ISCA Archive
http://www.isca-speech.org/archive

COST278 and ISCA Tutorial and Research Workshop
on Robustness Issues in Conversational Interaction
University of East Anglia, Norwich, UK
August 30-31, 2004

Packet Loss Modelling for Distributed Speech Recognition

Ben Milner and Alastair James

School of Computing Sciences,
University of East Anglia, Norwich, UK

{b.milner, a.james}@uea.ac.uk

Abstract

The evaluation of packet loss compensation techniques for distributed speech recognition requires an effective model of packet loss that is capable of reproducing the burst-like occurrence of loss. Several models have been applied to this task and are based on two or three state Markov chains or Markov models. This work reviews these models in terms of their channel characteristics such as the probability of packet loss and average burst length. Validation of the models is made against both GSM error patterns and a wireless LAN channel which demonstrates effective simulation. A series of speech recognition tests show that similar performance is obtained on the real and simulated channels using the packet loss models. Finally a set of model parameters is presented which allows testing across a range of channel conditions.

1. Introduction

The recent interest in developing robust algorithms for distributed speech recognition (DSR) has lead to considerable effort being made in the area of packet loss compensation. Some of these schemes operate at the bit level of the transmission channel and utilize error correction and detection methods to protect the packet [1,2,3]. In the event of uncorrectable bit errors the packet is then considered lost. Other schemes operate at the packet level where a packet is either delivered correctly or lost outright. In the event of loss, compensation methods fall broadly into two categories. The first attempts to restore the feature vector stream by estimating the value of lost vectors through techniques such as repetition or interpolation [4,5]. The second applies compensation either fully or in part at the decoding stage of the recogniser [6,7].

The methods of testing these packet loss compensation schemes vary considerably. Some systems are tested using real data such as GSM bit level error patterns which are configured at different levels of bit error rate [1,3]. More commonly testing is achieved using a packet loss model. In particular compensation methods have been tested using a 2-state Markov chain [2,5,6], the Gilbert model [7] or a 3-state Markov chain [3,4] to simulate packet loss.

The aim of this work is to compare these different models of packet loss and to examine their effectiveness at simulating different packet loss conditions. Section 2 describes the models under consideration which are based on either two or three state configurations of Markov chains or Markov models. The effectiveness of these models to simulate real packet loss profiles is examined in section 3. Validation on both a GSM channel and a wireless local area network (WLAN) channel are reported. Finally in section 4 speech recognition accuracy across these channels is compared for both the real and simulated channels.

2. Packet Loss Models

The typical environments over which DSR systems operate are mobile or IP networks. Both of these networks are considered unreliable and can result in packet loss. In mobile networks packet loss occurs when the received signal strength is weak and is associated with times of fading. Channel coding methods allow for the correction of certain proportions of bit errors but in severe conditions bit errors will manifest into uncorrectable errors and hence packet loss occurs. In IP networks the majority of packet loss occurs from congestion at network nodes. Transport layer protocols such as TCP (Transmission Control Protocol) provide for the retransmission of lost packets, but for real-time systems the delay they impose can be prohibitively long. Real-time systems usually employ UDP (User Datagram Protocol) as this has no retransmission facility but makes no correction for lost or delayed packets and hence loss can occur. The nature of both mobile and IP networks means that packet loss tends to occur in bursts. Periods of fading in mobile networks can occur for many hundreds of milliseconds and result in burst-like errors. IP network traffic has a burst-like nature which leads to burst-like packet loss in times of congestion.

To analyse the characteristics of a channel the most simple measurement is the average probability of packet loss or unconditional loss probability, α [8]. The burstiness of packet loss can be measured by the conditional loss probability that a packet is lost given that the previous packet was lost. This is more usefully expressed as the average burst length of packet loss, β. The remainder of this section examines three packet loss models in terms of their topologies and the packet loss characteristics which they can generate.

2.1. Two State Markov Chain

The two-state Markov chain model of packet loss uses state 1 to represent a packet being correctly received and state 2 to represent a packet being lost. Figure 1 illustrates the two-state Markov chain which shows that its character is governed by the self-loop probabilities of its two states Q and q.

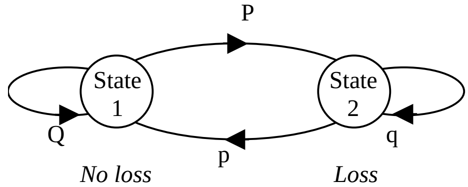

Figure 1: Two-state Markov chain

The average packet loss burst length, β_{2state}, produced by the model is governed by the state 2 self-loop probability, q, and is equal to its mean state duration which is given as

$$\beta_{2state} = \frac{1}{1-q} \tag{1}$$

The probability of packet loss, α_{2state}, relates to the steady-state conditions of the Markov chain. The set of steady state probabilities, $\boldsymbol{\pi}=\{\pi_1, \pi_2\}$, are governed by the transition matrix, **A**, through,

$$\boldsymbol{\pi} = \mathbf{A}\boldsymbol{\pi} \tag{2}$$

and are found from the normalized eigenvector of **A** corresponding to the largest eigenvalue [9]. The steady state probability of being in state 2, π_2, equals the overall probability of packet loss, α_{2state}, and is given,

$$\alpha_{2state} = \frac{P}{2-Q-q} \tag{3}$$

In configuring the model it is useful to specify the desired packet loss characteristics, α and β, and compute the required transition probabilities – from equations (1) and (3),

$$q = 1 - \frac{1}{\beta_{2state}} \quad \text{and} \quad Q = 1 - \frac{\alpha_{2state}}{(1-\alpha_{2state})\beta_{2state}} \tag{4}$$

2.2. Gilbert and Elliot Models

An extension to the two-state Markov chain is the Gilbert model which is a two-state Markov model [10]. State 1 of the model is termed the good state in which no packet loss occurs. State 2 is termed the bad state and has a probability of packet loss of *(1-h)*. Figure 2 illustrates the Gilbert model.

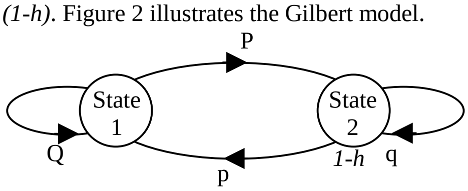

Figure 2: Gilbert model

An extension to the Gilbert model is the Gilbert-Elliott model which introduces a probability of loss *(1-k)* that is associated with the good state, where *(1-h) >> (1-k)* [11].

For the Gilbert model (where $k=1$) the overall probability of loss, $\alpha_{Gilbert}$, is given,

$$\alpha_{Gilbert} = \left(\frac{P}{2-Q-q}\right)(1-h) \tag{5}$$

and the average burst length, $\beta_{Gilbert}$, as,

$$\beta_{Gilbert} = \frac{1}{1-q(1-h)} \tag{6}$$

The three parameters of the Gilbert model, q, Q and h, cannot easily be estimated from packet loss data. Instead they can be expressed in terms of three other parameter (given as a, b and c) which can be measured easily [10].

The Gilbert model can also be represented as a 3-state Markov chain through expansion of state 2 into 2 separate states. Occupancy of state 2 indicates packet loss while occupancy of states 1 and 3 indicate no loss. Figure 3 illustrates this equivalence along with the set of 3-state transition probabilities.

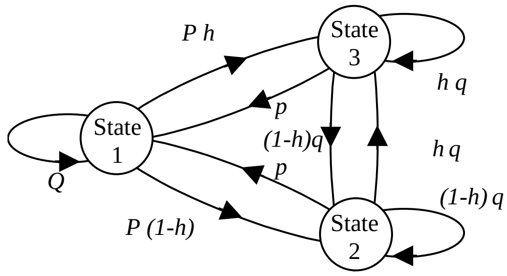

Figure 3: Three-state Markov chain Gilbert model

2.3. Three State Markov Chain

An alternative to the 3-state Markov chain equivalence of the Gilbert model is to remove the transitions which link states 1 and 3 and allow state 3 to have an independent self-loop probability, Q' – as illustrated in figure 4.

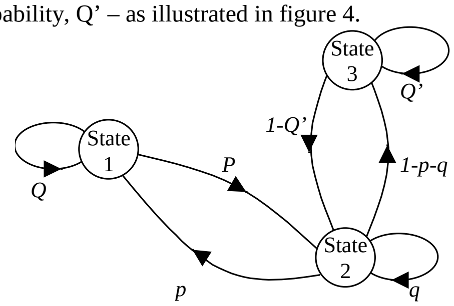

Figure 4: Three state packet loss model

The self-loop probability of state 3 in the Gilbert model (figure 3) is inversely related to the self-loop probability of state 2. Allowing an independent self-loop probability for state 3 enables two types of error free period to be defined which occur between periods of packet loss. Setting the self-loop probability to be high in state 1 and low in state 3 enables long duration periods of no loss to be modeled in state 1 while state 3 models short periods of no loss which occur in-between packet loss in burst-like conditions. This gives more control over the characteristics of packet loss and loss-free periods.

The average packet loss burst length, β_{3state}, of the three-state model is equal to the mean state duration of state 2,

$$\beta_{3state} = \frac{1}{1-q} \tag{7}$$

The overall probability of packet loss relates to the steady-state probability of being in state 2. This can be computed by finding the normalized eigenvector of the transition matrix associated with the largest eigenvalue, and is given,

$$\alpha_{3state} = \frac{(1-Q')(1-Q)}{Q(p+q+Q'-2) - Q'(1+p) - q + 2} \tag{8}$$

The transition probabilities of the model can be determined from four parameters estimated from training data,

- α = *overall probability of a packet being lost*
- β = *average packet loss burst length*
- N_1 = *average number of packets received consecutively between burst periods*
- N_3 = *average number of packets received consecutively between bursts in burst-like conditions*

Parameters β, N_1 and N_3 give the self-loop probabilities,

$$Q = 1 - \frac{1}{N_1}, \quad Q' = 1 - \frac{1}{N_3} \quad \text{and} \quad q = 1 - \frac{1}{\beta} \tag{9}$$

The remaining probability, p, is dependent on the self-loop probabilities and the overall probability of packet loss, α,

$$p = \frac{1-Q}{Q-Q'}\left[\frac{1-Q'}{\alpha} + q + Q' - 2\right] \tag{10}$$

3. Validation of Models

Using the three packet loss models this section compares real packet loss data from GSM and WLAN channels against simulated packet loss profiles.

3.1. Simulation of GSM DSR packet loss

A set of bit level error patterns (EP1, EP2 and EP3) for GSM data has been used in the DSR standardization work made by the ETSI Aurora group [1]. These three error patterns vary in terms of the carrier signal-to-noise ratio (SNR) with EP1 at 10dB, EP2 at 7dB and EP3 at 4dB. The Aurora standard specifies that MFCC vectors are compressed to 44 bits and transmitted in pairs along with a 4 bit cyclic redundancy check (CRC) to give a total of 92 bits. If any bit errors are detected in the packet then it is declared lost. Using this coding scheme a set of 100,000 DSR packets were created and according to the bit error patterns certain packets were lost when errors occurred. Tables 1, 2 and 3 show estimated model parameters for the 2-state, Gilbert and 3-state packet loss models for the three error patterns – EP1, EP2 and EP3.

	EP1	*EP2*	*EP3*
Q	0.9997	0.9791	0.8317
q	0.3770	0.5197	0.7298

Table 1: Two-state Markov chain parameters for EP1-3

	EP1	*EP2*	*EP3*
Q	0.99967	0.9794	0.8429
q	0.3765	0.5619	0.7685
1-h	1.0000	0.9250	0.9497

Table 2: Gilbert model parameters for EP1-3

	EP1	*EP2*	*EP3*
Q	0.99968	0.9798	0.8507
Q'	0.0000	0.3333	0.2000
q	0.1538	0.3372	0.6305
p	0.8462	0.6304	0.2982

Table 3: Three-state Markov chain parameters for EP1-3

To analyse the effectiveness of these models table 4 compares the real packet loss rate, α, and average burst length, β, with that simulated from the packet loss models. In each case a total of 100,000 packets were simulated.

	EP1		*EP2*		*EP3*	
	α	β	α	β	α	β
Real	0.0003	1.19	0.031	1.51	0.324	2.71
2-state	0.0003	1.22	0.031	1.48	0.318	2.70
Gilbert	0.0004	1.23	0.031	1.50	0.325	2.70
3-state	0.0004	1.18	0.031	1.51	0.330	2.71

Table 4: GSM DSR packet loss characteristics

The results show that all three packet loss models are able to reproduce the loss rate and average burst length characteristics of the real packet loss data. To extend the analysis, figure 5a compares the distribution of average burst lengths of packet loss for the real and simulated data using EP3. Similarly figure 5b shows the distribution of loss-free burst lengths using EP3.

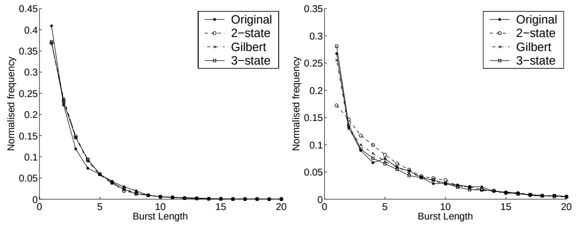

Figure 5: Distribution of real and simulated burst lengths for a) packet loss, b) loss-free periods on EP3

All three packet loss models show close similarity with the real packet loss data in terms of the distribution of packet loss burst lengths. However the distribution of loss-free periods shows that the 2-state Markov chain is less effective at reproducing the original profile than the Gilbert model and 3-state Markov chain. These models have more control over the distribution through the pair of loss free states (states 1 and 3).

3.2. Simulation of WLAN packet loss

A second set of real packet loss data is considered in this section based upon measurements taken using a wireless LAN connection. As in the previous section a training set of 100,000 packets was transmitted with lost packets noted leading to channel characteristics of $\alpha=0.139$ and $\beta=1.68$. From the resulting packet loss profile a set of model parameters for the 2-state, Gilbert and 3-state packet loss models of section 2 were calculated as shown in table 5.

	Q	*q*	*1-h*	*Q'*	*p*
2-state	0.9042	0.4060	-	-	-
Gilbert	0.8957	0.5607	0.7241	-	-
3-state	0.9363	0.4072	-	0.5652	0.3631

Table 5: WLAN parameters for packet loss models

As in the previous section, all three models of packet loss were able to accurately reproduce the packet loss rate and average burst length of the real packet loss data in a simulation of 100,000 packets. To further compare the real and simulated data, figures 6a and 6b show the distribution of packet loss burst lengths and loss-free burst lengths.

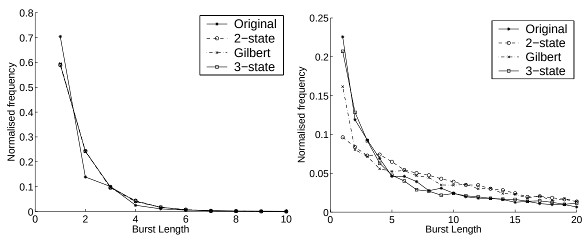

Figure 6: Distribution of a) packet loss, b) loss-free bursts

The distribution of loss-free bursts shows that the 3-state Markov chain is most effective at reproducing the characteristics of the real data. Both the Gilbert model and 2-state Markov chain produce too many longer duration periods of no loss at the expense of not producing enough short bursts. The independent duration of state 3 in the 3-state Markov chain allows more short duration loss-free bursts to be created.

3.3. Variability of packet loss

A series of tests measuring the characteristics of an IP network (between the University of East Anglia, Norwich, UK and a site in Taipei, Taiwan) revealed that both the packet loss rate and burst length was highly variable. Variations were related to the size of packet transmitted, the rate at which packets were sent, the time of day and also to seemingly random variations in the network. In a range of tests (each measured over a 24 hour period) the overall percentage of packet loss was below 10% but often peaked at over 50% for short periods of time. Similarly the long-term average burst length was below 5 packets, but could peak at over 20 packets. This high variation in channel conditions suggests that an alternative to performing speech recognition tests using one of the simulated channels is to define a set of packet loss characteristics. This enables recognition performance to be analysed across a range of different packet loss conditions. Based on the measurements above, a suitable range of channel conditions is probabilities of loss of $\alpha=0.1, 0.2, 0.3, 0.4, 0.5$ and average burst lengths of $\beta=1, 4, 8, 12, 20$ packets.

4. Speech Recognition Tests

To further analyse the effectiveness of the packet loss models this section compares speech recognition accuracy using both real and simulated channels. All experiments use the Aurora connected digits database which encompasses 4004 digit strings for testing. In the event of packet loss the pair of MFCC vectors in the packet are both lost.

4.1. GSM and WLAN

Table 6 compares speech recognition accuracy on both the real GSM error patterns and WLAN data with that simulated using the 2-state, Gilbert and 3-state models. In the event of packet loss nearest neighbour repetition of vectors is used to replace the value of lost vectors [1].

	EP1	*EP2*	*EP3*	*WLAN*
Real	99.0	98.9	94.4	97.6
2-state	99.0	98.9	95.0	98.5
Gilbert	99.0	99.0	94.7	98.5
3-state	99.0	98.9	94.7	98.5

Table 6: Recognition accuracy on real and simulated channels

For EP1 and EP2 recognition performance from the simulated and real channels are virtually identical. On EP3 the Gilbert and 3-state Markov chain give slightly closer recognition accuracy to the real channel than the 2-state Markov chain. On the WLAN channel all three models give identical performance which is 0.9% above that of the real channel. The loss characteristics of the WLAN channel are more variable than the error patterns which may partly explain the difference.

4.2. Packet loss profiles

This section performs recognition tests using the set of packet loss profiles suggested in section 3.3. Figure 7 shows surface plots which illustrates the effect of both the packet loss rate, α, and the average burst length, β, on recognition accuracy. In figure 7a no packet loss compensation has been applied while in figure 7b interpolation is used to estimate lost vectors [4].

Figure 7a shows that with no packet loss compensation as the burst length increases, for a given loss rate, the accuracy converges rapidly to

$$\textit{baseline accuracy} \times (1-\alpha)$$

meaning that word accuracy is sensitive to increases in α. However, if interpolation is applied, accuracy is less dependent on α provided β is small. Indeed, even at $\alpha=0.5$ baseline performance can be maintained provided $\beta < 4$.

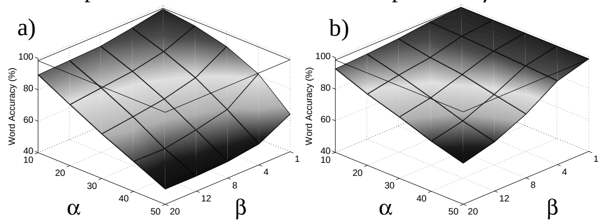

Figure 7: Accuracy with a) no compensation, b) interpolation

5. Conclusion

Three models of packet loss have been examined in terms of their packet loss rate and average burst length. Simulated packet loss profiles created by these model have been compared to real packet loss data taken from both GSM and WLAN channels. Analysis reveals that all models are able to reproduce the packet loss rate and burst lengths accurately. Comparison of average packet loss burst lengths and loss-free burst lengths reveals that for more lossy channels the 3-state Markov chain and Gilbert model are more effective than the 2-state Markov chain. Speech recognition tests comparing accuracy attained using the packet loss models with real packet loss data shows close similarity.

6. References

1. ETSI document - ES 202 212- STQ: DSR; Extended advanced front-end feature extraction algorithm, 2003.
2. C.B. Boulis et al, "Graceful degradation of speech recognition performance over packet-erasure networks", IEEE Trans. SAP, pp. 580-590, 2002.
3. Z-H. Tan, P. Dalsgaard and B. Lindberg, "OOV-detection and channel error protection for distributed speech recognition over wireless networks", Proc. ICSLP, 2003.
4. B.P. Milner, "Robust speech recognition in burst-like packet loss", Proc. ICASSP, 2001.
5. A. Gomez et al, "A source model mitigation technique for DSR over lossy packet channels", Proc. Eurospeech, 2003.
6. T. Endo, S. Kuroiwa and S. Nakamura, "Missing feature theory applied to robust speech recognition on IP networks", Proc. Eurospeech, 2003.
7. A. Bernard and A. Alwan, "Low-bitrate distributed speech recognition for packet-based and wireless communication", IEEE Trans. SAP, pp. 570-579, 2002.
8. J-C. Bolot and H. Crepin, "Analysis and control of audio packet loss over packet-switched networks", Proc. NOSSDAV, 1995.
9. A. Leon-Garcia, Probability and random processes for electrical engineering, Addison-Wesley, 1994.
10. E.N. Gilbert, "Capacity of a burst-noise channel", Bell System Technical Journal, pp. 1253-1265, 1960.
11. E.O. Elliott, "Estimation of error rates for codes on burst-noise channels", Bell System Technical Journal, pp. 1977-1997, 1963.