Subjective and Objective Speech Quality Evaluation under Bursty Losses

Lingfen Sun and Emmanuel C. Ifeachor
Department of Communication and Electronic Engineering
University of Plymouth, Plymouth PL4 8AA, U.K.
{L.Sun@plymouth.ac.uk; E.Ifeachor@plymouth.ac.uk}

Abstract – In VoIP applications, packet loss is a major source of speech impairment. Such a loss normally occures in bursts, rather than at random. In this paper, we report an investigation on how subjects perceive bursty losses and how current objective measurement methods, such as PSQM, MNB, EMBSD, and in particular, the latest ITU standard, PESQ correlate with subjective test results under burst loss conditions. Preliminary results show that the correlation of the objective measures with subjective MOS for all four algorithms is not sufficiently high under burst loss conditions. The MNB2 shows a slightly higher correlation than the other three. The new ITU algorithm, PESQ, displays an obvious sensitivity to bursty conditions compared to human subjects (it is more sensitive than subjects when loss burstiness is high and less sensitive when it is low).

Keywords -- Voice over IP, Speech Quality, Bursty Packet Loss, Objective Measurement, Subjective Measurement

I. INTRODUCTION

Packet loss is a major source of speech impairment in voice over IP (VoIP) applications. Such a loss may be caused by discarding packets in the IP networks due to congestion or by dropping packets at the gateway/terminal due to late arrival. The impact of packet loss on perceived speech quality depends on several factors, including loss pattern, codec type, packet size and loss locations. Research [1][2] has shown that packet losses in the Internet are temporally correlated, that is, they often occur in bursts rather than in a random pattern. It is therefore useful to study how subjects perceive bursty losses and how current objective measurement methods, such as PSQM [3], MNB [4], EMBSD [5] and, in particular, the latest ITU standard, PESQ [6] correlate with subjective test results (MOS) [7] under bursty loss conditions.

The work reported here was based on the G.729B [8][9] codec which is commonly used in VoIP applications. A 2-state Gilbert model was used in the simulation of bursty losses in IP networks. 15 different network loss conditions (a combination of different burstiness, packet size and loss locations) were chosen. 16 subjects took part in the subjective test. Four algorithms – PSQM, MNB, EMBSD and PESQ - were chosen for objective speech quality evaluation against subjective MOS. Preliminary results show that under bursty loss conditions, the correlation of the objective measures with subjective MOS for all four algorithms is not sufficiently high. The MNB2 shows a slightly higher correlation than the other three. The new ITU algorithm, PESQ, displays an obvious sensitivity to bursty conditions compared to human subjects (it is more sensitive than subjects when loss burstiness is high and less sensitive when it is low).

The remainder of the paper is organised as follows. In Section II, the simulation system and the experiment carried out in the study are introduced. In Section III, the main results and analysis are presented. Section IV concludes the paper.

II. SIMULATION SYSTEM AND EXPERIMENT

The block diagram of the system that was used in the study is depicted in Figure 1. It is a PC-based software system that allows the simulation of key processes in voice over IP. It enables the simulation of a variety of network conditions and objective measurement of their effects on perceived speech quality. The system includes a speech database, an encoder/decoder, a packet loss

simulator and an objective quality measurement module.

A 2-state Gilbert model was used to simulate packet loss (see Figure 2). In the figure, State 0 denotes 'received packet' condition and State 1 'dropped packet' condition. p is the probability that a packet will be dropped given that the previous packet was received. q is the probability that a packet will be dropped given that the previous packet was dropped. q is also referred to as the conditional loss probability (*clp*). The probability of being in State 1 is referred to as the unconditional loss probability (*ulp*). The *ulp* provides a measure of the average packet loss rate. It is given by: $ulp = p/(p+1-q)$. The *clp* and *ulp* are used in the study to characterise the loss behaviour of the network.

Fifteen different bursty loss conditions were chosen to cover cases of interest. They consist of combinations of unconditional loss probability (*ulp, 5%, 10% or 25%*), conditional loss probability (*clp, 20% or 60%*), packet size (2 or 4 frames/packet) and initial seeds to simulate different loss locations. The frame size for G.729 is 10ms. The reference speech file is about 10 seconds long and consists of four short sentences from two male and two female speakers. 15 different degraded speech files were generated for subjective and objective test.

For efficiency and to make it easier for people to participate in the listening tests, regardless of where they were, a VoIP MOS test website was created. All subjective tests were carried out via Internet. A total of 16 subjects (located on different floors within a building) participated in the MOS test. Most of them were Ph.D students with no previous MOS test experience. The participants used their own headphones to listen to the original and degraded speech files and were asked to give an opinion score between 1 to 5 (where 5 is excellent and 1 is bad) following the instructions on the web. The average score for a particular test material then gives the Mean Opinion Score (MOS).

III. RESULTS AND ANALYSIS

For each of the 15 conditions, objective measures were obtained using each of the four measurement algorithms (PSQM, PESQ, MNB and EMBSD) separately. As the simulated network conditions were related to packet loss only (end-to-end jitter was not induced), time-alignment was not required and so all four algorithms could be used to obtain an objective measure of speech quality. The scatter diagrams for the objective test results versus subjective MOS scores are illustrated in Figures 3 (a) to (d). The MNB produces two perceptual distances (MNB1 and MNB2) which are mapped onto a logistics value within the range [0, 1]. In our experiment, the results for MNB1 and MNB2 are similar and so only those for MNB2 are shown in Figure 3. The correlation for each condition, after mapping the objective measures with a 3rd order monotonically decreasing or increasing polynomial, was calculated to assess the objective algorithms against subjective MOS scores. To void a bias by our MOS test results, the correlation coefficients before and after the 3rd order polynomial mapping were both calculated and are shown in Table 1.

From Figure 3 and Table 1, it can be seen that, surprisingly, the PESQ does not have a better performance than the other objective methods under bursty packet loss conditions and that the MNB2 performs slightly better. The results were analysed further to understand the conditions under which the PESQ differs with subjective test results. The MOS scores from the subjective and objective tests using PESQ for all 15 samples (conditions) are illustrated in Figure 4. From the Figure, we noticed that the subjective MOS values in samples 2 and 3 were higher than that of the objective MOS values obtained from PESQ. Sample 2 is for 10% (*ulp*), 60% (*clp*) and packet size of 4, whilst Sample 3 is for 25% (*ulp*), 60% (*clp*) and packet size of 2. Both have almost one word missing due to very heavy bursty losses. In this case, PESQ is more sensitive than the subjects. On the other hand, there were three Samples (7, 11 and 14), where subjective MOS scores were slightly lower than for objective MOS scores. It was interesting to find that all three samples belonged to high loss rate conditions (20 or 25% of *ulp*), lower burstiness (20% of *clp*) and small packet sizes (2). The loss occurs evenly and

the speech sounds were annoying. In this case, PESQ is less sensitive than the subjects.

The ITU-T P.862 document [6] states that the "PESQ has *demonstrated* acceptable accuracy at factors such as packet loss and packet loss concealment *with CELP codecs*" and that the factors for which PESQ has *not* currently been validated include "packet loss and packet loss concealment with *PCM type codecs*" in which "PESQ appears to be more sensitive than subjects to front-end temporal clipping, especially in the case of missing words which *may not be* perceived by subjects. Conversely, PESQ *may be* less sensitive than subjects to regular, short time clipping (replacement of short sections of speech by silence). "

Although the codec used in our test is G.729 8 Kb/s CS-ACELP, which belongs to the CELP type codec with internal packet loss concealment, our test results for PESQ do not show acceptable accuracy under bursty loss conditions. The results are consistent with the conditions described in the P.862 document for packet loss and loss concealment with PCM type codecs.

IV. CONCLUSIONS

Our preliminary subjective and objective test results show that the four objective algorithms (PSQM, PESQ, MNB and EMBSD) do not have a sufficiently high correlation with subjective scores under network bursty loss conditions. The MNB2 has a slightly higher correlation than the other three algorithms. The PESQ only predicts the MOS score well for the average bursty loss cases. When loss burstiness is higher/lower, PESQ shows an obvious sensitivity to these conditions than human subjects. This highlights the need for a further modification/improvement of the latest ITU objective speech quality measurement algorithm to make it suitable for a variety of network bursty loss conditions. Possible modifications could be to incorporate loss patterns in the cognitive model to alleviate the impact of bursty losses.

REFERENCES

[1] J-C. Bolot, Characterizing End-to-end Packet Delay and Loss in the Internet, Journal of High-Speed Networks, Vol.2, No.3, PP.305-323, Dec. 1993

[2] M Yajnik, S. Moon, J. Kurose and D. Towsley, Measurement and Modelling of the Temporal Dependence in Packet Loss, Proceedings of IEEE INFOCOM 99, New York, March 1999

[1] ITU-T Recommendation P.861, Objective quality measurement of telephone-band (300-3400 Hz) speech codecs, February 1998

[2] S. Voran, Objective Estimation of Perceived Speech Quality – Part I: Development of the Measuring Normalizing Block Technique, IEEE Trans. on Speech and Audio Processing, Vol. 7, No.4. July 1999, pp. 371-382

[3] Wonho Yang, Enhanced Modified Bark Spectral Distortion (EMBSD): An Objective Speech Quality Measure Based on Audible Distortion and Cognition Model, Ph.D Dissertation, May 1999

[4] ITU-T Rec. P.862, Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs. February 2001

[5] ITU-T Rec. P.800, Methods for subjective determinatio of transmission quality, August 1996

[6] ITU-T Rec. G.729, Coding of Speech at 8 kbit/s Using Conjugate-Structure Algebraic-Code-Excited Linear-Prediction (CS-ACELP), March 1996

[7] ITU-T Rec. G.729 Annex B, A silence compression scheme for G.729 optimized for terminals conforming to Rec. V.70, Nov. 1996

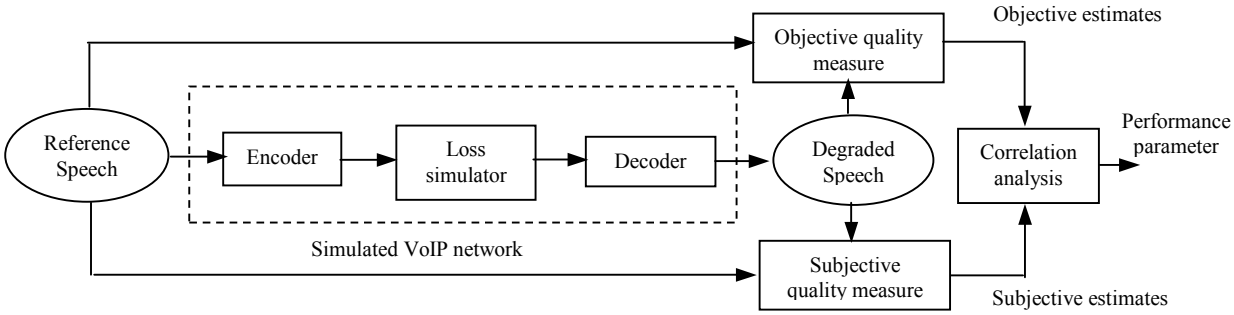

Figure 1. Objective and subjective speech quality evaluation system

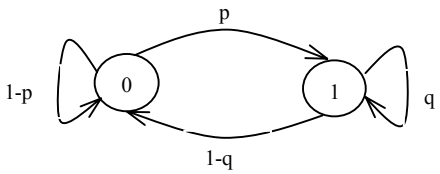

Figure 2. 2-State Gilbert model

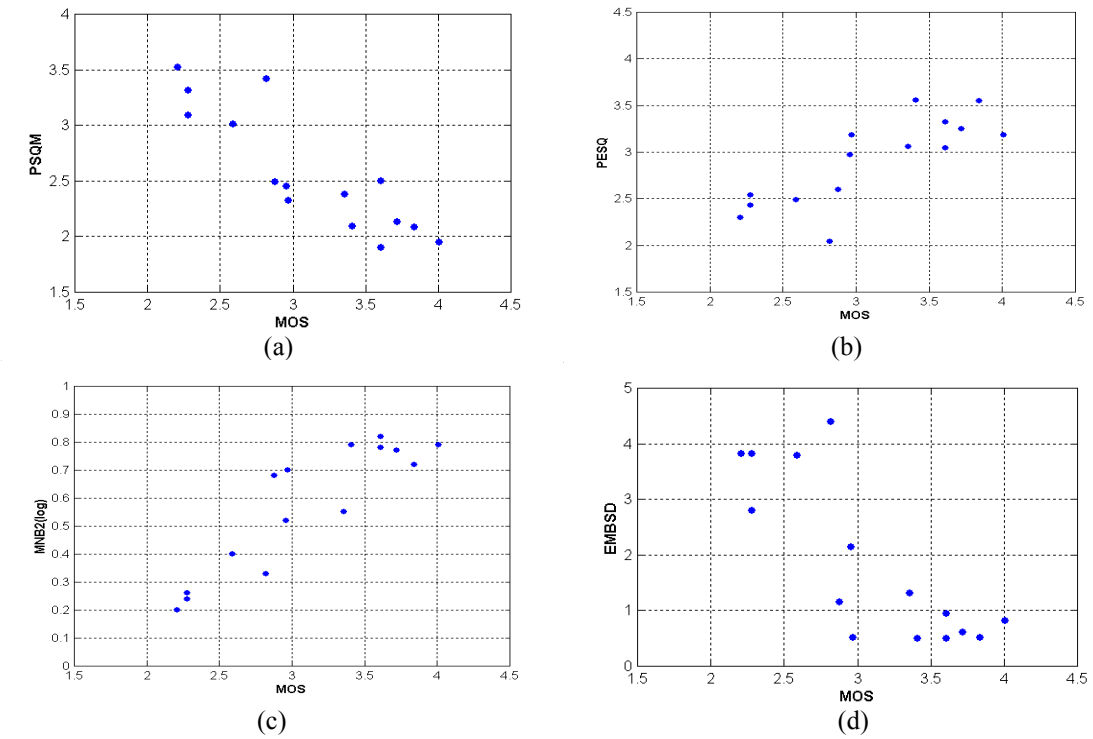

Figure 3. Scattered diagrams of objective test results versus subjective MOS scores
(a). PSQM; (b). PESQ; (c). MNB_2; (d). EMBSD

Table 1. Correlation between subjective MOS and objective measures

Algorithms	PESQ	PSQM	MNB_1	MNB_2	EMBSD
Correlation before mapping	0.81	-0.87	0.89	0.895	-0.80
Correlation after mapping	0.896	0.895	0.90	0.901	0.88

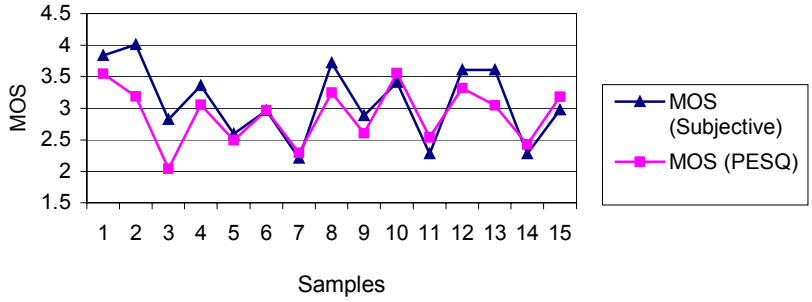

Figure 4. Objective (PESQ) and subjective MOS for 15 test samples