Classical Tamil Speech Enhancement with Modified Threshold Function using Wavelets

Indra. J[†], Kasthuri. N* and Navaneetha Krishnan. S*

Abstract – Speech enhancement is a challenging problem due to the diversity of noise sources and their effects in different applications. The goal of speech enhancement is to improve the quality and intelligibility of speech by reducing noise. Many research works in speech enhancement have been accomplished in English and other European Languages. There has been limited or no such works or efforts in the past in the context of Tamil speech enhancement in the literature. The aim of the proposed method is to reduce the background noise present in the Tamil speech signal by using wavelets. New modified thresholding function is introduced. The proposed method is evaluated on several speakers and under various noise conditions including White Gaussian noise, Babble noise and Car noise. The Signal to Noise Ratio (SNR), Mean Square Error (MSE) and Mean Opinion Score (MOS) results show that the proposed thresholding function improves the speech enhancement compared to the conventional hard and soft thresholding methods.

Keywords: Tamil, Speech enhancement, Wavelet thresholding, Subjective evaluation, Objective evaluation

1. Introduction

Speech enhancement plays a significant role in voice communications and speech processing applications [1]. Generally in daily life, speech signal can get contaminated by various types of environmental noises. As a result it suffers in terms of quality and intelligibility. The speech enhancement or noise cancellation techniques are used to achieve significant reduction in the background noise level thus producing pleasant speech and improving the intelligibility [2]. Speech enhancement techniques are usefully employed as a pre-processing stage while designing robust automatic speech and speaker recognition systems, low bit rate speech coders, voice communication systems and aids for hearing impaired under noisy environment. Different approaches have been proposed for speech enhancement such as Spectral subtraction, Hidden Markov Modeling, Signal subspace methods and Wavelet-based methods [3-12]. All these approaches are based on English and other European Languages. There has been limited or no such research in speech enhancement for the Classical Tamil Language. In this paper an attempt towards Tamil Speech enhancement is made. A noisy speech corpus for classical Tamil Language is developed. Then, the popular wavelet transform thresholds for speech enhancement are discussed. Then, speech enhancement is attempted using

modified threshold function and the performances are compared based on subjective and objective measures.

2. Development of Noisy Speech Corpus for Classical Tamil Language

Speech is the vocalized form of human communication for conveying information. Speech processing is the study of the speech signals and the processing methods of these signals. An effective Tamil noisy speech corpus is created for speech enhancement research [13]. Pure Tamil speech is taken from IIIT Hyderabad database [14].

Some of the Tamil sentences are as follows:

1. எனக்கு யார் மீதும் ஐயம் இல்லை.
2. காவிரி தென்னிந்தியாவின் முக்கிய ஆறு.
3. இந்த ஊர் மிகப்பழமையான கிராமம்.
4. இது முற்றிலும் ஒரு சார்புநிலைக்கட்டுரை.
5. வேறு நல்ல சொற்கள் இருந்தால் தெரிவிக்கவும்.

Ten basic noise signals under different environmental conditions taken from the Columbia University, Newyork database [15] are used. Noisy speech signal is generated by combining speech and noise signals based on SNA algorithm.

3 Wavelet Thresholding

3.1 Discrete Wavelet Transform

The Discrete Wavelet Transform (DWT) is a linear

---

†   Corresponding Author: Department of Electronics and Instrumentation Engineering, Kongu Engineering College, Erode, India. (indrajaganathan@gmail.com)
*   Department of Electronics and Communication Engineering, Kongu Engineering College, Erode, India.
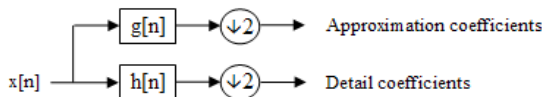   ({kasthurinehru, krishnanshanmugamn}@gmail.com)

Fig. 1. One level DWT

transformation which operates on a data vector whose length is an integer power of two. DWT transforms this vector into a numerically different vector of the same length. It separates data into different frequency components. It has recently been evolved as a powerful tool for removing noise from speech and image signals. The localizing property of the wavelets is helpful in thresholding and shrinking the wavelet coefficients that helps in separating the signal from noise.

The DWT of a signal x is calculated by passing it through a series of filters. First the samples are passed through a low pass filter with impulse response g resulting in a convolution of the two:

$$y[n] = (x * g)[n] = \sum_{k=-\infty}^{\infty} x[k]\, g[n-k] \qquad (1)$$

The signal is also decomposed simultaneously using a high pass filter h. The outputs giving the detail coefficients (from the high pass filter) and approximation coefficients (from the low pass). Fig. 1 shows one level DWT.

This constitutes one level of decomposition and can mathematically be expressed as follows:

$$y_{high}[k] = \sum_{n} x[n] * h[2k-n] \qquad (2)$$

$$y_{low}[k] = \sum_{n} x[n] * g[2k-n] \qquad (3)$$

where $y_{high}[k]$ and $y_{low}[k]$ are the outputs of the high pass and low pass filters, respectively, after sub sampling by 2.

The denoising by wavelet is quite different from traditional filtering approaches, because it is non-linear due to thresholding steps involved. Setting of a threshold criterion requires an accurate estimate of the additive noise level in the noisy speech. The higher frequency region of the wavelet coefficients mostly contains noise from which noise level is usually estimated. [16]. Fig. 2 shows the block diagram of speech enhancement using Discrete Wavelet Transform.

The basic steps in the denoising stage with the DWT are decomposition of the signal, thresholding, and recon-struction [17]. In the decomposition and reconstruction steps, the base and the level are the parameters of selection; while in the thresholding the parameters are the threshold and the rule of application [18].

3.2 Wavelet transform thresholds for enhancement

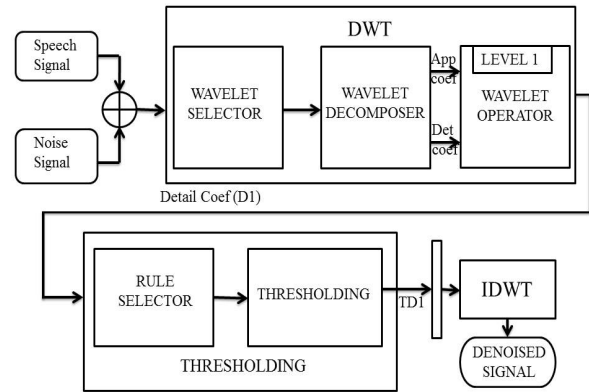The standard hard and soft thresholding functions used

Fig. 2. Block diagram of speech enhancement

in the wavelet-based methods are defined as the following equations respectively.

$$THR_H(Y,) = \begin{cases} Y, & |Y| \ge \lambda \\ 0, & |Y| < \lambda \end{cases} \qquad (4)$$

and

$$THR_S(Y,) = \begin{cases} Sign\,(Y)\,(|Y|\text{-}T), & Y \ge \lambda \\ 0, & |Y| < \lambda \end{cases} \qquad (5)$$

The denoised wavelet coefficients show discontinuity on threshold λ and $-\lambda$ in the hard thresholding approach. There is no discontinuity in the soft- thresholding method. But when the processed coefficient is greater than the threshold, a constant bias exists between the denoised and the original coefficients. In order to overcome the limitations in both the methods, a new approach is necessary. An improved noise reduction which retains the features of the original signal is possible by reducing the fixed bias of the reconstructed coefficients.

3.3 Proposed threshold function

New threshold function is proposed as modification from the conventional techniques.

$$Y = \begin{cases} Sign(X_{j,k}).exp(\text{-}\lambda/abs(|x|\text{-}\lambda^{2})), & |x| \ge |\lambda| \\ 0, & |x| < |\lambda| \end{cases} \qquad (6)$$

In the above equation Y is the denoised wavelet coefficient, $X_{j,k}$ is the noisy wavelet coefficient, λ is the subband adaptive threshold, X is the maximum wavelet coefficient of the subband under consideration. The exponential weighting function helps in reducing the fixed bias. It reduces the irregularities (noise) in the signal. The parameter x in the denominator of the exponential weighting function helps in further reducing the fixed bias of the reconstructed coefficient. Larger the absolute value of the denominator higher is the denoising coefficients value. Absolute value can account for the effects of fast-

varying noise components. The new threshold function better complies with the non-singular characteristic of the noise.

The value of threshold can be determined by using many ways:

A universal threshold for the Fast Wavelet Transform (FWT) was introduced [19] as

$$T = \sigma \sqrt{2\ln(N)} \tag{7}$$

For the Wavelet Packet Transform, the threshold value is

$$T = \sigma \sqrt{2\ln(N \log_2(N))} \tag{8}$$

A level dependent threshold was proposed [20] as

$$T_j = \sigma_j \sqrt{2\ln(N_j)} \tag{9}$$

4. Results and Discussion

The conventional and modified threshold functions are tested using the noisy speech corpus database developed. 4 sample speeches (by different speakers) combined with the three different noise samples, namely White Gaussian noise, Babble noise and Car noise at four different dBs (0db, 5db, 10db and 15db) are considered. Several wavelets from Daubechies, Symlets, Coiflet and Biorthogonal family are tried and one with the better output from each is presented. By trial and error to reach the best denoising effect, the level of decomposition is tried upto level 3. But level 1 yields better results and it is presented. After the third level, the obtained information is insufficient and the performance degrades.

4.1 Input speech signal combined with different noises

A sample sentence is as follows:
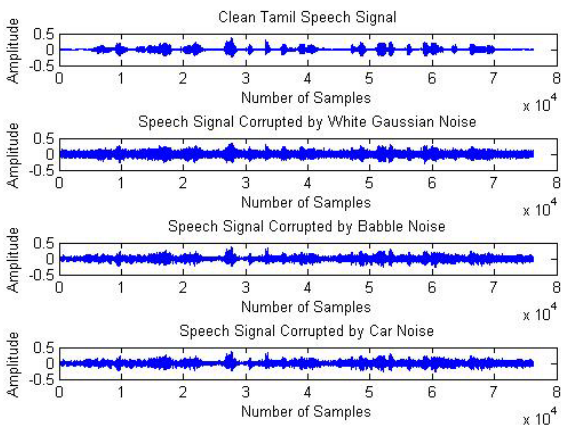
Fig. 3. Speech signal corrupted by noises (0dB)

"நீங்கள் உருவாக்கிய பேச்சுப்பக்கங்களில் சில கருத்துக்களை இட்டுள்ளேன்"

Fig. 3 shows speech signal combined with three different noises at 0dB.

4.2 Enhanced output from conventional threshold functions

Fig. 4 shows the clean speech, noisy speech and enhanced speech from White Gaussian noise (at 0dB) by existing threshold functions namely Hard and Soft Thresholding respectively.
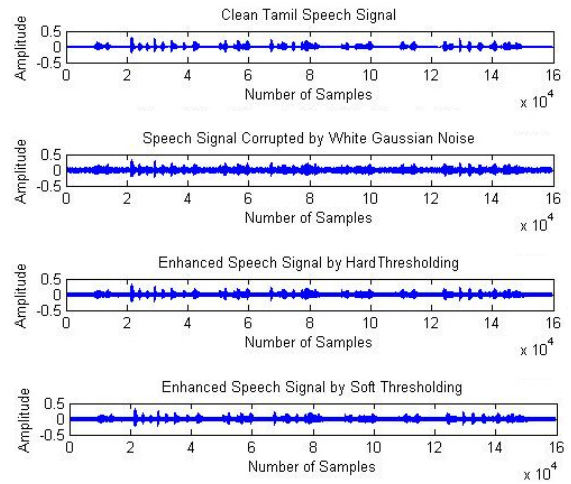
Fig. 4. Enhanced speech from white gaussian noise using hard and soft thresholding (0dB)

4.3 Enhanced output from new modified threshold function

Fig. 5 shows the clean speech, noisy speech and enhanced speech from White Gaussian noise (at 0dB) by the proposed threshold function. Similar results have been obtained for four sample speeches combined with three different noises at four different dBs'.

Fig. 5. Enhanced speech from white gaussian noise using new thresholding function (0db)

Table 1. Comparison of SNR values for conventional and modified threshold functions using db31, coif4, sym2 and bior1.1 wavelets
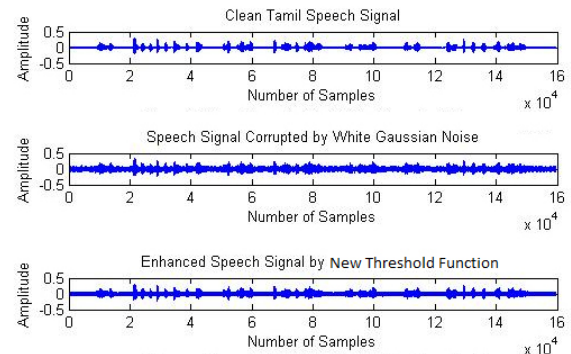
Wavelet types	Input Signal SNR (dB)		White Gaussian Noise			Babble			Car		
			Hard	Soft	NTF	Hard	Soft	NTF	Hard	Soft	NTF
db31	0	1	1.2512	2.8715	2.8834	0.0052	0.4570	0.4487	0.2806	0.2549	0.0242
		2	1.3059	3.0210	3.1834	0.0011	0.4570	0.3741	0.3160	0.2549	0.0341
		3	1.2609	3.0064	3.0011	0.0017	-0.012	0.0732	0.0032	0.1472	0.0514
		4	1.1966	2.9488	2.7142	0.0001	-0.006	0.0167	0.0032	0.1472	0.1991
	5	1	6.2629	7.9653	8.0053	4.9983	5.0006	5.4493	5.2732	5.2479	5.0239
		2	6.2292	7.9350	7.8143	4.7943	5.0006	5.0112	5.3046	5.2479	4.8147
		3	6.1930	7.9370	7.9463	4.8475	4.3415	5.7142	5.1973	5.9417	5.6641
		4	5.9130	6.0370	7.1263	4.8475	4.3415	4.7930	5.1973	5.9417	4.9883
	10	1	11.2201	12.916	12.552	9.9766	9.9786	9.9847	10.2500	10.227	10.021
		2	11.2315	12.904	12.812	9.9572	9.9786	9.4216	10.2688	10.227	9.4175
		3	11.1443	12.842	12.122	7.1542	9.9478	10.974	10.7519	10.964	10.759
		4	11.5143	12.642	12.622	7.1542	9.9478	9.3833	10.7519	10.964	9.4856
	15	1	16.1900	17.777	17.147	14.9084	14.950	14.957	15.1775	15.172	15.004
		2	16.2135	17.663	17.217	14.8535	14.950	13.941	15.1576	15.172	14.234
		3	16.0002	17.386	17.947	14.6749	14.712	16.008	15.6754	15.671	15.874
		4	16.0102	17.126	17.347	14.6749	14.712	13.317	15.6754	15.671	13.309
coif4	0	1	1.2569	2.8523	2.0022	0.0042	0.4599	0.2907	0.2595	0.1342	0.2907
		2	1.2046	1.0035	2.5422	-1.0484	0.4599	0.3340	0.2595	0.0226	0.3340
		3	1.1966	2.9488	2.7122	-0.9745	-1.214	-0.7861	0.0084	-0.418	-0.7861
		4	1.1966	2.9488	3.7722	-0.0987	0.0214	-0.7421	0.0004	-0.874	-0.7421
	5	1	6.2570	7.8785	6.9431	4.9921	4.9975	5.2778	5.2476	5.1283	5.2778
		2	6.2682	2.4995	7.4231	4.9834	4.9975	5.3162	5.2476	4.3112	5.3162
		3	6.1965	7.8839	7.8431	4.7158	4.3514	4.1528	4.0009	4.1492	4.1528
		4	5.1965	5.8839	8.0031	4.1541	4.8414	4.7128	5.1239	4.8442	4.7128
	10	1	11.216	12.916	12.941	9.9541	9.9541	10.237	10.214	10.108	10.237
		2	11.2417	8.1347	12.941	9.9317	9.9541	10.2605	10.214	8.4125	10.2605
		3	11.186	12.774	12.941	9.1574	8.9847	9.8475	9.6418	9.4155	9.8475
		4	10.288	11.774	12.941	9.1425	9.7447	8.7155	9.1218	8.1475	8.7155
	15	1	16.124	15.898	17.145	14.836	14.456	15.111	15.127	15.047	15.111
		2	16.0316	2.7401	17.145	14.7722	14.456	15.0887	15.127	12.415	15.0887
		3	16.043	17.424	17.145	14.5398	14.215	14.1429	14.759	13.645	14.1429
		4	16.143	17.324	17.245	14.3488	14.716	13.1419	14.249	12.745	13.1419
sym 2	0	1	1.2202	2.8659	2.9673	-0.0062	0.4645	0.4643	0.2951	0.2720	0.0217
		2	1.2599	2.8659	2.6673	-0.0051	-1.255	0.3208	0.3351	0.2947	0.0217
		3	1.2259	2.9045	2.9673	-0.0051	-1.255	0.0208	0.3351	0.2947	0.0318
		4	1.2259	2.9045	2.9673	-0.0987	0.0214	-0.0718	-0.7421	0.0004	-0.7189
	5	1	6.2079	7.8423	7.8671	4.9339	4.9680	5.3986	5.2310	5.2237	5.0070
		2	6.2230	7.8423	6.9412	4.9466	4.9708	5.2730	5.2829	5.2520	5.0070
		3	6.1522	7.7725	7.4652	4.9466	4.9708	5.3430	5.2829	5.2520	5.2786
		4	6.2122	7.4125	7.2671	4.1541	4.8414	4.7750	4.7128	5.1239	5.2908
	10	1	11.070	12.498	12.742	9.7498	9.8686	9.8767	10.034	10.291	9.9442
		2	11.126	12.498	12.742	9.7973	9.8835	9.6715	10.121	10.342	9.9442
		3	11.036	12.375	12.742	9.7973	9.8835	10.052	10.121	10.342	10.412
		4	11.038	12.485	12.867	9.1425	9.7447	9.3347	8.7155	9.1218	10.141
	15	1	15.696	16.542	16.648	14.214	14.744	14.596	14.464	14.744	14.704
		2	15.784	16.542	16.547	14.356	14.631	14.596	14.648	14.821	14.704
		3	15.633	16.246	16.748	14.356	14.631	15.751	14.648	14.821	14.854
		4	15.421	16.346	16.641	14.3488	14.716	13.198	13.1419	14.249	15.145
bior 1.1	0	1	1.2034	2.7637	2.8456	-0.0193	0.4498	0.4487	0.3223	0.2955	0.0250
		2	1.1951	2.7692	2.8456	-0.0030	-0.018	0.1749	0.3483	0.3037	-0.7516
		3	1.1963	2.8677	2.8456	-0.0074	-0.018	0.1749	-0.4153	-0.303	-0.1516
		4	1.2259	2.9045	2.9673	-0.0987	0.0214	-0.0718	-0.7421	0.0004	-0.7189
	5	1	6.0768	7.7891	7.5312	4.8047	4.8674	5.2571	5.1322	5.1333	4.9989
		2	6.0380	7.5366	7.5312	4.7943	4.8506	4.1587	5.1569	5.1324	4.1679
		3	6.0900	7.5300	7.5312	4.7563	4.8506	4.1587	4.7819	4.8324	4.5479
		4	6.1522	7.7725	7.8671	4.1541	4.8414	4.7750	4.7128	5.1239	5.2908
	10	1	10.644	11.529	12.512	9.2909	9.5200	9.5798	9.5806	9.7167	9.8549
		2	10.638	11.609	12.512	9.2812	9.4869	8.6374	9.6018	9.7017	3.7498
		3	10.628	11.562	12.512	9.7586	9.4869	8.6374	9.1428	9.3017	9.7498
		4	11.036	12.375	12.742	9.1425	9.7447	9.3347	8.7155	9.1218	10.141
	15	1	14.594	14.545	17.658	13.452	14.021	14.241	13.190	13.690	14.337
		2	14.564	14.533	17.658	12.970	13.561	13.849	13.204	13.640	13.672
		3	14.528	14.473	17.658	13.481	13.561	13.849	13.714	14.142	14.672
		4	15.633	16.246	16.648	14.3488	14.716	13.198	13.1419	14.249	15.145

Table2. Comparison of MSE values for conventional and modified threshold functions using db31, coif4, sym2 and bior1.1 wavelets

Wavelet types	Input Signal SNR(dB)		White Gaussian Noise			Babble			Car		
			Hard	Soft	NTF	Hard	Soft	NTF	Hard	Soft	NTF
db31	0	1	1.2399	1.0035	1.0001	6.3436	5.7101	5.7100	5.9468	5.9820	5.2115
		2	0.0012	7.7617	1.0023	0.0016	5.7101	5.7121	0.0016	5.9820	5.2335
		3	7.4758	5.0100	1.0012	0.0016	5.7101	1.4111	5.7518	0.7415	5.1445
		4	9.4128	6.0100	3.0000	0.0106	1.4640	1.4123	5.3618	5.4115	3.1248
	5	1	1.4899	1.0135	1.0123	2.0068	1.8089	2.0041	1.8837	1.8947	1.9840
		2	3.7494	2.5315	1.0123	4.9743	1.8089	2.0033	4.9760	1.8947	1.8410
		3	2.4013	1.6071	1.0024	3.2415	2.4159	3.5311	4.1247	4.1528	2.6640
		4	1.8113	1.7411	2.0025	2.2415	2.6182	2.1312	4.2137	3.9528	3.6741
	10	1	4.7899	3.2141	3.1245	6.3780	6.3612	6.3788	5.9887	6.0204	5.8561
		2	1.1850	8.0626	3.1245	1.5732	6.1612	6.3781	1.5732	6.0204	5.8412
		3	7.6793	5.1943	3.1245	6.4158	6.9582	5.4121	6.4198	6.3418	5.9741
		4	6.6793	5.4223	4.1244	5.4158	8.6101	5.3120	6.7588	6.4418	6.1674
	15	1	1.5252	1.0582	1.2112	2.0488	2.0257	2.0284	1.9257	1.9279	2.0124
		2	3.7630	2.6951	1.0411	4.9760	2.0257	2.0274	4.9760	1.9279	2.0067
		3	2.4988	1.8245	1.5411	3.7592	2.9847	3.4152	4.9760	4.6248	4.5234
		4	4.1581	1.9245	1.0101	3.6192	2.4844	3.4177	4.8760	4.7248	3.1214
coif4	0	1	1.3219	1.0063	1.0031	6.3376	5.7063	5.7041	5.9329	5.9757	6.3010
		2	0.0012	2.8715	1.0032	0.0016	6.5432	5.7041	0.0016	5.9757	5.4674
		3	7.5873	7.5873	1.0035	0.4126	2.5432	0.7084	0.0035	4.9757	4.4101
		4	9.4128	6.0100	3.0006	0.0106	1.4640	1.4111	5.3618	5.4115	3.1429
	5	1	1.5019	1.0083	1.0042	2.0097	1.8101	2.0011	1.8817	1.8949	1.9900
		2	3.7158	7.9904	1.0043	4.9760	2.1486	2.0036	4.9760	1.8949	2.7415
		3	2.3994	1.6269	1.0047	4.1547	3.1486	5.0044	4.7153	3.8949	4.1468
		4	1.8113	1.7411	2.0021	2.2415	2.6182	2.1322	4.2137	3.9528	3.6412
	10	1	4.7942	3.2410	3.2298	6.4111	6.3772	6.4133	6.0064	6.0383	6.3198
		2	1.1823	12.865	3.2298	1.5730	5.4713	6.4121	1.5736	6.0383	4.2399
		3	7.6053	5.2765	3.2296	3.2154	5.4713	6.4211	4.3251	4.0383	3.9361
		4	6.6793	5.4223	4.1242	5.4158	8.6101	5.3188	6.7588	6.4418	6.1888
	15	1	5.2549	5.6572	5.2141	2.0832	2.8412	2.0131	1.9554	1.9482	2.0741
		2	3.9260	17.591	5.2147	4.9760	1.8729	2.0133	4.9548	3.9482	3.4899
		3	2.4857	1.8086	5.2145	4.1529	3.8729	5.0144	4.9760	1.9482	1.4874
		4	4.1581	1.9245	1.0101	3.6192	2.4844	3.4112	4.8760	4.7248	3.1199
sym 2	0	1	1.2290	1.0325	1.0455	6.3527	5.7004	5.7011	5.9262	5.9845	6.3133
		2	0.0016	7.9199	1.8448	0.0016	4.2142	0.0187	0.0016	0.0015	6.3774
		3	7.5363	5.1204	1.8647	3.1248	3.0054	1.2022	1.2032	1.7519	5.1461
		4	0.0016	7.9199	1.8548	0.0016	4.2142	0.0199	0.0016	0.0015	6.3784
	5	1	1.5190	1.0425	1.0362	2.0368	1.8310	2.0211	1.9021	1.9052	2.0011
		2	3.7547	2.5715	2.3174	4.9760	2.1726	5.0088	4.9760	4.6955	4.6799
		3	2.4240	1.6692	2.3171	4.1586	6.5142	4.3166	3.1425	3.4127	4.6774
		4	3.7547	2.5715	2.3172	4.9760	2.1726	5.0187	4.9760	4.6955	4.6763
	10	1	4.9578	3.5690	3.4234	6.7198	6.5263	6.5495	6.2941	6.2119	6.4241
		2	1.2142	8.6485	4.1422	1.5736	5.4723	1.6814	1.5736	1.5242	5.2498
		3	7.8725	5.7820	4.1621	3.1475	4.9741	3.1899	0.0152	3.4159	5.2461
		4	1.2142	8.6485	4.1523	1.5736	5.4723	1.6847	1.5736	1.5242	5.2997
	15	1	1.7089	1.4062	1.3725	2.4039	2.2013	2.1237	2.6292	2.1278	2.1314
		2	4.1537	3.3161	6.4861	4.9760	3.4185	5.4841	4.9760	5.7425	1.8147
		3	2.7318	2.3741	6.4862	8.4816	7.7861	4.7574	3.1215	3.5136	6.7741
		4	4.1537	3.3161	6.4864	4.9760	3.4185	5.4412	4.9760	5.7425	1.8745
bior 1.1	0	1	1.2355	1.0074	1.1800	6.3720	5.7210	5.7841	5.8900	5.2964	6.3684
		2	0.0016	7.9199	1.8641	0.0016	4.2142	0.1241	0.0016	0.0015	6.3712
		3	7.5363	5.1204	1.8647	3.1248	3.0054	1.2141	1.2032	1.7519	5.1474
		4	0.0012	0.0016	2.4121	0.0016	5.7321	0.0841	0.0016	0.0015	2.3484
	5	1	1.5655	1.0084	1.1208	2.0983	1.8907	2.0674	1.9549	1.9454	2.0000
		2	3.7547	2.5715	2.3171	4.9760	2.1726	5.0100	4.9760	4.6955	4.6874
		3	2.4240	1.6692	2.3172	4.1586	6.5142	4.3214	3.1425	3.4127	4.6571
		4	3.9181	1.0084	2.3254	4.9760	4.4516	5.1541	4.9760	4.8204	1.3741
	10	1	5.4692	4.3920	3.8746	7.4688	6.9881	7.0741	6.9868	6.7712	6.5541
		2	1.2142	8.6485	4.1522	1.5736	5.4723	1.6321	1.5736	1.5242	5.2314
		3	7.8725	5.7820	4.1528	3.1475	4.9741	3.1741	0.0152	3.4159	5.2874
		4	1.3584	1.0933	3.7510	1.5736	7.2697	1.7784	1.5736	1.6854	4.6514
	15	1	2.2024	2.2278	2.4143	2.4039	2.2013	2.1341	3.0427	2.7127	2.3179
		2	4.1537	3.3161	6.4822	4.9760	3.4185	5.4984	4.9760	5.7425	1.8487
		3	2.7318	2.3741	6.4821	8.4816	7.7861	4.7964	3.1215	3.5136	6.7121
		4	5.5005	5.5400	4.8418	4.9760	3.4178	6.9471	4.9760	6.8053	2.8841

Table 3. Comparison of MOS values for conventional and modified threshold functions using db31, coif4, sym2 and bior1.1 wavelets

Wavelet types	Input Signal SNR (dB)		White Gaussian Noise			Babble			Car		
			Hard	Soft	NTF	Hard	Soft	NTF	Hard	Soft	NTF
db31	0	1	2.4	2.5	2.8	1.9	2.2	2.1	1.8	2.2	1.9
		2	2.3	2.6	2.9	1.7	2.3	2.0	1.6	2.1	1.8
		3	2.2	2.8	2.7	1.6	2.2	2.2	1.7	2.2	1.7
		4	2.4	2.9	3.1	1.9	2.5	2.3	1.7	2.2	2.5
	5	1	2.5	2.6	2.8	1.6	1.9	2.7	1.5	1.8	2.6
		2	2.4	2.5	2.7	1.5	1.6	2.5	1.4	1.6	2.6
		3	2.1	2.4	2.5	1.4	1.8	2.4	1.3	1.4	2.5
		4	2.1	2.5	2.7	1.3	1.7	2.7	1.3	1.4	2.7
	10	1	2.5	2.6	2.6	1.7	2.5	2.7	1.6	2.3	2.4
		2	2.2	2.4	2.4	1.6	2.4	2.5	1.5	2.1	2.5
		3	2.3	2.5	2.3	1.8	2.6	2.6	1.4	2.2	2.4
		4	2.4	2.4	2.4	1.8	2.5	2.7	1.4	2.2	2.4
	15	1	2.6	2.8	2.8	1.6	2.4	2.6	1.7	2.4	2.5
		2	2.4	2.7	2.7	1.5	2.2	2.4	1.6	2.3	2.4
		3	2.1	2.8	2.6	1.4	2.5	2.3	1.7	2.4	2.4
		4	2.4	2.9	2.7	1.4	2.4	2.5	1.7	2.4	2.3
coif4	0	1	2.1	2.4	2.6	1.7	2.0	1.8	1.9	2.2	2.4
		2	2.2	2.3	2.6	1.4	1.8	1.7	1.7	2.3	2.4
		3	2.1	2.4	2.4	1.5	1.7	1.6	1.7	2.4	2.3
		4	2.3	2.2	2.5	1.3	1.4	1.6	1.8	2.0	2.3
	5	1	2.2	2.4	2.7	1.9	2.0	2.1	1.8	2.3	2.3
		2	2.3	2.2	2.4	1.7	1.8	1.9	1.4	2.6	2.3
		3	2.2	2.3	2.5	1.6	1.7	1.4	1.5	2.5	2.2
		4	2.1	2.2	2.4	1.3	1.4	1.6	1.4	2.4	2.2
	10	1	2.4	2.5	2.6	1.6	1.8	2.3	1.9	2.2	2.5
		2	2.6	2.4	2.5	1.3	1.6	2.2	1.4	2.1	2.6
		3	2.3	2.3	2.4	1.4	1.4	2.1	1.3	2.2	2.4
		4	2.4	2.5	2.3	1.6	1.6	2.2	1.6	2.2	2.4
	15	1	2.5	2.6	2.4	1.8	1.5	2.1	1.7	2.4	2.6
		2	2.2	2.7	2.3	1.6	1.6	2.0	1.3	2.3	2.5
		3	2.5	2.6	2.3	1.5	1.6	2.2	1.2	2.1	2.6
		4	2.5	2.7	2.4	1.8	1.9	2.4	1.7	2.4	2.5
sym 2	0	1	2.3	2.5	2.3	1.8	1.9	2.6	1.6	1.8	2.4
		2	2.2	2.5	2.4	1.6	1.9	2.4	1.5	1.8	2.5
		3	1.9	2.4	2.3	1.5	1.7	2.3	1.4	1.3	2.0
		4	1.8	2.4	2.2	1.4	1.8	2.5	1.3	1.4	2.5
	5	1	2.4	2.2	2.6	1.6	1.6	1.9	1.7	1.9	2.1
		2	2.6	2.2	2.5	1.3	1.5	1.7	1.6	1.9	2.2
		3	1.8	2.1	2.1	1.4	1.7	1.6	1.4	1.8	2.3
		4	1.9	2.1	2.2	1.6	1.9	1.8	1.4	1.6	2.4
	10	1	2.2	2.1	2.6	1.4	1.8	2.4	1.8	1.7	2.0
		2	2.1	2.1	2.4	1.2	1.9	2.3	1.7	1.7	2.2
		3	2.1	2.0	2.2	1.3	1.7	2.2	1.5	1.6	1.8
		4	2.3	2.0	2.3	1.6	1.9	2.4	1.9	2.0	2.3
	15	1	2.2	2.3	2.4	1.5	1.7	2.3	1.5	1.6	2.1
		2	2.4	2.3	2.3	1.1	1.8	2.2	1.3	1.5	2.0
		3	2.2	2.4	2.1	1.0	1.9	2.1	1.2	1.4	2.1
		4	2.6	2.4	2.1	2.0	2.1	2.6	1.8	1.9	2.5
bior 1.1	0	1	1.9	2.2	2.3	1.6	1.9	2.1	1.4	1.6	1.6
		2	1.7	2.1	2.4	1.5	1.8	2.0	1.3	1.5	1.5
		3	1.3	2.0	2.3	1.5	1.9	2.1	1.3	1.5	1.3
		4	1.4	1.5	1.5	1.4	1.5	2.2	1.4	1.4	1.5
	5	1	2.1	2.1	2.7	1.7	1.8	1.6	1.6	1.8	2.0
		2	2.1	1.8	2.5	1.5	1.7	1.5	1.2	1.7	2.1
		3	2.0	1.4	2.5	1.5	1.8.	1.6	1.2	1.7	2.0
		4	2.2	1.7	2.1	1.6	1.5	1.8	1.6	1.6	2.7
	10	1	1.9	1.9	2.3	1.9	1.7	2.2	1.5	1.9	2.1
		2	1.3	1.8	2.4	1.4	1.8	2.4	1.4	1.7	2.4
		3	1.4	1.4	2.0	1.4	1.9	2.7	1.4	1.7	2.1
		4	2.5	2.6	2.4	1.7	1.8	2.1	1.7	1.9	2.2
	15	1	1.8	2.1	2.2	1.8	1.8	2.1	1.4	2.1	2.5
		2	1.4	2.0	2.1	1.7	1.5	2.5	1.3	2.1	2.3
		3	1.7	1.8	2.2	1.7	1.4	2.7	1.3	2.1	2.4
		4	2.9	2.8	2.4	1.8	1.9	2.3	1.8	2.0	2.5

4.4 Performance measures

To evaluate the performance of the proposed threshold functions, objective and subjective measures are considered.

The most popular objective measure, which gives account of the signal power improvement related to the noise power is SNR [21]. It is defined as the power ratio between speech signal (meaningful information) and background noise (unwanted information)

$$SNR = 10 \log_{10} (P_{signal} / P_{Noise}) \qquad (10)$$

Mean Square Error (MSE) is defined as to be the average power of the difference between the enhanced speech and clean one [22]. It can be obtained by

$$MSE = \sum \{ |s(n) - s(n)|^2 \} \qquad (11)$$

The Mean Opinion Score (MOS) provides a numerical measure of the quality of human speech [23]. The scheme uses subjective tests (opinionated scores) that are mathematically averaged to obtain a quantitative indicator of the system performance. To determine MOS, a number of listeners rate the quality of test sentences by hearing test. Based on the perceived quality of enhanced speech, listener gives a rating for each sentence as follows: (1) Bad (2) Poor (3) Fair (4) Good (5) Excellent. The MOS is the arithmetic mean of all the individual scores, and can range from 1 (worst) to 5 (best). For evaluation, a group of 10 listeners are used.
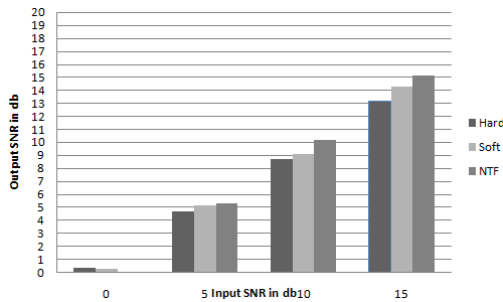
Fig. 6. Comparison of SNR values using bior 1.1 wavelet for white gaussian noise
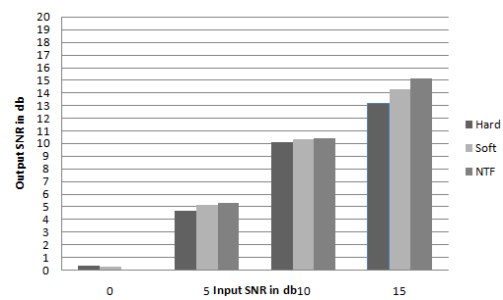
Fig. 7. Comparison of SNR values using bior 1.1 wavelet for babble noise

4.5 Performance comparison

Four sample speeches combined with three different noises at four different dB levels (0, 5, 10, and 15) were taken as input. The enhanced output from four different wavelets, db31, coif4, sym2 and bior1.1 from the Daubechies, Coiflets, Symlets and Bioorthogonal families respectively are presented.

The comparison of conventional and modified threshold functions based on SNR values using the above four wavelets is shown in Table 1. Fig. 6 to 8 show the comparison of SNR values using bior 1.1 wavelet for the three different noises.

Table 2 shows the comparison of conventional and modified threshold functions based on MSE values using the above four wavelets. The comparison of MSE values using coif4 wavelet for the three different noises is shown from Fig. 9 to 11.
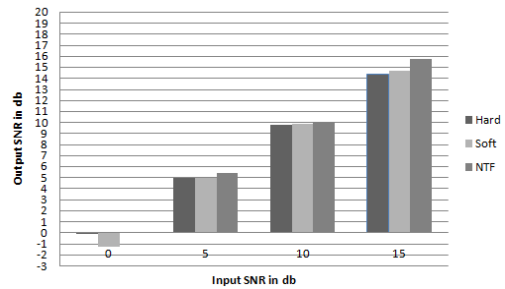

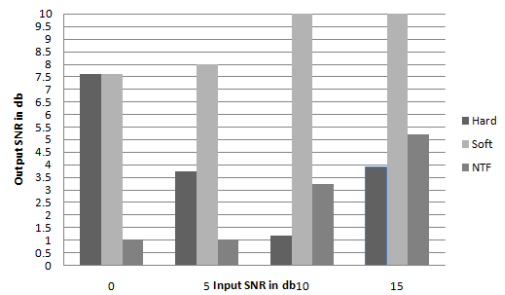**Fig. 8.** Comparison of SNR values using bior 1.1 wavelet for car noise

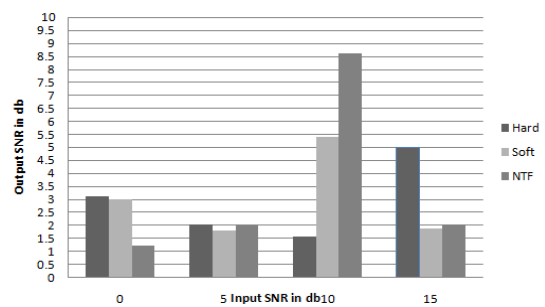**Fig. 9.** Comparison of MSE values using coif4 wavelet for white gaussian noise

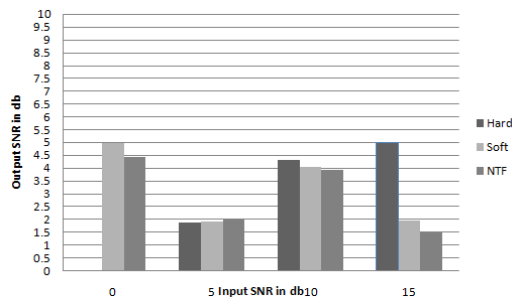**Fig. 10.** Comparison of MSE values using coif4 wavelet for babble noise

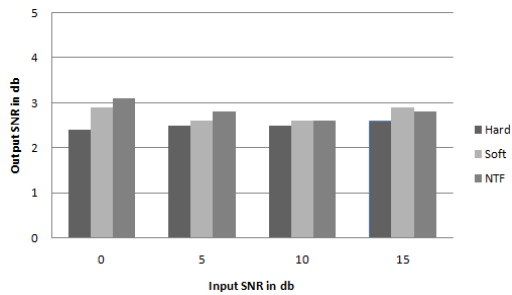**Fig. 11.** Comparison of MSE values using coif4 wavelet for car noise

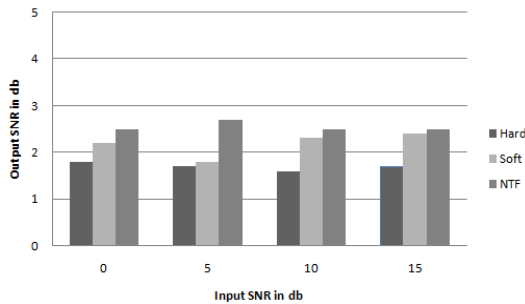**Fig. 12.** Comparison of MOS values using db31 wavelet for white gaussian noise

Fig. 13. Comparison of MOS values using db31 wavelet for babble noise
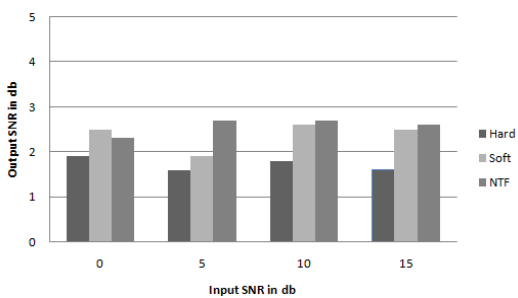

Fig. 14. Comparison of MOS values using db31 wavelet for car noise

The comparison of conventional and modified threshold functions based on MOS values using the above four wavelets is shown in Table 3. Fig. 12 to 14 show the comparison of MOS values using dB31 wavelet for the three different noises.

5. Conclusions and Future Work

In this paper speech enhancement using wavelets is investigated. New modified threshold function is presented and different wavelets are adopted for speech enhancement. Speech enhancement is carried out using conventional and new modified threshold functions. A comparison of new modified threshold function with the conventional soft and hard thresholding techniques is investigated. The presented method is tested with the different subjective and objective measures such as Signal to Noise Ratio, Mean Square Error and Mean Opinion Score. New modified threshold function and db31 wavelet show an improvement in SNR, MSE and MOS values. Tamil is one of the ancient classical languages and is spoken by 8 million people all over India. This proposed speech enhancement approach can be used as a preprocessing stage in Tamil speaker recognition systems. In future, wavelet packet thresholding may be used to obtain enhanced speech from noisy speech. Also various other real world noises may be included for testing in future.

References

[1] RichaTyagi, SunitaMaithani, "A Hybrid Speech Enhancement System Based On Wavelet Denoising", in *Proceedings of 13th Oriental COCOSDA Workshop*, 2010.

[2] Sachin Singh, Manoj Tripathya & R. S. Ananda, "Subjective and Objective Analysis of Speech Enhancement Algorithms for Single Channel Speech Patterns of Indian and English Languages", *IETE Technical Review,* Vol. 31, No. 1, 2014

[3] J.R. Deller, J.H.L. Hansen, J.G. Proakis, Discrete-Time Processing of Speech Signals, Second ed. IEEE Press. 2000.

[4] S.F. Boll, "Suppression of acoustic noise in speech using spectral subtraction", *IEEE Trans. Acoust. Speech Signal Process*. ASSP-27, pp. 113-120, 1979.

[5] M. Berouti, R. Schwartz, J. Makhoul, "Enhancement of speech corrupted by acoustic noise", in *Proc. IEEE Internat. Conf. on Acoust. Speech Signal Process. (ICASSP),* Washington DC, pp. 208-211, 1979.

[6] S. Kamath, P. Loizou, "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise", in *Proc. IEEE Internat. Conf. Acoust. Speech Signal Process.* (*ICASSP*), Orlando, Florida, 2002.

[7] Y. Ghanbari, M. Karami, "Spectral subtraction in the wavelet domain for speech enhancement", *Internat. J. Software Inf. Technol.* (IJSIT) 1 (1), pp. 26-30, 2004.

[8] Y. Ghanbari, M. Karami, B. Amelifard, "Improved multiband spectral subtraction method for speech enhancement", in *Proc. 6th IASTED Internat. Conf. on Signal Image Process*. USA, pp. 225-230, 2004.

[9] H. Sameti, H. Sheikhzadeh, Deng, Li, R.L. Brennan, "HMM-based strategies for enhancement of speech signals embedded in nonstationary noise", *IEEE Trans. Speech AudioProcess*. 6 (5), pp. 445-455, 1998.

[10] M. Klein, P. Kabal, "Signal subspace speech enhancement with perceptual post-filtering" in *Proc. IEEE Internat. Conf. Acoust. Speech Signal Process. (ICASSP)* 1, pp. 537-540, 2002.

[11] H. Sheikhzadeh, H. R. Abutalebi, "An improved wavelet based speech enhancement system", in *Proc. 7th Eur. Conf. Speech Comm. Technol. (Euro Speech),* Aalborg, Denmark, September 2001.

[12] J. Seok, K. Bae, "Speech enhancement with reduction of noise components in the wavelet domain", in *Proc. IEEE Internat. Conf. Acoust. Speech Signal Process. (ICASSP)* 2, pp. 1323-1326, 1997.

[13] J. Indra, N. Kasthuri, "Development of Noisy Corpus Database for Classical Tamil Language", *International Journal of Applied Engineering Research*, 2014, Vol. 9, no. 23, pp. 23039-23048, 2014

[14] K. Prahallad, E. Naresh Kumar, V. Keri, S. Rajendran and A. W Black., "The IIIT-H Indic Speech Databases," in *Proceedings of Inter speech,* Portland, 2012. OR. Available: http://speech.iiit.ac.in/index. php/research-svl/69.html

[15] Noise signals is referred from Columbia University-Newyork http://www.ee.columbia.edu/~dpwe/sounds/noise/

[16] D. L. Donoho and I. M. Johnstone," Ideal spatial adaptation by wavelet shrinkage," *Biometrica*, vol. 81, no. 3, pp. 425-455, 1994.

[17] D. L. Donoho, "De-noising by soft thresholding," *IEEE Trans. Inform. Theory*, vol. 41, pp. 613-627, May 1995.

[18] D. M. Ballesteros, "Procesamiento digital de señalesutilizando Matlab y Simulink. Chapter: Transformada Wavelet Discreta". Ed. Orcas, pp. 61-83, 2010.

[19] D. L. Donoho, I. M. Johnstone, "Threshold selection for wavelet shrinkage of noisy data", *16th Annual International Conference of the IEEE*, pp. A24-A25, 1994.

[20] I. M. Johnston, B. W. Silverman, "Wavelet threshold estimators for data with correlated noise", J. Roy. Statist. Soc. Ser. B 59, pp. 319-351, 1997.

[21] V. Turbin, N. Faucheur, "Estimation of speech quality of noise reduced signals", in *Proceedings of Online workshop meas. speech audio quality network*, 2007.

[22] Bernhard Wieland, thesis (October 2009) "Speech Signal Noise Reduction with Wavelets", pp. 55-56

[23] ITU-T. P. 835. "Series P: Telephone transmission quality, telephone installations, local line networks: methods for objective and subjective assessment of quality", 2003.

J. Indra She received her B.E Degree in Electrical & Electronics Engineering from Bharathiyar University, Tamilnadu, India in 2003 and M.E. Degree in VLSI Design from Anna University, Chennai, Tamilnadu, in 2007. She is currently working as an Assistant Professor (Selection Grade) in the Department of Electronics & Instrumentation Engineering at Kongu Engineering College. Her research interests include Speech signal processing, VLSI and Embedded Systems.

N. Kasthuri She obtained her Doctoral Degree in Information and Communication Engineering from Anna University, Chennai, Tamilnadu, India in 2008. She received her B.E Degree in Electronics and Communication Engineering from A.C tech Karaikudi and M.E. Degree in Applied Electronics from the Bharathiyar University, Tamilnadu, India in 1990 and 1999 respectively. She is currently working as a Professor in the Department of Electronics and Communication Engineering, Kongu Engineering College, India. Her research interests cover Signal Processing, Speech Signal Processing and Embedded Systems.

S. Navaneetha krishnan completed his B.E Degree in Electronics and Communication Engineering and M.E degree in Applied Electronics from Anna University, Chennai in 2013 and 2015 respectively. Now he works in Tata Consultancy Services.