IJournals: International Journal of Software & Hardware Research in Engineering
ISSN-2347-4890
Volume 4 Issue 6 June, 2016

VARIATION IN NOISE PARAMETER ESTIMATES FOR BACKGROUND NOISE CLASSIFICATION

Author: Aarti Neema[1]; Sujeet Mishra[2]

Affiliation: Ms. Aarti Neema M.Tech Student EC Department SIMS, Indore[1];
Mr. Sujeet Mishra Head of EC Department SIMS, Indore[2];
aarti_neema@yahoo.com[1]; sujeetmishra84@gmail.com[2]

ABSTRACT

In this paper, we investigate variation in speech parameter estimates which can be used to classify environmental noise for grouping a large range of environmental noise into a reduced set of classes of noise with similar type of speech characteristic parameters. One hundred original noises from environment were recorded with the help of a microphone connected to personal computer & stored as a noise database in memory of the computer. Built-in programs for Linear predictive coding (LPC) and Real cepstral parameter (RCEP) have been used while user defined program was written in MATLAB for Mel Frequency Cepstral coefficient (MFCC) in MATLAB to estimate variation in speech parameters which may be utilized for speech analysis through any one of the soft computing techniques viz. neural networks, fuzzy logic, genetic algorithms or a combination of these. Twenty five samples each of four commonly encountered environmental noises (o2car1-o2car25, o3office1-o3office25, o4market1-o4marke25 & o5train1-o5train25) i.e. 100 noises in total have been considered in our study for estimation of three coefficients viz. Mel Frequency Cepstral coefficient, Linear predictive coding and real cepstral parameter. Our experimental results show that Mel Frequency Cepstral Frequencies are robust features for finding out variation in noise parameter estimates. Twenty seven filter banks were used and filter bank output along with power spectrum was obtained in MATLAB. By experimentation through trial & error method, it was found that while considering average of second highest & third highest MFCC coefficients, the noise parameter estimates varied by at most 1% only when internet noise samples were compared to those of original noise samples.

Index Terms- Mel Frequency Cepstral Coefficient (MFCC), Linear Predictive Coding (LPC),

1. INTRODUCTION

Since over two decades, several algorithms and techniques have been proposed by many researchers regarding classification of environmental noise using parameters such as power spectral density (PSD), zero crossing rate (ZCR), line spectral frequency (LSF) and log area ratio (LAR) coefficients but none of the techniques have proven to be highly effective because of their own inherent limitations associated with each technique so far. Recently, different research groups have carried out studies on new methods and algorithms for environmental noise classification but in current paper, we have tried to explore noise parameter estimation variants for speech analysis. In our day-to-day life, we encounter different types and levels of environmental acoustical noises like train noise, office noise, market noise etc. In various speech analysis and processing systems such as speech recognition, speaker verification and speech coding, the unwanted noise signals are picked up along with the speech signals which often cause degradation in the performance of communication systems [1]. After modification of processing according to the type of background noise, the performance can be enhanced which requires noise classification based on speech parameter estimation and characterization. Background noise classifier can be used in various fields as, speech recognition and coding being the main ones. Acoustic features can be made adaptable to the type of environmental noise by choosing the most appropriate set to ensure separability between phonetic classes. Since low cost DSP's are increasingly becoming popular, therefore, the next generation of speech coders and intelligent volume controllers are likely to include classification modules in order to improve robustness to environmental/ background noise [2] .

2. ENVIRONMENTAL NOISE CLASSIFICATION METHODOLOGY

The type of methodology that can be adopted for environmental noise classification through parameter estimation variants is based on exploring any one or a few of the environmental noise parameters viz .Linear Predictive Coding, Mel-cepstral based parameters, Real Cepstrum based parameters, line spectral frequencies coefficients, log area ratio coefficients, zero crossing rate and power spectral density [3]. From these noise parameters, we have explored and analyzed two main parameters Linear predictive coding, Mel frequency cepstral coefficients and one allied parameter i.e. real cepstrum parameter for internet noise samples as well as original recorded samples in this paper. Noise database created can be explored on basis of noise classes as follows:

- ***Automobiles noise class (ANC)*:** Cars, trucks, buses, trains, ambulance, police cars etc
- ***Babble noise class (BNC)*:** Cafeteria, sports, stadium, office etc
- ***Factory noise class (FNC)*:** Tools such as drilling machines, power hammer etc.
- ***Street noise class (SNC)*:** Shopping mall, market, busy street, bus station, gas station etc.
- ***Miscellaneous noise class (MNC)*:** Aircraft noise, thunder storm etc

Out of these noise classes, only three noise classes have been considered viz. car & train noise from automobile noise class (ANC), office noise from babble noise class (BNC) and market noise from street noise class (SNC).

3. SPEECH PARAMETER ANALYSIS

The variants of speech parameters have been analyzed by acoustic-phonetic approach after spectral analysis. The first step in speech processing is feature measurement which provides an appropriate spectral representation of the characteristics of the time-varying speech signal by filter bank method implemented in MATLAB. Signal representation of internet downloaded and original car noise is as follows:

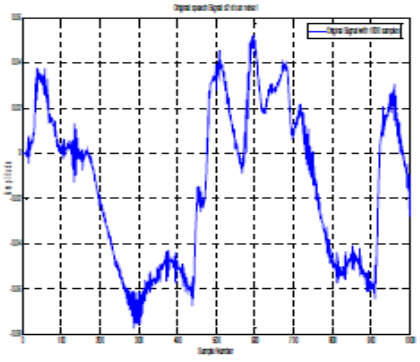

Fig1: Internet Car noise signal (s2car1) representation in MATLAB

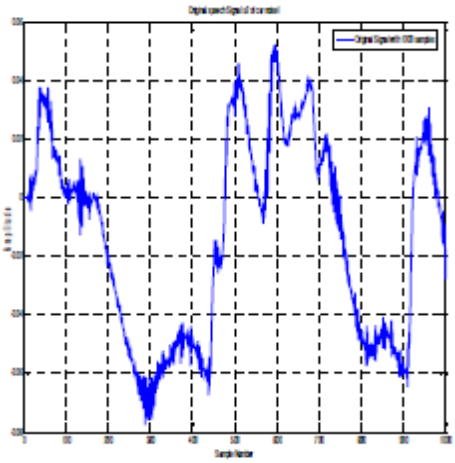

Fig2: Original Car noise signal (o2car1) representation in MATLAB

The most common type of filter bank used for speech analysis is the uniform filter bank for which the center frequency, fi, of the ith band pass filter is defined as

Fi = Fs/N i 1 < i < Q,

where Fs is the sampling rate of the speech signal, and N is the number of uniformly spaced filters required to span the frequency range of the speech [4]. The actual number of filters used in the filter bank, Q, of our work satisfies the relation

Q < N / 2 < 54/2 < 27

with equality meaning that there is no frequency overlap between adjacent filter channels, and with inequality meaning that adjacent filter channels overlap..The digital speech signal, s(n), was passed through a bank of 27 band pass filters whose coverage spans the frequency range of interest in the signal (e.g., 100-3000 Hz for telephone-quality signals, 100-8000 Hz for broadband signals) & output in MATLAB is as follows [5]

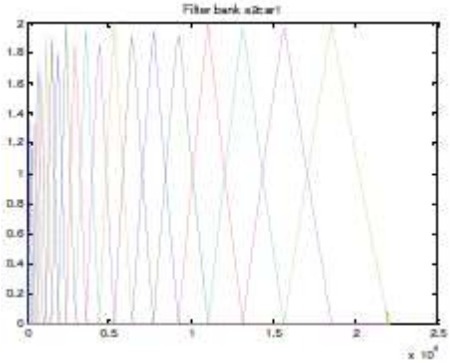

Fig3: Filter-bank output of Internet Car noise signal (s2car1) in MATLAB

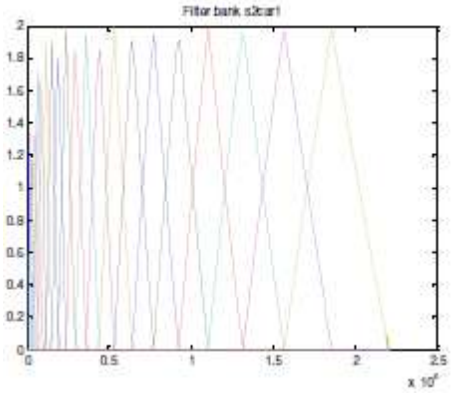

Fig4: Filter-bank output of Original Car noise signal (o2car1) in MATLAB

Power spectrum output of all noises were obtained in MATLAB and that of car noise obtained is as follows-

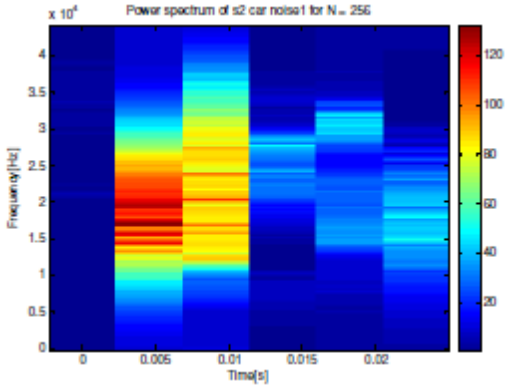

Fig5: Power spectrum output of Internet Car noise signal (s2car1) in MATLAB

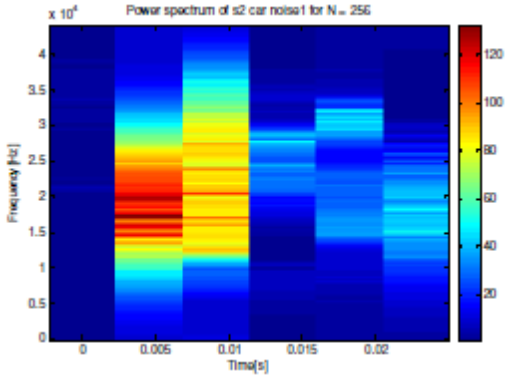

Fig6: Power spectrum output of Original Car noise signal (o2car1) in MATLAB

4. SPECTRAL MODELS USED FOR ENVIRONMENTAL NOISE CLASSIFICATION

Following models are widely used for environmental noise classification:

7.1 MFEE MODEL

MFCC is based on human hearing perceptions which cannot perceive frequencies over 1Khz. In other words, in MFCC is based on known variation of the human ear's critical bandwidth with frequency [5]. MFCC has two types of filter which are spaced linearly at low frequency below 500 Hz and logarithmic spacing above 500Hz. A subjective pitch is present on Mel Frequency Scale to capture important characteristic of phonetic in speech. The overall process of the MFCC is shown in Figure

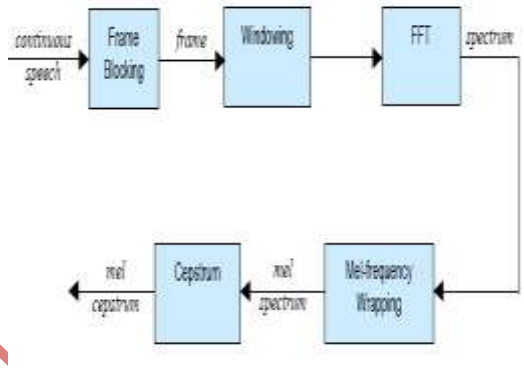

Fig7: Block Diagram of MFCC Process

Steps in MFCC extraction are as follows:

4.1.1 FRAME BLOKING

Framing is the first applied to the speech signal of the speaker. The signal is partitioned or blocked into N segments (frames) [7].

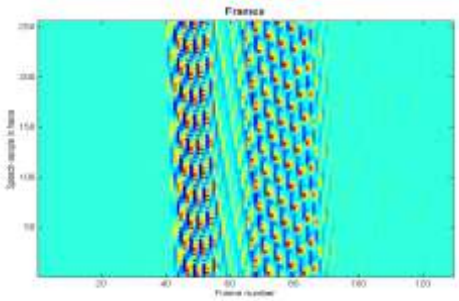

Fig8: Framing of Signal

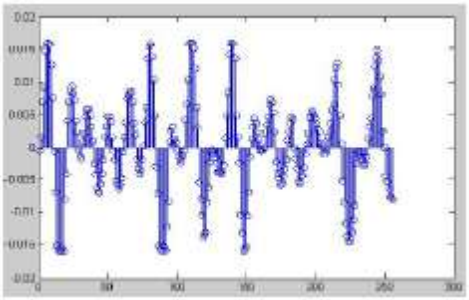

Fig9: Frame of Internet Car noise signal (s2car1) in MATLAB

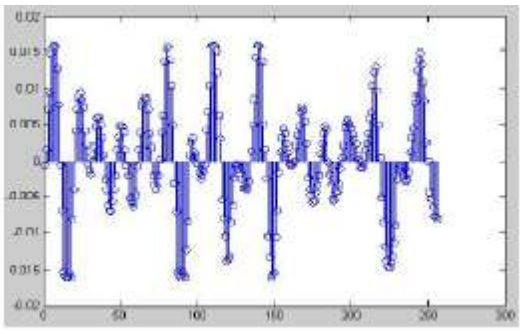

Fig10: Frame of Original Car noise signal (o2car1) in MATLAB

4.1.2 WINDOWING

The next step in the processing is to window each individual frame so as to minimize the signal discontinuities at the beginning and end of each frame. Hamming window is used as window shape by considering the next block in feature extraction processing chain and integrates all the closest frequency lines. The Hamming window equation is given as: If the window is defined as W (n), 0 ≤ n ≤ N-1 where N = number of samples in each frame, then the result of windowing is the signal

$$y_l(n) = x_l(n)w(n) \qquad 0 \leq n \leq N-1$$

$y_l(n)$ = Output signal
$x_l(n)$ = input signal
w (n) = Hamming window,

Typically the *Hamming* window is used, and then the result of windowing signal is shown below:

$$W(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right) \qquad 0 \leq n \leq N-1$$

Use of the window function reduces the frequency resolution by 40%, so the frames must overlap to permit tracing and continuity of the signal.

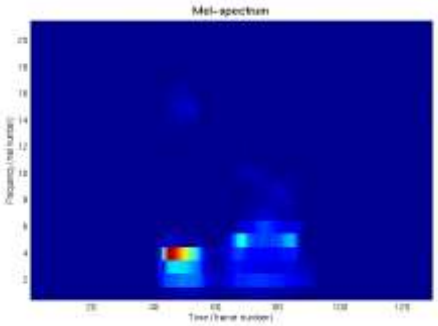

Fig11: Windowing of Signal

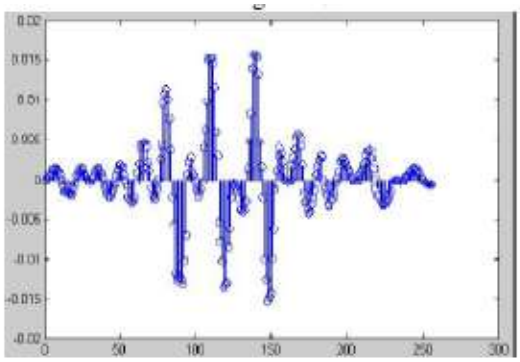

Fig12: Internet Car noise signal (s2car1) windowed data after Hamming in MATLAB

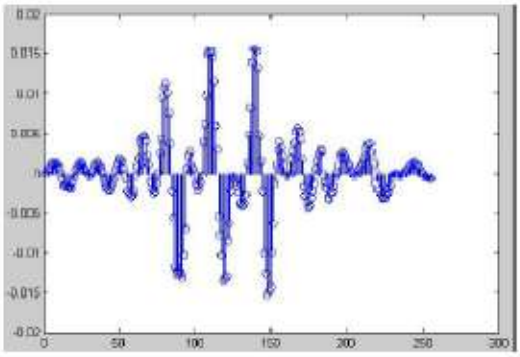

Fig13: Original Car noise signal (o2car1) windowed data after Hamming in MATLAB

4.1.3 FAST FOURIER TRANSFORM

Next step is the Fast Fourier Transform which converts each frame of N samples in time domain to frequency domain. To convert each frame of N samples from time domain into frequency domain The Fourier Transform is to convert the convolution of the glottal pulse U[n] and the vocal tract impulse response H[n] in the time domain. This statement supports the equation below:

$$Y(w) = FFT[h(t) * X(t)] = H(w) * W(n)$$

If X (w), H (w) and Y (w) are the Fourier Transform of X (t), H (t) and Y (t) respectively.

FFT is used for doing conversion from the spatial domain to the frequency domain. Fourier transformation is a fast algorithm to apply Discrete Fourier Transform (DFT), on the given set of Nm samples shown below:

$$D_k = \sum_{n=0}^{N-1} D_n e^{\frac{-j2\pi kn}{N}}$$

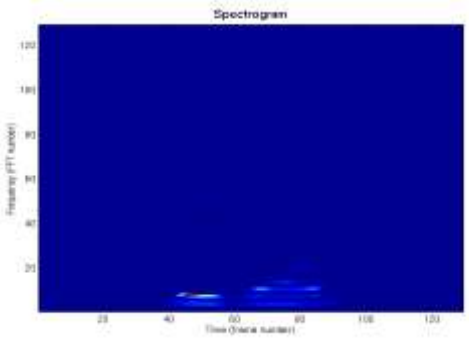

Fig14: FFT of Signal

4.1.4 MEL-FREQUENCY WRAPPING

The spectrum obtained from the above step is Mel Frequency Wrapped; the major work done in this process is to convert the frequency spectrum to Mel spectrum. The process of obtaining Mel-cepstral coefficients involves the use of a Mel scale filter bank. The spectral coefficients of each frame are then converted to Mel scale after applying a filter bank. The Mel-scale is a logarithmic scale resembling the way that the human ear perceives sound.

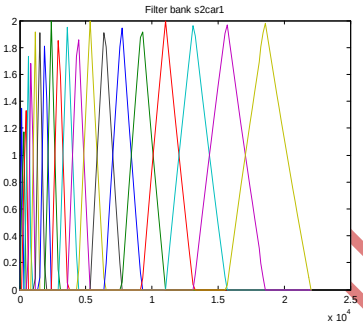

Fig15: MEL-SPACED FILTER BANK (K =20) PLOT FOR ORIGINAL CAR NOISE IN MATLAB

After that the following equation is used to compute the Mel for given frequency f in HZ:

$$m_f = 2595 \log_5 [\frac{f}{700}+1]$$

Thus, with the help of Filter bank with proper spacing done by Mel scaling it becomes easy to get the estimation about the energies at each spot and once this energies are estimated then the log of these energies also known as Mel spectrum Hence, first 13 coefficients are calculated using DCT and higher are discarded.

4.1.5 CEPSTRUM

In this final step, we convert the log Mel spectrum back to time. The result is called the Mel frequency Cepstrum coefficients (MFCC). We can calculate what is called the mel-frequency Cepstrum, Cn,

$$C_n = \sum_{k-1}^{k} (log\ D_k)\ cos\ [m\ (k-\frac{1}{2})\frac{\pi}{k}]$$

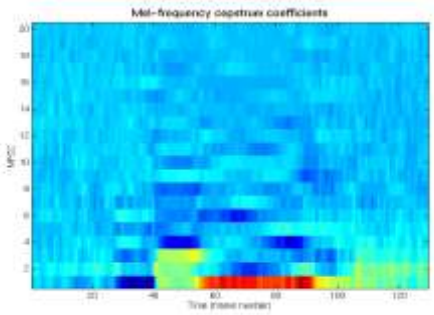

Fig16: Frequency Wrapping of Signal

Where m = 0, 1… k- 1

Where Cn represents the MFCC and m is the number of the coefficients here m=13 so, total number of coefficients extracted from each frame is 13.

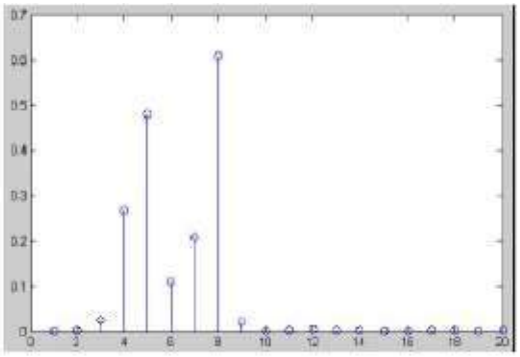

Fig17: Mel Spectral Coefficients of Internet Car noise signal (s2car1) in MATLAB

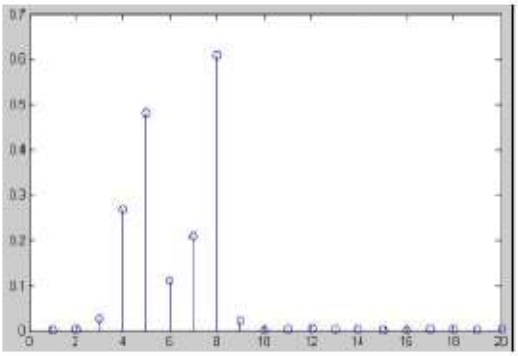

Fig18: Mel Spectral Coefficients of Original Car noise signal (o2car1) in MATLAB

4.1.6 PROCESS VISUALIZATION

In this section, we will visualize the results obtained from some of the main parts of the feature extraction process, having recently viewed the details of the procedure. Following figure will serve as the audio signal intended for the analysis in this case.

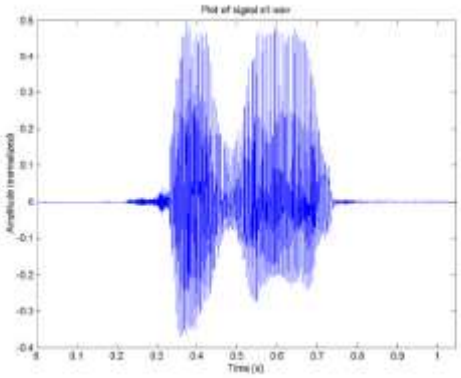

Fig19: DIGITAL AUDIO SIGNAL

This audio sample represents the audio signal recorded. It will serve as the input of the feature extraction in order to visualize the results of the processing involved.

4.2 LPC MODEL

The choice of signal features is usually based on previous knowledge of the nature of the signals to be analyzed. Noise synthesis based on LPC model is comparable to vocal tract of human throat.

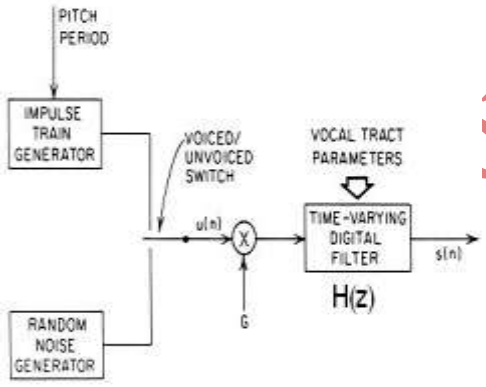

Fig20: LPC MODEL IN HUMAN THROAT

The object of linear prediction is to form a model of a Linear Time Invariant (LTI) digital system through observation of input and output sequences [8]. The basic idea behind linear prediction is that a noise sample can be approximated as a linear combination of past noise samples. By minimizing the sum of the squared differences (over a finite interval) between the actual noise samples and the linearly predicted ones, a unique set of predictor coefficients can be determined.

If u (n) is a normalized excitation source and being scaled by 'G', the gain of the excitation source, then LPC model is the most common form of spectral analysis models on blocks of noise (noise frames) and is constrained to be of the following form, where

H (z) is a pth order polynomial with z-transform and the coefficients a_1, a_2, …, a_p are assumed to be constant over the noise analysis frame

$$H(z) = 1 + a_1 z^{-1} + a_2 z^{-2} + a_3 z^{-3} + \ldots + a_p z^{-p}$$

Here the order 'p' is called the LPC order. If 'N' is the number of samples per frame and 'M ' is the distance between the beginnings of two frame, then for a given noise sample at time 'n'; S(n), can be approximated as a linear combination of the past 'p' noise samples, such that

$$s(n) = a_1 s(n-1) + a_2 s(n-2) + \ldots + a_p s(n) \quad (1)$$

Where the coefficients a_1, a_2… a_p are assumed constant over the noise analysis frame. We convert eq. (1) to an equality by including an excitation, G u (n), giving

$$s(n) = \sum a_i s(n-i) + Gu(n), \qquad 1 \le i \le p \quad (2)$$

Where u (n) is a normalized excitation and G is the gain of the excitation. By expressing eq (2) in the z-domain, we get the relation as follows in (3)

$$S(z) = \sum a_i z^{-i} S(z) + G\,U(z), \; 1 \le i \le p \qquad (3)$$

Leading to the transfer function as given in (4)

$$H(Z) = \frac{S(Z)}{G\,U(Z)} = \frac{1}{P} = \frac{1}{H(Z)}$$

Because noise signals vary with time, this process is done on short chunks of the noise signal, which are called frames. Usually 30 to 50 frames per second give intelligible noise with good compression. When applying LPC to audio at high sampling rates, it is important to carry out some kind of auditory frequency warping, such as according to mel or Bank frequency scales.

4.3 RCEP MODEL

As per theoretical point of view, the Cepstreal logarithm of the magnitude of few cepstral coefficients and setting the remaining coefficients to zero, it is possible to smooth the harmonic structure of the spectrum. Cepstral coefficients are therefore very convenient coefficients to represent the speech spectral envelope [6]. Hence, the following function calculates the real Cepstrum of the signal x.

$$y = \frac{1}{2\pi} \int_{-\pi}^{\pi} log\ |X(e^{j\omega t}|)\, e^{-j\omega t}\, d\omega$$

This denotes the Fourier Transform used for the separation of two signals convolved with each other based technique for determining a Harmonics valid technique for determining the sensitive to both noise and jitter for a large part of the noise or jitter. Thus real Cepstrum block gives the real Cepstrum way to define the prediction filter.

5. RESULT ANALYSIS

RESULTS OBTAINED IN MATLAB FOR FIVE SAMPLES OF FOUR INTERNET NOISES

A) MFCC

SAMPLES	CAR	OFF ICE	MAR KET	TRA IN
S1	0.7606	1.1829	0.8646	0.0271
S2	0.8497	0.2051	0.4135	0.5599
S3	0.1915	0.6141	0.8271	0.1922
S4	0.5952	0.7134	0.0903	0.9966
S5	0.6787	0.2297	0.1616	0.9129

B) LPC

SAMPLES	CAR	OFF ICE	MAR KET	TRA IN
S1	0.2164	0.5474	0.1504	0.6579
S2	0.1270	0.5195	0.1527	0.6629
S3	0.2298	0.2179	0.1558	0.7030
S4	0.0988	0.1775	0.1181	0.6006
S5	0.1835	0.2018	0.1645	0.6627

C) RCEP

SAMPLES	CAR	OFF ICE	MAR KET	TRA IN
S1	0.0011	0.0007	0.0006	0.0012
S2	0.0009	0.0010	0.0003	0.0005
S3	0.0003	0.0009	0.0001	0.0013
S4	0.0004	0.0001	0.0007	0.0008
S5	0.0000	0.0000	0.0002	0.0017

D) AVERAGE OF COEFFICIENT

MFCC COEEFICIENT	CAR (S1-S5)	OFF ICE (S1-S5)	MAR KET (S1-S5)	TRA IN (S1-S5)
C1	16.613	0.978	2.040	1.572
C2	0.978	1.074	1.302	1.154
C3	2.040	1.787	1.748	1.687
C4	1.572	1.397	1.377	1.437
C5	2.101	1.331	1.718	1.594

LPC COEEFICIENT	CAR (S1-S5)	OFF ICE (S1-S5)	MAR KET (S1-S5)	TRA IN (S1-S5)
C1	1.000	1.000	1.000	1.000
C2	0.497	0.590	0.298	1.432
C3	0.546	0.430	0.553	0.005
C4	0.281	0.172	0.619	0.769
C5	0.543	0.099	0.209	0.731

RECP COEEFICIENT	CAR (S1-S5)	OFF ICE (S1-S5)	MAR KET (S1-S5)	TRA IN (S1-S5)
C1	7.997	8.522	8.808	8.433
C2	0.004	0.000	0.799	0.001
C3	0.004	0.005	0.000	0.003
C4	0.004	0.002	0.002	0.003
C5	0.002	0.006	0.004	0.001

RESULTS OBTAINED IN MATLAB FOR FIVE SAMPLES OF FOUR ORIGINAL NOISES

A) MFCC

SAMPLES	CAR	OFF ICE	MAR KET	TRA IN
S1	0.7503	1.1929	0.8749	0.0219
S2	0.8399	0.2091	0.4179	0.5629
S3	0.1924	0.6541	0.8418	0.1969
S4	0.5953	0.7234	0.0916	0.9993
S5	0.6887	0.2397	0.1713	0.9279

B) LPC

SAMPLES	CAR	OFF ICE	MAR KET	TRA IN
S1	0.2267	0.5584	0.1518	0.6749
S2	0.1287	0.5185	0.1687	0.6636
S3	0.2399	0.2279	0.1564	0.7180
S4	0.0998	0.1785	0.1219	0.6017
S5	0.1848	0.2188	0.1658	0.6377

C) RECP

SAMPLES	CAR	OFF ICE	MAR KET	TRA IN
S1	0.0181	0.0015	0.0146	0.0029
S2	0.0012	0.0130	0.0017	0.0175
S3	0.0133	0.0017	0.0191	0.0025
S4	0.0012	0.0191	0.0014	0.0168
S5	0.0120	0.0063	0.0115	0.0029

D) AVERAGE OF COEFFICIENT

MFCC COEEFICIENT	CAR (S1-S5)	OFFICE (S1-S5)	MARKET (S1-S5)	TRAIN (S1-S5)
C1	16.723	19.669	19.297	18.984
C2	0.981	1.151	1.412	1.244
C3	2.170	1.827	1.751	1.696
C4	1.586	1.399	1.146	1.517
C5	2.241	1.411	1.729	1.599

LPC COEEFICIENT	CAR (S1-S5)	OFFICE (S1-S5)	MARKET (S1-S5)	TRAIN (S1-S5)
C1	1.000	1.000	1.000	1.000
C2	0.499	0.596	0.388	1.572
C3	0.636	0.520	0.561	0.016
C4	0.294	0.183	0.729	0.839
C5	0.613	0.119	0.218	0.881

RECP COEEFICIENT	CAR (S1-S5)	OFFICE (S1-S5)	MARKET (S1-S5)	TRAIN (S1-S5)
C1	7.998	8.632	8.814	8.513
C2	0.114	0.005	0.829	0.018
C3	0.0013	0.0185	0.0017	0.143
C4	0.104	0.016	0.182	0.016
C5	0.0013	0.146	0.011	0.121

6. CONCLUSION

From simulation results in Matlab, it was found that volume controller performed better for MFCC model as compared to other two models (LPC & RCEP) of original noise database as compared to internet noise database. This was due to the fact that classification accuracy (based on classification confusion matrix) was the highest for MFCC model parameter estimate viz 95% as compared to other two models (LPC & RCEP) viz. 94.05% as found earlier. It has been observed that the classification accuracy is 1% higher for Original noise dataset w.r.t. Internet noise dataset in classifier as well as Intelligent Volume Controller [3]

Finally, intelligent volume controller was implemented using active noise control & volume controller results were compared for internet noise data set and original noise data sets for above three models of noise parameter estimates independently. We have achieved classification accuracies of upto 93.01%, 94.05% and 95.00% using RCEP, LPC and MFCC based feature sets, respectively. Also, improvement in intelligent volume controller was achieved in terms of noise attenuation level of upto 0.05db, 0.012db and 0.017 db using RCEP, LPC and MFCC based feature sets, respectively. Thus, the highest classification accuracy upto 95.00% in noise classifier and maximum improvement in terms of noise attenuation level upto 0.017 db in intelligent volume controller was attained using MFCC based feature set through active noise control.

The real challenge for designing Intelligent Volume Controller (IVC) is generally based on background noise classification accuracy which is hard to achieve.

7. REFERENCES

[1] Schafer, R. and Rabiner, L.Digital Representation of Speech Signals.. Proceedings of the IEEE 63 (1975): 662-677.

[2] Gray, R.M.Vector Quantization.. IEEE ASSP Magazine 1 (1984): 4-29.

[3] Schafer, R. and Rabiner, L Systems for Automatic Formant Analysis of Voiced Speech..Journal of the Acoustical Society of America 47 (1970): 634-648.

[4] Tokhura, Y. A weighted cepstral distance measure for speech recognition..IEEE Transactions on acoustics, speech and signal processing 35 (1987): 1414-1422.

[5] Fujimura, O.Analysis of nasal consonants.. Journal of the Acoustical Society of America 34 (1962): 1865- 1875.

[6] Hughes, G. and Halle, M. .Acoustic Properties of Stop Consonants.. Journal of the Acoustical Society of America 30 (1957): 07-116.

[7] Atal, B.S. Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification. Journal of the Acoustical Society of America 55 (1974): 1304-1312.

[8] Furui, Sadaoki. Digital Speech Processing, Synthesis, and Recognition. New York: Marcel Dekker, 2001