Degradation of Speech Recognition Performance over Lossy Data Networks

Dimitris Pratsolis
Department of Electronic and Computer Engineering
Technical University of Crete
Chania, Greece
pratsdim@telecom.tuc.gr

Nikos Tsourakis
Department of Electronic and Computer Engineering
Technical University of Crete
Chania, Greece
ntsourak@telecom.tuc.gr

Vassilis Digalakis
Department of Electronic and Computer Engineering
Technical University of Crete
Chania, Greece
vas@telecom.tuc.gr

ABSTRACT
In this work we investigate the effects of lossy data networks on the speech recognition performance, utilizing a stock information corpus. Within the context of the current paper we present the models of the lossy channel that were used in order to perform the specific simulations. The resulted voiceprints, after the transmission though the lossy channels, were fed to an Automatic Recognition System (ASR) in order to calculate the degradation in the performance. The specific procedure can be very helpful in extracting information that can be used for the design and the parameter tuning of the underlying data network.

Categories and Subject Descriptors
C.2.1 [**Computer-Communication Networks**]: Network Architecture and Design – *wireless communication, packet-switching networks, network communications.*

General Terms
Algorithms, Measurement, Performance

Keywords
Speech recognition performance; Lossy data networks; Gilbert-Elliot model; Three-state Markov model.

1. INTRODUCTION
The trend of the continuously increasing use of data communication is expanding to the mobile wireless world, as it has already taken place in the world of landline communications. The need for data access is growing and so is the demand for new applications, especially multimedia, that take advantage of the offered medium.

As wireless data networks evolve, the design of new communication protocols increases in size and complexity. The proper evaluation of the current networks provides the basis for the optimization of future protocols. A number of techniques are available for modeling and simulating the channel conditions of the underlying data network. The most common techniques include simulation, analysis of empirical data and analytical models [1].

Gilbert [5] appears to be the first to present a burst error model utilizing a Markov or multi-state model, a work that was later extended by Elliot [6] and Cain and Simpson [4]. Higher state Markov models that represent error distributions were also described by Blank and Trafton [3]. Others followed a different approach with the identification of the statistical distribution of gaps, e.g. hyperbolic distributions [9] and Pareto distributions that model inter-error gaps [2]. Lewis and Cox [7] concluded that there is a strong positive correlation between adjacent gaps in measured error distributions. In IP networks the packet loss, typically caused by congestion, is modeled with similar techniques.

In our work we focus on packet-based IP networks, as they are becoming an attractive alternative especially for wireless voice communications.

The volume of the required data in order to perform simulations and the availability of the network being tested, which may even be under design or deployment, often prohibit the utilization of real traces. One can consequently generate synthetic traces (e.g. from voiceprints) that simulate different conditions of the network and perform the necessary experiments. This is the procedure that we chose to follow and will describe in the following sections.

2. SIMULATION MODELS
In the context of our work we incorporated two error models, namely the Gilbert-Elliot model and a Three-state Markov model.

2.1 Gilbert-Elliot Model
The Gilbert-Elliot (GE) [5][6] channel model is a two-state Markov model, which is widely used to simulate the bursty packet loss behavior. This channel model has been shown to be able to effectively capture the bursty packet loss behavior of the Internet and wireless channels.

Furthermore, the special structure of the Markov model makes it analytically tractable. The two states of the GE model are denoted as G (good) and B (bad), as illustrated in Figure 1. In state G, packets are assumed to be received correctly and timely, whereas in state B, packets are assumed to be lost.

WMuNeP'07, October 22, 2007, Chania, Crete Island, Greece.

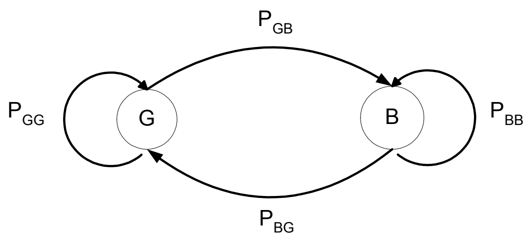

Figure 1. Gilbert-Elliot model state transition diagram

This model can be described by the transition probabilities P_{GB} from state G to B and P_{BG} from state B to G. Typically, the average packet loss probability (P_B) and the average burst length (L_B) must be defined in order to describe the bursty packet loss behavior of the GE channel model, which are given by the following equations:

$$P_B = P_{GB} / (P_{GB} + P_{BG}), \quad L_B = 1 / P_{BG}$$

2.2 Three-state Markov Model

The two-state Markov model presented earlier is generally able to capture only short bursts of the error sequence, for example 1 to 3 in length. In real data networks however consecutive losses appear in longer bursts that may extend to tens of seconds. We therefore utilize a three-state Markov model that can capture both very short duration consecutive loss events and longer lower density events.

The corresponding error model is represented by a three-state discrete-time Markov chain [8], as the one shown in Figure 2. Errors over the channel occur in the states long bad (LB) and short bad (SB), while the good (G) state is error-free. The difference between the long bad LB and short bad SB states is the time correlation of errors: LB corresponds to long bursts of errors and SB to short ones.

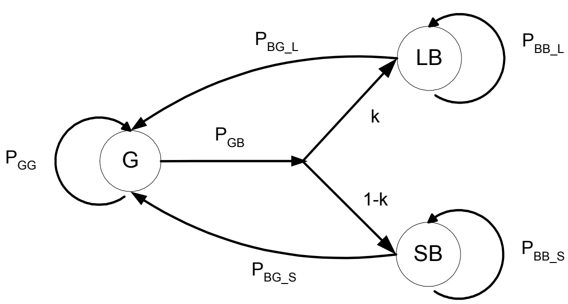

Figure 2. Three-state Markov model transition diagram

The parameter k is the probability that the Markov chain moves to state LB given that it leaves state G; k also represents the probability that an error burst is long, or, in other terms, the fraction of long bursts over the total number of error bursts.

Let P denote the transition probability matrix for the packet error process, which is given from the equation below:

$$P = \begin{pmatrix} P_{GG} & P_{GB\_S} & P_{GB\_L} \\ P_{BG\_S} & P_{BB\_S} & 0 \\ P_{BG\_L} & 0 & P_{BB\_L} \end{pmatrix} = \begin{pmatrix} 1-P_{GB} & k \cdot P_{GB} & (1-k)P_{GB} \\ P_{BG\_S} & 1-P_{BG\_S} & 0 \\ P_{BG\_L} & 0 & 1-P_{BG\_L} \end{pmatrix}$$

Note that no transitions are allowed between states LB and SB and the model can be fully incorporated when parameters P_{GB}, P_{BG_S}, P_{BG_L} and k are defined.

3. EVALUATION PROCESS

The word error rate (WER) is a commonly used metric to evaluate speech recognizers. A different error metric and in some cases more appropriate, is the natural language error rate (NLER). Generally, in speech recognition applications, we usually are interested in the interpretation of a spoken utterance rather than its accurate transcription. The NLER is simply defined as the number of NL errors occurred during the examination of the utterances, divided by the number of the reference utterances expressed with the following equation:

$$\%NLER = (NLerrors / utterances) \cdot 100\%$$

During the evaluation process we calculated the NL error rate after processing a test set derived from a stock information application, which consisted of 1000 utterances (8KHz, 8 Bit, PCM). We applied the error models discussed earlier, using different configuration parameters and simulated different strategies of the ASR system for handling transmission errors. The configuration parameters for each model are the transition probabilities presented earlier. The recognizer can incorporate three different strategies for a missing packet:

1. Replacing the missing packet with silence.
2. Ignoring the missing packet.
3. Replacing the missing packet with the previous one.

Each state in the error models corresponds to a time slot of 20 ms, which is associated with the transmission of a packet of 160 bytes. Thus, a packet transmission is successful only if the error model is in state G for all slots needed for the packet to be transmitted, while it fails otherwise. Each utterance is therefore split into packets of 160 bytes, and the error models determine which one of them will be received by the ASR.

In the Gilbert-Elliot model two parameters must be defined, namely the packet loss probability P_B and the average burst length L_B. For each parameter, the range of possible values is provided and each pair constitutes the current configuration of the model. The parameter values used in our simulations are summarized in the following table:

TABLE I. GILBERT-ELLIOT MODEL CONFIGURATION

Parameter	Min	Max	Step
P_B	1%	15%	1%
L_B	1	4	1

After executing 50 repetitions for each configuration, the number of different experiments derived from the specific model simulations is:

$$(values\ of\ P_B) \cdot (values\ of\ L_B) \cdot (ASR\ strategies) \cdot (repetitions) = 15 \cdot 4 \cdot 3 \cdot 50 = 9000$$

We used a similar process for the Three-State Markov model, where P_{GB}, P_{BG_S}, P_{BG_L} and k are the configuration parameters with a range presented in the table below:

TABLE II. THREE-STATE MARKOV MODEL CONFIGURATION

Parameter	Min	Max	Step
P_{GB}	1%	5%	1%
P_{BG_S}	35%	35%	0
P_{BG_L}	15%	15%	0
k	5%	20%	5%

Similarly, the number of experiments obtained from each configuration and for 50 repetitions is 3000.

For each experiment, we simulated the packet loss for all 1000 sentences of the test set and fed the resulting sentences to the ASR system in order to calculate the NL error rate. We should note that the system could recognize among 500 unique stock names uttered in different contexts.

4. EVALUATION RESULTS

In this section we will present the graphs obtained from the simulations with the two models. The specific results are compared with the NL error rate obtained from the original clean test set (baseline), which yields a NLER of 4.64%.

4.1 Gilbert-Elliot Model Evaluation Results

Initially, four graphs that correspond to the Gilbert-Elliot model will be presented. The NL error rate associated with each one of the three ASR strategies discussed earlier is depicted in Figure 3, 4 and 5. The specific error rate, as an average from all repetitions, is calculated with respect of the packet loss probability (P_b) in the x-axis and the average burst error length (L_b). Each graph contains four plots that correspond to the different values of L_b. When L_b equals to 1 it is guarantied that no consecutive error packets will be encountered in the packet error burst as P_{BB}=0.

The strategy of replacing the lost packets with the preceding one yields to the best results, as in most cases there is a strong correlation of adjacent speech samples and consequently of adjacent speech packets. The worst results were obtained using the strategy of replacing the missing packets with silence.

In all configurations we calculated the amount of lost packets, which is more or less the same.

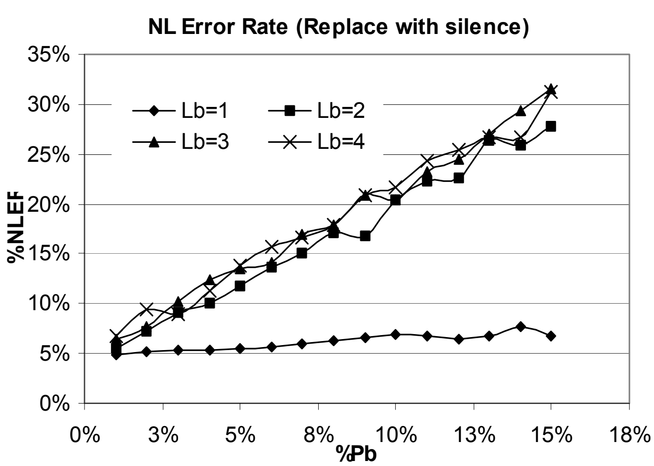

Figure 3. NL Error Rate for the first ASR strategy

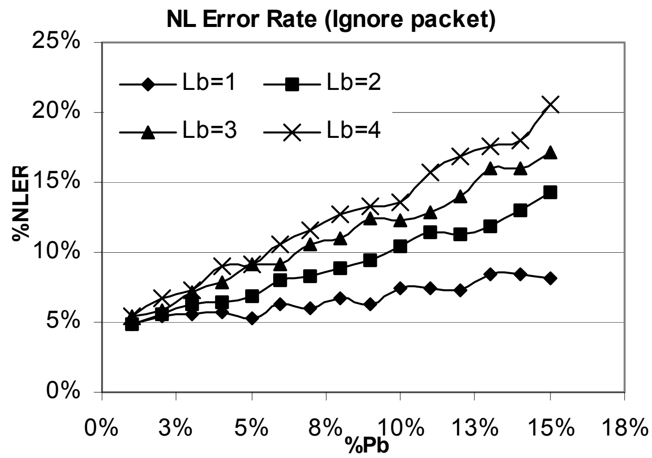

Figure 4. NL Error Rate for the second ASR strategy

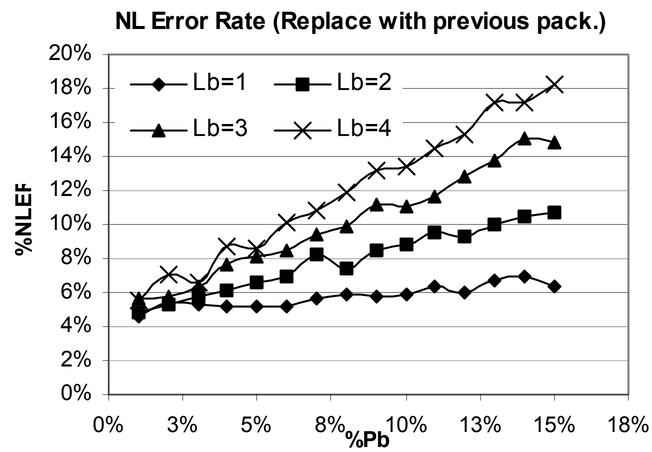

Figure 5. NL Error Rate for the third ASR strategy

4.2 3-state Markov Model Evaluation Results

For the three-state Markov model the NL error rate associated with each one of the three ASR strategies is depicted in Figure 6, 7 and 8. The specific error rate is calculated with respect to the probability P_{GB} in the x-axis and the parameter k. Each graph contains four plots that correspond to the different values of k.

We should note that when k equals 0, the specific three-state Markov model becomes a Gilbert-Elliot model and no transitions to state LB take place.

Again the third ASR strategy yields to the best results and the calculated amount of lost packets is similar in the four configurations, which differ in the way they are distributed within the waveform.

The histogram in Figure 9 corresponds to P_{GB}=2% and k=0.05 and presents the real effects of applying the model on the waveforms. In the x-axis the number of consecutive lost packets is depicted while the y-axis represents the total number of the corresponding error bursts that occurred after the application of the model (the average number from all repetitions). For the specific configuration we encountered 228 lost triplets and 137 lost quadruplets. The total number of packets in the test set is 79190.

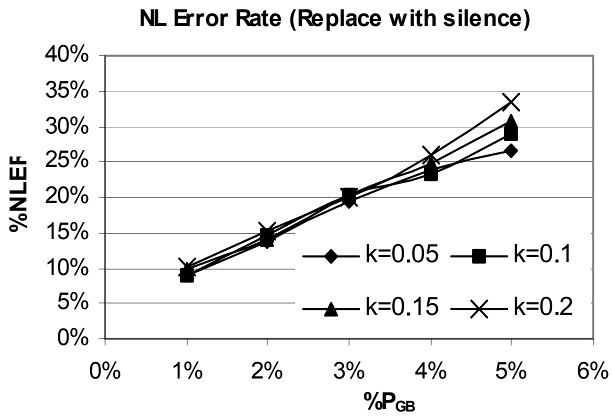

Figure. 6. NL Error Rate for the first ASR strategy

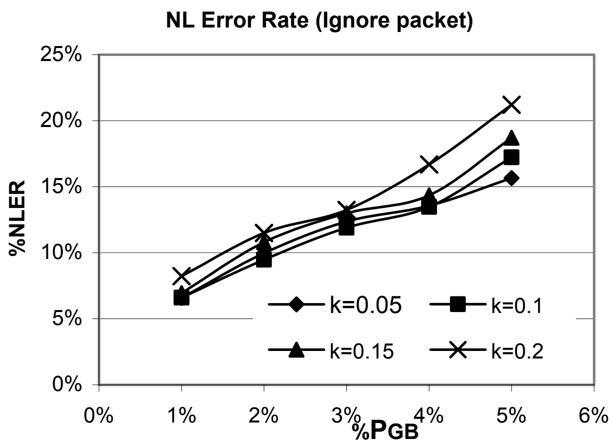

Figure 7. NL Error Rate for the second ASR strategy

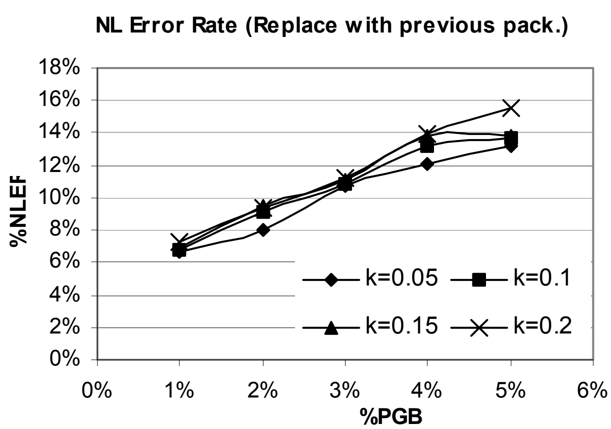

Figure 8. NL Error Rate for the third ASR strategy

5. Conclusions

In this paper we presented the effects of packet loss for ASR applications. As we can see from the presented analysis we come up with a graceful degradation of the performance as the packet loss ratio increases.

Comparing the results from the two models, we can conclude that the Gilbert-Elliot model definitely offers smaller NL error rates for the same packet loss probability. For example for P_B=P_{GB}=5% and using the strategy of replacing the lost packet with the previous one, we yield to NLER between 5%-10% for the GE model and 14%-16% for the three-state Markov model.

One can utilize specific models that better simulate existent data networks (GPRS, 3G etc). Ideally, the analysis should be performed on real voiceprints collected in the ASR server, after the transmission over the data network. The analysis would therefore provide an objective measurement of the expectations for a new speech recognition deployment.

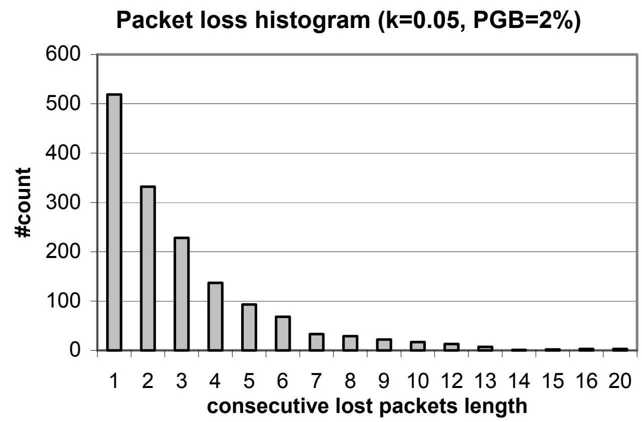

Figure 9. Packet loss histogram (k=0.05, P_{GB}=2%)

6. ACKNOWLEDGMENTS

The corpus was provided by Dialogos Speech Communications.

7. REFERENCES

[1] Almudena Konrad et al, “A Markov-Based Channel Model Algorithm for Wireless Networks”, Journal on Wireless Networks, Vol. 9, No. 3, pp. 189-199, May 2003.

[2] Berger J. M., Mandelbrot B. A New Model for Error Clustering in Telephone Circuits. IBM J R&D July 1963.

[3] Blank H. A, Trafton P. J., A Markov Error Channel Model, Proc Nat Telecomm Conference 1973.

[4] Cain J. B., Simpson R. S., The Distribution of Burst Lengths on a Gilbert Channel, IEEE Trans IT-15 Sept 1969.

[5] E. N. Gilbert, “Capacity of a burst-noise channel”, Bell Syst. Tech. J., Vol. 39, pp. 1253-1265, September 1960.

[6] E. O. Elliot, “Estimates of error rates for codes on burst-noise channels,” Bell Syst. Tech. J., Vol. 42, pp. 1977-1997, September 1963.

[7] Lewis P, Cox D., A Statistical Analysis of Telephone Circuit Error Data. IEEE Trans COM-14 1966.

[8] M. Bottigliengo et al, “Short-term Fairness for TCP Flows in 802.11b WLANs”, Proceedings of the IEEE Infocom 2004, Hong Kong, March 7-11, 2004.

[9] Mertz P., Statistics of Hyperbolic Error Distributions in Data Transmission, IRE Trans CS-9, Dec 1961.