7891	3.0807
Mixture of pop music, babble and car	Daubachies10	1.0922	2.7713	4.1224	5.541	7.6975
	Daubachies40	0.5249	2.4197	3.6372	5.8159	8.3292
	Symlet18	1.0948	2.6932	4.214	5.9636	7.2837
	Coiflet5	1.1106	2.831	4.1155	5.9364	7.2314
	BiorSpline6.8	0.9124	2.3825	3.6248	5.5308	7.7161
	Wiener	−1.0559	−1.2944	0.0416	2.9951	6.3158
	Spectral Sub.	−0.7107	−0.9886	0.326	3.5461	6.6824
	Log-MMSE	−2.0703	−2.9673	−2.5849	−0.441	3.6587

frequency domain method for reduction of additive uncorrelated noises from a speech pattern (Boll 1979). Log-spectrum based MMSE is described by Ephraim and Malah (1985) after simple MMSE. This algorithm assumes a Gaussian model for the complex spectral amplitudes of both speech and noise. It gives the optimum estimate of the log-spectrum of the clean speech signal.

Wavelet is a mathematical function that is used to divide a given function into different scale components. It breaks the signal into a shifted and dilated version of a short term waveform called the mother wavelet. It has high frequency-resolution in low bands and low frequency-resolution in high bands. Hence, it is very helpful in various fields of signal processing and widely used for signal analysis. The wavelet transform $W(s, \tau)$ for a signal $x(t)$ is defined as:

$$W(s, \tau) = \frac{1}{\sqrt{s}} \int x(t)\psi\left(\frac{(t-\tau)}{s}\right)dt \quad (2)$$

where $s > 0$ and $\tau \in R$, $x(t)$ is the input noisy speech signal. $\psi(t)$ is mother wavelet function and satisfies the orthogonal condition. It is localized in time and frequency domain. In the mother wavelet S is scaling parameter and determining the width of the mother wavelet. τ is a translation parameter and gives the center of mother wavelet. The selection of an appropriate mother wavelet plays an important role in analysis and depends on the application.

Table 2 Performance parameter output PESQ values for various mixed highly non-stationary noises

Noises	Enhancement techniques	PESQ				
	Noisy input SNR	−25	−20	−15	−10	−5
Mixture of pop music and exhibition	Daubachies10	1.7975	2.0329	2.2488	2.4738	2.7376
	Daubachies40	1.7284	1.832	2.0069	2.2336	2.5601
	Symlet18	1.7044	1.9052	2.086	2.3345	2.6612
	Coiflet5	1.836	2.0081	2.1695	2.4699	2.7791
	BiorSpline6.8	1.9892	2.0991	2.315	2.541	2.7968
	Wiener	1.7767	1.9217	2.1954	2.4082	2.6749
	Spectral Sub.	1.1578	1.4453	1.7408	1.9675	2.2118
	Log-MMSE	1.5915	1.6447	1.8735	2.1185	2.3852
Mixture of restaurant and train	Daubachies10	1.7615	1.9077	2.0698	2.2444	2.4677
	Daubachies40	1.6325	1.8544	2.0365	2.1887	2.3944
	Symlet18	1.7213	1.8758	2.0431	2.2281	2.4639
	Coiflet5	1.8037	1.9412	2.0847	2.2437	2.4669
	BiorSpline6.8	1.8945	2.0108	2.0918	2.2821	2.4945
	Wiener	1.2845	1.5261	1.8245	2.1448	2.4077
	Spectral Sub.	0.4118	1.0711	1.6169	1.9773	2.2154
	Log-MMSE	1.1598	1.2703	1.6724	1.9332	2.2222
Mixture of pop music, babble and train	Daubachies10	1.7391	1.8591	2.009	2.1764	2.3469
	Daubachies40	1.663	1.859	2.0239	2.198	2.3962
	Symlet18	1.572	1.7777	1.9983	2.1564	2.3318
	Coiflet5	1.6767	1.8412	2.0795	2.1976	2.3647
	BiorSpline6.8	1.9029	1.965	2.0767	2.2166	2.374
	Wiener	0.8696	1.3534	1.5609	1.9926	2.3889
	Spectral Sub.	1.1354	0.8954	1.3852	1.7674	2.1551
	Log-MMSE	1.2801	1.3824	1.5719	1.8125	2.1835
Mixture of pop music, babble and car	Daubachies10	1.7253	1.8303	1.9812	2.1824	2.3524
	Daubachies40	1.4639	1.6782	1.8575	2.0642	2.2426
	Symlet18	1.5906	1.7349	1.9946	2.1449	2.3337
	Coiflet5	1.7398	1.8269	2.0582	2.1971	2.3737
	BiorSpline6.8	1.8431	1.947	2.0521	2.2178	2.3966
	Wiener	0.8801	1.2662	1.6529	1.957	2.3763
	Spectral Sub.	0.4528	0.8988	1.3882	1.8005	2.1994
	Log-MMSE	1.09	1.3062	1.5997	1.8362	2.1937

Various basis functions have been proposed, including Harr, Morlet, Maxican, Daubechies, bi-orthogonal etc. The Daubachies10, Daubachies40, Symlet18, Coiflet5, BiorSpline6.8 mother wavelets are used for decomposition of detailed and approximation coefficients in the proposed work. The five levels in wavelet decomposition are given in Fig. 1. The five level detailed coefficients are recovered for same number of samples as in input speech. Now these detailed coefficients D1–D5 are given to binary mask threshold function for removing noise coefficients. Block diagram of the proposed procedure is given in Fig. 2. The given binary mask decision is applied for all five levels detailed and approximated coefficients. After applying binary mask decision on coefficients we get denoised coefficients. Now these denoised detailed coefficients are added with approximated coefficients and Inverse Wavelet Transform (IWT) is obtained to get denoised speech signal.

To obtain clean speech patterns from the noisy speech patterns, the estimated noise spectrum is subtracted from the noisy speech pattern, which is represented as:

$$x(f,t) = y(f,t) - n(f,t) \tag{3}$$

where, $n(f, t)$ denotes the noise, and $x(f, t)$ and $y(f, t)$ denotes the enhanced Hindi speech and noisy speech spectrum respectively. Where t, f indicates the frame index and channel or frequency bin index, respectively.

In Eq. 3, clean speech spectrum is computed by subtraction of estimated noise spectrum. The calculated noisy

Table 3 Performance parameter output Cepstrum distance values for various mixed highly non-stationary noises

Noises	Enhancement Techniques	Cepstrum distance				
	Noisy input SNR	−25	−20	−15	−10	−5
Mixture of pop music and exhibition	Daubachies10	6.1996	6.361	6.2503	6.0682	5.6613
	Daubachies40	7.2455	6.9871	6.7037	6.3263	5.8292
	Symlet18	7.6177	7.4867	7.1014	6.6419	6.0306
	Coiflet5	6.365	6.5225	6.3806	6.1382	5.7466
	BiorSpline6.8	4.5873	4.6849	4.6623	4.5788	4.5158
	Wiener	8.1715	7.5538	6.8791	6.2485	5.5266
	Spectral Sub.	9.3655	9.2189	8.9318	8.3762	7.7894
	Log-MMSE	7.8548	7.2522	6.6295	6.1953	5.6946
Mixture of restaurant and train	Daubachies10	5.2703	5.2349	4.97	4.4912	3.8801
	Daubachies40	6.1322	5.7765	5.2137	4.5874	3.8916
	Symlet18	6.4269	6.0712	5.4772	4.7418	3.9931
	Coiflet5	5.4057	5.363	5.0551	4.5421	3.9129
	BiorSpline6.8	3.835	3.8154	3.7565	3.4698	3.1544
	Wiener	6.1549	5.5898	4.9348	4.3541	3.996
	Spectral Sub.	7.6726	7.1117	6.4517	5.7344	5.0821
	Log-MMSE	5.8469	5.3723	5.0276	4.8026	4.5186
Mixture of pop music, babble and train	Daubachies10	5.4376	5.4302	5.2164	4.8905	4.3959
	Daubachies40	6.2534	5.9033	5.3457	4.7055	4.0426
	Symlet18	6.686	6.3405	5.9395	5.328	4.5737
	Coiflet5	5.5322	5.4596	5.313	4.9865	4.4278
	BiorSpline6.8	3.9342	3.8526	3.7872	3.6585	3.5218
	Wiener	7.1264	6.4947	5.6874	4.9701	4.3908
	Spectral Sub.	8.0359	7.6151	6.9885	6.3509	5.6932
	Log-MMSE	6.6447	5.9583	5.3032	4.901	4.4906
Mixture of pop music, babble and car	Daubachies10	5.4494	5.4378	5.1803	4.917	4.4628
	Daubachies40	6.3906	6.0126	5.6154	5.1621	4.5909
	Symlet18	6.6996	6.3663	5.8557	5.3392	4.6811
	Coiflet5	5.5193	5.4616	5.2638	4.9735	4.5007
	BiorSpline6.8	3.9629	3.8503	3.7709	3.6867	3.5692
	Wiener	7.2559	6.6097	5.7733	5.0296	4.4336
	Spectral Sub.	8.03	7.6601	7.0413	6.4095	5.7365
	Log-MMSE	6.7736	6.0816	5.4915	4.9831	4.6085

speech spectrum is accurate and very effective in terms of speech quality and intelligibility (Scalart and Filho 1996; Hu and Loizou 2007). The Eq. 4, priori SNR is calculated by using speech and noise signals (Feng 2015).

On the basis of this estimated noisy spectrum, a binary mask is constructed. A clean speech channel is selected on the basis of ideal binary mask. The ideal term is indicated that a priori information of the target signal is used. To calculate the binary mask SNR criterion is used as (Kim and Loizou 2010):

$$SNR(f,t) = 10\log_{10}\frac{|s(f,t)|^2}{|n(f,t)|^2} \quad (4)$$

where, $s(f, t)$, $n(f, t)$ represent clean speech and noise signals respectively. The noise signal is calculated frame by frame. The binary mask (BM) is calculated by using SNR criterion. This is given as (Hazrati and Loizou 2012):

$$BM(f,t) = \begin{cases} 1 & if\ SNR(f,t) > Threshold \\ 0 & Otherwise \end{cases} \quad (5)$$

The value of threshold is set to −6 dB, which is located around the center of performance.

3 Simulation conditions

Wiener, Spectral Subtraction, log-MMSE and wavelet transform (Daubachies10, Daubachies40, Symlet18, Coiflet5, BiorSpline6.8) based method are compared with

Table 4 Performance parameter SII index values for various mixed highly non-stationary noises

Noises	Enhancement techniques	Speech Intelligibility Index (SII)				
	Noisy input SNR	−25	−20	−15	−10	−5
Mixture of pop music and exhibition	Daubachies10	0.0115	0.0859	0.2253	0.3381	0.4731
	Daubachies40	0.0115	0.0852	0.2251	0.3385	0.4733
	Symlet18	0.0115	0.0852	0.2249	0.3383	0.4733
	Coiflet5	0.0115	0.0854	0.2249	0.3384	0.4734
	BiorSpline6.8	0.0115	0.0861	0.2254	0.3389	0.4735
	Wiener	0.0092	0.0664	0.1807	0.2886	0.4021
	Spectral Sub.	0.0112	0.0735	0.1964	0.3149	0.4403
	Log-MMSE	0.0034	0.0472	0.1485	0.2418	0.3411
Mixture of restaurant and train	Daubachies10	0	0	0.0252	0.0826	0.1728
	Daubachies40	0	0	0.0252	0.0829	0.1736
	Symlet18	0	0	0.0252	0.0829	0.1734
	Coiflet5	0	0	0.0252	0.0828	0.1733
	BiorSpline6.8	0	0	0.0252	0.0829	0.1737
	Wiener	0	0	0.0177	0.0558	0.1378
	Spectral Sub.	0	0	0.0195	0.063	0.1544
	Log-MMSE	0	0	0.0145	0.0406	0.1047
Mixture of pop music, babble and train	Daubachies10	0	0	0.0272	0.0556	0.1274
	Daubachies40	0	0.0019	0.0274	0.0554	0.1278
	Symlet18	0	0.0019	0.0274	0.0553	0.1278
	Coiflet5	0	0.0019	0.0278	0.0554	0.1278
	BiorSpline6.8	0	0.0019	0.0273	0.0556	0.1278
	Wiener	0	0.0004	0.0125	0.0373	0.0975
	Spectral Sub.	0	0.0003	0.018	0.0442	0.1084
	Log-MMSE	0	0.0002	0.0046	0.0182	0.0701
Mixture of pop music, babble and car	Daubachies10	0.0002	0.0021	0.0237	0.0527	0.1371
	Daubachies40	0.0002	0.0022	0.0239	0.0527	0.1367
	Symlet18	0.0002	0.0022	0.0239	0.0525	0.1368
	Coiflet5	0.0002	0.0021	0.0239	0.0526	0.1367
	BiorSpline6.8	0.0002	0.0022	0.0239	0.0528	0.1375
	Wiener	0	0.0005	0.0133	0.0403	0.1072
	Spectral Sub.	0	0.0015	0.0165	0.048	0.1195
	Log-MMSE	0	0.0004	0.0043	0.0184	0.0735

proposed method for performance evaluation. The clean speech pattern of Hindi language [taken from IIIT-H Indic speech database (Prahallad et al. 2012)] has been added with four different types of noise patterns [taken from NOIZEUS AURORA database (Hirsch and Pearce 2000)] for noisy speech generation. These four types of noises (pop music, babble, car, train, and restaurant) are added each other and with clean Hindi speech patterns at different levels of signal to noise ratio (SNR) ranging from −25 to −5 dB. These mixed noise patterns are exhibition + pop music, restaurant + train, pop music + train + babble and pop music + babble + car. These four mixed noise groups are used for quality and intelligibility evaluation of Hindi speech patterns in terms of performance measure parameters SNR, PESQ, SII and Cepstrum distance. All algorithms were implemented in MATLAB 7.1.

4 Results and discussion

With the aim of improving quality and intelligibility of mixed highly non-stationary noisy Hindi speech pattern, the four performance parameters are used and output values of those parameters are given in Tables 1, 2 and 3.

The output SNR values from various methods are given in Table 1. The maximum output SNR values given by coiflet5, BiorSpline6.8 and symlet18 wavelet transform at various levels of input SNR for all types of noises.

The higher values of PESQ parameter is given by BiorSpline6.8 wavelet transform at all level of input SNR values. These output PESQ values in Table 2 shows the maximum improvement of intelligibility and quality in enhanced Hindi speech pattern.

The lower Cepstrum distance value shows higher output PESQ values and maximum improvement in quality of

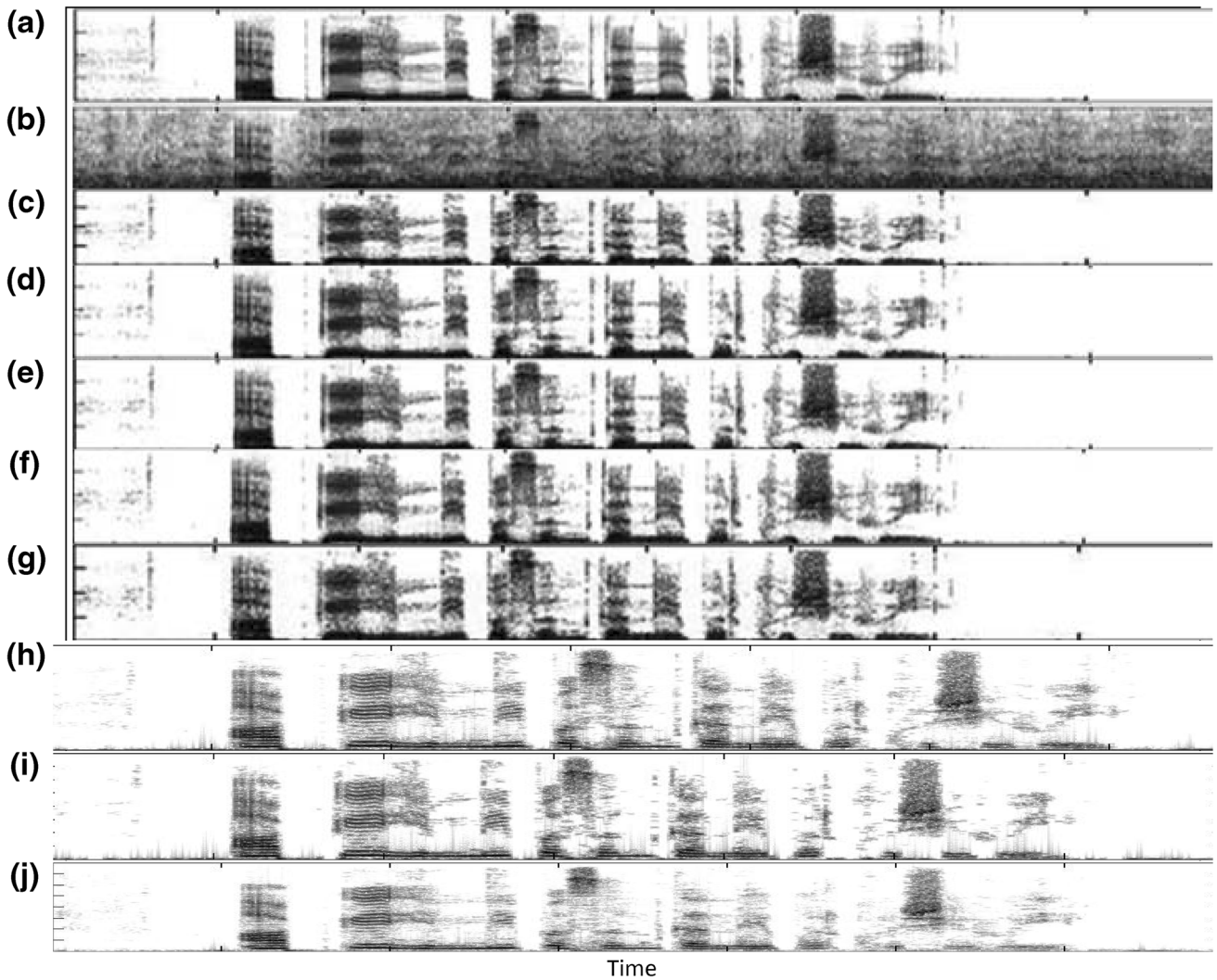

Fig. 3 Spectrograms of enhancement of single-channel speech (variation of frequency w.r.t. time): **a** clean, **b** mixed noisy speech (speech + pop music + babble + train), **c** Db10, **d** Db40, **e** Symlet18, **f** Coiflet5, **g** Bior 6.8, **h** Wiener, **i** Spectral Sub. **j** Log-MMSE

speech. The Table 3 shows all output Cepstrum distance measure values and minimum values are given by BiorSpline6.8 wavelet transform.

MOS parameter is used for speech intelligibility measure. The results for MOS values are given in Table 4. The improvement in MOS is increased as input SNR level is increased. The improvement in intelligibility can also be compared on the basis of various spectrograms given in Fig. 3. The noisy and enhanced spectrograms of Hindi speech are given by different methods and clear spectrogram is given by BiorSpline6.8 wavelet transform. The maximum listening quality of the enhanced output spectrum is given by the proposed method.

5 Conclusion

This paper presents a binary mask threshold function based BiorSpline6.8 wavelet transform method to enhance the speech quality and intelligibility of mixed highly non-stationary low SNR noises Hindi speech pattern. A comparative study is also done in this paper, which shows the performance of the conventional methods and wavelet based algorithms for enhancement in mixed noises single channel Hindi speech patterns. Wavelet domain methods show the higher improvement in quality and intelligibility measuring parameters in comparison to other spectral methods. The BiorSpline6.8 wavelets transform domain method give maximum improvement in speech quality and intelligibility parameters like PESQ and output SNR. BiorSpline6.8 wavelets transform method shows the maximum improvement in terms of performance measure parameters. In addition to that, the spectrograms also support same results and therefore the proposed method BiorSpline6.8 is more suitable for reduction of mixed highly non-stationary noises of negative SNR from noisy speech pattern in comparison to other speech enhancement methods.

References

Aggarwalet, R., et al. (2011). Noise reductions of speech signal using wavelet transform with modified universal threshold. *International Journal of Computer Application, 20*(5), 14–19.

Bahoura, M., & Rouat, J. (2001). Wavelet speech enhancement based on the teager energy operator. *IEEE Signal Processing Letters, 8*, 10–12.

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.

Donoho, D. L. (1995). De-noising by soft-thresholding. *IEEE Transactions on Information Theory, 41*, 613–627.

Ephraim, Y., & Malah, D. (1984). Speech enhancement using a minimum mean square error short time spectral amplitude estimator. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 32*(6), 1109–1121.

Ephraim, Y., & Malah, D. (1985). Speech enhancement using a minimum mean square error log-spectral amplitude estimator. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 23*(2), 443–445.

Feng, D., et al. (2015). Sparse HMM-based speech enhancement method for stationary and non-stationary noise environments. In *IEEE international conference on acoustics, speech and signal processing (ICASSP)*, Australia.

Ghanbari, Y., & Reza, M. (2006). A new approach for speech enhancement based on the adaptive thresholding of the wavelet packets. *Speech Communication, 48*, 927–940.

Hamid, Md. E., et al. (2013). Single channel speech enhancement using adaptive soft-thresholding with bivariate EMD. In *ISRN signal processing* (Vol. 8).

Hansen, J., & Clements, M. (1991). Constrained iterative speech enhancement with application to speech recognition. *IEEE Transactions on Signal Processing, 39*(4), 795–805.

Hazrati, O., & Loizou, P. C. (2012a). Tackling the combined effects of reverberation and masking noise using ideal channel selection. *Journal of Speech, Language, and Hearing Research, 55*, 500–510.

Hazrati, O., & Loizou, P. (2012b). Tackling the combined effects of reverberation and masking noise using ideal channel selection. *Journal of Speech, Language, and Hearing Research, 55*, 500–510.

Hirsch, H. G., & Pearce, D. (2000). The Aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions. In *ISCA ITRW ASR2000*, Paris, France, September 18–20, 2000. http://www.utdallas.edu/~loizou/speech/noizeus/.

Hu, Y., & Loizou, P. C. (2007). A comparative intelligibility study of single-microphone noise reduction algorithms. *Journal Acoustic Society of America, 22*, 1777–1786.

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

Kim, G., & Loizou, P. C. (2010). A binary mask based on noise constraints for improved speech intelligibility. In *Interspeech ISCA*, Japan.

Lim, J., & Oppenheim, A. V. (1979). Enhancement and bandwidth compression of noisy speech. *Proceedings of the IEEE, 67*(12), 1586–1604.

Loizou, P. C. (2007). *Speech enhancement theory and practice*. USA: CRC Press.

Paliwal, K. K., Schwerin, B., & Wojcicki, K. K. (2011). Role of modulation magnitude and phase spectrum towards speech intelligibility. *Speech Communication, 53*(3), 327–339.

Paliwal, K. K., Schwerin, B., & Wojcicki, K. (2012). Speech enhancement using a minimum mean-square error short-time spectral modulation magnitude estimator. *Speech Communication, 54*(2), 282–305.

Prahallad, K., Kumar, E. N., Keri, V., Rajendran, S., & Black, A. W. Interspeech-2012. (http://speech.iiit.ac.in/index.php/research-svl/69.html).

Sanam, T. F., & Shahnaz, C. (2012a). Teager energy operation on wavelet packet coefficients for enhancing noisy speech using a hard thresholding function. *Signal Processing: An International Journal, 6*(2), 22.

Sanam, T. F., & Shahnaz, C. (2012b). Enhancement of noisy speech based on a custom thresholding function with a statistically determined threshold. *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 acoustics, speech, signal processing (pp. 629–632).

Shao, Y., & Chang, C. (2007). A generalized time–frequency subtraction method for robust speech enhancement based on wavelet filter banks modeling of human auditory system. *IEEE Transactions on Systems, Man, and Cybernetics, 37*(4), 877–889.

Sheikhzadeh, H. & Abutalebi, H. R. (2001). An improved wavelet-based speech enhancement system. In *EUROSPEECH* (pp. 1855–1858).

Singh, S., Tripathy, M., & Anand, R. S. (2014). Single channel speech enhancement for mixed non-stationary noise environments. *Advances in Signal Processing and Intelligent Recognition Systems, 64*, 545–555.

Singh, S., Tripathy, M., & Anand, R. S. (2015). A wavelet based method for removal of highly non-stationary noises from single-channel hindi speech patterns of low input SNR. *International Journal of Speech Technology, 18*(2), 157–166.

Sumithra, A. (2009). Performance evaluation of different thresholding methods in time adaptivewavelet based speech enhancement. *IACSIT, 1*(5), 42–51.

Tabibian, S., Akbari, A., & Nasersharif, B. (2009). A new wavelet thresholding method for speech enhancement based on symmetric Kullback–Leibler divergence. In *14th international computer conference (CSICC)* (pp. 495–500).

Wang, J., & Zhang, C. (2005). Noise reduction in speech based on bark scaled wavelet packet decomposition and teager energy operator. *Signal Processing, China, 21*, 44–47.

Weiss, M., Aschkenasy, E., & Parsons, T. W. (1974). Study and the development of the INTEL technique for improving speech intelligibility. Technical Report NSC-FR/4023, Nicolet Scientific Corporation.

Wiener, N. (1949). *Extrapolation, interpolation and smoothing of stationary time series with engineering applications*. Cambridge, MA: MIT Press.

Wojcicki, K., & Loizou, P. C. (2012). Channel selection in the modulation domain for improved speech intelligibility in noise. *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.

Yu, G., Bacry, E., & Mallat, S. (2007). Audio signal denoising with complex wavelets and adaptive block attenuation. In *Proceedings of IEEE international conference on acoustics, speech and signal processing (ICASSP)* (Vol. 3, pp. 869–872).

Zhao, H., et al. (2011). An improved speech enhancement method based on teager energy operator and perceptual wavelet packet decomposition. *Journal of Multimedia, 6*(3), 308–315.

Zhou, B. (2010). An improved wavelet-based speech enhancement method using adaptive block thresholding. In *IEEE international conference on acoustic, speech, signal processing (ICASSP)*.