A Time Domain Based Efficient Block Decision Algorithm for Audio coders

I Zakir Ahmed, Vijaya Yajnanarayana and Nirmal Kumar Sancheti
Motorola India Electronics Limited, Bangalore INDIA
Tel: +91-80-26010000, E-mail: a16709@motorola.com

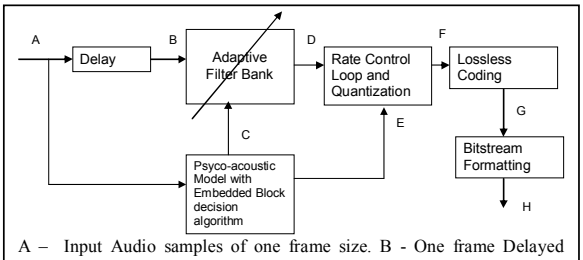
Abstract— **In typical audio encoders the block decision is done either using time-domain techniques like energy computation or frequency domain techniques like Temporal Noise Shaping (TNS) [1], [2]. The time-domain energy computation based decisions are less effective for detecting many of the stringent scenarios presented by test cases like castanets and fatboy. The frequency domain based algorithms have better decision making capabilities, however they are inherently complex as they require the computation of the FFT, additionally in case of TNS the computation of LPC (Linear Prediction coding) in the frequency domain. An improved time-domain technique with better block decision capability compared to TNS and with lesser computational complexity is proposed in this paper.**

I. Introduction and Background

This paper is organized as follows: Section-I talks about the importance of block decision in enhancing the quality of a perceptual audio encoder. The computational advantage of having the time-domain based block decision algorithms over frequency domain techniques are explored in Section-II. We propose a novel time-domain based algorithm whose performance is superior to existing time domain techniques and frequency domain based techniques like TNS. The proposed algorithm is presented in Section-III. In Section IV we discuss the tests carried out and the results of experiments conducted followed by conclusion.

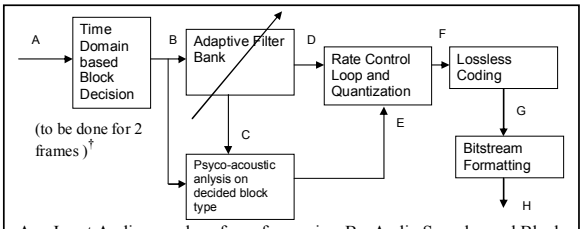
The block-decision algorithm is helpful for making crucial decision to switch the block length used by the encoder for representing signal in transform domain (Typically a Filter Bank). Most of the audio encoders use 2 block lengths. One length usually equal to the input audio frame data length called a "LONG BLOCK" and another shorter length called a "SHORT BLOCK". For example, in case of AAC encoder [2], the long block length is 2048 and a short block length is 256.

This is important because, coding of frames which has huge surges (transients in time) especially at the end of the analysis blocks, when analyzed using longer time-windows (LONG BLOCKS) by the perceptual encoders like MP3, AAC etc, spreads the quantization noise in time domain uniformly, this gives rise to a phenomenon called pre-echoes [3]. This drastically reduces the quality of encoded high-fidelity audio. Pre-echo distortion can arise in transform coders using perceptual coding rules. Pre-echoes occur when a signal with a sharp attack begins near the end of a transform block, immediately following a region of low energy. This situation can arise when coding recordings of percussive instruments such as the triangle, the glockenspiel, or the castanets. For a block-based algorithm, when quantization and encoding are performed in order to satisfy the masking thresholds associated with the block average spectral estimate, time-frequency uncertainty dictates that the inverse transform will spread quantization distortion evenly in time throughout the reconstructed block, This results in unmasked distortion throughout the low-energy region preceding in time the signal



A – Input Audio samples of one frame size. B - One frame Delayed Audio Samples. C - Block Length Decision to switch Filter Bank Length . D - Spectral representation of audio frame data. E - Psyco-acoustic parameters. F - Quantized Spectral samples adhering to both the bits budget and psyco-parameters. G - Lossless encoded representation. H - Compressed bitstream.

Fig. 1. Audio Encoder with frequency domain based block decision module function of applied field.



A – Input Audio samples of one frame size. B - Audio Samples and Block length decided. C - Spectral representation re-use for psycho analysis. D - Spectral representation of audio frame data. E - Psyco-acoustic parameters. F - Quantized Spectral samples adhering to both the bits budget and psyco-parameters. G - Lossless encoded representation. H - Compressed bitstream.

Fig. 2. Audio Encoder with Time domain based block decision module

† To accommodate START/STOP window types [2]

attack at the decoder. This can be reduced by choosing transform block size to be sufficiently small (minimal coder delay, e.g., 2–5 ms) near the transient signal[3]. Thus the block decision module in a typical perceptual audio coder plays a crucial role in deciding the quality of the audio encoder.

II. Overview of Design of Encoders Based on Time Domain and Frequency Domain Based Block Decision Algorithms

A typical audio encoder with a frequency domain based block decision algorithm is shown in Fig-1. Usually the block-decision is embedded within the psyco-acoustic analysis block [2],[3]. Based on the block length decided, the Filter-Bank [3] length is adaptively switched [1]. However there are 2 major disvantages because of this:

(1)The psyco-acoustic analysis is carried out on both the long and short block lengths [1]. This needs computation of long and short length FFT's, and also evaluation of psyco-acoustic parameters for both lengths. As a result of this, the analysis carried out on the block-length not decided becomes redundant.

(2)In such a design the psyco-acoustic analysis is carried out on a frame of audio data which looks ahead one frame, or in other words input to the Filter Bank needs to delayed by one frame. This delay compensation to the Filter Bank will add to both the computational complexity and extra storage.

However, with the time-domain based block-decision, the design can be greatly simplified as indicated in Fig-2. With this design the delay compensation is removed and both the psyco-acoustic analysis and Filter Bank operates on the same frame of audio data, and more importantly on the decided block length. This design gives the flexibility of reusing the Filter Bank output for pysco-acoustic analysis. Further optimizations could also be explored with this design.

III. Proposed Algorithm

We propose a 2-stage algorithm for transient detection to make a choice between LONG and SHORT. The algorithm computes the first and the second derivates of the audio signal of longest block length used by the audio coder.

$$Y(i) = X(i+1) - X(i) \tag{1}$$

Where, X(i) is the input audio signal.

$$Z(i) = Y(i+1) - Y(i) \tag{2}$$

Gaussian Low-pass filtering is done on the input signal to remove noise which favors block decision towards SHORT. The filtering followed by the second derivate of the signal is implemented as a simple 7-tap LOG (Laplacian Of a Gaussian [4]) filter. Based on the experiments on ISO test cases, for optimal classification the variance of the filter σ is chosen to be 0.8. The design of the filter is as given below.

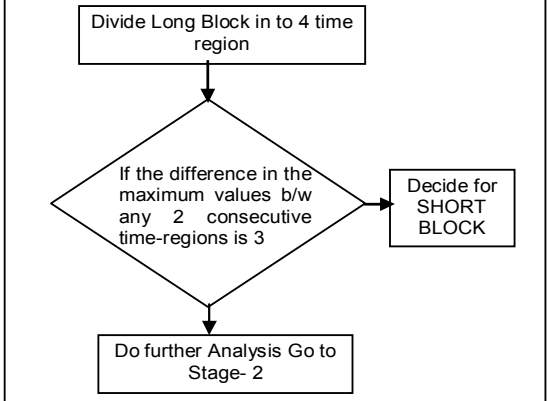


Fig 3: The Stage-1 algorithm.

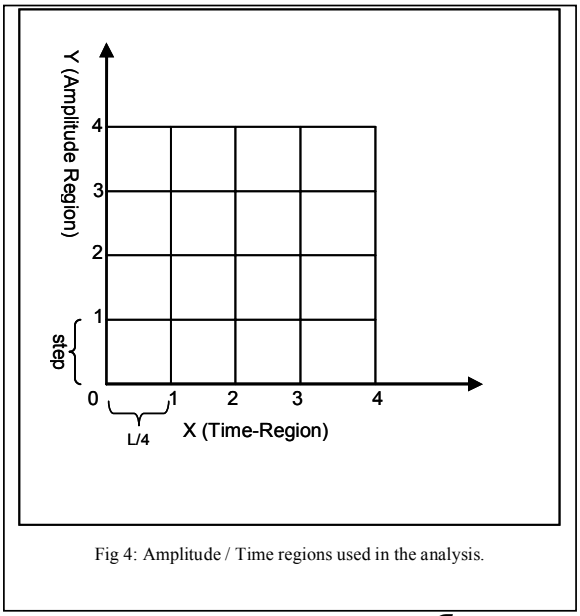


Fig 4: Amplitude / Time regions used in the analysis.

$$h(i) = -\left(\frac{(i-4)^2 - \sigma^2}{\sigma^4}\right) e^{\frac{-(i-4)^2}{2\sigma^4}} \tag{3}$$

for i=0, 1,2,3,4,5,6

$$z(i) = \sum_{n=0}^{6} h(n)x(i-n) \tag{4}$$

for i=0 to L-1.

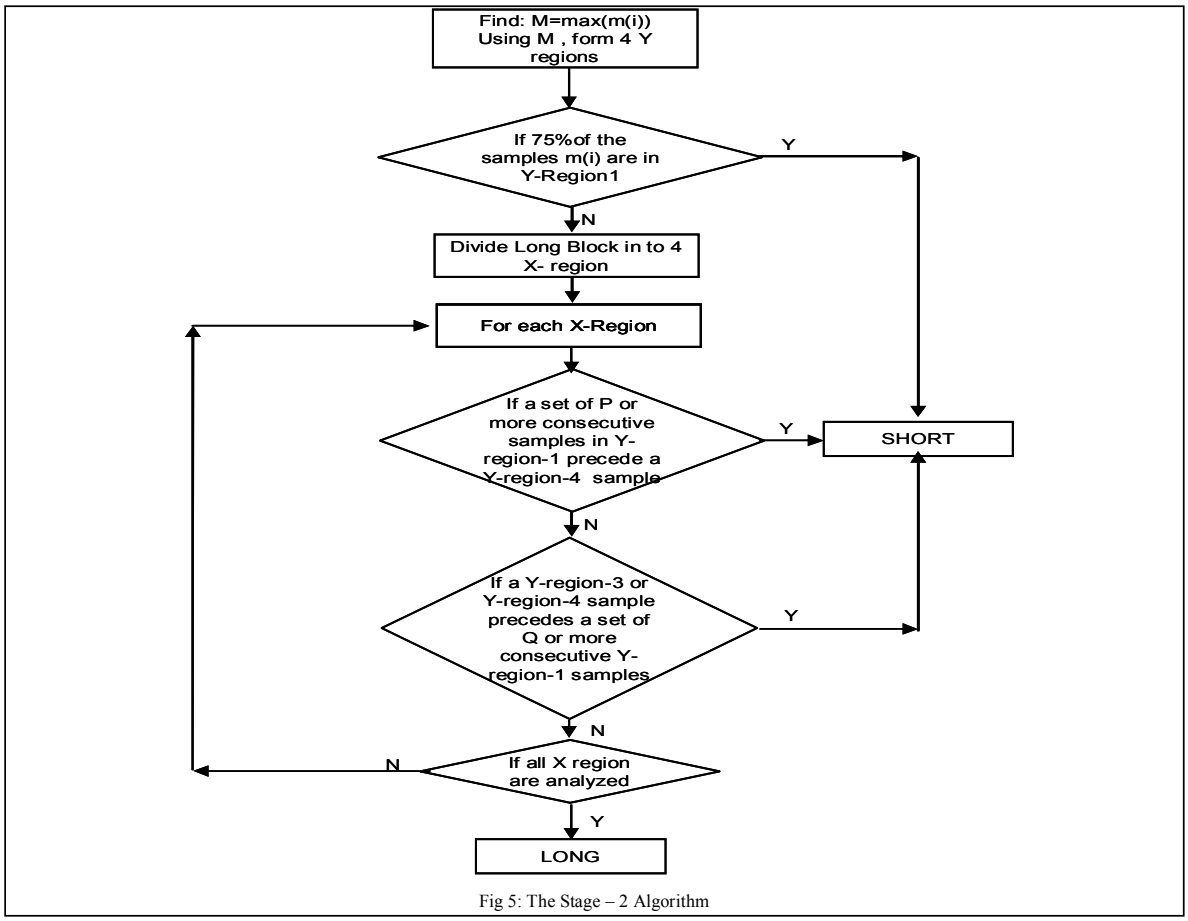


Fig 5: The Stage – 2 Algorithm

Where L is the length of the longest block used by the audio encoder.The normalized sequence m(i) denoting the amplitudes of zero-crossings are computed as below.

$$m(i) = \begin{cases} \dfrac{|z(i-1) - z(i)|}{2\max(x(i))} & \text{If at i, there exists a zerocrossing} \\ 0 & \text{Otherwise} \end{cases} \quad (5)$$

For i=1,2,….L-1.

First stage of the two stage transient detection algorithm is shown in **Fig 3**. Here we compute the amplitude of the zero-crossings as per Eqn 5. We divide the set of L samples into 4 equal time-regions each containing L/4 samples, These are called X-region. The maximum amplitude of the zero-crossing (M) is computed. This maximum value is divided by 4 to obtain the "step". We can see that the $m(i)$ can lie in any of the following 4 regions which is (0-step, step-2*step, 2*step-3*step, 3*step-M). These are called Y-regions. X and Y regions are shown in the **Fig 4**. The maximum value of amplitude $m(i)$ in each of the 4 time-regions is computed. If the difference in any of the consecutive time regions is greater than 3 Y-regions (refer **Fig 4**), we conclude that a huge surge has occurred and the current block is decided to be coded as SHORT. If not, a second stage of analysis is carried out.

In the second-stage; If more than 75% of the samples of m(i) are in Y-region 1[1] then we decide for SHORT block, else for each of the X-regions the following 2 conditions are evaluated, and if any one them is satisfied then we conclude that the block is SHORT . If for all the X-regions, following 2 conditions fail then we conclude as LONG.

If a sample m(i) in Y-region-4 is preceded by a consecutive run of P-samples in Y-region-1

If a sample m(i) in Y-region-4 is followed by a consecutive run of Q samples in Y-region-1.

[1] The region [0-step] is region-1, [step-2*step] is region-2, [2*step-3*step] is region-3 and [3*step-M] is region-4

The value of P and Q are selected such that P and Q fall within the forward and backward temporal masking characteristics of the human ear respectively [3], [7]. For a sampling rate of 44100, we select P=80 and Q=150.

Note that the simple transients like the one given in Fig 6 can be detected in stage 1.On the other hand some classical signals like the one shown in Fig 7 cannot be detected in stage-1 algorithm. So we proposed to do stage-2 processing as indicated in Fig 5. Note that signals shown in Fig 7 are not detected by currently available time-domain energy computation based algorithms.

IV. Tests and Results

Test methodology employed to test the proposed algorithm is as shown in **Fig 8**. Note that we used a standard AAC decoder to decode the two encoded AAC content, one in which TNS (Powerful and popular technique) is used for block decision and another which uses the proposed algorithm. The decoded audio signals are compared against the original signal using the "opticomm" tool to measure Perceptual objective quality through Objective Difference Grade (ODG) score. ODG score is a score between 0 to -4, with -4 indicating the worst and 0 indicating the closest to the original. Objective measurements for evaluating the perceptual audio quality are conducted as per ITU standard BS-1387 "Method for objective measurements of perceived audio quality" [5]. This standard recommends "opticomm" for evaluating Perceptual Evaluation of Audio Quality (PEAQ). The details pertaining to OPTICOM and ODG scores can be found in [6]. The test was carried out for all the ISO recommended test clips and results are tabulated. **Fig 9** and **Fig 10** compares the proposed and TNS algorithm for average and peak ODG scores for various ISO test cases.

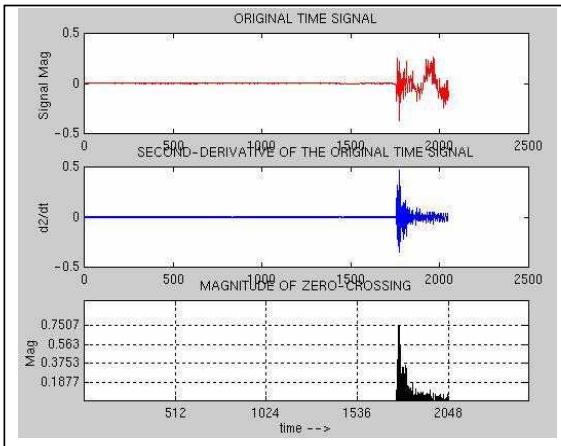


Fig 6: Transient detected in stage-1Algorithm , Here : Sampling-Freq = 44.1Khz, P=80 and Q=150.

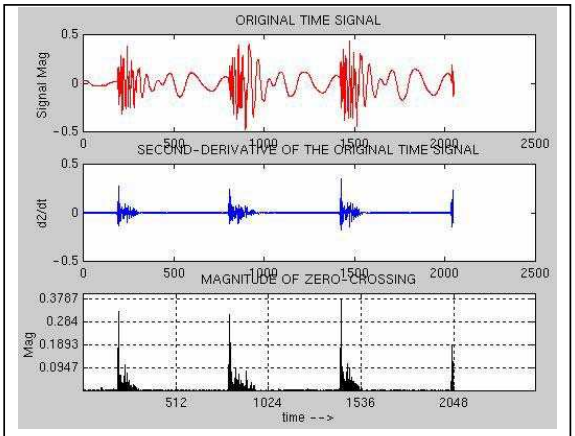


Fig 7 : Transient Detected in Stage-2 Algorithm , Here : Sampling-Freq = 44.1Khz, P=80 and Q=150.

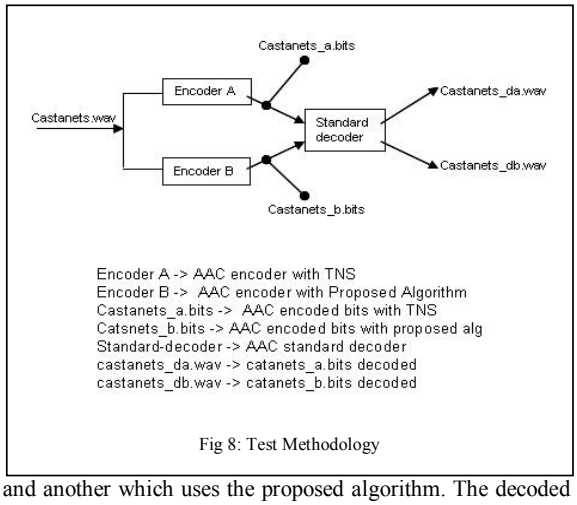


Fig 8: Test Methodology

V. Conclusion

A superior time-domain based block-decision algorithm is proposed in paper, which detects transients effectively compared to the frequency domain based approaches like the TNS and other existing time-domain based approaches. This is evident from the results of the experiments conducted on all the recommended ISO test cases. The advantage of having the time-domain based approach significantly reduces the computational complexity of the encoder, as per the discussions in section-II.

VI. Reference

[1] J. Herre, J. D. Johnston, "Enhancing the Performance of Perceptual Audio Coders by Using Temporal Noise Shaping (TNS)", lOlst AES convention, Los Angeles 1996, Preprint 4384.

[2] International Standard ISO/IEC 13818-7 "Information technology - Generic coding of moving pictures and associated audio information - Part 7 Advanced Audio Coding (AAC)".

[3] Ted Painter and Andreas S. Spanias, "Perceptual Coding of Digital Audio," Proceedings of the IEEE, Vol. 88, No.4, pp451-513, Apr. 2000.

[4] Rafael C. Gonzalez, Richard E. Woods "Digital Image Processing" Second Edition, Addison Wesley Publishing Company ISBN : 0-201-50803-6 Pages : 582-585
[5] ITU BS 1387 - Method for objective measurements of perceived audio quality.
[6] http://www.opticom.de for "opticom" tool used for perceptual audio quality measurements.
[7] Mark Kahrs , Karlheinz Brandenburg " Applications of Digital Signal Processing To Audio and Acoustics" , Kluwer Academic Publisers , ISBN 0-7923-8130-0. Pages : 44-45.

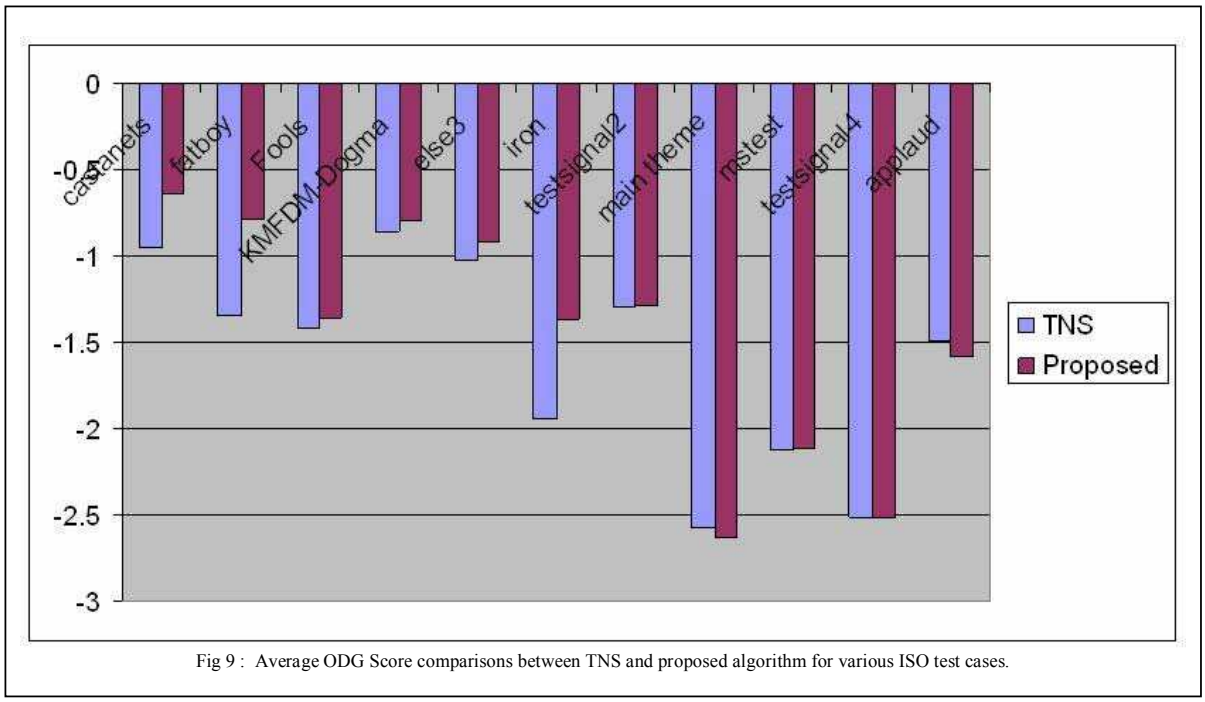


Fig 9 : Average ODG Score comparisons between TNS and proposed algorithm for various ISO test cases.

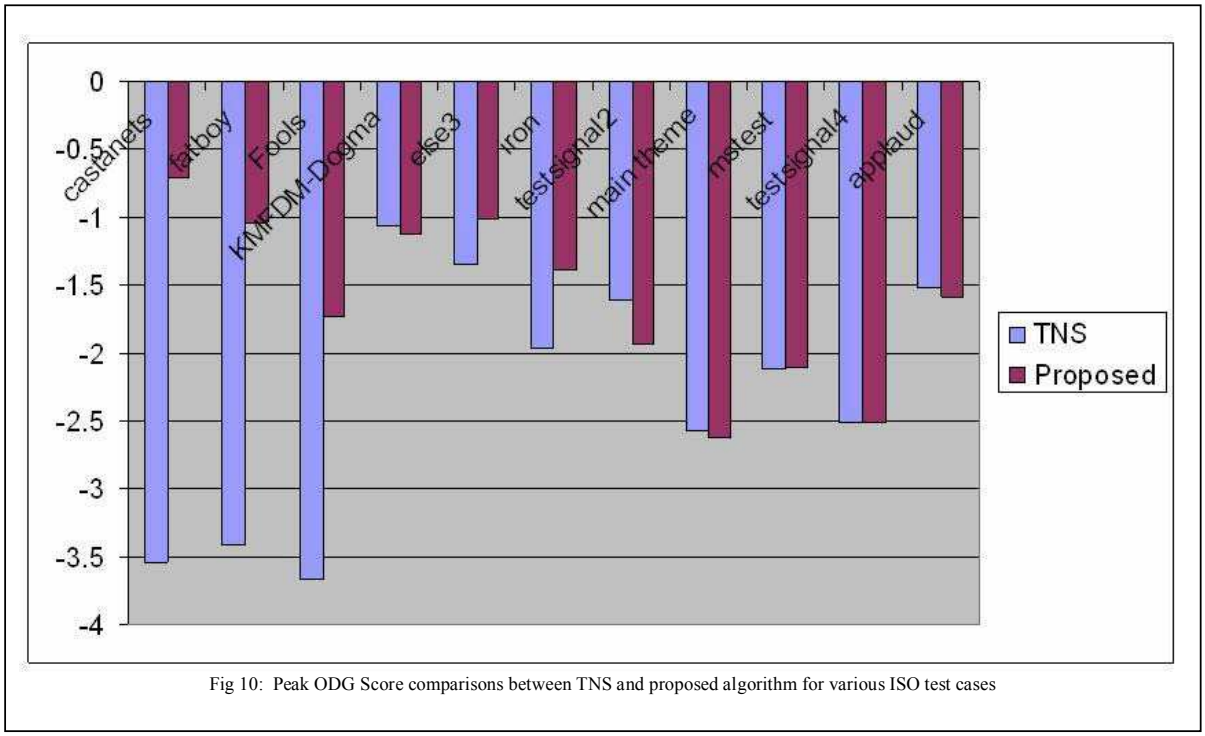


Fig 10: Peak ODG Score comparisons between TNS and proposed algorithm for various ISO test cases