Chapter 4

MATLAB Implementation and Performance Evaluation of Transform Domain Methods

The simulation work is carried out to understand the functionality and behavior of all transform domain methods explained in chapter 3. An automatic VAD based on magnitude spectral distance proposed in [1] is integrated with STSA methods. The MATLAB simulation work is concreted and converged by preparing a MATLAB GUI (Graphical User Interface). This GUI can be used to simulate any transform domain algorithm for different noise conditions. The performance comparisons of various methods based on spectrographic analysis and objective tests are reported in this chapter. Various objective measures are available to evaluate speech enhancement techniques and they are described in brief here. The IEEE standard database NOIZEUS (noisy corpus) is used to test algorithms [18]. The database contains clean speech sample files as well as real world noisy speech files at different SNRs and noise conditions like airport, car, restaurant, train, station etc. The GUI also includes evaluation of algorithms using objective measures. The basic wavelet de-noising methods are also implemented in MATLAB and objective measures are obtained and compared with the STSA methods. The limitations and present implementations of these methods are also mentioned.

4.1 MATLAB Implementation -STSA Techniques

Eight important STSA algorithms viz. magnitude spectral subtraction (MSS) proposed by Boll [2], power spectral subtraction (PSS) proposed by Boll [2], Berouti spectral subtraction (BSS) proposed by Berouti [3], multi-band spectral subtraction (MBSS) proposed by Kamath [4], Wiener Scalart (WS) proposed by Scalart [5], maximum likelihood (ML) proposed by McAulay and Malpass [6], minimum mean square error with spectral amplitude (MMSE-SA) and minimum mean square error with log spectral amplitude (MMSE-LSA) proposed by Ephrahim and Malah [7,8] are simulated in the MATLAB environment. The sampling rate of the speech signal used in all the experiments carried out here is 8 KHz. The Hamming window of 25ms (200 samples) with 40% (10ms) overlap is selected. The FFT and IFFT are calculated using 256 points radix-2 algorithms. The general flow chart is shown in figure 4.1.

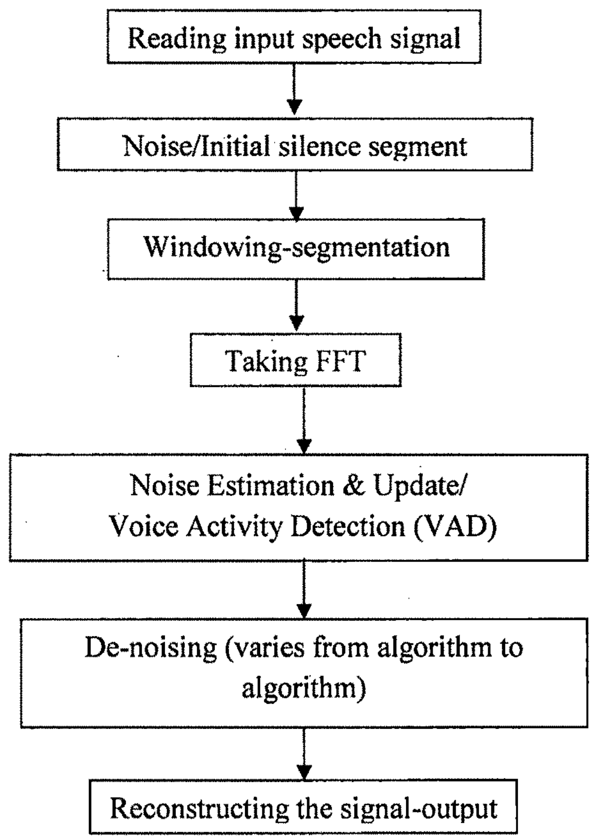


Fig. 4.1 A flow chart showing general implementation of STSA algorithms

The MSS and PSS methods described by equations 3.13, 3.14 and 3.15 and block diagram in fig. 3.7 are implemented with some additional features suggested in literature to improve the performance. The spectral magnitude is averaged over three successive frames (one past, current and one next) before applying the spectral subtraction. This will smooth the spectrum and helps to reduce the musical noise in the enhanced speech [1]. Before applying the half wave rectification to the speech frames; the residual noise reduction is applied by considering minimum spectral component from the minimum of three: the current clean estimate of spectral component, past frame noisy smoothed spectral component and next frame noisy smoothed spectral component [1]. For non-speech (silence) frame the spectral floor is applied to maintain the floor noise in the enhanced speech which will reduce the listener fatigue. The flow chart is shown in figure 4.2.

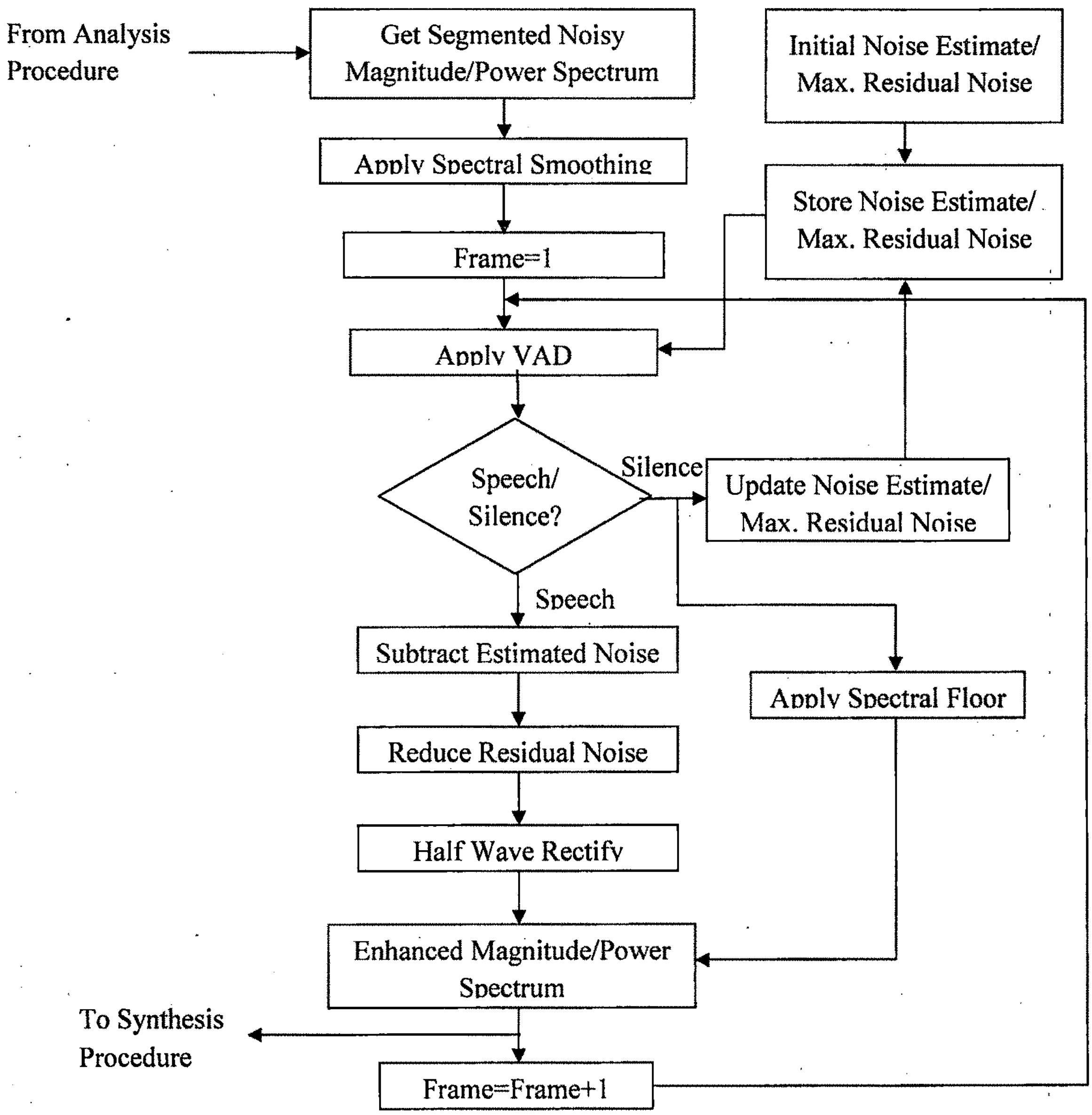


Fig. 4.2 Flow chart for MSS and PSS implementation

The BSS represents general spectral subtraction and its implementation is described by the block diagram shown in figure 4.3. The spectral floor parameter β is taken 0.03. The value of over subtraction factor α is adapted according to SNR values as per equation 3.18. The value of parameters α_0 and s is set considering the SNR varies between -5 to 20dB and the value of α lies between 1 and 3. These parameter settings are subjectively found optimal values for wide range of SNR values, except for very low SNR values below 0dB [3].

The MBSS method is described by figure 3.8. For simplification in implementation and comparison with other spectral subtraction algorithms the parameters α and β are set to same values as described earlier. The parameter δ is set to the originally prescribed values [4] for different frequency bands.

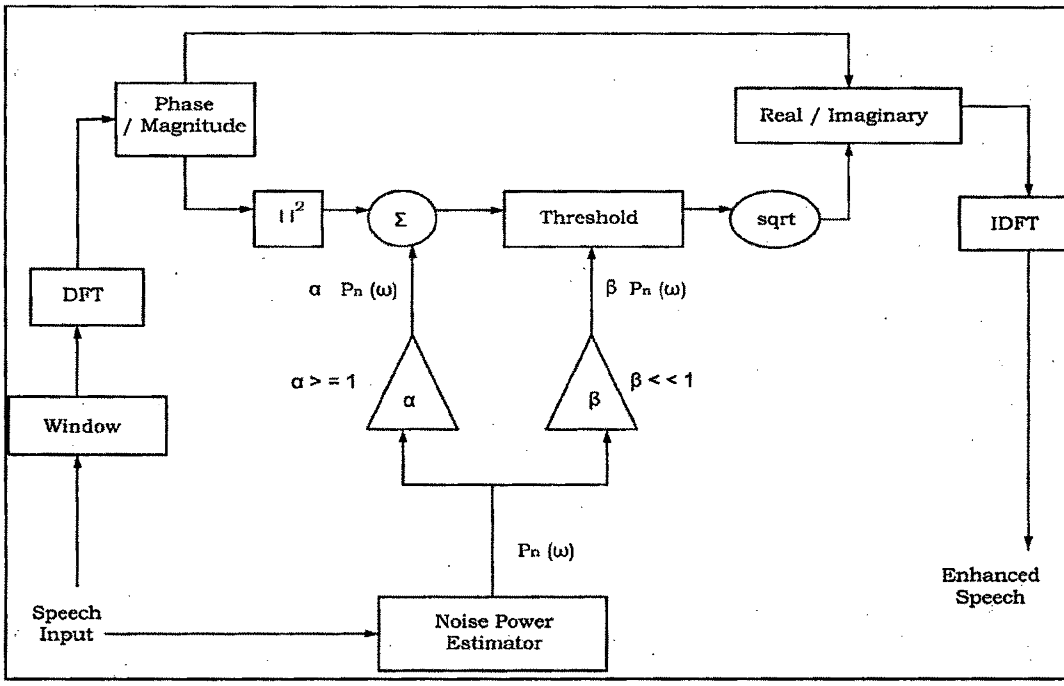


Fig. 4.3 A block diagram for BSS implementation

The Wiener filter implementation described by equation 3.27 is non-causal as it requires evaluating the *a priori* SNR. The *a priori* SNR is estimated using decision direct rule described by equation 3.28 which estimate it from *a posteriori* SNR of previous and current frames. The optimum value of smoothing constant η is taken 0.99 [6]. For first frame the *a posteriori* SNR is assumed unity and which is obvious.

The implementation of ML method is similar to that of MSS and PSS except the spectral subtraction equation is replaced by equation 3.30. For implementation of MMSE SA and LSA methods equation 3.31 and 3.32 are used along with decision directed rule to estimate *a priori* SNR.

In the implementation of all the above algorithms initial silence period of around 0.25

second (10 frames) in recorded speech is assumed. From this the initial noise estimate is derived by computing the mean (average) value of spectral components $|\widehat{D}(K)|$ and the variance of the spectral components $\lambda_D(K)$ during the initial silence period. These two are updated in every non-speech frame detected by VAD. One more parameter called noise smoothing factor σ is used during updating noise estimate and update. It is initialized to 9 for optimum smoothing [1]. The noise estimate update for frame t is described by following equations.

$$|\widehat{D^t}(K)| = \frac{\sigma|\widehat{D^{(t-1)}}(K)| + |Y^t(K)|}{\sigma + 1} \tag{4.1}$$

$$\lambda_D{}^t(K) = \frac{\sigma\lambda_D{}^{(t-1)}(K) + |Y^t(K)|^2}{\sigma + 1} \tag{4.2}$$

The VAD used is magnitude spectral distance type. It operates on a framed data. The terms involved here are explained as follows.

"Signal" is the current frame's magnitude spectrum and it is input to VAD, which is to be labeled as noise or speech. "Noise" is noise magnitude spectrum template (estimation), "noise counter" is the number of immediate previous noise frames, "noise margin" (default 3) is the spectral distance threshold. The noise margin is fixed to 3, which is the threshold value for comparison with the SNR of the current frame. "Hangover" (default 8) is the number of noise segments after which the "Speech flag" is reset (goes to zero). "Noise flag" is set to one if the segment is labeled as noise. "Dist" is the mean spectral distance. Spectral distance is calculated by using the SNR formula

$$Spectral\ distance = log_{10}(signal) - log_{10}(noise) \tag{4.3}$$

Mean of this spectral distance value is the "Dist" value. This "Dist" value is the real SNR value of the current frame. This value is compared with the noise margin and if this value is lesser than noise margin then the "Noise flag" is said to one and "Noise counter" is incremented by one.

If the "Dist" value is greater than the "Noise Margin" then the "Noise flag" is set to zero and the noise counter is reset (i.e., zero). If the "Noise counter" value is greater than the "Hangover period" then the speech flag is reset to zero and if vice versa then the "Speech Flag" is set (i.e., one). Its implementation is shown in figure 4.4 by means of a flow chart.

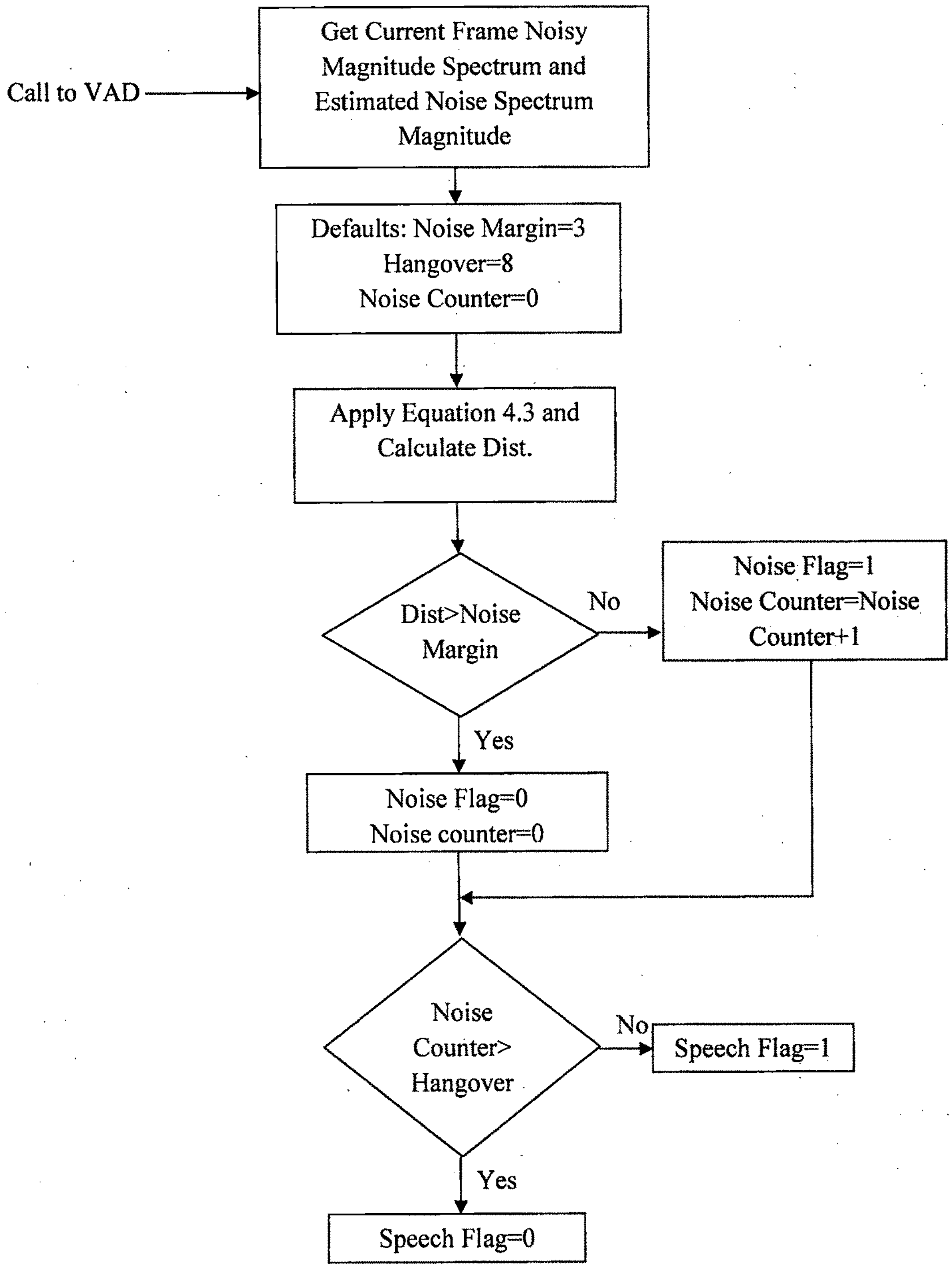


Fig. 4.4 Flow chart for magnitude spectral distance VAD implementation

4.2 Spectrographic Results of Simulation

The spectrograms of enhanced speech from the enhancement methods under comparison are plotted in figure 4.5 [9, 10], in order to compare the noise suppression capabilities based on presence of residual and musical noise in the enhanced speech. The spectrogram used for comparison is narrowband spectrogram obtained using Hamming window of 32ms (256 points) with 50% overlap and 256 point DFT. Figure 4.5(top panel) and (bottom panel) shows the spectrograms of clean and noisy speech sentence corrupted by 0dB white noise respectively. Figure 4.6 shows the spectrograms of enhanced speech by various algorithms as indicated.

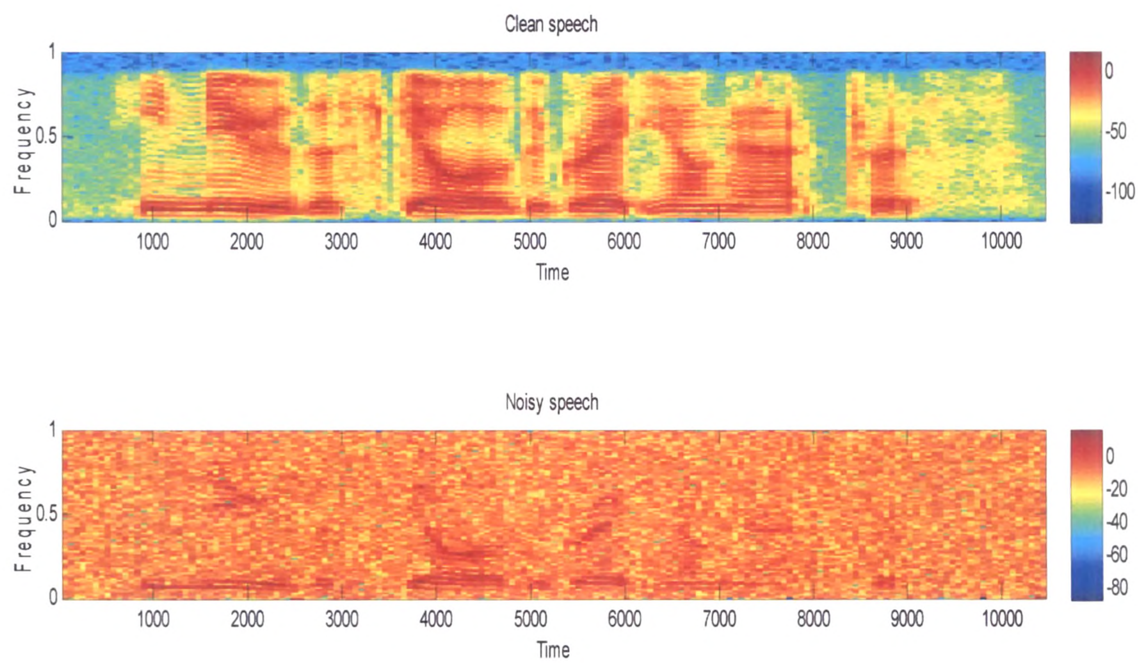
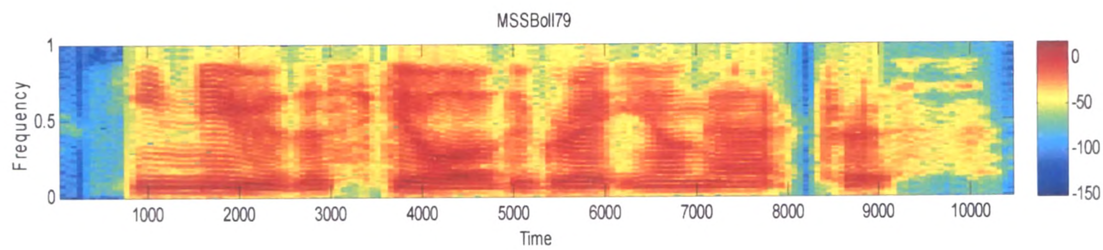
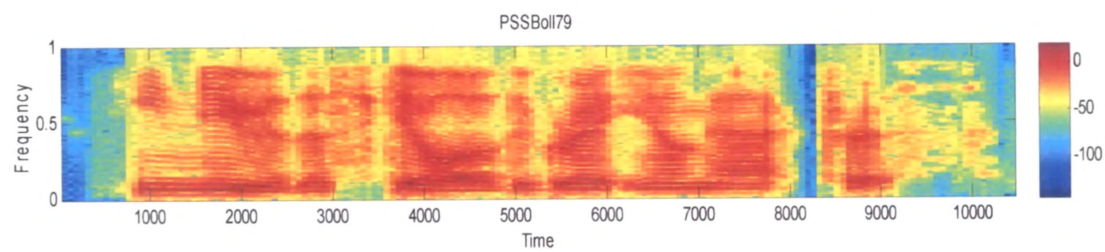
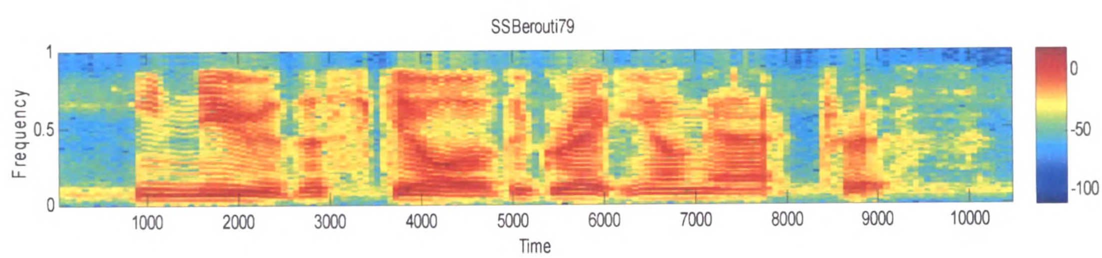
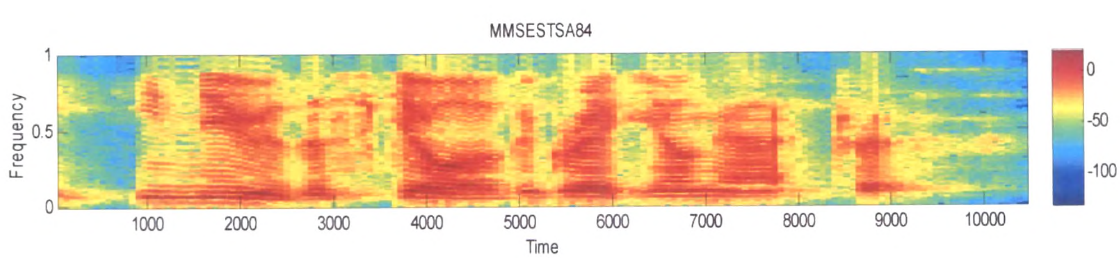


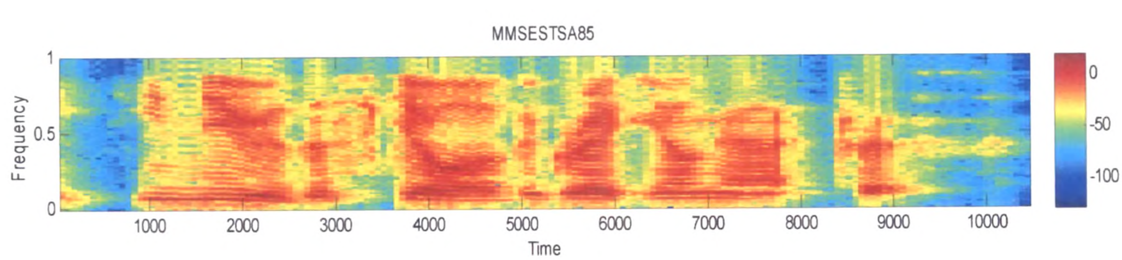
Fig. 4.5 Spectrogram of clean speech signal containing sentence 'He knew the skill of the great young actress' (top panel) and spectrogram of the signal subjected to 0dB white noise (bottom panel)


MSSBoll79
Frequency
Time
1
0.5
0
1000 2000 3000 4000 5000 6000 7000 8000 9000 10000
0
-50
-100
-150


PSSBoll79
Frequency
Time
1
0.5
0
1000 2000 3000 4000 5000 6000 7000 8000 9000 10000
0
-50
-100


SSBerouti79
Frequency
Time
1
0.5
0
1000 2000 3000 4000 5000 6000 7000 8000 9000 10000
0
-50
-100


MMSESTSA84
Frequency
Time
1
0.5
0
1000 2000 3000 4000 5000 6000 7000 8000 9000 10000
0
-50
-100


MMSESTSA85
Frequency
Time
1
0.5
0
1000 2000 3000 4000 5000 6000 7000 8000 9000 10000
0
-50
-100

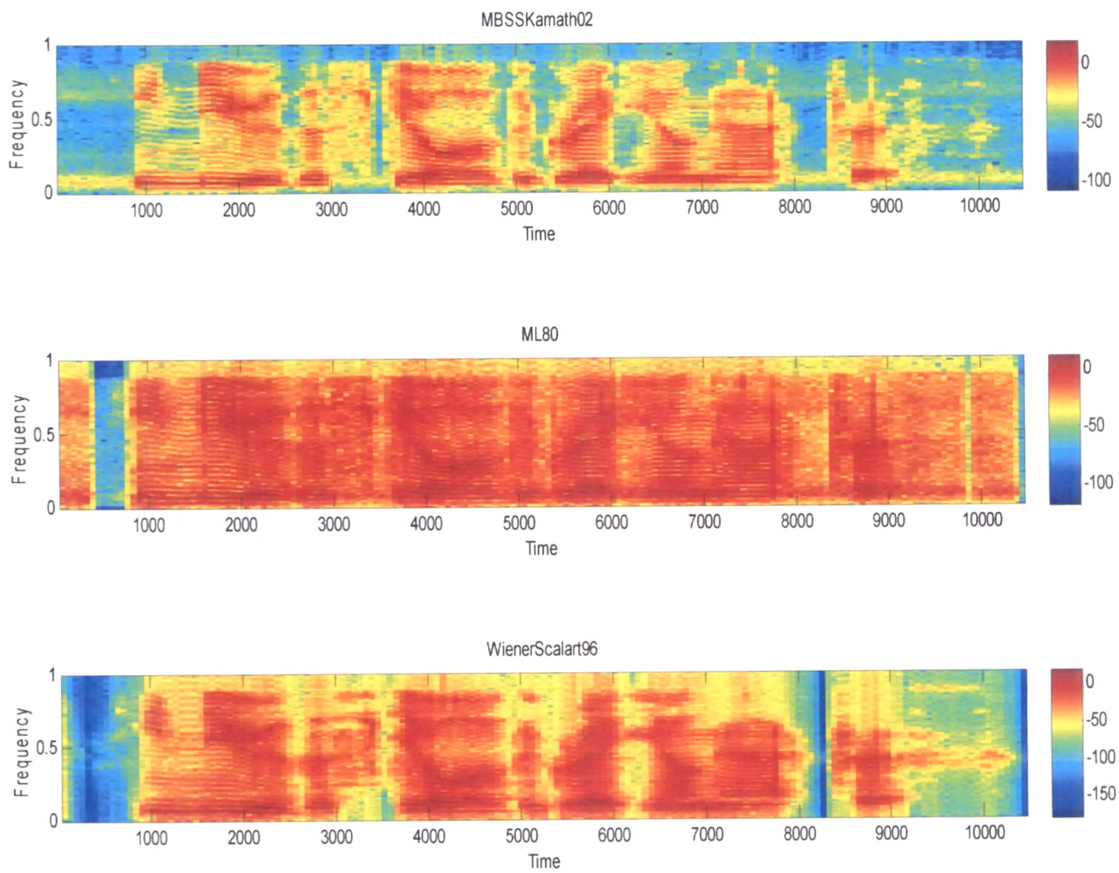


Fig. 4.6 Spectrogram of enhanced speech signal using various algorithms indicated on the head of each panel

The spectrographic analysis shows that the speech enhanced by MSS, PSS and BSS have random dots in the spectrogram compared to MBSS method. The random dots in the spectrogram represent sharp spectral peaks in the enhanced speech and contribute to musical noise. Also if we compare the results with original spectrogram the MBSS is more nearer to original. Hence in spectral subtraction category the MBSS is performing best. In statistical modeling method the ML method is worst while the MMSE-STSA85 (MMSE-LSA) gives best result. The Wiener filter method also gives less random dots but slightly more distortion in spectrogram (results in more residual noise or speech distortion) compared to MBSS and MMSESTSA85 methods. The formal listening also backs the results obtained. The MMSE-LSA method and MBSS methods give optimized performance compared to other methods in terms of residual and musical noise trade off. The MMSE-LSA is found the best from these two from listening point of view. A more useful judgement is obtained using objective measures described in the following section.

4.3 The NOIZEUS Database for Performance Evaluation

NOIZEUS is a noisy speech corpus recorded in Center for Robust Speech Systems, Department of Electrical Engineering, University of Texas at Dallas, Richardson, Texas to facilitate comparison of speech enhancement algorithms among research groups. The noisy database contains 30 IEEE sentences [14] produced by three male and three female speakers (five sentences /speaker), and was corrupted by eight different real-world noises at different SNRs. It is available at [18] and researchers can download it free of cost.

Thirty sentences from the IEEE sentence database were recorded in a sound proof booth using Tucker Davis Technologies (TDT) recording equipment. The IEEE database was selected because it contains phonetically balanced sentences with relatively low word-context predictability. The 30 sentences were selected from the IEEE database so as to include all phonemes in the American English language. The sentences were originally sampled at 25 KHz and down sampled to 8 KHz. A subset of the sentences recorded is given in Table 4.1. To simulate the receiving frequency characteristics of telephone handsets, the speech and noise signals were filtered by the modified Intermediate Reference System (IRS) filters used in ITU-T P.862 [16] for evaluation of the PESQ measure. Noise was artificially added to the speech signal as follows. The IRS filter was independently applied to the clean and noise signals. The active speech level of the filtered clean speech signal was first determined using method B of ITU-T P.56 [17]. A noise segment of the same length as the speech signal was randomly cut out of the noise recordings, appropriately scaled to reach the desired SNR level, and finally added to the filtered clean speech signal.

Filename	Speaker	Gender	Sentence Text
sp01.wav	CH	M	The birch canoe slid on the smooth planks.
sp02.wav	CH	M	He knew the skill of the great young actress
sp03.wav	CH	M	Her purse was full of useless Trash
sp04.wav	CH	M	Read verse out loud for pleasure
sp05.wav	CH	M	Wipe the grease off his dirty face
sp06.wav	DE	M	Men strive but seldom get rich
sp07.wav	DE	M	We find joy in the simplest things
sp08.wav	DE	M	Hedge apples may stain your hands green
sp09.wav	DE	M	Hurdle the pit with the aid of a long pole
sp10.wav	DE	M	The sky that morning was clear and bright blue
sp11.wav	JE	F	He wrote down a long list of items
sp12.wav	JE	F	The drip of the rain made a pleasant sound
sp13.wav	JE	F	Smoke poured out of every crack
sp14.wav	JE	F	Hats are worn to tea and not to dinner
sp15.wav	JE	F	The clothes dried on a thin wooden rack
sp16.wav	KI	F	The stray cat gave birth to kittens
sp17.wav	KI	F	The lazy cow lay in the cool grass
sp18.wav	KI	F	The friendly gang left the drug store
sp19.wav	KI	F	We talked of the sideshow in the circus
sp20.wav	KI	F	The set of china hit the floor with a crash
sp21.wav	TI	M	Clams are small, round, soft and tasty
sp22.wav	TI	M	The line where the edges join was clean
sp23.wav	TI	M	Stop whistling and watch the boys march
sp24.wav	TI	M	A cruise in warm waters in a sleek yacht is fun
sp25.wav	TI	M	A good book informs of what we ought to know
sp26.wav	SI	F	She has a smart way of wearing clothes
sp27.wav	SI	F	Bring your best compass to the third class
sp28.wav	SI	F	The club rented the rink for the fifth night
sp29.wav	SI	F	The flint sputtered and lit a pine torch
sp30.wav	SI	F	Let us all join as we sing the last chorus

Table 4.1 Sentences from the NOIZEUS speech corpus used in quality evaluation

Noise signals were taken from the AURORA database [15] and included the following recordings from different places: babble (crowd of people), car, exhibition hall, restaurant, street, airport, train station, and train. The noise signals were added to the speech signals at SNRs of 0, 5, 10, and 15dB. The NOIZEUS speech corpus is used in the objective quality evaluation of STSA based speech enhancement algorithms and it is described in the next section.

4.4 Objective Evaluation of STSA Algorithms

Eight STSA algorithms are evaluated using objective measures SSNR, LLR, WSS and

PESQ. The evaluation is done using NOIZEUS database. The MATLAB function that implements and returns the values of the SSNR, LLR, WSS and PESQ is available at [19] and it is widely accepted by researchers for quality evaluation of their speech enhancement algorithms [12,13]. The reason for using the above mentioned code for evaluation is to maintain authenticity, consistency and compatibility with results obtained by other researchers. The measures have been observed over 0-10dB range of SNRs with all eight types of colored noises included in NOIZEUS database. Each algorithm is evaluated here as well as in all future cases on all 30 phonetically balanced speech sentences from NOIZEUS data base corrupted by 3 different SNR values (0, 5 and 10dB) in all 8 colored noise environments. So for one algorithm the number of test runs are 30speech sentences * 3SNRs * 8Noise types = 720. In addition to speech sentences corrupted by colored noise included in the database; a synthesized white noise added to clean speech sentences of NOIZEUS database in 0-10dB SNR range is also used to test the algorithms. This adds another 90 test runs on one algorithm. Hence each algorithm has been tested for total of 810 different conditions. This is sufficient to reflect the real life scenario in which almost all speech communication systems have to work. The results are tabulated in tables 4.2 to 4.10.

AIRPORT NOISE	0 dB				5 dB				10 dB			
	SSNR	WSS	LLR	PESQ	SSNR	WSS	LLR	PESQ	SSNR	WSS	LLR	PESQ
MSSBoll79	-1.8033	98.1232	0.944	1.7485	0.1593	80.9384	0.7754	2.1597	2.1429	64.6463	0.6044	2.5111
PSSBoll79	-3.1179	82.7364	0.8976	1.663	-1.7308	69.4178	0.7467	2.1239	-0.0603	57.764	0.5867	2.4847
SSBerouti79	-3.6373	82.9816	0.8967	1.8111	-2.0258	66.2091	0.7119	2.1437	-0.1931	51.4494	0.5261	2.4931
ML80	-3.9171	75.6016	1.0743	1.2804	-3.5451	64.3518	1.0479	1.549	-3.1767	53.5394	1.0235	1.7805
MMSESTSA84	-3.0485	86.4321	0.9334	1.8364	-1.3994	68.1463	0.7522	2.239	0.341	52.4128	0.5691	2.5556
MMSESTSA85	-2.4197	97.1345	1.0146	1.8019	-0.7883	78.9599	0.8287	2.2261	0.99	61.751	0.6462	2.5643
WienerScalart96	-1.4812	123.103	1.2835	1.5979	0.0496	101.768	1.0489	2.0812	1.8138	78.9998	0.8351	2.4489
SSMultibandKamath02	-3.3835	80.3375	0.8987	1.787	-1.7897	64.7142	0.7196	2.1499	0.0325	50.8774	0.5464	2.4947

Table 4.2 Objective quality evaluation with airport noise

CAR NOISE	0 dB				5 dB				10 dB			
	SSNR	WSS	LLR	PESQ	SSNR	WSS	LLR	PESQ	SSNR	WSS	LLR	PESQ
MSSBoll79	0.8033	80.1232	0.944	1.7485	1.2048	67.5299	0.8189	2.1252	3.19	56.7015	0.6279	2.5632
PSSBoll79	-2.1179	72.6456	0.8976	1.663	-1.4238	61.8076	0.8297	2.0597	0.254	52.6771	0.6421	2.4307
SSBerouti79	-2.8973	72.9816	0.8967	1.8111	-1.5631	62.0198	0.7977	2.0808	0.1389	48.8516	0.5826	2.4489
ML80	-2.9171	65.6016	1.0743	1.2804	-2.8386	61.5944	1.0412	1.5508	-2.6228	54.196	1.049	1.7026
MMSESTSA84	-1.0485	66.4321	0.8231	2.0808	-0.647	54.1602	0.7559	2.2373	1.0084	42.6779	0.5838	2.6155
MMSESTSA85	-0.4197	67.5299	0.9052	2.0597	-0.0036	63.0102	0.8231	2.2368	1.7412	50.5242	0.6545	2.653
WienerScalart96	0.1389	103.103	1.2835	1.5979	1.2449	93.088	1.1127	1.9947	2.9627	74.2115	0.9052	2.5361
SSMultibandKamath02	-2.3835	70.3375	0.8987	1.787	-1.3157	57.5471	0.7686	2.0931	0.4256	45.6784	0.579	2.4836

Table 4.3 Objective quality evaluation with car noise

STREET NOISE	0 dB				5 dB				10 dB			
	SSNR	WSS	LLR	PESQ	SSNR	WSS	LLR	PESQ	SSNR	WSS	LLR	PESQ
MSSBoll79	-1.351	84.6089	1.0555	1.663	0.2108	68.916	0.8763	2.0632	2.1157	56.1325	0.6477	2.4959
PSSBoll79	-3.027	71.8156	1.0212	1.6219	-1.5639	60.713	0.8514	2.0325	-0.0251	51.0499	0.6398	2.4324
SSBerouti79	-3.3585	75.1916	1.0318	1.7536	-1.7424	60.468	0.8265	2.0736	-0.0168	48.2113	0.599	2.4432
ML80	-3.8069	66.8665	1.1807	1.3986	-3.3169	58.655	1.1216	1.5591	-3.0933	51.687	1.0913	1.7089
MMSESTSA84	-2.6324	72.1776	1.0089	1.8595	-1.1443	58.033	0.8167	2.1747	0.4857	46.73	0.6183	2.5199
MMSESTSA85	-1.9643	84.2573	1.079	1.8572	-0.5892	68.0733	0.8875	2.1911	1.0071	55.3345	0.6774	2.5461
WienerScalart96	-0.972	118.2388	1.3972	1.6009	0.2629	95.8322	1.1828	1.9547	1.7746	74.9996	0.8795	2.4198
SSMultibandKamath02	-3.0104	70.6299	0.9921	1.7235	-1.4438	58.6753	0.7972	2.0858	0.2581	47.3427	0.5989	2.4732

Table 4.4 Objective quality evaluation with street noise

TRAIN NOISE	0 dB				5 dB				10 dB			
	SSNR	WSS	LLR	PESQ	SSNR	WSS	LLR	PESQ	SSNR	WSS	LLR	PESQ
MSSBoll79	-1.995	79.6182	1.2669	1.5185	0.2759	65.3372	1.0021	2.024	2.0976	52.8069	0.7394	2.4006
PSSBoll79	-3.5103	71.2996	1.2267	1.6337	-1.7645	59.2825	1.0313	1.985	-0.3135	48.9557	0.7648	2.3338
SSBerouti79	-3.7293	72.7156	1.1909	1.6997	-1.8556	58.9354	0.9815	2.0019	-0.2832	46.9488	0.7073	2.3307
ML80	-4.3353	65.4476	1.3645	1.3839	-3.5039	56.0807	1.2954	1.5387	-3.3991	50.1037	1.1885	1.6934
MMSESTSA84	-2.9603	65.7305	1.172	1.7879	-1.0325	51.5249	0.9158	2.1436	0.4313	41.7044	0.6774	2.4606
MMSESTSA85	-2.2732	75.85	1.2313	1.764	-0.3563	60.3413	0.9752	2.1757	1.1985	48.9939	0.7452	2.5194
WienerScalart96	-1.2984	109.7372	1.5581	1.4948	0.5225	87.0401	1.2741	1.9865	2.1049	69.9746	1.0154	2.4287
SSMultibandKamath02	-3.4043	71.5309	1.152	1.668	-1.5685	58.1421	0.9543	2.0134	-0.0054	46.7996	0.6989	2.3709

Table 4.5 Objective quality evaluation with train noise

BABBLE NOISE	0 dB				5 dB				10 dB			
	SSNR	WSS	LLR	PESQ	SSNR	WSS	LLR	PESQ	SSNR	WSS	LLR	PESQ
MSSBoll79	-2.1658	99.8102	1.0416	1.6979	-0.0676	84.5867	0.8295	2.1014	2.0667	63.6597	0.6222	2.4865
PSSBoll79	-3.178	80.2496	0.9505	1.7806	-1.7223	70.5785	0.7779	2.109	-0.2454	58.0147	0.6136	2.4399
SSBerouti79	-3.8347	84.3648	0.981	1.7724	-2.0109	68.0814	0.7689	2.1053	-0.3475	52.8799	0.5664	2.4539
ML80	-3.9253	74.3201	1.0864	1.3516	-3.4184	64.8609	1.0413	1.5392	-3.1517	54.6316	1.0314	1.7431
MMSESTSA84	-3.2938	84.7417	1.0154	1.8368	-1.4556	69.2224	0.8002	2.1987	0.2845	50.5584	0.5969	2.5374
MMSESTSA85	-2.6609	98.5622	1.1069	1.8157	-0.8334	81.7387	0.8885	2.1765	0.8869	60.5833	0.6794	2.5458
WienerScalart96	-1.8441	127.917	1.3919	1.6144	-0.0682	106.0967	1.1065	2.0202	1.7351	80.4268	0.8561	2.4259
SSMultibandKamath02	-3.5064	80.0677	0.9553	1.7772	-1.8025	66.3579	0.7644	2.1148	-0.1396	52.2343	0.5767	2.4656

Table 4.6 Objective quality evaluation with babble noise

STATION NOISE	0 dB				5 dB				10 dB			
	SSNR	WSS	LLR	PESQ	SSNR	WSS	LLR	PESQ	SSNR	WSS	LLR	PESQ
MSSBoll79	-1.355	87.7056	1.0241	1.6836	0.747	71.2343	0.7916	2.1252	2.8388	59.5597	0.6034	2.553
PSSBoll79	-3.1837	79.5555	0.98	1.6856	-1.5158	62.519	0.7625	2.1065	0.1521	54.6333	0.6167	2.4464
SSBerouti79	-3.5612	80.1046	0.9741	1.7731	-1.7323	62.9343	0.741	2.1472	0.0173	49.5866	0.5592	2.4684
ML80	-3.8867	72.506	1.11	1.3706	-3.0294	61.1479	1.001	1.5762	-2.766	54.0446	1.0038	1.707
MMSESTSA84	-2.6902	73.681	0.9704	1.8226	-1.0064	59.0474	0.7407	2.2417	0.6919	45.6896	0.5755	2.6019
MMSESTSA85	-2.1629	85.1952	1.041	1.8299	-0.4963	68.7516	0.8068	2.2265	1.4095	53.9997	0.651	2.619
WienerScalart96	-1.0319	120.2635	1.35	1.5471	0.5771	97.8194	1.0681	1.9713	2.5065	75.8354	0.8545	2.4995
SSMultibandKamath02	-3.2386	75.9053	0.9541	1.7445	-1.5342	59.8868	0.7329	2.1477	0.2864	47.1549	0.5675	2.4862

Table 4.7 Objective quality evaluation with station noise

EXHIBITION NOISE	0 dB				5 dB				10 dB			
	SSNR	WSS	LLR	PESQ	SSNR	WSS	LLR	PESQ	SSNR	WSS	LLR	PESQ
MSSBoll79	-1.5704	89.9787	1.2204	1.604	0.5896	81.0013	0.9169	2.049	2.4415	66.3214	0.7154	2.4528
PSSBoll79	-3.0198	71.932	1.2158	1.6095	-1.5056	64.2678	0.9535	2.0437	0.0304	57.8017	0.7363	2.3891
SSBerouti79	-3.5836	81.1314	1.2256	1.6753	-1.8016	67.1195	0.9347	2.034	-0.0964	53.8661	0.684	2.3953
ML80	-3.4806	66.6221	1.3068	1.3764	-2.9098	59.2169	1.2174	1.5708	-2.8403	51.8378	1.1681	1.6954
MMSESTSA84	-2.8118	76.5844	1.1343	1.7032	-1.0386	64.3108	0.8679	2.1156	0.6413	51.2957	0.6976	2.4996
MMSESTSA85	-2.1916	88.072	1.1912	1.6525	-0.3228	74.2186	0.9222	2.114	1.3673	59.7444	0.7604	2.5295
WienerScalart96	-1.112	124.0819	1.4734	1.3924	0.6991	100.6981	1.1933	1.9494	2.325	81.3709	1.0114	2.4211
SSMultibandKamath02	-3.1278	73.9685	1.1429	1.6363	-1.4479	62.3126	0.8799	2.0504	0.2523	50.913	0.6583	2.4288

Table 4.8 Objective quality evaluation with exhibition noise

RESTAURANT NOISE	0 dB				5 dB				10 dB			
	SSNR	WSS	LLR	PESQ	SSNR	WSS	LLR	PESQ	SSNR	WSS	LLR	PESQ
MSSBoll79	-2.1791	97.2494	0.9992	1.6782	-0.2867	79.0998	0.7826	2.113	1.8766	62.717	0.6157	2.4885
PSSBoll79	-3.1735	80.1378	0.9106	1.7692	-2.0154	68.8552	0.7432	2.0829	-0.3076	57.4583	0.5764	2.4577
SSBerouti79	-3.792	82.0692	0.9269	1.7892	-2.2321	66.2487	0.7272	2.0831	-0.3805	51.5288	0.5373	2.4705
ML80	-3.8993	72.75	1.0709	1.3825	-3.9688	62.7957	1.1046	1.5372	-3.4398	52.6755	1.0534	1.7371
MMSESTSA84	-3.3229	84.9662	0.9705	1.7367	-1.7239	67.2118	0.7577	2.1413	0.1966	50.8995	0.5764	2.5284
MMSESTSA85	-2.7603	97.998	1.0605	1.6748	-1.1707	78.324	0.8522	2.1295	0.7521	60.1569	0.6649	2.5248
WienerScalart96	-2.0164	122.7877	1.356	1.4891	-0.4357	98.5382	1.0725	1.9893	1.4381	76.1611	0.8663	2.4148
SSMultibandKamath02	-3.5731	80.4985	0.9269	1.7802	-2.0399	66.1375	0.7322	2.0886	-0.2064	51.9146	0.5575	2.4774

Table 4.9 Objective quality evaluation with restaurant noise

WHITE NOISE	0 dB				5 dB				10 dB			
	SSNR	WSS	LLR	PESQ	SSNR	WSS	LLR	PESQ	SSNR	WSS	LLR	PESQ
MSSBoll79	-1.1132	83.3377	1.7319	1.5769	0.5497	73.6445	1.4188	1.8581	2.9451	64.4099	1.0996	2.3998
PSSBoll79	-2.94	81.7164	1.7683	1.6055	-1.6546	70.093	1.4921	1.9465	-0.1395	60.4332	1.2165	2.3313
SSBerouti79	-3.0658	90.7983	1.7465	1.6889	-1.4241	78.0174	1.4524	2.0059	0.0974	63.7341	1.1594	2.3402
ML80	-3.0383	80.8097	1.781	1.5117	-2.2353	71.5198	1.5827	1.7821	-1.8503	65.438	1.4659	1.9233
MMSESTSA84	-2.1017	77.7371	1.5934	1.7619	-0.6197	65.2738	1.2853	2.1548	1.2022	51.0057	0.9809	2.5624
MMSESTSA85	-1.7984	88.4065	1.6884	1.7397	-0.3158	75.3228	1.3865	2.1378	1.6449	59.4232	1.077	2.5553
WienerScalart96	-0.4152	131.7448	1.9454	1.4284	0.9242	110.4736	1.6645	1.7615	2.8012	85.6904	1.3557	2.3741
SSMultibandKamath02	-2.6404	82.0069	1.6006	1.5931	-1.2496	70.8528	1.3271	1.9842	0.232	58.3821	1.0682	2.366

Table 4.10 Objective quality evaluation with white noise

For comparison purpose the results of SSNR, WSS, LLR and PESQ for all conditions are shown in the form of bar chart in figures 4.7 to 4.11 respectively. The SSNR value for MSS and Wiener filter is higher under all test condition as compared to other methods. The WSS score is lower for MMSE STSA methods for most of the cases which reveals that the speech enhanced by these methods has lesser spectral distortion. In some cases the ML and MBSS methods give lower WSS but they have low SSNR in comparison with MMSE STSA methods. From LLR comparison the MMSE STSA algorithms have value less than one for most cases. Ideally LLR should be zero. The PESQ score above 2.5 is desirable from the noise perception and speech quality point of view. In this regards the MMSE STSA algorithms work satisfactorily. Hence it is concluded here that from all eight STSA algorithms the MMSE STSA algorithms are performing better compared to any other algorithm. Now from MMSE STSA 84 and MMSE STSA 85 algorithms the use of MMSE STSA85 algorithm is recommended for any future enhancement as it follows the LSA (Log Spectral Attenuation) characteristics of human ear.

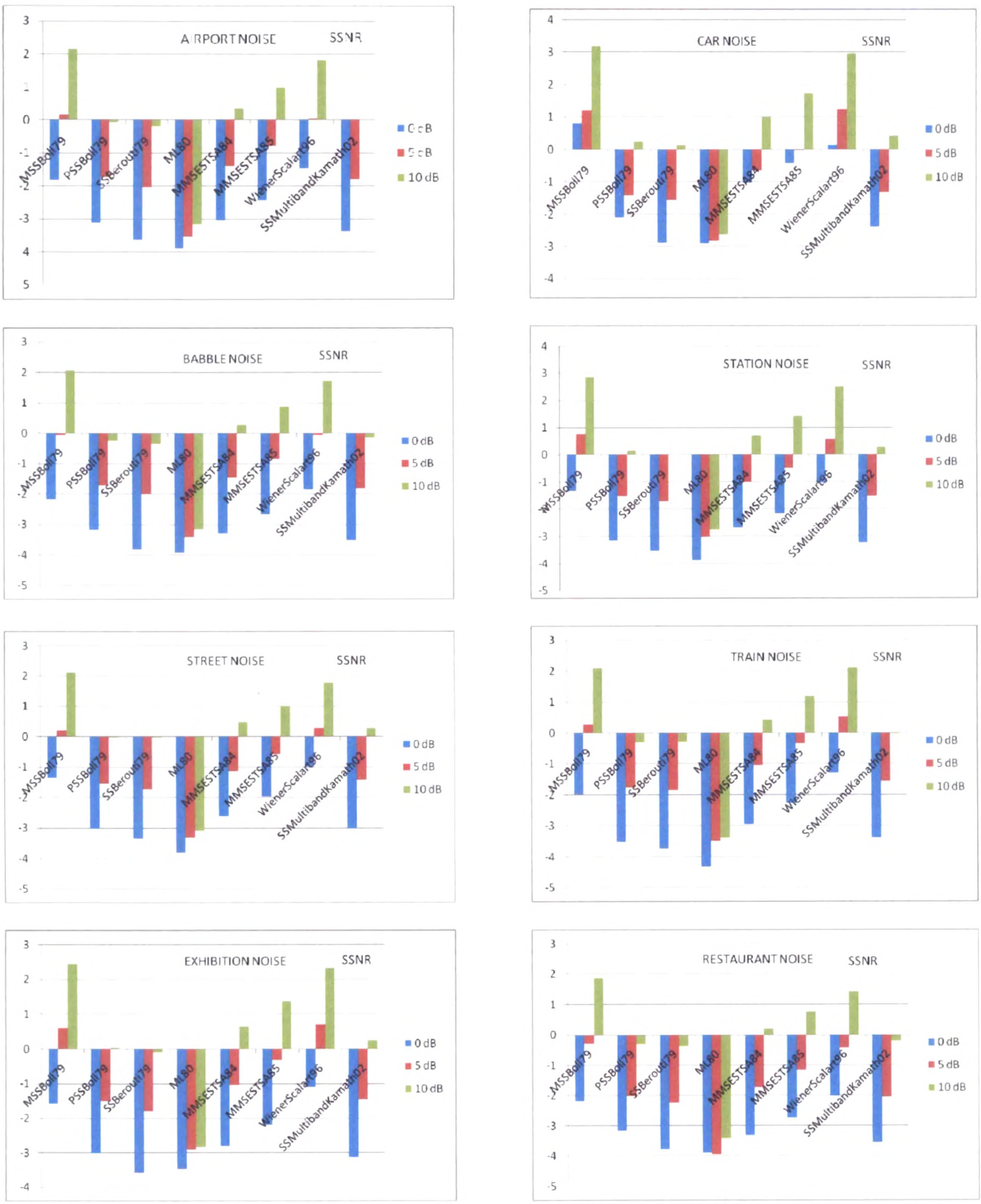


Fig. 4.7 SSNR comparison of STSA algorithms over NOIZEUS database

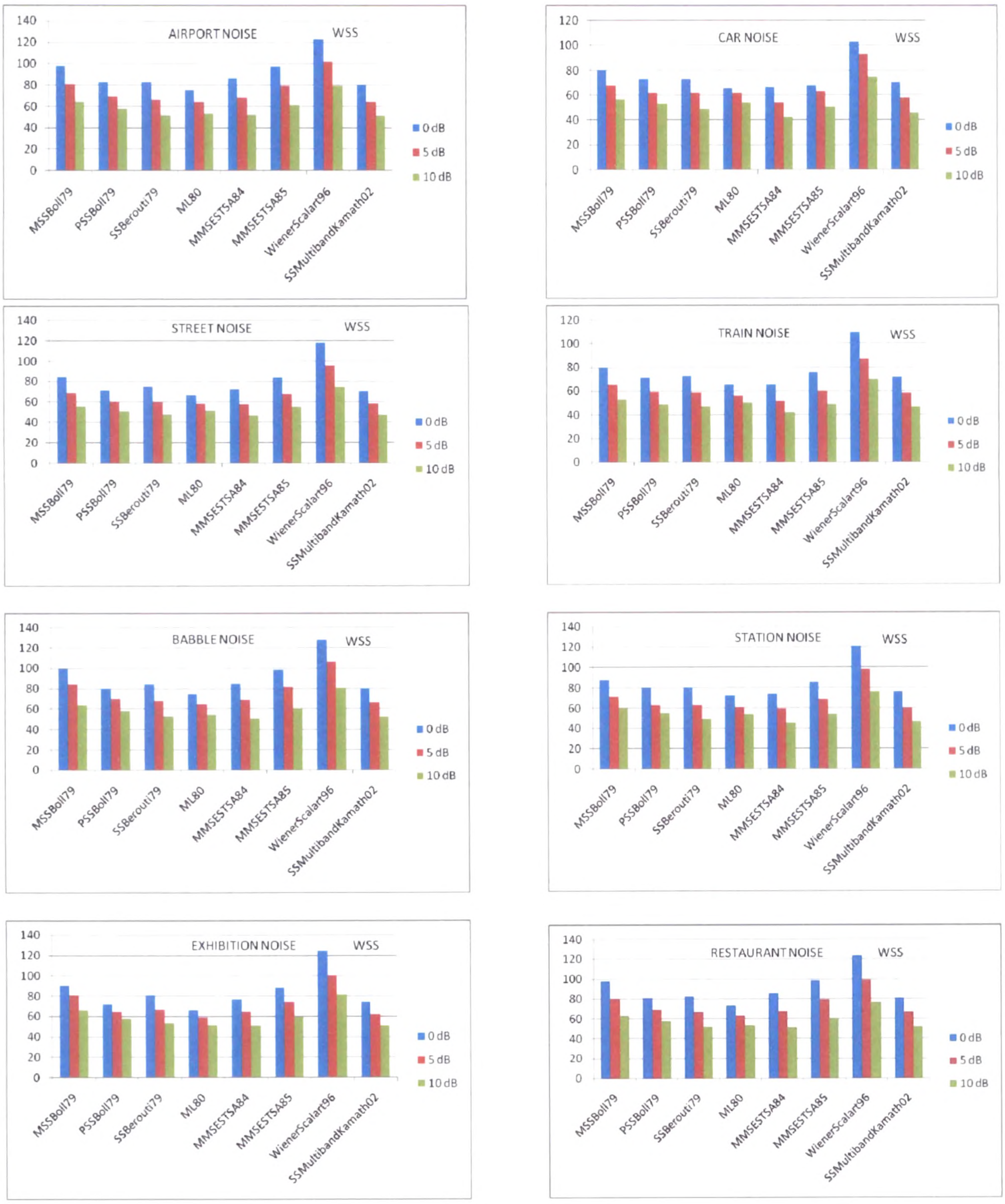


Fig. 4.8 WSS comparison of STSA algorithms over NOIZEUS database

Fig. 4.9 LLR comparison of STSA algorithms over NOIZEUS database

Fig. 4.10 PESQ comparison of STSA algorithms over NOIZEUS database

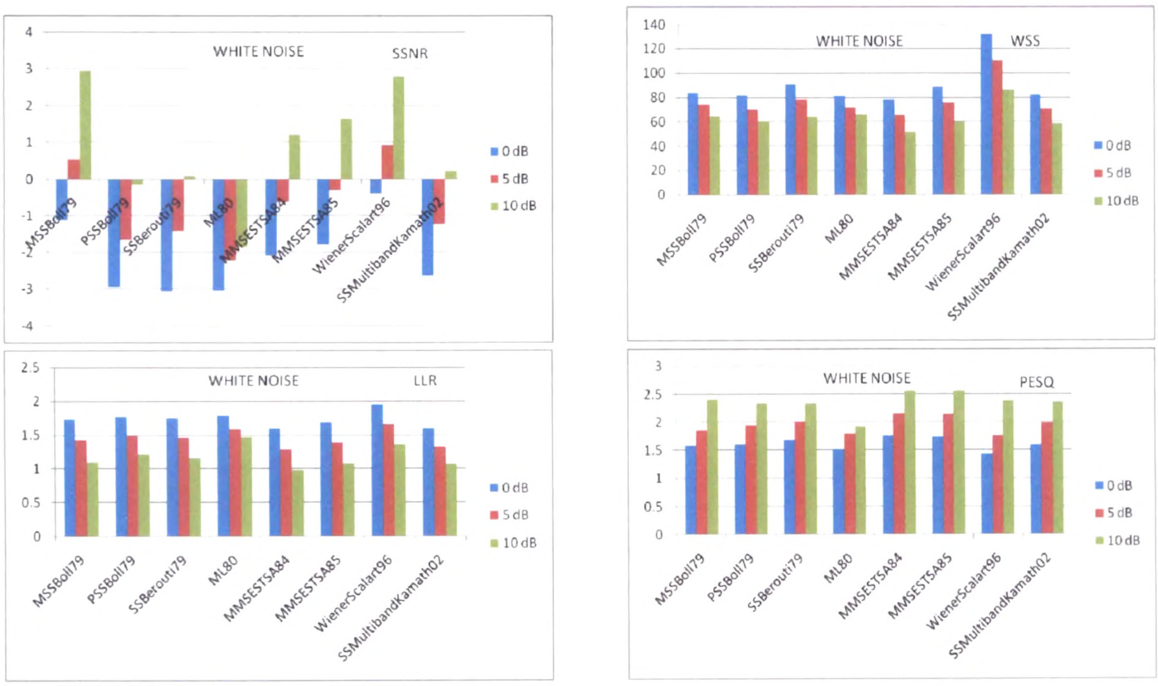


Fig. 4.11 Objective evaluation of STSA algorithms under white noise

4.5 Graphical User Interface (GUI)

To consolidate the simulation of STSA algorithms with white and colored noise a MATLAB GUI [11] is designed and it is depicted in figure 4.12.

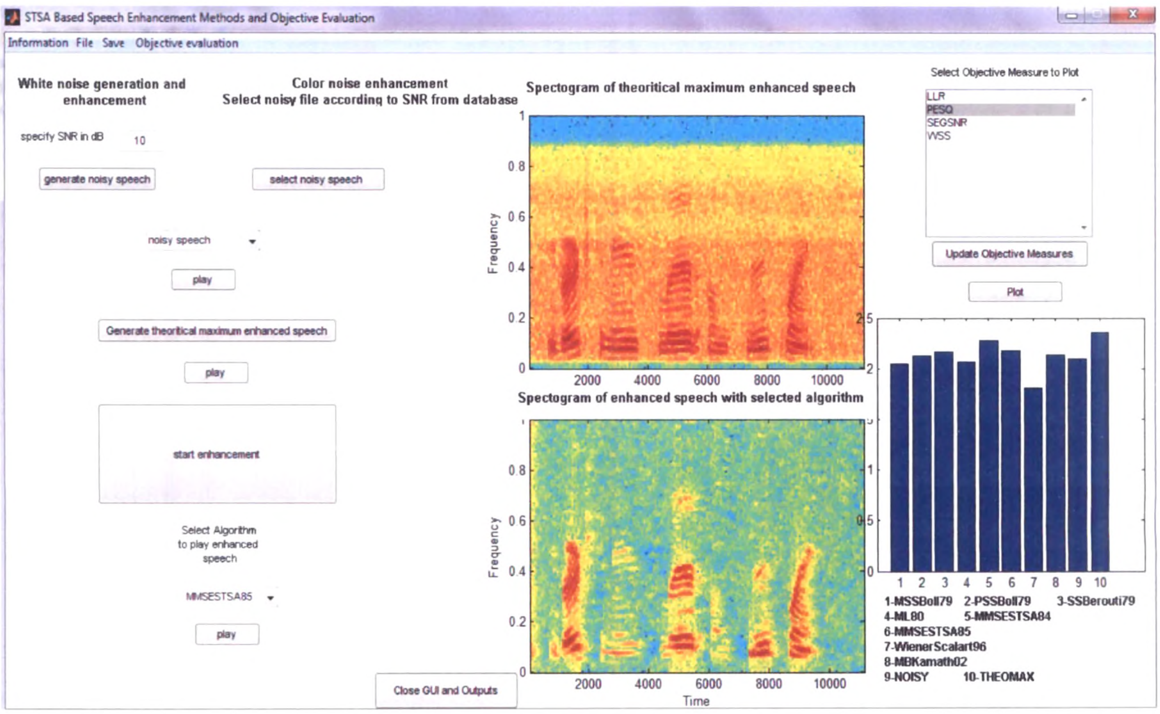


Fig. 4.12 MATLAB GUI for STSA algorithms

The important features of GUI are as follows.

1. It allows selecting a clean speech file from NOIZEUS database or any other .wav file at 8 KHz sampling frequency and 16bits/sample resolution.
2. The user can specify the SNR in dB for white noise which is added to clean speech file to generate the noisy file or the noisy file with particular SNR and type from NOIZEUS database.
3. The spectrograms of clean and noisy files are displayed and the file can be played to have listening experience. The eight different STSA algorithms can be applied to noisy file and the spectrogram of enhanced speech signal is displayed in GUI as well as in separate window for storage purpose.
4. The enhanced speech signal can be played as well as it can be saved in .wav file by specifying the name of output file.
5. The GUI also supports the objective evaluation using SSNR, WSS, LLR and PESQ scores.

6. The results can be displayed in tabular and bar graph forms. For reference the spectrograms and objective measures are also displayed for theoretical maximum limit (obtained by combining clean magnitude and noisy phase) as well as for noisy speech. The snapshot of the developed GUI is shown in figure 4.12. [1]

4.6 Implementation of Wavelet De-noising Methods

The wavelet de-noising using hard and soft thresholding with universal and SURE level dependent thresholds described in section 3.6 is implemented in MATLAB with different mother wavelets (Daubechies 20, Coiflets 4 and Symlet 20 with level 3). The objective evaluation results with white noise over SNRs 0 dB to 10 dB are summarized in table 4.11.

WHITE NOISE	0 dB				5 dB				10 dB			
	SSNR	WSS	LLR	PESQ	SSNR	WSS	LLR	PESQ	SSNR	WSS	LLR	PESQ
UNI_H_DB20	-2.8905	160.0671	5.775	0.6488	-1.0113	140.841	5.4235	0.9801	1.0555	118.8326	4.9883	1.3446
UNI_H_SYM20	-2.8715	159.5993	5.7625	0.6519	-0.9866	140.3427	5.3847	0.9738	1.1052	118.2422	4.9159	1.3666
UNI_H_COIF4	-2.9599	254.0805	3.7652	0.7467	-1.1484	208.4518	3.5026	1.0651	0.8767	168.5184	3.1939	1.4492
UNI_S_DB20	-3.0632	160.3541	5.8567	1.3663	-1.481	142.1192	5.5544	1.4404	0.1617	120.6072	5.1556	1.6331
UNI_S_SYM20	-3.049	160.0692	5.8474	1.342	-1.4585	141.4369	5.5194	1.4182	0.1893	119.821	5.0981	1.627
UNI_S_COIF4	-3.1417	258.2193	3.8408	1.3886	-1.6097	214.5133	3.626	1.4529	-0.021	175.2163	3.3394	1.6603
SURE_H_DB20	-3.121	97.7551	2.7404	1.6331	-0.8038	73.1521	1.5973	1.9518	1.9458	58.0661	1.1924	2.2562
SURE_H_SYM20	-3.1384	96.0477	2.5555	1.6397	-0.8195	72.9212	1.5699	1.952	1.9662	57.802	1.2075	2.2572
SURE_H_COIF4	-3.1715	86.0584	1.9675	1.6295	-0.9351	69.9765	1.3421	1.9366	1.8849	56.7398	1.1165	2.2503
SURE_S_DB20	-2.2239	100.9366	3.0619	1.9025	0.1356	77.5271	1.7359	2.2349	2.7272	61.6227	1.2222	2.5328
SURE_S_SYM20	-2.2133	99.7078	2.8725	1.904	0.1685	76.8788	1.6873	2.2413	2.7609	61.2674	1.2223	2.5414
SURE_S_COIF4	-2.2821	90.8677	2.1646	1.8536	0.0487	74.0614	1.3256	2.1825	2.6355	60.0887	1.047	2.4916

Table 4.11 Objective quality evaluation of wavelet de-noising methods

For comparison purpose same results are shown in bar chart form in figure 4.13. The results are very poor compared to STSA algorithms especially at low SNRs. The results with SURE soft thresholding are somewhat comparable to STSA methods. However, the results explain the reason for non popularity of wavelet de-noising for speech enhancement. The poor performance also encountered in colored noise conditions.

[1] A paper entitled "Performance Evaluation of STSA based Speech Enhancement Techniques for Speech Communication Systems" is presented in National conference on Wireless Communication and VLSI design (NCWCVD-2010) Organized by GEC, Gwalior and IEEE MP Subsection in March 2010.

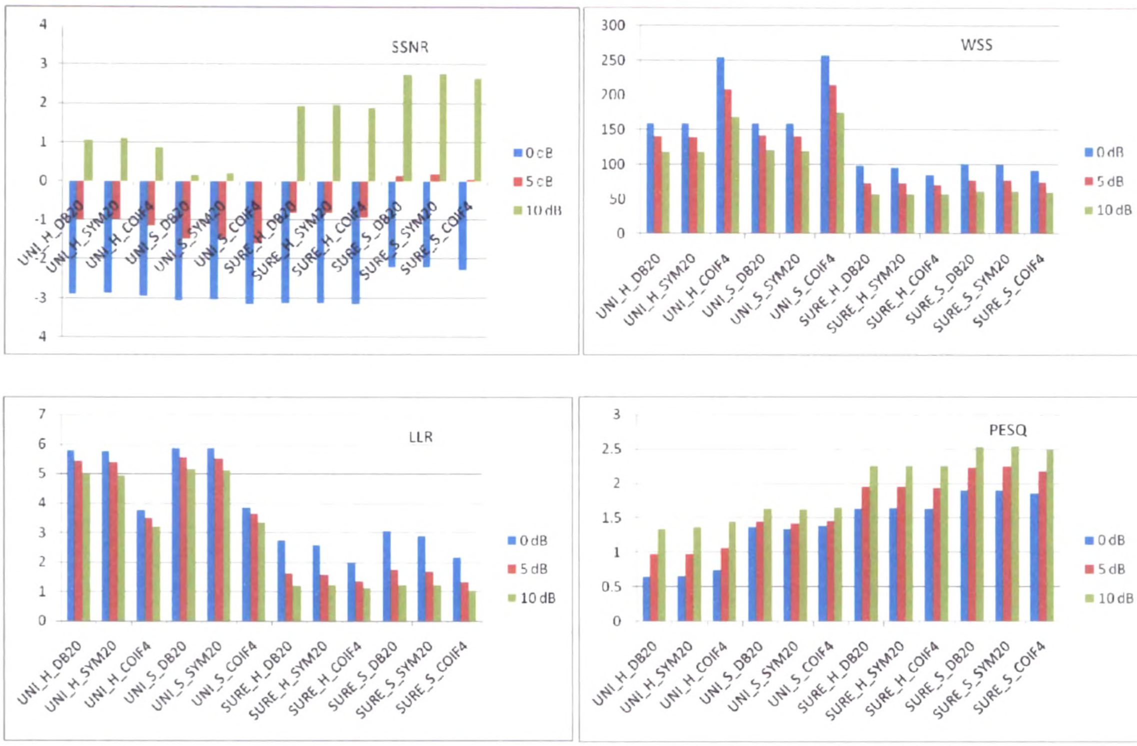


Fig. 4.13 Objective evaluation of wavelet algorithms under white noise

4.7 Summary

The MATLAB simulation of STSA and wavelet de-noising techniques along with their objective evaluations are described in this chapter. The details for implementation are shown using flow charts and block diagrams. The objective evaluation is performed by finding SSNR, LLR, WSS and PESQ scores for all algorithms under different noise conditions. The NOIZEUS database is utilized for evaluation. The MATLAB GUI is prepared which can simulate the STSA algorithms and also evaluates them. The discussion of objective evaluation results has concluded that the MMSE STSA85 algorithm is superior compared to all other STSA algorithms. The performance is found consistent in both white and colored noise environments. However, the performance is not satisfactory at low SNR conditions. The wavelet de-noising is not found much successful and feasible for real time speech enhancement systems. It is recommended here to shift the focus to other domains. The relative spectral analysis (RASTA) is novel approach for speech enhancement and it is described in the next chapter and performance is compared with the STSA algorithms.