Sindh Univ. Res. Jour. (Sci.Ser.), Vol: 34-B (2) 117 - 119 (2002)

SENSITIVITY OF COMPRESSED SPEECH SIGNAL ENERGIES IN RELATION TO HUMAN PERCEPTION

Asadullah Shah , Asma khan and Sadaf Khan

Department of Computer Science , Isra University, Hyderabad.

Abstract

Speech signal energy is one of the criteria upon which higher and lower sensitivity parts of a signal can be identified. The lower sensitivity parts of a signal generally constitute less than 10% of the talk spurts. Experiments have shown that elimination of the low sensitivity signals produce little effect on perceptual speech quality .This will save the channel bandwidth up to 10%, on top of bandwidth savings, achieved by detecting silence periods of talk.

Keywords: Sensitivities, Compressed Speech, Human Perception.

Introduction

There are several objective [1-3] and subjective [4] methods that are used to measure sensitivities of speech signal. The signal energy measure is one of the objective measure in time domain. For compression at low bit rates the speech signal is divided into frames of 20 to 30 ms. The energy of each of the segments (frames) is calculated by squaring the samples of that segment and adding them up is given in [5]. The results indicate that as soon as threshold values are set at 100×10^3 against energies of each segment, the dropped frames percentage is reached to 8%. In this paper results of various threshold settings are presented. Also objective and subjective measures are presented.

Experimental approach to observe the sensitivities

In order to see the level of energies for each frame of speech and then drop out the frames (packets) at certain energy levels a threshold value is set. As soon as the energy of any particular frame is observed below the set threshold value then that frame is dropped out. The dropped packet is reconstructed [6] to mitigate the dropped packet effects. The Figure 1, shows that as soon as the threshold is set to 100 thousand, the dropped packets are reached to 8%. Increasing the threshold will further set larger packet losses obviously, but for all practical purposes this loss is enough to consider for the required experimentations.

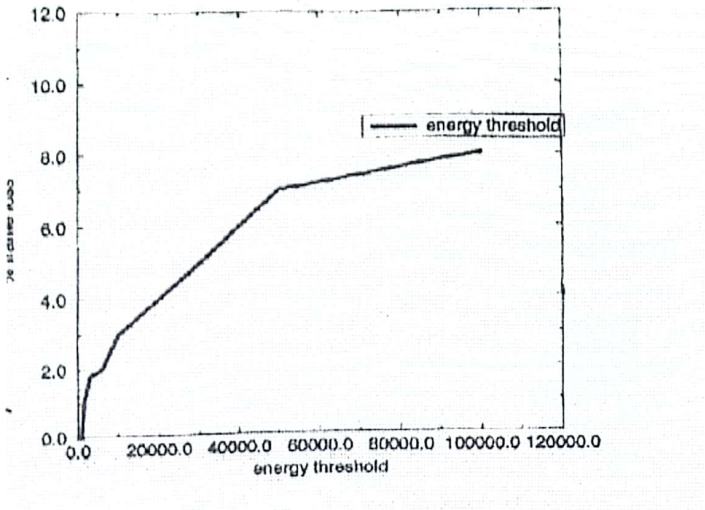


Figure:1 Energy threshold verses packet losses

Results and Observation

The experimentation is carried out over many speech files compressed at rates as low as 6.4kb/s [7]. The graph in Figure 1 indicates that the sensitivity of each packet or segment varies according to the content of energy present in the speech signal. Therefore some of the packets are more sensitive than others. Further the findings reveal that

(a) As long as frames having energies below 10^5 thresholds are discarded, the effect of loss would be minimal. That is the maximum capability of this criterion can be achieved as high as 8% loss rate while quality of speech will still be close to fair as indicated by the MOS results in Table 1.

(b) The SNRseg results, as shown in Table 2, also indicates that at 8% loss, all the speech utterances indicate very slow decay in this objective measure. This further indicates that the quality gets deteriorated, but slowly.

Threshold *10^6	File (a)	File (b)	File(c)	File (d)	File (e)	Frame loss
0.0005	3.5	3.5	3.5	3.5	3.5	0.005
0.001	3.5	3.5	3.5	3.5	3.5	0.008
0.0015	3.5	3.5	3.5	3.5	3.5	0.012
0.003	3.0	3.0	3.0	3.5	3.5	0.019
0.006	3.0	3.0	3.0	3.5	3.5	0.029
0.01	2.7	3.0	3.0	3.5	3.5	0.035
0.03	2.7	3.0	3.0	3.0	3.0	0.054
0.05	2.5	2.7	2.5	2.7	3.0	0.07
0.10	2.5	2.7	2.5	2.0	2.5	0.08

Table: 1 MOS results for energy sensitivity

Threshold*10^6	File (a)	File (b)	File(c)	File (d)	File (e)	Frame loss
0.0005	6.1	7.9	4.9	5.2	5.0	0.005
0.001	6.1	7.9	4.9	5.1	5.0	0.008
0.0015	6.0	7.9	4.9	5.1	4.8	0.012
0.003	6.0	7.8	4.9	5.0	4.8	0.019
0.006	5.7	7.3	4.8	5.0	4.8	0.029
0.01	5.7	7.3	4.8	5.0	4.4	0.035
0.03	5.6	7.2	4.7	4.9	4.3	0.054
0.05	5.5	7.2	4.5	4.9	4.3	0.07
0.10	5.4	6.6	4.4	4.6	4.3	0.08

Table 2: SNRseg for energy criterion.

CamScanner

Discussion

The experimentation results for perceptual sensitivities were carried out. For this purpose the compressed speech at 6.4kb/s PRELP coding algorithm was chosen. The energies of each of the packet were calculated and a threshold was set to discard those packets that carry the energies below the set threshold. In the beginning the threshold was set to a very low values as is shown in Figure 1: at that threshold value the packet losses were negligible. Slowly increasing the threshold caused some percentage of packets started dropped out. Raising the threshold to 0.10 Million, the packet dropped that were having their energies below it were 8% of the total packets of all five speech files used for this experimentations. As shown in Table 1 &2, the Mean Opinion score and a subjective measure and SNRseg an objective measure, both of these are slowly decreasing. This indicates that quality of speech is also decreasing. As for any practical purpose, 8% loss or above can not be reconstructed, therefore, this threshold and packet loss can be treated sufficient for packet discarding experimentations.

Conclusion

Energy of speech signal as a packet discarding criterion can be used for various experimentations. The speech packets are discarded if their energies are below the 0.10M value will causing very little degradation in quality. This also shows that full utilization of this criterion can provide 8 % less requirement of the services of the system because of their being perpetually less important.

References

[1]. Augustine JR., Gray H,. and Markel JD. "Distance Measures for Speech Processing" IEEE Trans.On ASSP, 24(5):380-391, Oct. 1976.

[2]. Kitawaki N., Nagabuchi H., and Itoh K. "Objectice Quality Evaluation for Low Bit Rate Speech Coding Systems", IEEE Jour.On Selected Area in Communicationss,6 (2) : 242-247, Feb.1988.

[3]. Kitawaki N., Itoh K., Honda M., and Kakehli K. " Comparision of Objective Speech Quality Measures for Voiceband Codecs", In Proc. of ICASSP, volume 2, pahes 1000-1003, Masashino Electical Communications Lab, Tokyo, Jap, 1982.

[4]. Kitawaki N., Honda M., and Itoh K. " Speech Quality Assesment Methods for Speech coding Systems". IEEE Communications Magazine, 22(10):26-33, Oct. 1984.

[5]. Asadulllah Shah and Sadaf Khan " Energy of speech signals as a packet discarding criterion", Mehran University Research Journal of Engineering and Technology, Volume 21, No3, P 127-132. July, 2002.

[6]. S. Atungsiri, R. Soheili, A. Kondoz, and B. Evans. "Effective Lost Speech Frame Reconstruction for CELP Coders" In Proc. of EUROSPEECH, volume 2, P. 599-602, Genova, Italy, Sep. 1991.

[7]. A. Kondoz, J. Horos, R.Suddle, and B.G. Evans. "Pulsed Residual Excited Linear Prediction". IEE Proc.: Vision Image and signal Processing, 142(2):105-110, April 1995

CamScanner