ABSTRACT

In packetized voice systems, a possible overload scheme is to decrease the bit rate when the traffic load increases past a threshold. This paper continues the research into such systems with particular emphasis on the interaction between the traffic performance and voice quality. The adaptive bit rate strategy can be an effective overload control but at the possible expense of voice quality. The object of the paper is to see the effect of traffic-sensitive bit rates upon voice quality. The analysis uses a spectral approach.

1. INTRODUCTION AND OVERVIEW

In packetized voice systems, a possible overload scheme is to decrease the bit rate when the traffic load increases past a threshold. Such schemes are considered, e.g., in [1] and [2] and the references therein. This paper continues the research into such systems with particular emphasis on the interaction between the traffic performance and voice quality. The adaptive bit rate strategy can be an effective overload control but at the possible expense of voice quality. The object of the paper is to see the effect of traffic-sensitive bit rates upon voice quality.

The problem considered here is the interaction between the queueing system/overload control strategy and the coding algorithm and speech perception. This paper ties these disparate phenomena together in an overall model. The analysis uses a spectral approach to analyze the effect of the queueing system/overload control upon voice quality.

The rationale for choice of voice quality measure is given in Section 2. Most importantly, it must quantitatively reflect subjective quality tests. A discussion of the relation between subjective quality tests and objective quality measures is given. Secondarily, but crucial to the present analysis, is that the chosen measure be analyzable in terms of the queuing model.

The methodology for analyzing the effect of reduced bit rates is discussed in Section 3. We combine a queueing analysis with consideration of a voice quality measure to characterize the voice quality degradation. A state dependent queueing model reflects the traffic sensitive bit rates. Then the queueing model is combined with the voice quality measure by using their spectral characteristics as a common denominator. In this way both the traffic performance and voice quality are jointly analyzed.

Multiple bit dropping levels are considered in Section 4. Concluding remarks are given in Section 5.

2. VOICE QUALITY

As mentioned, a modeling approach is being presented to relate voice quality to traffic load. The modeling is intended to be consistent with subjective quality test results and be able to extrapolate beyond the specific parameters of those tests.

We combine a queueing analysis with consideration of a distortion measure to characterize the voice quality degradation. We use a scaled $L_2$ norm of the difference of the log spectra as the distance between two speech segments $S_1$ and $S_2$:

$$d(S_1, S_2) = k \left\| \log(F_1) - \log(F_2) \right\|_2$$

$$= k \left\{ \frac{1}{W} \int_0^w \left[ \log F_1(\omega) - \log F_2(\omega) \right]^2 d\omega \right\}^k$$

(2.1)

where $F$ is the spectrum of $S$, and $w$ is a maximum frequency of interest. The parameter $k$ is a normalizing factor to be discussed. There are a number of other distortion measures (some better than others for specific applications; see [3], [4], [5]). The approach used is not strictly the use of an objective measure, as opposed to a subjective
measure, so that there is a degree of robustness in the choice of distortion measure.

A good discussion of some of the characteristics of objective and subjective measures is given in [6]. The dilemma of using subjective testing for system design, briefly stated, is as follows:

1. Subjective quality testing is necessary for determining the perceptual effects of new modes of voice degradation.

2. Subjective quality testing is time consuming and does not lead naturally to the functional relationships needed for system design and optimization.

3. Objective measures, appropriately chosen, are well suited to system design but cannot be generally expected to track subjective measures well over a wide variety of conditions.

The solution proposed here is to use an objective measure but one which is anchored to subjective results. An objective measure is chosen which available evidence indicates has a reasonable relationship to subjective results. The connection with subjective results is strengthened via incorporation of those results into the objective measure. Specifically, the distortion measure for variable bit rates is normalized using the parameter \( k \) by results for steady bit rates so that it replicates those results when the rate is constant. That is, the norm is scaled so that it equals the difference in mean opinion scores* corresponding to steady bit rates \( r_2 \) and \( r_1 \). Thus, the model is anchored to known results.

We distinguish this incorporation of subjective results from two other procedures:

1. Tuning up of a given objective measure to improve correlation with subjective results. This may be appropriate in some cases but our approach uses subjective results as a more integral ingredient of the objective measure.

2. Taking all the existing subjective results and forcing an objective measure to replicate those results. This is little more than curve fitting and would not yield much predictive power. We use the minimum amount of subjective results so that the resulting objective measure can predict the rest (and thus be validated as a predictor for new conditions).

3. EFFECT OF BIT DROPPING ON THE SPECTRA VIA A QUEUEING MODEL

To illustrate the approach, consider the following state dependent model. When the queue size \( q \) (not including the packet in service) reaches the threshold \( Q \), the bit rate drops from \( r_1 \) to \( r_2 \). When \( q \) drops below \( Q \), the \( r_1 \) rate is resumed. Note that in this first illustrative model, only one lower bit rate is used (as opposed to multiple thresholds) and no hysteresis is used on the thresholding. We refer to continual time periods when \( q \geq Q \) and \( q < Q \) as brisk and slack periods, respectively. The analysis assumes that the brisk and slack periods are not too long, in particular, not so long that the effect of switching between them is possibly worse than a steady \( r_2 \) rate.

During the brisk periods there is a lowered bit rate while during the slack periods there is considered to be essentially no degradation. That is, the \( r_1 \) rate is used as a benchmark. A simple way to see the effect of the bit dropping on the noise is to consider it as a random switch. During the brisk period, additional quantization noise is let through a switch. Thus the noise is multiplied by a random function which equals zero for a random time with mean \( t_b \) and which is one for a random time with mean \( t_s \). \( t_b \) and \( t_s \) are the mean slack and brisk periods, respectively. Then, since this 0–1 random function is (to a good approximation) independent of the noise, the effect is to multiply the noise autocorrelation function by its autocorrelation function.† The composite power spectrum is then determined by Fourier transform. Then the spectrum of interest is

\[
L(\omega) = \Gamma[R(\tau)N(\tau)]|H(\omega)|^2
\]  

(3.1)

where \( \Gamma \) means Fourier transform, \( R(\tau) \) is the autocorrelation of the random switch, \( N(\tau) \) is the autocorrelation function of the lowered bit rate noise, and \( H(\omega) \) is an auditory transfer function. The distortion measure becomes

\[
d = k \left\{ \frac{1}{\pi} \int_0^\pi \log(\{S(\omega)+\Gamma[R(\tau)N(\tau)]\}|H(\omega)|^2) \right. \\
\left. - \log(S(\omega)|H(\omega)|^2)^\frac{1}{2} d\omega \right\}^a
\]  

(3.2)

* The mean opinion score is probably the simplest and most commonly used subjective voice quality measure. See Appendices E and F of [7] for more discussion and further references.

† We are investigating the effect of multi-users on a single user.
where $S(\omega)$ is an averaged or long term speech spectrum. The last approximation, using $\ln(1+x) \approx x$ (for relatively low noise) and renormalizing, shows most clearly how the bit dropping effect enters. This approximation can get crude for large bit rate reductions.

3.1 Brisk and Slack Periods

Reference 8 discusses the calculation of $t_b$ and $t_s$ for state dependent service and arrival rates. Such a model can be used with a higher service rate when in the brisk period. This assumes that bit dropping is done on packets served during the brisk period. The case of dropping bits on packets arriving during the brisk period can be modelled (approximately) with a state dependent arrival rate. The latter case is not as quick to react in overloads. See [9] and the references therein for queueing models of superpositions of packet voice streams (Section II of [9] describes several models).

3.2 Autocorrelation Function

To gain some insight, consider modelling the brisk and slack periods as independent exponential random variables with means determined from the queueing analysis. (The exponentiality is an approximation). The autocorrelation function of the 0-1 function is

$$R(\tau) = \frac{t_s}{(t_s + t_e)^2} \left[ t_s + t_e e^{-0 \cdot \tau / t_s - \tau} \right]$$

which

$$R(\tau) = f^2 + f(1-f) e^{-\alpha \tau}$$

where

$$f = \frac{t_s}{t_s + t_e}$$

is the fraction of time at the reduced bit rate.

We then get

$$\Gamma[R(\tau) N(\tau)] = f^2 \eta(\omega) + f(1-f) \eta(\omega + (t_e t_s)^{-1})$$

for insertion into (3.2), where $\eta$ is the spectrum of the quantization noise $N$.

Equation (3.5) shows how the variable bit rate effect is affected by the time scale of the switching. In particular, for fixed $f$, $R(\tau)$ monotonically increases with $t_s$ and

$$\lim_{t_s \to 0} R(\tau) = f^2.$$  

The result for very small $t_s$ is physically intuitive. For very small $t_s$ and $t_e$ (we are keeping $f$ fixed in this discussion), the sampling of the noise is so fast that it is like multiplying the noise amplitude by a factor of $f$ or the spectrum by $f^2$.

4. ANALYSIS OF MULTIPLE LOWER BIT RATES

We shall extend the model to dropping the bit rate to two lower levels, from $r_1$ to $r_2$ and $r_3$. It will be clear how to generalize the model to any number of levels. While we are doing this, we will also generalize the queueing model to any model for which we can conveniently obtain transition probabilities between states.

Let $x$ be the queue size (now including the packet in service) and define the following:

- Slack period = time continually in states $(0, X_1, X_2)$
- Brisk period = time continually in states $(X_1, X_2)$
- Very brisk periods = time continually in states $(X_2, \infty)$

Thus, $X_1$ and $X_2$ are the thresholds for dropping the bit rate to $r_2$ and $r_3$, respectively.

Define a continuous time Markov process with states 0, 1, and 2 corresponding to the slack, brisk, and very brisk periods, respectively. Let $p_i$ be the stationary probability* that $x(t) = i$ and let $p_{ij}$ be the transition probability

$$p_{ij}(\tau) = P[x(t+\tau) = i | x(t) = j]$$

Then, the autocorrelation function needed for the analysis is

$$R(\tau) = \sum_{i=0}^{2} \sum_{j=0}^{2} p_i p_j (\tau) \left[ \sum_{j=1}^{2} \sum_{i=1}^{2} p_{ij} (\tau) \right] + \alpha \sum_{i=1}^{2} \sum_{j=1}^{2} p_i p_j (\tau) \left[ \sum_{j=2}^{2} \sum_{i=2}^{2} p_{ij} (\tau) \right]$$

* Satisfaction of conditions for existence of these probabilities is implicitly assumed.
where $a \geq 2$ represents the nonlinear effect of reducing bit rates (the mean opinion score decreases faster than linearly as a function of decreased bit rate; see [10]).

For any queueing system for which the above stationary and transition probabilities are available (calculable), the relevant autocorrelation function can be calculated.

Remark. This assumes that the noise spectra have the same shape for the $r_2$ and $r_2$ bit rates. This is an approximation since it is known that as the number of bits increases, the noise becomes more white ([11], p. 185). A simple approximation for multiple bit dropping levels, which can use more than one noise spectrum, is to superimpose the effect of single levels.

5. CONCLUDING REMARKS

Preliminary results using this approach to predict subjective quality tests of mean opinion scores showed reasonably good results. The general approach allows use of different distortion measures and queueing models. For example, some applications might use frequency weighting with the $L_2$ measure. The general approach may also be applicable to some digital speech interpolation (DSI) systems not using packet format. Section 6.4.4 of [7] gives a brief discussion of DSI with references, as well as additional references on varying the bit rate.

ACKNOWLEDGEMENT

I thank M. R. Aaron and N. S. Jayant for their comments.

REFERENCES


