Analysis of a TASI System Employing Speech Storage

MICHAEL S. NAKHLA, MEMBER, IEEE, AND DAVID H. A. BLACK

***Abstract*—This paper presents a performance analysis of a PLC-1 private line voice concentrator which uses speech interpolation to increase the capacity of transmission facilities. The PLC-1 employs speech storage. As a result it can be applied to relatively small trunk groups where statistics of loading patterns are particularly unfavorable. Speech impairments can be categorized as delay, gap modulation, and clipping. A mathematical model based on queueing theory is presented for the evaluation of the statistics of these impairments. The model can be used to determine obtainable TASI advantage for various system configurations and loading conditions. It is shown that a TASI advantage of two is achievable with 24 transmission facilities. The veracity of the queueing model is established using simulation techniques.**

I. INTRODUCTION

THE PLC-1 private line voice concentrator utilizes the principle of speech interpolation to increase the capacity of analog transmission facilities. Its primary application is to long haul private line trunks between dispersed business locations. Cost savings resulting from the reduction in private line leasing costs will generally pay for the PLC-1 terminals in one or two years. A more complete description is given in a companion paper [1].

The increased capacity is obtained by eliminating transmission of the silent intervals normally present in conversational speech. The presence of valid signals is detected by a signal classifier and the associated trunk is connected (or assigned) to the first free transmission facility. At the distant receiver the facility is routed to the appropriate trunk for the duration of the speech burst. Routing information must therefore be transmitted to the distant receiver.

In the PLC-1, one transmission facility is reserved for this signaling function. This "signaling facility" is heavily protected by means of cyclic code error detection and a variety of recovery schemes. Failure of the signaling facility results in a rapid switchover to an alternative facility.

The signaling facility transmits data at 2.4 kbits/s and carries all speech burst assignments, signaling and other types of data. The limited capacity of the signaling facility frequently results in queueing of assignment requests. There are, therefore, two queueing processes associated with the transmission of each burst.

1) Transmission facility queue.
2) Signaling facility queue.

TASI systems commonly clip the initial portion of a speech burst when a facility or assignment message is not immediately available. In the PLC-1 speech bursts are stored in a buffer memory until both a transmission facility and an assignment slot are available. In this way, a subjectively preferable variable delay is substituted for a clip. This paper analyzes the delay variations for a range of equipment configurations and recommends the concentration ratio for acceptable performance.

Manuscript received June 26, 1981. This paper was presented at the International Conference on Communications, Denver, CO, June 1981.

The authors are with Bell-Northern Research, Ottawa, Ont., Canada K1Y 4H7.

II. SIGNALING FACILITY

A standard 2.4 kbit/s DPSK modem similar in performance to a 201B is used to transmit full duplex data over the four wire signaling facility. Several different types of data exist including

1) speech burst assignments;
2) signaling;
3) maintenance commands;
4) analytical reports.

Data are organized in 20 ms frames. Framing bits are included to allow the receiver to synchronize to the data stream. Each frame may contain one of the above types of data, speech burst assignments (SBA's) being the most common.

When a speech burst arrives, an active interrupt is sent to the assignment controller and simultaneously the speech burst is written into a buffer memory. It is stored in the memory until

1) a free transmission facility is found;
2) a free frame in the signaling facility is found;
3) the SBA has been transmitted.

Subsequently, the speech burst is read out on the transmission facility without loss. It has, however, experienced a delay, which could range from as little as 40 ms to hundreds of milliseconds. Subjective tests have confirmed the tolerance of the ear to the variable delays introduced by the system, provided they remain within a limited range. For the PLC-1 the requirement has been set that the variation in delay, or "gap modulation," shall have a probability of exceeding 300 ms for less than 10 percent of the time.

An upper limit of 1 s storage for any speech burst is placed by the software, to ensure that the buffer memory capacity is allocated fairly. After one second of storage, a speech burst will be clipped. A further criterion, therefore, is that clips due to this, or other processes, should have less than 2 percent probability of exceeding 50 ms.

The subsequent analysis establishes the number of facilities required to meet the two criteria over a range of input trunks.

III. PERFORMANCE ANALYSIS

In this section two models are described for the analysis of speech delay statistics, a mathematical model based on queueing theory and a simulation model. The queueing model neglects the situation where more than one speech burst

0090-6778/82/0400-0780$00.75 © 1982 IEEE

originated by the same trunk may be stored simultaneously. The simulation assumes the same traffic model but does not exclude multiple speech burst queueing. The results verified that the queueing model is sufficiently accurate to determine obtainable TASI advantage for various system configurations and loading conditions.

A. Queueing Model

Each of the "off-hook" trunks alternates randomly between intervals of talk spurt and pause with density function in the form [2]-[4]

$$P_s(t) = \frac{1}{T_s} e^{-(t/T_s)} \quad (1)$$

$$P_p(t) = \frac{1}{T_p} e^{-(t/T_p)} \quad (2)$$

where T_s and T_p are the mean speech and pause intervals, respectively. The speech activity factor defined as

$$\alpha = \frac{T_s}{T_s + T_p} \quad (3)$$

is typically in the range $0.35 \leqslant \alpha \leqslant 0.45$.

Speech bursts in the buffer are stored according to their order of arrival. The queue discipline is first-in first-out. Therefore, when a facility is available, the speech burst at the head of the queue is the one to be serviced. If the buffer is empty, the available facility becomes idle until a speech burst appears on one of the trunks. At any given time, a speech burst is not allowed to occupy more than one second of storage in the buffer. If at the time of a speech burst arrival, no facilities or buffers are available, a freeze-out condition is encountered where the speech is lost until storage in the buffer is available. The following analysis considers the peak periods where all the trunks are assumed to be off-hook and capable of demanding service.

The combination of the trunks/facilities/buffer can be viewed as a birth-and-death process where the customers are the speech bursts originated by the N trunks and the servers are the M facilities ($M \leqslant N$). In order to make the analysis tractable, we will assume that the number of customers in the system at time t equals the number of trunks associated with speech bursts being served or waiting for service at time t. In other words, we neglect the probability that a trunk demands service while a speech burst belonging to the same trunk is still in the system. The effect of this assumption on the delay statistics will be quantified using the simulation model.

Using (1) and (2), the birth and death rates can be written in the form

$$\lambda_K = \frac{N-K}{T_p} \qquad 0 \leqslant K \leqslant N-1 \quad (4)$$

$$\mu_K = \frac{K}{T_s} \qquad 1 \leqslant K \leqslant M \quad (5a)$$

$$= \frac{M}{T_s} \qquad M \leqslant K \leqslant N. \quad (5b)$$

The statistical equilibrium state probabilities $P_K[N]$, that is, the probability there are K speech bursts in the system being served or awaiting service, can be calculated as follows [5]:

$$P_K[N] = \binom{N}{K} \gamma^K P_0[N] \qquad 0 \leqslant K \leqslant M \quad (6a)$$

$$= \frac{N!}{(N-K)!M!M^{K-M}} \gamma^K P_0[N] \quad M \leqslant K \leqslant N \quad (6b)$$

where

$$\gamma = \frac{\alpha}{1-\alpha}$$

and

$$P_0^{-1}[N] = \sum_{K=0}^{M} \binom{N}{K} \gamma^K + \sum_{K=M+1}^{N} \frac{N!\gamma^K}{(N-K)!M!M^{K-M}}. \quad (6c)$$

The arriving distribution function $\Pi_K[N]$, probability that an arriving speech burst finds K speech bursts in the system, is given by

$$\Pi_K[N] = P_K[N-1].$$

For a system with M facilities, the probability that an arriving speech burst requires a discretionary buffer P_{buf} is the probability that the speech burst finds the system in state $E_K \geqslant M$, i.e.,

$$P_{\text{buf}} = \sum_{K=M}^{N-1} P_K[N-1]. \quad (7)$$

Using (6b)

$$P_{\text{buf}} = \frac{(N-1)!}{M!} P_0[N-1] \sum_{j=0}^{N-M-1} \frac{\gamma^{j+M}}{(N-1-j-M)!M^j}. \quad (8)$$

A plot of P_{buf} versus the speech activity factor α is shown in Fig. 1. It should be noted that in defining the system configuration one of the facilities is reserved for signaling.

The probability that the waiting time in the buffer exceeds t can be written as

$$P[D \geqslant t] = \sum_{K=M}^{N-1} P[D \geqslant t \mid E_K] \Pi_K[N]$$

$$= \sum_{K=0}^{N-M-1} P[D \geqslant t \mid E_{K+M}] \Pi_{K+M}[N] \quad (9)$$

where $P[D \geq t \mid E_K]$ is the conditional probability that a speech burst is delayed more than t given that at the time of its arrival it finds the system in state E_K.

For a birth-and-death process with exponentially distributed service time and order-of-arrival service, it can be shown that [5]

$$P[D \geq t \mid E_{K+M}] = e^{-M\tau} \sum_{j=0}^{K} \frac{(M\tau)^j}{j!} \tag{10}$$

where $\tau = t/T_s$, and T_s is the mean service time which is equivalent to the mean speech duration.

Using (6), (9), and (10)

$$P[D \geq t] = e^{-M\tau} \frac{(N-1)!}{M!} P_0[N-1] \cdot \sum_{K=0}^{N-M-1} \frac{\gamma^{K+M}}{(N-1-K-M)!M^K} \cdot \sum_{j=0}^{K} \frac{(M\tau)^j}{j!}. \tag{11}$$

A plot of the delay distribution function $P[D \geq t]$ is shown in Fig. 2 for different system configurations.

The gap modulation distribution function is given by

$$P[G > t] = P[W_2 - W_1 > t] \tag{12}$$

where W_1 and $W_2 > W_1$ are the waiting times associated with any two consecutive speech bursts

$$\begin{aligned} P[W_2 - W_1 > t] &= P[W_1 = 0]P[W_2 > t] \\ &+ \lim_{\epsilon \to 0^+} \sum_{\eta=\epsilon}^{\infty} P[W_1 = \eta]P[W_2 > T + \eta] \\ &= \{1 - P[D > 0]\} P[D > t] \\ &+ \int_0^{\infty} P[D > \eta + t] \left. \frac{dP[D > x]}{dx} \right|_{x=\eta} d\eta \end{aligned} \tag{13}$$

where $P[D \geq t]$ can be calculated using (11).

A comparison between the delay and gap modulation distribution functions is shown in Fig. 3.

B. Simulation and Comparisons

To check the accuracy of the queueing model, a simulation package has been developed. The input parameters to the simulator are

1) number of trunks;
2) number of facilities;
3) speech activity factor;
4) mean speech duration;
5) number of simulated speech-pause intervals. These intervals are generated according to the random distribution given in (1) and (2).

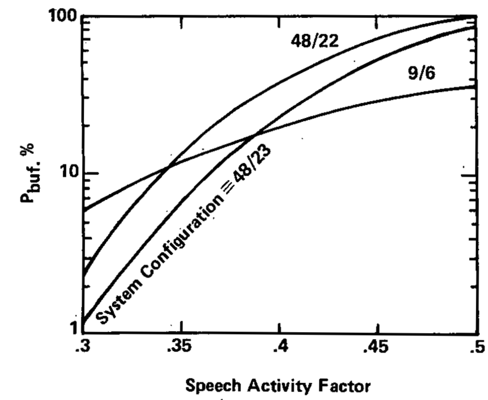

Fig. 1. Probability of buffer usage.

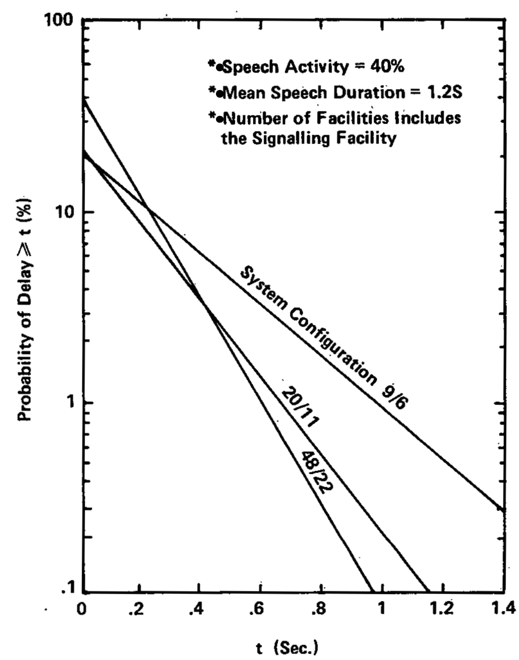

Fig. 2. Delay probability distribution function.

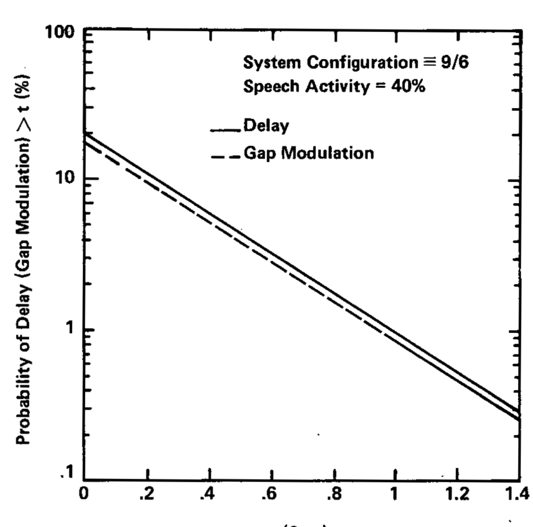

Fig. 3. Comparison between delay and gap modulation distribution functions.

The model is monitored at the following discrete time events associated with each speech burst:

1) arrival to the system;
2) buffer entry;
3) buffer depletion;
4) service completion.

The simulation runs involve measurements of the holding time in the buffer and the variation in the pause interval between two consecutive speech bursts originated by the same channel. To ensure that the chosen number of speech-pause intervals is sufficiently high and at the same time increase the accuracy of the results, the simulation experiment is repeated several times (defined by the program's user) and the cumulative average statistics are computed.

Fig. 4 shows a comparison between the delay statistics obtained using the queueing model and the simulated results. In the simulation experiment 300 speech-pause intervals/trunk were randomly generated. The mean speech duration was assumed to be 1.2 s and the speech activity factor equal to 40 percent. The statistical distribution shown is the average of ten independent simulation runs. The slight discrepancy between the results of the two models is primarily due to the fact that in deriving (11), we have neglected the situation where a trunk demands service while a speech burst originated by the same trunk is still in the system (being serviced or waiting in the buffer). Of course, this is an optimistic assumption which tends to lower the delay probabilities as can be verified by Fig. 4.

The developed models can also be used to determine the number of facilities which are required to meet a given performance criterion. As an example Fig. 5 specifies the minimum number of facilities required to ensure that the probability of a gap modulation greater than ±300 ms is less than 10 percent. Three cases are considered corresponding to the conditions where 0, 5, and 10 percent of trunks contain continuous data.

The foregoing comparison between the statistics obtained using the queueing model and the simulation results establishes the veracity of the queueing model. This promotes its use for large systems where simulation runs require excessive CPU time or in cases where impairment probability is relatively small which makes it impractical and inaccurate to find the statistical distribution via simulation.

C. Freezeout

Freezeout or front-end clipping can arise in two ways.

1) No speech burst is permitted more than T_{max} seconds storage. When this is exceeded the buffer is overwritten and front-end clipping results.

2) When the buffer memory is fully utilized, a channel becoming active cannot be served and is frozen out until memory becomes available.

The probability distribution of the first type of freezeout can be calculated as follows.

Let T be a random variable equal to the freezeout time. Then $T > t$ if

1) the length of the observed speech burst $> t + T_{max}$;
2) the observed speech burst is delayed in the buffer

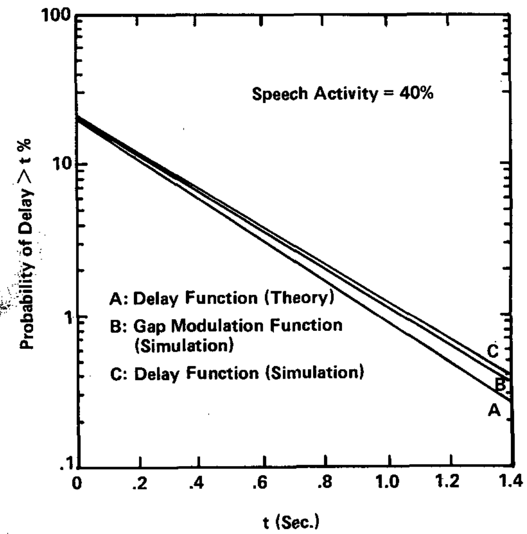

Fig. 4. Delay and gap modulation distribution functions.

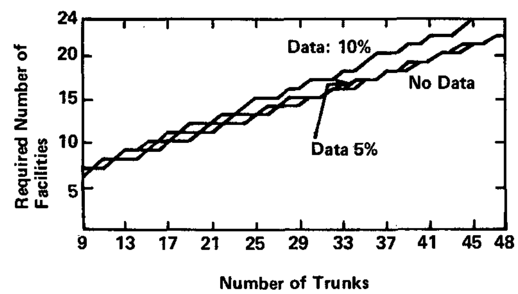

Fig. 5. Number of facilities required to ensure that the probability of a gap modulation greater than ±300 ms is less than 10 percent.

more than $t + T_{max}$. Therefore,

$$P[\text{freezeout} > t] = e^{-(T_{max}+t)/T_s} P[D > t + T_{max}] \quad (14)$$

where $P[D > t + T_{max}]$ is given by (11).

From (14), it is obvious that the probability of this type of freezeout is inversely proportional to the system size. Theoretical and simulated results for the 9/6 system are shown in Fig. 6.

The probability distribution of the second type of freezeout is difficult to calculate for two main reasons.

1) The buffer memory is dynamically allocated according to the delayed speech bursts' durations and waiting times. Conceptually, this represents a queueing system with nonstationary number of servers.

2) As will be shown the freezeout probability is relatively small, which makes it impractical to find the statistical distribution via simulation.

An alternative solution is to calculate an upper bound for the freezeout probability where every speech burst in the buffer is assumed to utilize the maximum allowable number of packets. This bound turns out to be low enough that we will not be concerned with the actual distribution.

Let S be the total number of packets in the buffer memory and S_{max} be the maximum number of packets a single speech burst is permitted to utilize. An observed speech burst may be clipped if it finds the system in state E_K where $K \geqslant [M + S/S_{max} - 1]$ and M is the number of facilities. However, if

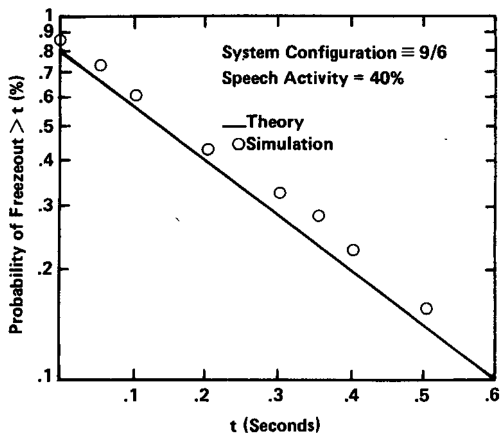

Fig. 6. Probability distribution function of front end clipping.

$K \leqslant [M + S/S_{\max} - 1]$ then the probability that the arriving speech burst will be frozen (due to buffer capacity limitation) is zero, i.e.,

$$P[\text{freezeout} \geqslant 0] \leqslant \sum_{K=M+S/S_{\max}-1}^{N-1} \pi_K[N] = \sum_{K=M+S/S_{\max}-1}^{N-1} P_K[N-1] = \frac{(N-1)!}{M!} P_0[N-1] \cdot \sum_{K=N+S/S_{\max}-1}^{N-1} \frac{\gamma^K}{M^{K-M}(N-1-K)!} \tag{15}$$

where $\gamma = (\alpha/1 - \alpha)$ and $P_0[N-1]$ is given by (6c).

For example, if $N = 47, M = 24, S = 256$, and $S_{\max} = 32$, then

$$P[\text{freezeout} \geqslant 0] < 0.053 \text{ percent.}$$

D. Speech Assignment Delay

A single speech burst assignment update takes one frame and consists of a trunk and a facility word. If more than one speech burst simultaneously request an assignment frame, the additional bursts are delayed. This assignment delay adds to the total speech delay.

$$D_{\text{total}} = D_{\text{fixed}} + D_F + D_A$$

where

D_{fixed} fixed processing delay

D_F facility search time

D_A assignment word search time.

$$P[D_{\text{total}} > t] = \int_0^\infty P[D_A > t - \eta - D_{\text{fixed}}] \cdot \frac{dP[D_F > \tau]}{d\tau}\bigg|_{\tau=\eta} d\eta \tag{16}$$

where the probability distribution function $P[D_F > \cdot]$ is given by (11). In the following, a queueing model is described to calculate the assignment delay distribution function $P[D_A > \cdot]$.

The speech burst assignment can be presented by a single server queueing model with the following average birth and death rates:

$$\lambda_K = \frac{M-K}{T_S} \qquad 0 \leqslant K \leqslant M-1 \tag{17a}$$

$$\mu_K = \frac{1}{T_f} \qquad 1 \leqslant K \leqslant M \tag{17b}$$

where M is the number of facilities.

The probability that the assignment delay exceeds t can be written as

$$P[D_A > t] = \sum_{K=0}^{M-1} P[D_A > t \mid E_K] \Pi_K[M] \tag{18}$$

where $\Pi_K[M]$ is the arriving customer's distribution function

$$\Pi_K[M] = \frac{\gamma^K/(M-1-K)!}{\sum_{j=0}^{N-1} \gamma^j/(M-1-j)!}$$

$$\gamma = T_f/T_S \qquad 0 \leqslant K \leqslant M-1. \tag{19}$$

If the system is in state E_K, then the waiting time of an arriving customer is given by

$$W_K = A \qquad K = 0$$
$$= A + T_f(K-1) \qquad K > 0$$

where A is a random variable uniformly distributed between 0 and T_f. Therefore, the conditional probability $P[D_A \geqslant t \mid E_K]$ that a customer waits beyond t when placing a request given that he finds K other customers in the system is

$$P[D_A > t \mid E_{K \neq 0}] = 0 \qquad t < (K-1)T_f$$
$$= K - t/T_f \qquad (K-1)T_f \leqslant t \leqslant KT_f$$
$$= 0 \qquad t \geqslant KT_f. \tag{20}$$

If $K = 0$ then

$$P[D_A > t \mid E_0] = 1 - t/T_f \qquad 0 \leqslant t \leqslant T_f$$
$$= 0 \qquad t \geqslant T_f.$$

The probability of the total delay is calculated using (16) and substituting from (11) and (18)-(20).

Fig. 7 shows the effect of the system size on the 10 percent probability point of the delay distribution function (value of t such that $P[D_{\text{total}} \geqslant t] = 10$ percent). As expected D_F dominates for small systems (where overload is more frequent) and D_A becomes increasingly important as system size increases.

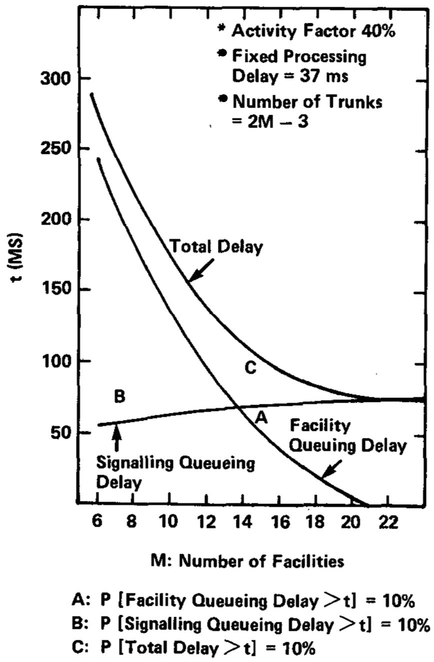

Fig. 7. Statistics of speech delay.

IV. DISCUSSION

The analysis shows that freezeout probability is so low as to cause imperceptible speech impairments. The dominant effects therefore are due to the delays introduced by the speech buffer.

The variable delays result in "gap modulation," i.e., the interval between successive speech bursts is modified. It becomes important to more closely consider the make-up of a speech burst. This is largely determined by the sensitivity and hangover of the speech detector. To minimize gap modulation effects it is essential to use a relatively long hangover to bridge short gaps in the speech so that a speech burst consists of entire phrases and sentences. It has been found that the ear is tolerant to relatively large variations in arrival times of such speech bursts, while very noticeable distortion would arise if individual phonemes were modulated. Subjective tests have established that the mean gap modulation should be in the range of one to two times the speech detector hangover to remain imperceptible. Assuming a 200 ms hangover, a reasonable specification is that the gap modulation should not exceed 300 ms more than 10 percent of the time.

Fig. 7 shows that the PLC-1 will meet this specification with the recommended loading rules, e.g., configurations of 48/22, 20/11, and 9/6. Informal listening tests performed on early PLC-1 terminals have confirmed that speech and conversation quality is excellent under these conditions.

V. CONCLUSION

A queueing model has been described for the evaluation of PLC-1 delay, gap modulation and freezeout statistics. The accuracy of the model was verified using simulation techniques. The analysis shows that freezeout is negligible and that the delay and gap modulation resulting from speech storage remain within the specification over a wide range of configurations provided the loading rules are followed. While performance improves with larger systems, entirely acceptable speech quality may be obtained with nine trunks on six facilities. The signaling facility technique employed by the PLC-1 ensures that speech quality is independent of hook state signaling activity. Subsequent analytical efforts will concentrate on further refinement of the speech models.

REFERENCES

[1] D. H. A. Black, "PLC1: A TASI system for small trunk groups," in *Proc. Int. Conf. Commun.*, June 1981.

[2] P. T. Brady, "A technique for investigating on-off patterns of speech," *Bell Syst. Tech. J.*, vol. 54, pp. 1–23, Jan. 1975.

[3] A. C. Norwine and O. J. Murphy, "Characteristic time intervals in telephone conversation," *Bell Syst. Tech. J.*, vol. 17, pp. 281–291, Apr. 1938.

[4] J. T. Wang and M. T. Liu, "Analysis and simulation of the mixed voice/data transmission system for computer communication," in *Proc. Nat. Telecommun. Conf.*, Dallas, TX, Nov. 1976, pp. 42.3–ff.

[5] R. Cooper, *Introduction to Queueing Theory.* New York: Macmillan, 1972.

[6] S. J. Campanella, "Digital speech interpolation techniques," *Comsat Tech. Rev.*, vol. 6, pp. 127–157, 1976.

[7] J. Elder and J. F. O'Neill, "A speech interpolation system for private networks," in *Proc. Nat. Telecommun. Conf.*, 1978, pp. 14.6–1ff.

★

Michael S. Nakhla (S'73–M'75) received the M.A.Sc. and Ph.D. degrees from the University of Waterloo, Waterloo, Ont., Canada, in 1973 and 1975, respectively, both in electrical engineering.

From 1967 to 1971, he worked as Instructor in the Department of Applied Mathematics, Faculty of Engineering, Cairo University, Cairo, Egypt. From 1971 to 1975, he was Research Assistant in the Department of Electrical Engineering, University of Waterloo. During 1975 he was a Postdoctoral Fellow with the Department of Electrical Engineering, University of Toronto, Toronto, Ont., Canada. Since 1976 he has been with Bell-Northern Research, Ottawa, Ont., Canada where he is presently Manager of the IC modeling and design aids group. His research interests include computer-aided circuit analysis and design, communication theory, and fiber optic communication systems.

★

David H. A. Black was born in Belfast, Northern Ireland. He received the Master's degree in electronic engineering from Queen's University, Belfast, in 1966.

In early 1967 he joined Northern Electric R&D (subsequently Bell-Northern Research), Ottawa, Ont., Canada, where he worked on the development of PCM channel banks and transmission equipment. Design responsibilities included PCM codecs, control logic, and power supplies. Since 1976 he has been Manager of Digital Terminal Development, specializing in speech interpolation systems. Currently he is responsible for PLC-1 system and hardware design.