Impact of Packet Loss Location on Perceived Speech Quality

L. F. Sun, G. Wade, B. M. Lines, E. C. Ifeachor
Department of Communication and Electronic Engineering,
University of Plymouth,
Drake Circus, Plymouth PL4 8AA, United Kingdom,
{ L.F.Sun@jack.see.plym.ac.uk, j.wade@plymouth.ac.uk, B.Lines@plymouth.ac.uk, E.Ifeachor@plymouth.ac.uk}

Abstract – In VoIP applications, packet loss can have a major impact on perceived speech quality. The impact is affected by factors such as packet loss size, loss pattern and loss locations. In this paper, we report an investigation into the impact of loss location on perceived speech quality and the relationships between convergence time and loss location for three different codecs (G.729, G.723.1 and AMR) using perceptual-based objective measurement methods (PSQM+, MNB and EMBSD). Our results show that loss location has a severe effect on perceived speech quality. The loss at unvoiced speech segments has little impact on perceived speech quality for all codecs. However, the loss at the beginning of voiced segments has the most severe impact on perceived speech quality. The convergence time depends on the speech content (voiced/unvoiced). For unvoiced segments, the convergence time is stable whereas for voiced segments it varies but has an upper bound at the end of the segment. Our method allows a more accurate measurement of the exact effect of packet loss on perceived speech quality. This could help in the development of a perceptually relevant packet loss metric, which could be valuable in non-intrusive VoIP measurements.

Keywords -- Voice over IP, Packet loss, Speech quality, Objective perceptual measurement, Codecs, Concealment performance

I. INTRODUCTION

Packet loss is a major source of speech impairment in voice over IP (VoIP) applications. Such a loss could 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, and packet loss size [1][2]. It may also depend on the location of loss within the speech.

In modern codecs (e.g. G.729, G.723.1 and Adaptive Multi-Rate, AMR codec), internal concealment algorithms are used to alleviate the effects of packet loss on perceived speech quality [3][4][5]. When a loss occurs the decoder derives the parameters for the lost frame from the parameters of previous frames to conceal the loss. The loss also affects subsequent frames because the decoder takes a finite time (the convergence time) to resynchronise its state to that of the encoder. Recent research has shown that for some codecs (e.g. G.729) concealment works well for a single frame loss, but not for consecutive or burst losses [1], and that the convergence times are dependent on speech content. Further, the effectiveness of a concealment algorithm is affected by which part of speech is lost (e.g. voiced or unvoiced). For example, it has been shown that concealment for G.729 works well for unvoiced frames, but for voiced frames it only works well after the decoder has obtained sufficient information [6]. Further, the decoder fails to conceal the loss of voiced frames at an unvoiced/voiced transition. Thus, the location of packet loss in relation to different parts of speech is important.

In most studies [1][6], the analysis of concealment performance and convergence times is based on the mean square error (MSE) and signal-to-noise ratio (SNR) criteria (with subjective or perceptual-based objective methods only used to assess overall quality under stochastic loss simulations). The perceptual impact of concealment algorithms or convergence times for different loss locations is still unknown. It is important to understand the effects of loss location and loss pattern on perceived speech quality, for different types of codec, to allow a more accurate measurement of voice quality. This requires the use of perceptual-based objective methods in the analysis. This could be helpful in setting up more efficient speech recovery system and for the development of perceptually relevant packet loss metrics which could be valuable in non-intrusive VoIP measurement.

The IETF has recently proposed a set of new metrics for packet loss [2]. This includes loss constraint distance (i.e. distance threshold between two losses) and “noticeable” loss rate (i.e. percentage of lost packets with loss distances smaller than loss constraint distance). For the same loss rate, different loss patterns may have different effects on perceived speech. In VoIP applications, the loss constraint is related to the convergence times of the decoder. However, it is still unclear how to determine the loss constraint threshold and whether (or how) the threshold is related to codec type, burst size or speech.

The aims of the study reported in this paper are two fold: (1) to investigate the impact of loss location on perceived speech quality and hence the concealment performance of codecs, and (2) to investigate the relationships between convergence times and loss locations/speech content, codec type or loss size.

The work reported here is based on three codecs – two existing codecs (G.729B [13] and G.723.1) and a new codec (AMR [7][14]) for VoIP. Three major perceptual distance measurement algorithms (PSQM/PSQM+ [8][9], MNB [10][11] and EMBSD [12]) are used for perceptual performance analysis for different loss location. Each algorithm quantifies perceptual quality, but has a different range of perceptual distance.

The results show that the loss location has a severe effect on perceived speech quality. The loss at unvoiced speech segments has little impact on perceived speech quality for all three codecs. However, the loss at the beginning of voiced segments has the most severe impact on perceived speech quality. The extent of the impact depends on the size of the burst loss and codec type. The convergence time depends on the speech content. For unvoiced segments the convergence time is stable whereas for voiced segments it varies but constrained by the duration of the segment.

The remaining sections of the paper are structured as follows: Section II presents a brief overview of the codecs used and their concealment algorithms. The perceptual distance measurement algorithms (PSQM/PSQM+, EMBSD and MNB) are summarised briefly in Section III. The simulation system is described in Section IV, the experiments, results and their analysis are given in Section V. Section VI concludes the paper.

II. CODECS AND THEIR INTERNAL CONCEALMENT

A. *Codec types - G.729, G.723.1 and AMR*

The G.729 CS-ACELP (Conjugate Structure Algebraic Codebook Excited Linear Prediction, 8 Kbps) and G.723.1 (MP-MLQ/ACELP: Multipulse excitation with a maximum-likelihood-quantizer/Algebraic Codebook Excited Linear Prediction, Dual rate: 5.3/6.3 Kbps) are both standardized by the ITU and have been used in VoIP applications. The AMR (Adaptive Multi-Rate, ACELP) speech codec was developed by ETSI and has been standardized for GSM. It has been chosen by 3GPP as the mandatory codec. The AMR is a multi-mode codec with 8 narrow band modes with bit rates between 4.75 to 12.2 Kb/s. Mode switching can occur at any time (frame-based). AMR speech codec represents a new generation of coding algorithms which are developed to work with inaccurate transport channels. The flexibility on bandwidth requirements and the tolerance in bit errors of AMR codecs are not only beneficial for wireless links , but are also desirable for VoIP applications.

The three codec types belong to CELP analysis-by-synthesis hybrid codec. At each speech analysis frame, the speech signal is analysed to extract the parameters of the CELP model (Linear Prediction, or LP filter coefficients, adaptive and fixed codebooks' indices and gains). For stability and efficiency, LP filter coefficients are transformed into Line Spectral Frequencies, or LSF's for transmission. These parameters are then encoded and transmitted. At the decoder, the parameters are decoded and speech is synthesized by filtering the reconstructed excitation signal throuth the LP synthesis filter.

The major differences between the three codecs lie in the excitation signals, the partitioning of the excitation space (the algebraic codebook), delay and the way in which the coefficients of the filter are represented. For example, the G.729 uses two stage codebook structures for LSP parameters and gets the name "conjugate structure".

The frame sizes for the three codecs are 10 ms (80 samples at 8 kHz sampling) for G.729, 20 ms (160 samples) for AMR and 30 ms (240 samples) for G.723.1. They all have voice activity detection and silence suppression processing. The frames are classified as normal speech frame, SID (Silence Insertion Description) frame and null frame (non-transmitted frame).

B. *Codec Internal Concealment*

All three codecs have built-in concealment algorithms, which can interpolate the parameters for the loss frames from the parameters of the previous frames. For example, for the G.729 the concealment algorithm works in accordance to the following steps:

- The line spectral pair coefficients of the last good frame are repeated
- The adaptive and fixed codebook gain are taken from the previous frame but are damped to gradually reduce their impact.
- If the last reconstructed frame was classified as voiced, the fixed codebook contribution is set to zero. The pitch delay is taken from the previous frame and is repeated for each following frame. If the last reconstructed frame was classified as unvoiced, the adaptive codebook contribution is set to zero and the fixed codebook vector is randomly chosen.

III. PERCEPTUAL SPEECH QUALITY MEASURE – PERCEPTUAL DISTANCE

Perceptual distance is used to measure the perceptual difference between a reference speech signal and a degraded speech signal. It normally includes a perceptual model and a cognition model to mimic the process in the human's hearing perceptual process. Various perceptual speech quality measurement algorithms exist with different perceptual or cognition models.

PSQM (Perceptual Speech Quality Measurement) developed by KPN has been adopted as ITU-T Recommendation P.861 for assessing the speech quality for codecs [8]. PSQM+ was proposed by KPN to improve the performance of PSQM for loud distortions and temporal clipping [9]. PSQM/PSQM+ can generate a perceptual distortion value for each frame (32 ms for 8 kHz sampling, with 50% overlapping) and the overall PSQM/PSQM+ value is calculated for the whole test sentence via different weighting factors for silence or non-silence frames. As PSQM+ provides a more accurate measure of perceived speech quality under frame loss situations, we have chosen it for overall perceived speech quality and perceptual distance calculation for each frame.

The MNB (Measuring Normalizing Blocks) developed by the US department of Commerce [10][11], is included as an Appendix in ITU-T P.861 Recommendation. The MNB does not generate a distortion value for each frame since each MNB is integrated over frequency or time internals.

EMBSD (Enhanced Modified Bark Spectral Distortion) was developed by Temple University in USA [12]. It estimates speech distortion in the loudness domain taking into account the noise masking threshold in order to include only audible distortions in the calculation of the distortion measure. As EMBSD only takes into account the non-silence frame for the final perceptual distortion calculation, the setting of the threshold of silence or non-silence will affect the final result.

In the paper, MNB and EMBSD are used for the overall quality measurement.

IV. Simulation system

In order to investigate the impact of packet loss location on perceived speech quality, and the relationships between convergence time and loss location, we set up a simulation system. This includes speech encoder/decoder, loss simulation, perceptual quality measure and convergence time analysis, as shown in Figure 1. For codecs, we have a choice of G.729, G.723.1 and AMR. The standard 16 bit, 8 kHz sampled speech signal is processed by the encoder first. Then the parameter-based bit stream is sent to the decoder without frame losses (speech quality degradation in this case is only due to codec). The bitstream is also sent to the loss simulation module where the loss position and frame loss size can be selected. After loss simulation the bit stream is processed by the decoder to obtain the degraded speech signal with loss. The overall perceptual speech quality is measured between the reference speech signal and the degraded speech signal with loss by calculating the perceptual distance values using the PSQM+, MNB and EMBSD algorithms. The perceptual distance for each frame is also measured between the degraded speech without loss and the degraded speech with loss using PSQM+ for the analysis of convergence time. This eliminates coding impairment from the computation. The convergence time is also calculated using the normal Mean Square Error (MSE) method [1].

Loss simulation for each codec differs from the loss specification in the codecs. For G.729, if a parameter byte in the bit stream is set to zero, the frame is treated as a loss by the decoder and concealment is initiated automatically. For AMR, there is an extra byte for the transmit/receive frame type. For a lost frame, there is only a need to set the type as a BAD/ERASED frame. For G.723.1, a loss location mark file is created and serves as the input to the decoder.

V. Experiments and analysis of results

A. *Loss location and perceived speech quality*

In the first experiment, the impact of loss position on the overall perceptual speech quality or the performance of concealment under different loss locations is investigated. The PSQM+, MNB and EMBSD perceptual distance values are calculated for the whole test speech sentence (about 6 seconds), while only one loss is produced each time and the loss position moves smoothly from left to right. The move is one frame each time and the frame size is decided by the codec chosen. At each loss location , the frame loss size can change by one, two, three or four frames to simulate different packet size or burst loss size.

The waveform for the first talkspurt for the test sentence "Each decision show (s)" is shown in Figure 2. It consists of four voiced segments - V (1) to V (4) corresponding to the vowels 'i', 'i', 'ə' and 'au'. The voiced segments are separated by unvoiced segments.

The overall perceptual distance values for PSQM+, MNB and EMBSD for G.729 are shown in Figures 3, 4 and 5, respectively. The values (using PSQM+) for G.723.1 (6.3 Kb/s) and AMR (12.2 Kb/s and 4.75 Kb/s mode) are shown in Figures 6, 7 and 8. In all the figures, the horizontal scales are in the unit of frames. As the frame sizes are 10, 20 and 30ms for G.729, AMR and G.723.1, respectively, the total number of frames for the test segments shown are 134, 67 and 45.

Examination of Figure 3 shows that the perceptual distance value varies between 1.4 and 2.4 as the loss location moves from left to right. In the PSQM+, a change in perceptual distance indicates a change in perceptual speech quality (the smaller the distance, the better the perceived quality). Similar changes in perceived speech quality can also be seen for the MNB (Figure 4) and EMBSD (Figure 5), as well as for the different codecs (Figure 6, 7 and 8). It is evident that the same loss condition (one packet loss for the whole test speech segment) causes an obvious variation in overall perceived speech quality, but the variation is dependent on speech content. A loss at unvoiced speech segments shows little impact on perceived speech quality (almost the same perceptual distance values as for no-loss cases). However, a loss at voiced segments has different effects on perceived speech quality depending on its location within the voiced segment. At the beginning of a voiced segment, it has the most severe impact (the peaks in the figures). At the end of voiced segments, the impact is small. In the middle voiced segments, perceptual distances change depending on the codec and frame loss size. For example, for the G.729 one-frame loss (Figure 3), the perceptual distance value reaches its peak when the loss is at the beginning of voiced segments. Then, as the loss position moves to the right (for each voiced segment), the perceptual distance rapidly returns to the minimum value, showing a good convergence performance for voiced segments 1, 2 and 3. For voiced segment 4, the value varies depending on the speech content. As the frame loss size increases, the perceptual distance increases.

We explain this phenomenon from two perspectives:

(i). From the perspective of the codec or concealment algorithms

In the case of a loss at the beginning of voiced segment, as the previous frame is clearly an unvoiced frame or an unvoiced/voiced transition frame. The concealment algorithm will conceal the loss using the filter coefficients and the excitation for an unvoiced sound. It causes the lost frame to be concealed using the unvoiced features. In other words, during the unvoiced to voiced transition period, the shape of the vocal tract is in transition (not stable), and the LP filter coefficients will change rapidly for each frame. The excitation signal is also changing from unvoiced to voiced. The concealment algorithm can not conceal properly for the loss at this transition stage.

For a loss during the stationary part of a voiced segment, the concealment algorithm will conceal the current frame with the gain further reduced from the previous frame (adaptive codebook gain). The line spectral pair coefficients (or LP filter coefficients) of the last good frame are repeated. In other words, the vocal tract is at a stable stage (after the transition) and keeps the same shape. The LP filter coefficients are very stable during this stage. If the pitch delay does not change much within a short time period, a small loss can be concealed perfectly using the parameters of the previous frames. However, when there is an increase in burst loss size or frame size, it is difficult to conceal the losses adequately. The concealment performance degrades depending on the features in the voiced segments.

(ii). From the perspective of the perceptual quality measurement algorithms

The signal energy is very important for the overall perceived speech quality for all the perceptual algorithms. If a reference signal frame has a large signal energy (e.g. the beginning of a voiced segment), and the degraded signal has a very small energy (due to improper concealment), this will cause a significant increase in the perceptual distance. For a loss during the voiced segment, the degraded signal will normally have a rather large energy. Perceptual distance will vary for different loss size and loss location.

For different codecs (G.729, G.723.1 and AMR), the perceived speech quality shows large variations due to differences in the frame sizes. The perceptual distances using PSQM+ for the three codecs for a loss at the beginning of voiced segment 4 is summarized in Table 1 (including perceptual distances for no-loss cases).

Table 1: Perceptual distance using PSQM+

Codec Type	No-loss	1-frame	2-frame	3-frame	4-frame
G.729 (8 Kb/s)	1.36	1.62	1.83	2.11	2.42
G.723.1 (6.3 Kb/s)	1.51	1.79	2.84	3.54	4.03
AMR (12.2Kb/s)	0.98	1.35	1.6	2.06	2.45
AMR (4.75Kb/s)	1.92	2.17	2.42	2.81	3.34

From Table 1, it can be seen that the AMR (12.2 Kb/s) has the best perceptual quality and the AMR (4.75 Kb/s) the worst for no-loss cases. For a one-frame loss, the quality sequences remain the same. For a two-frame loss, the G.723.1 has the worst quality while AMR (12.2 Kb/s) remains the best. For three-frame and four-frame loss, G.729 and AMR (12.2 Kb/s) have similar perceptual quality, while G.723.1 remains the worst.

Of the three perceptual measurement methods (PSQM+, MNB and EMBSD), the PSQM+ provides perceptual distance values for most parts of the speech segment. The EMBSD and MNB only show the variations in perceived speech quality for frames with high energy. A loss at the unvoiced or voiced segments with small energy (see Figure 2) has no impact on perceived speech quality (flat line area in Figures 4 and 5). This is due to the different processing methods for silence and non-silence frames in the perceptual quality measurement algorithms. For EMBSD, the perceptual distance for an entire test speech segment is obtained by averaging over all non-silence frames (which are defined as the frames with the energy of the reference speech and the degraded speech both above their preset thresholds). For a loss at short and small energy voiced segments (e.g. voiced segment 1), the degraded speech with a loss has a limited energy. This is not taken into account by the EMBSD in the overall perceptual distance calculation and causes a flat area in Figure 5 (e.g. for voiced segments 1 and 3). A similar phenomenon exists for the MNB. The PSQM+ also classifies the frames as silence or non-silence. But it calculates all perceptual distances for silence or non-silence frames and uses different weighting factors for the overall perceptual distance calculation. Thus PSQM+ (Figure 3) also gives the perceptual distance value for a loss during small energy.

B. Convergence time with loss location

The second experiment was carried out to analyze the convergence time and its relationship to speech content or loss position. The convergence time is calculated by comparing the difference between the degraded signal without loss and the degraded signal with loss (as shown in Figure 1). First the MSE method [1] is used to calculate the convergence time for each loss position for a speech waveform such as that shown in Figure 2. Here the convergence time is defined as the first good frame received after a burst of lost frames until the frame with its MSE value below a threshold (1% of the maximum MSE value seen so far). The convergence time for G.729 is shown in Figures 9, in units of frames (10ms/frame). From the figure, we can see that the convergence times are almost the same for different loss sizes. It shows a good linear relationship for loss at the voiced segments. It is at a maximum at the beginning of the voiced segments and decreases gradually to a minimum at the end of the voiced segments. The convergence time for a loss at the unvoiced segments appears stable. Similar results were also obtained for the AMR and G.723.1 codecs. It seems that the convergence time is only related to the speech content and not to codec and frame loss size.

We analyze further the convergence time based on perceptual distance. We measured the frame-based PSQM+ values between degraded speech without loss and degraded speech with loss. We choose two voiced segments in Figure 2. One with only voiced part (V(2) in Figure 2) and another one with the adjacent unvoiced part (V(4) in Figure 2). We change loss positions from the beginning to the end of the waveforms. The perceptual distance variation curves for selected loss positions are shown in Figure 10 and 11, in the unit of frames (here it is the frame of PSQM+ calculation, which is 32ms frame size with 50% overlapping resulting in 16 ms real frame size). Curves 1 to 5 (Figure 10) and 1 to 12 (Figure 11) correspond to the loss position from left to right. The loss position for each curve corresponds to the first non-zero point in the curve. The duration of the frames with non-zero (or over a threshold) perceptual distance is related to the convergence time.

From Figures 10 and 11, we can see that if a loss occurs during a voiced segment, then the convergence time is almost the remainder of the length of that voiced segment from the loss point (curve 1 to 5 in Figure 10 and curve 6 to 12 in Figure 11). The perceptual distance itself changes significantly with changes in the location of loss while the influence of the loss seems only limited to the voiced segment. The convergence times are almost the same as for a loss at unvoiced parts (curves 1 to 5 in Figure 11). The PSQM+ curves vary in a similar way. This explains the linear relationship of the convergence time during the voiced segments and flat variation during the unvoiced segments as shown in Figure 9. PSQM+ variation curves also show the overall PSQM+ values for the different loss position. We also tested other voiced segments and obtained similar results. The convergence time is more closely related to speech content and less affected by frame loss size and codec type. The convergence time is constrained by the duration of the voiced segments.

VI. CONCLUSIONS

We have investigated the impact of loss positions on perceived speech quality and the relationships between the convergence time and loss locations. Preliminary results show that a loss at unvoiced speech segment has almost no obvious impact on perceived speech quality. However, a loss at the beginning of voiced segments has the most severe impact on perceived speech quality. We have explained this effect from both the perspectives of the concealment and objective perceptual measurement algorithms. The impact of loss position on perceived speech or the concealment performance of three modern codecs (G.729, G.723.1 and AMR) have also been compared and analyzed. Three different perceptual speech quality measurement algorithms (PSQM+, MNB and EMBSD) are compared for the purpose of loss location analysis. We have analyzed the convergence times for different loss locations and different codecs by taking into account the normal MSE and perceptual PSQM+ measure. The results show that the convergence time is affected mainly by speech content (e.g. it is very stable within unvoiced segment whereas it varies but constrained by the duration of the voiced segments).

This work should help to fully understand the real impact of packet loss on perceived speech quality and the features of the convergence time in order to set the real loss constraint distance between the losses. This could be help for the development of a perceptually relevant packet loss metric, which could be valuable in non-intrusive VoIP measurements or to set up more efficient speech recovery systems.

Further research will focus on a more extensive analysis of the impact of packet loss on speech content.

ACKNOWLEDGEMENT

We are grateful to the Speech Processing Lab of the Electrical and Computer Engineering Department at Temple University, especially Dr. Wonho Yang and Prof. Robert Yantorno, for providing us with the Enhanced Modified Bark Spectral Distortion (EMBSD) software to evaluate the performance of concealment in this paper.

We are grateful to WWG/Acterna for sponsorship.

REFERENCES

[1] J. Rosenberg. G.729 Error Recovery for Internet Telephony. Project Report, Columbia University, May 1997

[2] R. Koodli and R. Ravikanth, One-way Loss Pattern Sample Metrics <draft-ietf-ippm-loss-pattern-03.txt>, Internet Draft, Internet Engineering Task Force, July 2000

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

[4] ITU-T Recommendation G.723.1, Dual Rate Speech Coder for Multimedia Communication Transmitting at 5.3 and 6.3 kbit/s, March 1996

[5] 3G TS 26.091, AMR Speech Codec; Error Concealment of Lost Frames

[6] H. Sanneck and N. Tuong Long Le, Speech Property-Based FEC for Internet Telephony Applications, Proceedings of the SPIE/ACM SIGMM Multimedia Computing and Networking Conference 2000, San Jose, CA, January 2000

[7] Q. Xie, S. Gupta, Error Tolerant RTP Payload Format for AMR <draft-xie-avt-et-rtp-amr-00.txt>, Internet Draft, Internet Engineering Tasks Force, October 2000

[8] ITU-T Recommendation P.861, Objective quality measurement of telephone-band (300-3400 Hz) speech codecs, February 1998
[9] ITU-T Contribution COM 12-20-E, Improvement of the P.861 Perceptual Speech Quality Measure, KPN Research, Netherlands, Dec. 1997
[10] 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
[11] S. Voran, Objective Estimation of Perceived Speech Quality – Part II: Evaluation of the Measuring Normalizing Block Technique, IEEE Trans. on Speech and Audio Processing, Vol. 7, No.4. July 1999, pp. 383 -390
[12] W. Yang, Enhanced Modified Bark Spectral Distortion (EMBSD): An Objective Speech Quality Measurement Based on Audible Distortion and Cognition Model, Ph.D Dissertation, May 1999, Temple University, USA
[13] ITU-T Recommendation G.729 Annex B, A silence compression scheme for G.729 optimized for terminals conforming to Recommendation V.70, November 1996
[14] ETSI EN 301 704 V7.2.1 (2000-04), Digital cellular telecommunications system (Phase 2+); Adaptive Multi-Rate (AMR) speech transcoding (GSM 06.90 version 7.2.1 Release 1998)

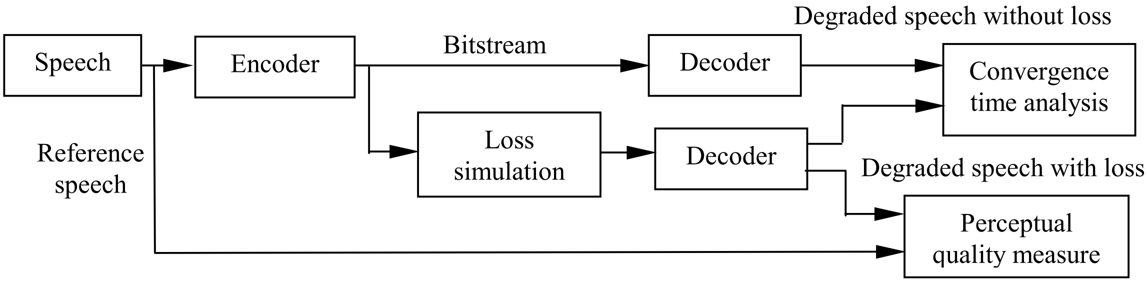

Figure 1: Structure of the simulation system

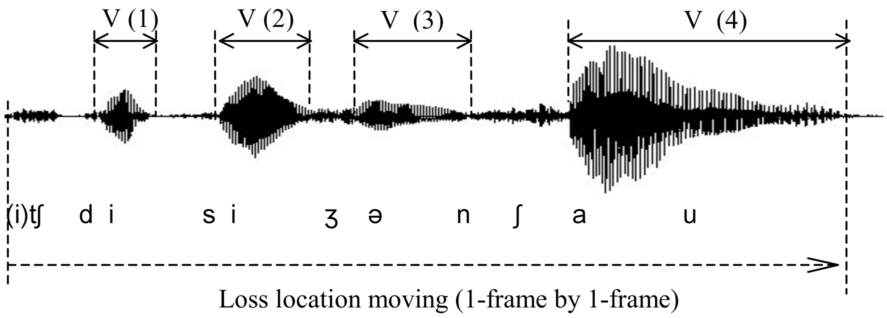

Figure 2: Speech waveform for the 1st talkspurt of test sentence
(The sentence is "_each decision show(s)_". V(1) to V(4) corresponds to 4 voiced segments)

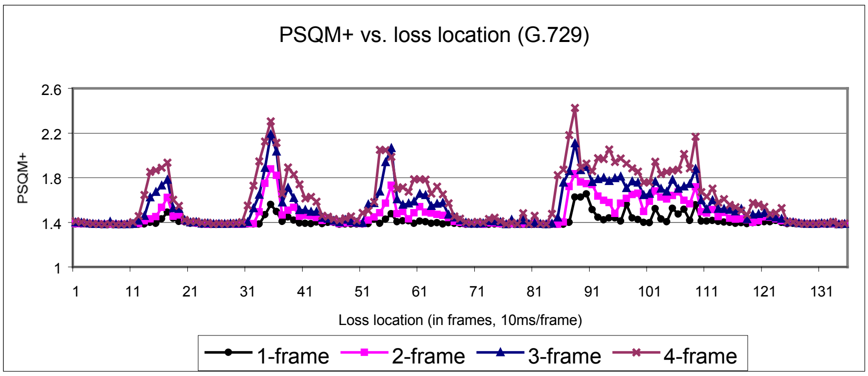

Figure 3: Overall PSQM+ values vs. loss location for G.729

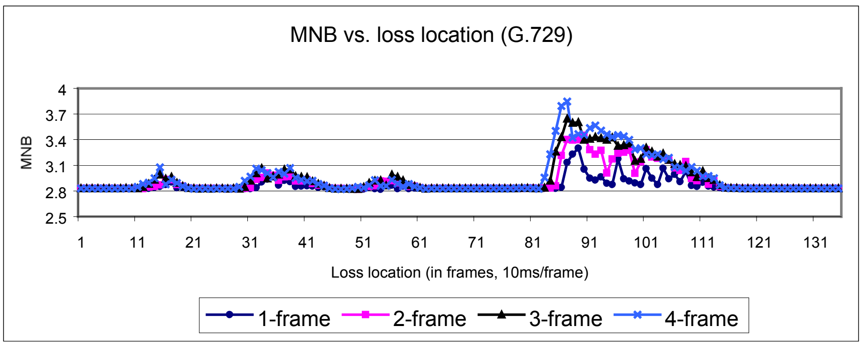

Figure 4: Overall MNB value vs. loss location for G.729

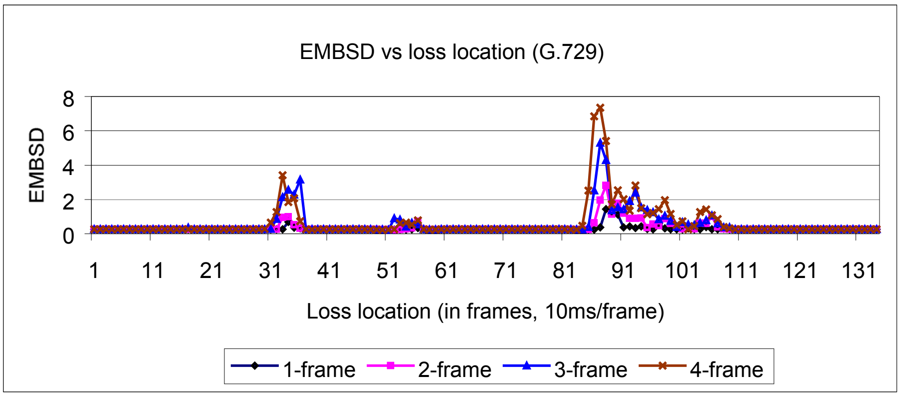

Figure 5: Overall EMBSD value vs. loss location for G.729

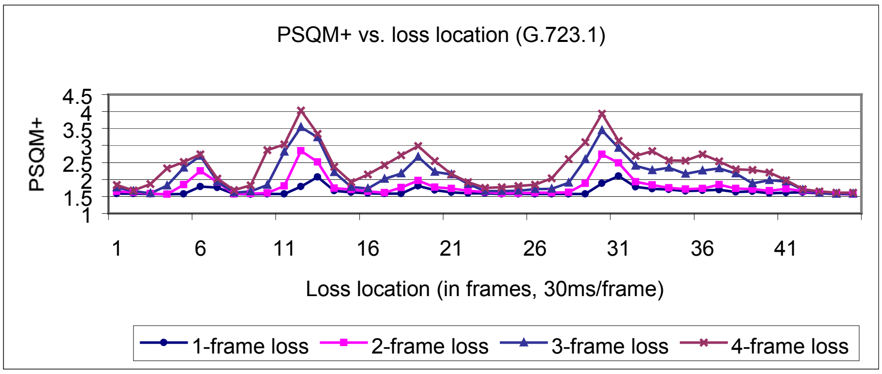

Figure 6: Overall PSQM+ value vs. loss location for G.723.1 (6.3 Kb/s)

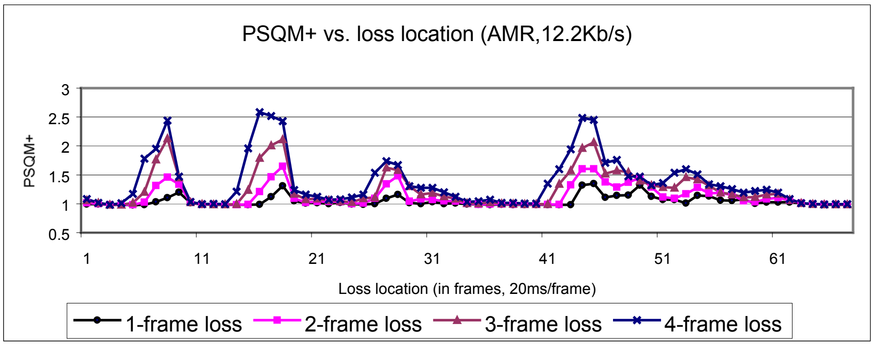

Figure 7: Overall PSQM+ value vs. loss location for AMR (12.2 Kb/s)

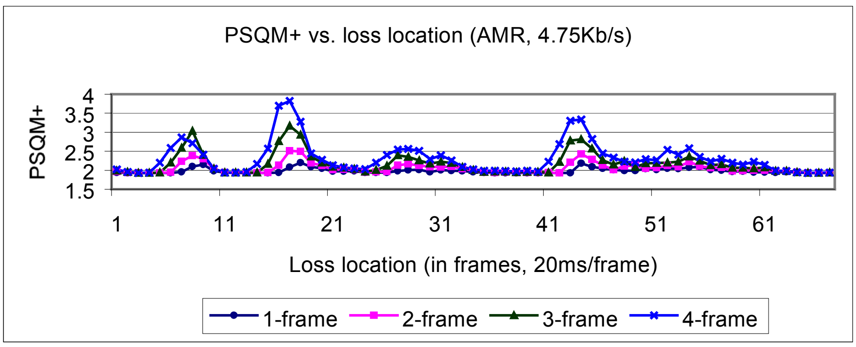

Figure 8: Overall PSQM+ values vs. loss location for AMR (4.75Kb/s)

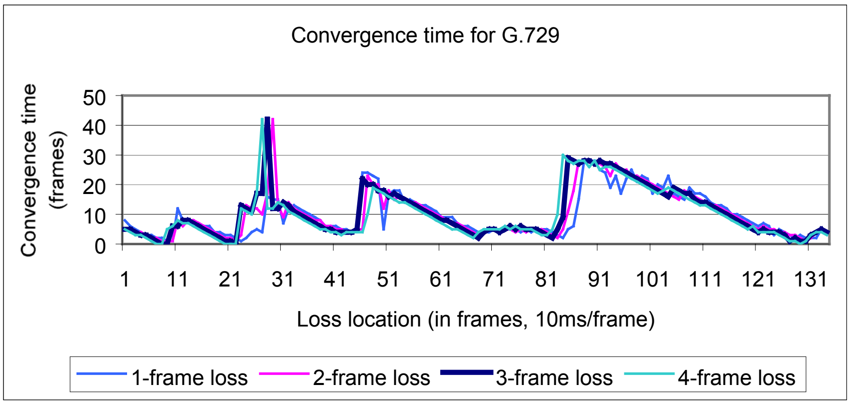

Figure 9: Convergence time vs. loss location for G.729

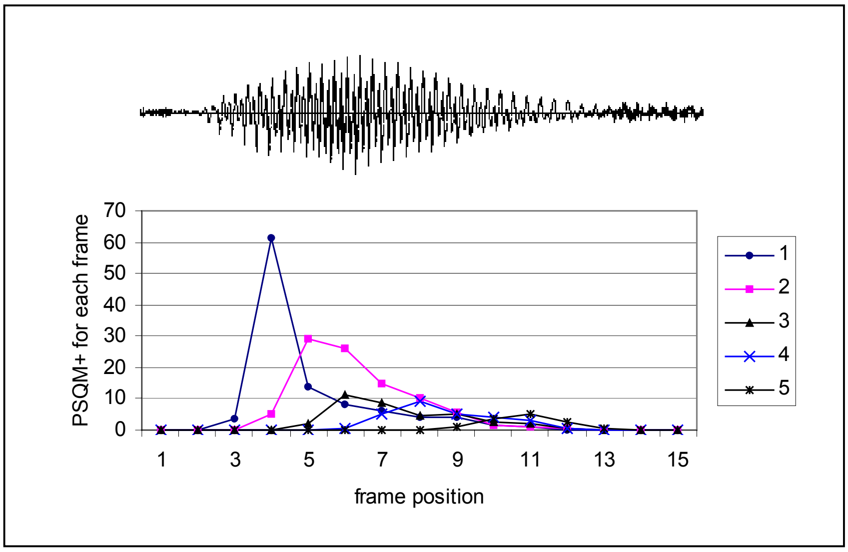

Figure 10: PSQM+ for voiced segment 2 (G.729, 2-frame loss)
(Curves 1 – 5 correspond to 5 loss locations from left to right)

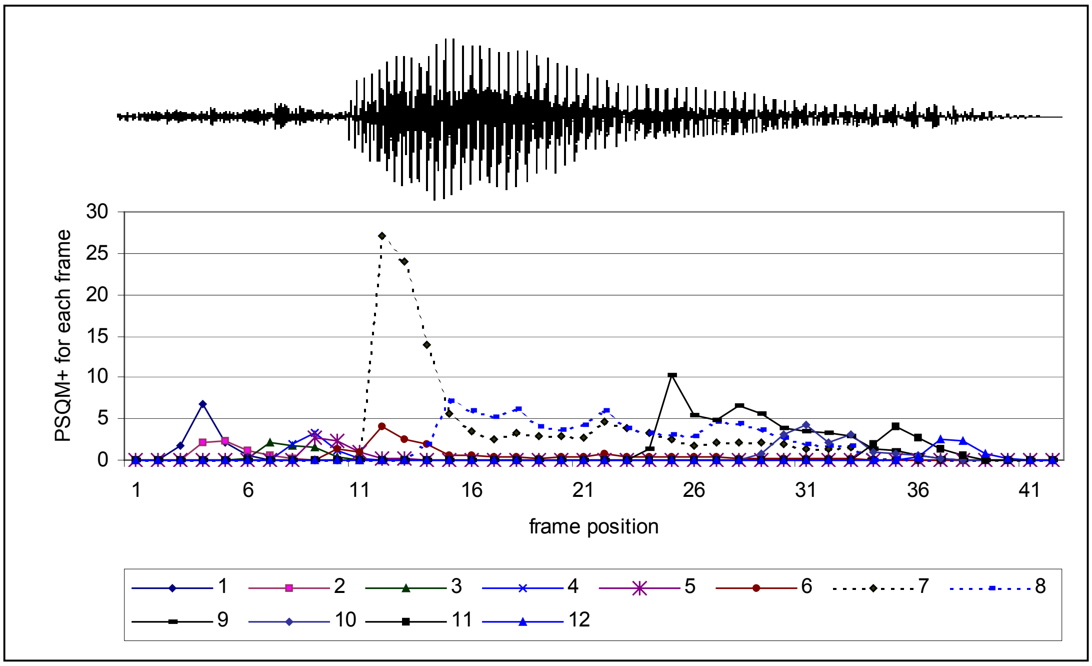

Figure 11: PSQM+ for voiced segment 4 (G.729, 2-frame loss)
(Curves 1 to 12 correspond to 12 loss locations from left to right)