ELECTRO-ACOUSTIC FACTORS DETERMINING TELEPHONE SPEECH QUALITY

by

J. Gruber, J. Bhullar and G. Williams
Bell-Northern Research
P.O. Box 3511, Stn. C
Ottawa, Ontario, Canada K1Y 4H7

ABSTRACT

In this paper the electro-acoustic factors affecting the quality of telephone speech are categorized as: - listening factors such as loudness, noise, frequency response, and listener echo, - talking factors such as talker echo and sidetone, and - conversational factors such as delay. Listening factors in particular are emphasized and qualitatively interpreted in terms of acoustic pressures at the ear relative to the thresholds of hearing and pain. Typical telephone connections are then considered in which a number of the above factors are manifested; this is done in the context of a network simulation facility used by Bell-Northern Research to simulate various telephone network impairments for the purposes of objective and subjective evaluation. Grade-of-service results that relate customer opinion of transmission quality to the level of a transmission parameter are also discussed. These include listener echo and delay, and subjective-equivalence modelling of speech-correlated digital noise in terms of continuous analog noise.

SOMMAIRE

Dans ce document, les facteurs électro-acoustiques exerçant un effet sur la qualité des signaux vocaux téléphoniques transmis sont divisés en trois classes : les facteurs liés à l'écoute, comme la force des sons, le bruit, la réponse en fréquence et l'écho de la personne à l'écoute, les facteurs liés au parler, comme l'écho de la personne qui parle et l'effet local, et enfin les facteurs liés à la conversation comme le délai d'attente. L'auteur met l'accent plus particulièrement sur les facteurs liés à l'écoute ; ces derniers font l'objet d'une interprétation qualitative en termes de pression acoustique imposée à l'oreille par rapport aux seuils de l'audition et de la douleur.

Aux fins d'illustration, l'auteur examine ensuite des communications téléphoniques typiques où un certain nombre des facteurs susmentionnés se manifestent. Pour ce faire, il utilise une installation de simulation de réseau de la Société de recherches Bell-Northern pour simuler différentes réductions de la qualité de transmission pour faire une évaluation objective et subjective. L'auteur traite aussi de résultats quant au niveau du service, en établissant un rapport entre l'opinion du consommateur sur la qualité de la transmission et le niveau d'un paramètre de transmission. Ces résultats comprennent l'écho de la personne à l'écoute, le délai d'attente et l'établissement d'un modèle subjectif et équivalent du bruit de type numérique lié à la conversation en termes de bruit analogue continuel.

1.0 INTRODUCTION

In order to extend the range over which humans can communicate verbally, it is necessary to effectively open the acoustic face-to-face communication path, and bridge the resulting gap by an electrical network (i.e., the telephone network). This gives rise to an electro-acoustic communication path. The electro-acoustic components of the path correspond to the telephone sets, which on the talking side convert acoustic pressures into electrical signals to be sent into the network, and conversely on the listening side convert received electrical signals from the network into acoustic pressures.

One of the concerns of a telephone company is to have a thorough understanding of electro-acoustic factors such as attenuation, noise, distortion, echo, delay and sidetone etc., and the way in which they affect telephone speech quality[1]. This is essential in order that satisfactory service can be provided to customers on an ongoing basis as the network continually evolves as a result of new technology, economic considerations, and the need to provide new services.

It is useful to classify electro-acoustic factors according to the way in which they affect the listening, talking and conversational modes of communication, as indicated in Table 1[2]. Listening factors include continuous noise which is usually associated with the analog communication systems, and noise which is correlated with the speech signal amplitude as in digital systems; results pertaining to the subjective nature of these noise components will be presented later. Another listening factor of significance is frequency response distortion, an example of which is also discussed in terms of listener echo. Listener echo is an electrical reverberation (analogous to acoustic room reverberation) which can arise in (but is not peculiar to) telephone connections employing digital technology, if such connections are not properly engineered.

TABLE 1

CLASSIFICATION OF ELECTRO-ACOUSTIC FACTORS

- LISTENING FACTORS :
  - Loudness Loss
  - Noise
    - Continuous (Analog)
    - Correlated (Digital)
  - Listener Echo
  - Frequency Response
- TALKING FACTORS :
  - Talker Echo
  - Sidetone
- CONVERSATIONAL FACTORS:
  - Delay
  - Asymmetry

A prime example of an electro-acoustic factor affecting the talking mode of communication is talker echo. If a speech signal reflected or echoed from the distant end of the connection has sufficient volume and is sufficiently delayed (due to a long connection), it can have an inhibiting effect on the talker. Thus such echo signals must be carefully controlled.

Finally, the conversational mode of communication involves a dynamic interaction between the two participants. A classical electro-acoustic factor which can

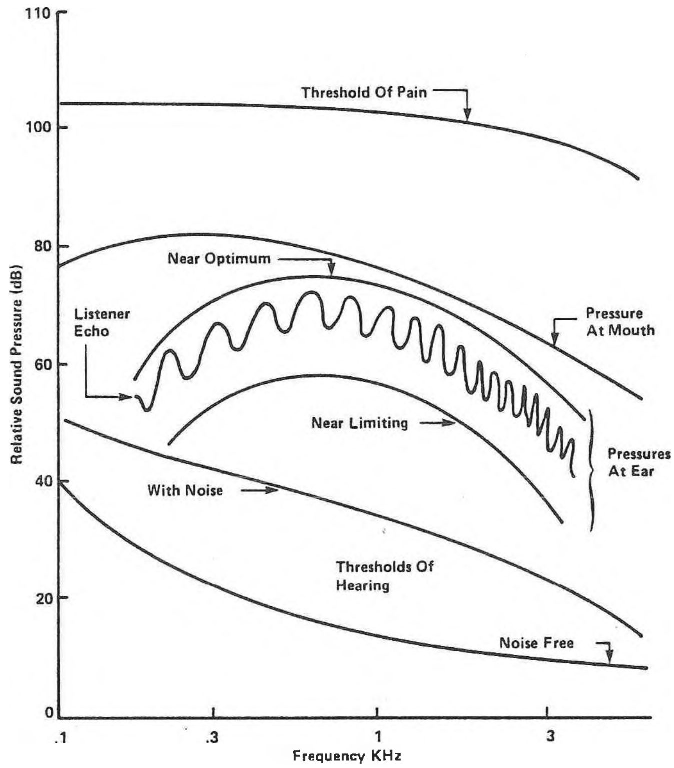


Figure 1 Putting Listening Factors In Context

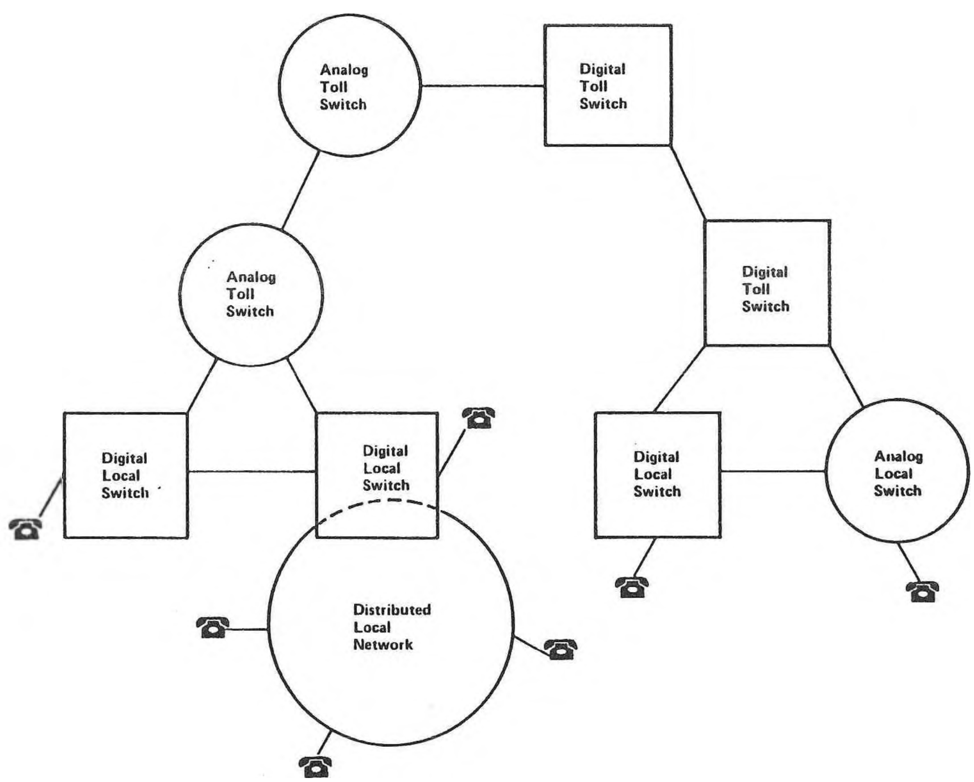


Figure 2 General Mixed Analog/Digital Network

influence this dynamic interaction is significant round trip delay; for example, on 2-hop satellite connections such delays can exceed 1 second.

Although all of the above as well as other electro-acoustic factors are important as far as telephone network performance is concerned, in order to limit the scope of this paper, only certain listening factors will be emphasized in what follows.

It is useful to interpret listening factors in terms of the way in which they are perceived at the listener's ear. This is illustrated in Figure 1 which presents relative sound pressure level at the ear versus frequency[2]. All sound pressures ordinarily fall between the threshold of pain on the high pressure side, and the noise free threshold of hearing on the low pressure side. The effect of noise generally has the undesirable effect of making low level speech signals difficult to hear and thereby effectively raises the threshold of hearing. For this basic reason the telephone network is engineered to control both received speech and noise levels.

Frequency response distortion is also evident from the two smooth convex curves in Figure 1, in that the received sound pressures do not differ by a constant amount from the transmitted pressures at all frequencies. This is largely because of imperfections in the telephone sets and the connections (called local loops) between the sets and the switching offices. The ripple response in Figure 1 is another example of frequency distortion caused by electrical reverberation, or the listener echo phenomenon referred to earlier. The subjective effect is a hollow or "rain barrel" sound, and will be an item of further discussion in this paper.

Thus Figure 1 illustrates that for satisfactory quality or grade-of- service, telephone connections should be engineered such that noise levels are suitably low, speech levels are near optimum, and frequency distortion is not excessive.

2.0 SPECIFIC LISTENING FACTORS AFFECTING LONG DISTANCE TELEPHONE NETWORK PERFORMANCE

2.1 General

Figure 2 is a simplified illustration of the telephone network showing the local and toll levels, as well as the more recent local distributed switching network. This figure indicates that the switching offices may be either analog (with electro-mechanical crosspoints) or digital (with electronic crosspoints). Similarily the connections or trunks joining the offices may be either analog or digital. The current analog/digital stage of network evolution gives rise to some specific electro-acoustic factors affecting telephone network performance, one of which is discussed in what follows.

2.2 The Nature of Analog and Digital Noise

In order to determine the performance of a mixed network, it is necessary to characterize the combination of continuous noise from the analog portion and the speech-correlated noise from the digital portion. A basic difference between these two noise components is that continuous analog noise is perceptible during both speech and silence, whereas correlated digital noise is only perceptible during speech, and only if the digital noise level is high enough.

Basic threshold of perceptibility tests for speech-correlated digital noise have been carried out by Bell-Northern Research and other organizations[2,3].

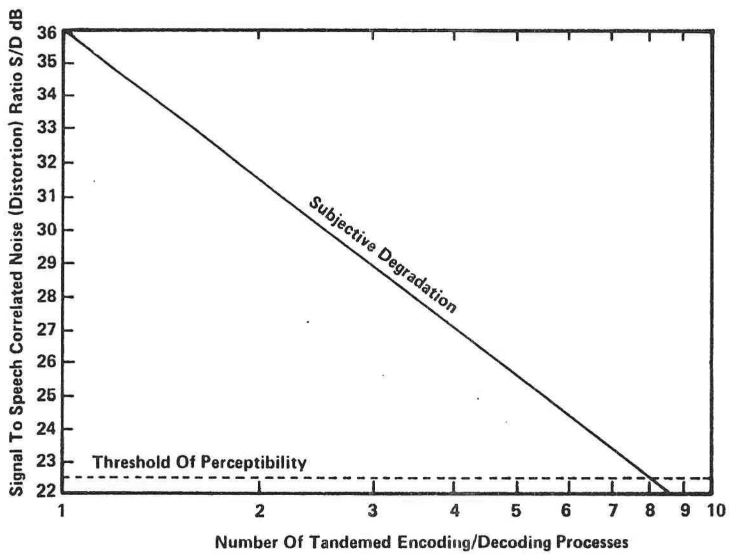


Figure 3 Degradation Of Signal To Speech Correlated Noise Ratio With Number Of Tandemed Encoding/Decoding Processes (For 64kb/s PCM With μ-Law Companding)

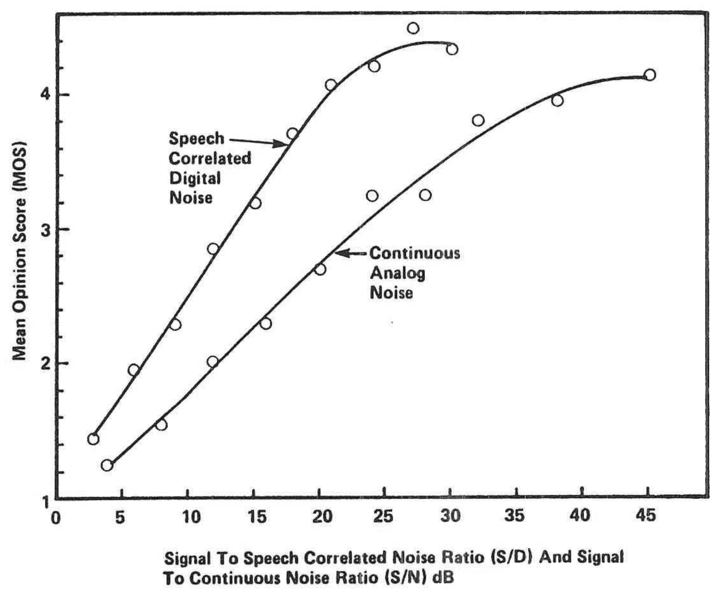


Figure 4 Subjective Test Results For Speech Correlated Digital Noise And Continuous Analog Noise

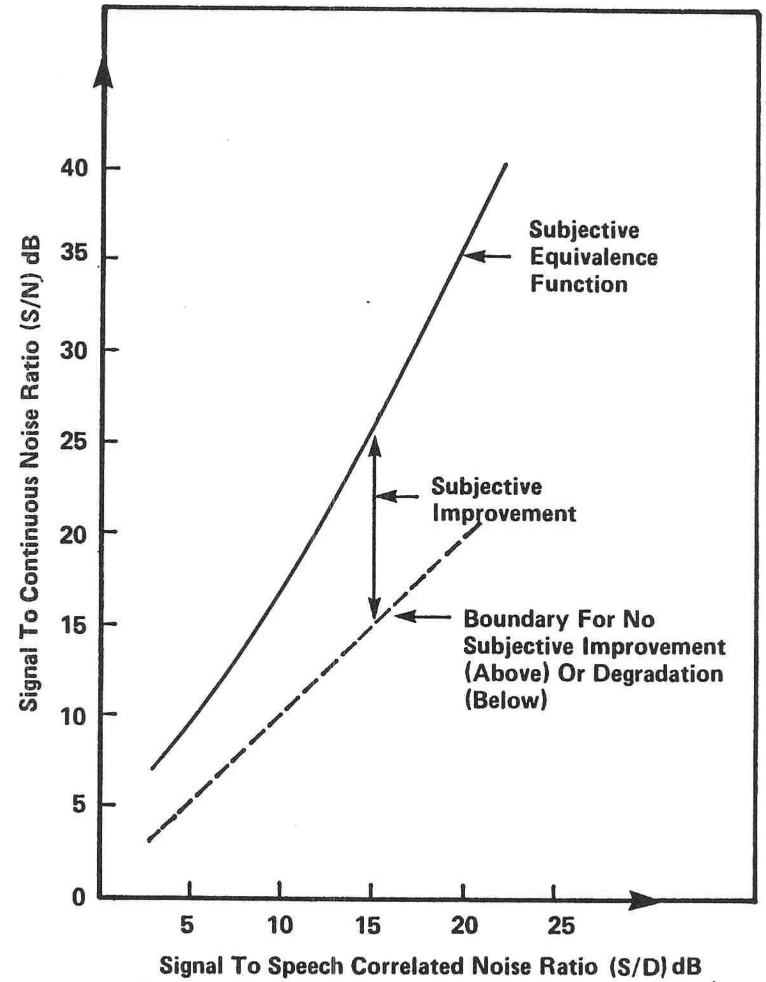


Figure 5 Signal To Continuous Noise Ratios Which Are Subjectively Equivalent To Signal To Speech Correlated Noise Ratios

These tests were conducted with subjects listening to speech plus various levels of speech correlated noise simulated by a Modulated Noise Reference Unit (MNRU)[2]. The results indicate that the perceptible threshold corresponds to a speech signal to distortion (or speech signal to speech-correlated noise) ratio (S/D) of between 21 and 24 dB.

The relevance of this to a mixed network having analog switching and digital transmission for example, is that speech signals are converted from analog to digital (A/D) and from digital to analog (D/A) by terminal equipment called channel banks[5]. In North America the encoding (A/D) and decoding (D/A) process corresponds to 64 kb/s Pulse Code Modulation (PCM) with u-law companding[8]. A single encoding/decoding process of this type produces a S/D ratio of about 36 dB, which is well above the threshold of perceptibility.

A basic question of interest to a telephone network planner is: how many tandem encoding/decoding processes of this nature can be permitted in a connection before the speech correlated digital noise becomes perceptible? Based on extensive studies at Bell Laboratories[4,6], Figure 3 indicates about eight processes, since at that number the perceived S/D ratio has degraded from 36 dB, to the perceptible threshold of about 22 or 23 dB.

Another question of interest is: how do perceptible levels of correlated digital noise compare subjectively with levels of continuous analog noise, and how should the two be combined in a mixed analog/digital network? To assist in answering this question appropriate subjective tests have been carried out by Bell-Northern Research and other organizations[3,4]. In these tests subjects were asked to rate the quality of pre-recorded segments of speech in the presence of various levels of continuous analog noise, and of speech in the presence of various levels of speech-correlated digital noise (again the speech correlated noise was simulated by an MNRU[2]).

The ratings are based on a 5 point scale similar to the one which follows, and can be obtained from the subjects by means of a question such as:

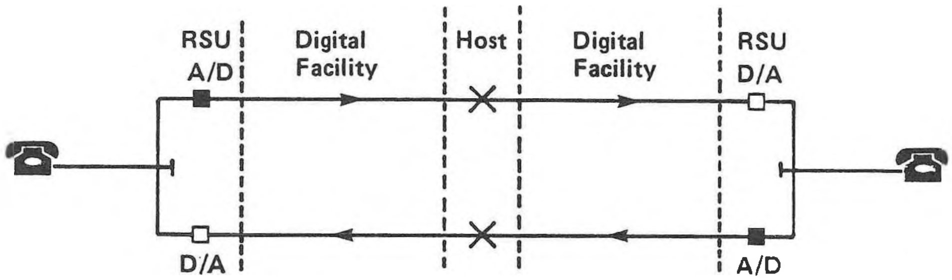
> Which of these five words comes closest to describing the quality of the connection?
>
> - Excellent (E), Good (G), Fair (F), Poor (P), Bad (B).

(The subjective testing experimental arrangement is similar to Figure 8).

Based on these tests the results shown in Figure 4 were obtained which show Mean Opinion Score (MOS) versus speech signal to distortion ratio (S/D) for speech-correlated digital noise, and speech signal to noise ratio (S/N) for continuous analog noise. (Note MOS is a weighted average rating across all test subjects for each test condition; weights of 5, 4, 3, 2 and 1 are assigned to E, G, F, P and B respectively).

Note in Figure 4 that for a given MOS, the S/N ratio for continuous noise must be higher than the S/D ratio for speech-correlated noise, to account for the subjective effect of continuous noise presence regardless of signal level or signal presence. This gives rise to the notion of levels of continuous noise which are subjectively equivalent to levels of speech-correlated noise. This subjective-equivalence can be obtained for various ratings in Figure 4, and is plotted in Figure 5.

(a)



4-Wire Path Established By An RSU–HOST–RSU Connection

(b)

(c)

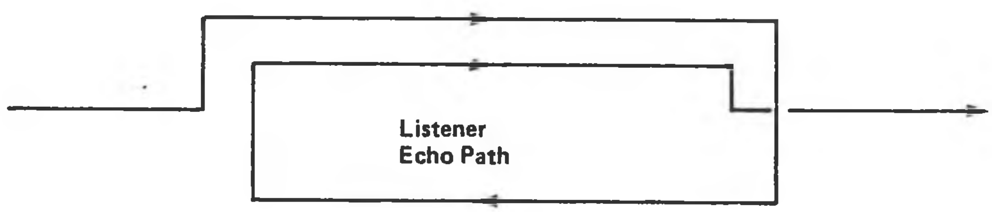


Figure 6 Local Network Connection And Associated Echo Signals

Figure 5 illustrates that equal levels of speech-correlated and continuous noise (S/D = S/N) are not perceived to be subjectively equivalent, but rather the S/N ratio must be significantly higher than the S/D ratio for the same subjective effect. Figure 5 also provides the basis for combining speech-correlated digital and continuous analog noise in a mixed analog/digital network.

3.0 SPECIFIC LISTENING FACTORS AFFECTING THE DISTRIBUTED LOCAL NETWORK

3.1 General

Brief mention of the distributed local network was made earlier in reference to Figure 2. This is a very important contemporary area in the evolution of the telephone network, since the introduction of digital transmission and switching makes it possible to provide local service more economically[7,9]. This is accomplished by converting part of the existing local network to digital transmission which can then support many customers per pair of copper wires. The digital transmission lines are terminated by a digital host switch in the central office and by smaller digital machines called remote switching units (RSU) situated near customer premises, as indicated in Figure 6(a)[10].

Connections of this nature provide highly satisfactory grade-of-service, when engineered in accordance with the considerations discussed presently.

3.2 Echo in the Local Distributed Network

Figure 6(a) illustrates the topology of the local distributed network in which customers are served by means of remote switching units connected to the digital host switch in the central office by digital lines[10,11]. An important parameter is the distance between remote switching units and the host switch, since this distance translates directly into round trip delay, which in turn can lead to a more noticeable echo impairment.

In Figure 6(a) talker A's signal will ordinarily be partially reflected by less than ideal return losses at the four to two-wire interface (at end B of the connection); this gives rise to the talker echo signal in Figure 6(b). If in addition this talker echo signal is again reflected at the four to two-wire interface at end A of the connection, then listener B will also hear an echo of talker A's signal; this gives rise to the listener echo signal in Figure 6(c).

As noted earlier listener echo in a telephone connection is analogous to acoustic room reverberation, and distorts the frequency response of the connection as suggested by the ripple response shown earlier in Figure 1. The subjective effect of this impairment is a potentially bothersome hollow sound not unlike what would be experienced by talking with one's head in a rain barrel.

In what follows emphasis will be placed on an investigation of this impairment based on both an objective laboratory simulation of the appropriate telephone network connections, followed by subjective evaluation of these connections.

From a practical standpoint such an investigation is necessary in order to be able to answer important network planning questions. For example; from a grade-of-service point of view, up to what distance can local network distribution be applied? This is relevant since longer distances correspond to longer round trip delays, which make echo signals more perceptible.

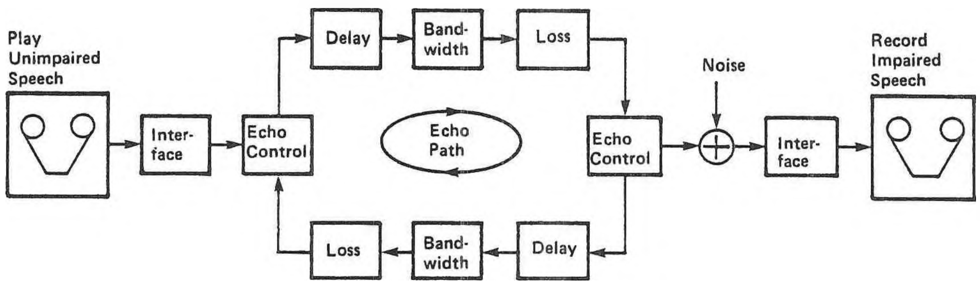


Figure 7 Preparation Of Test Conditions Using Network Simulation Facility

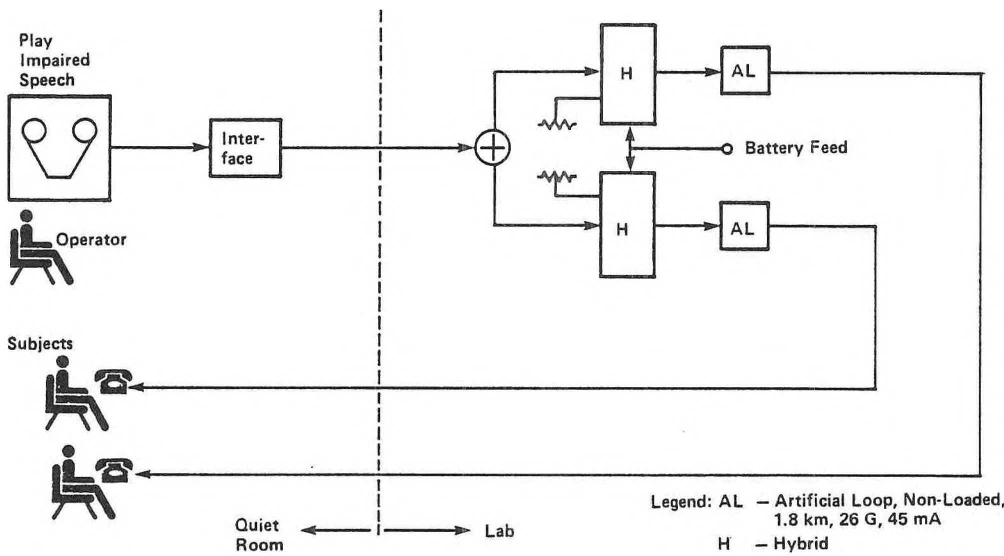


Figure 8 Arrangement For Subjective Tests

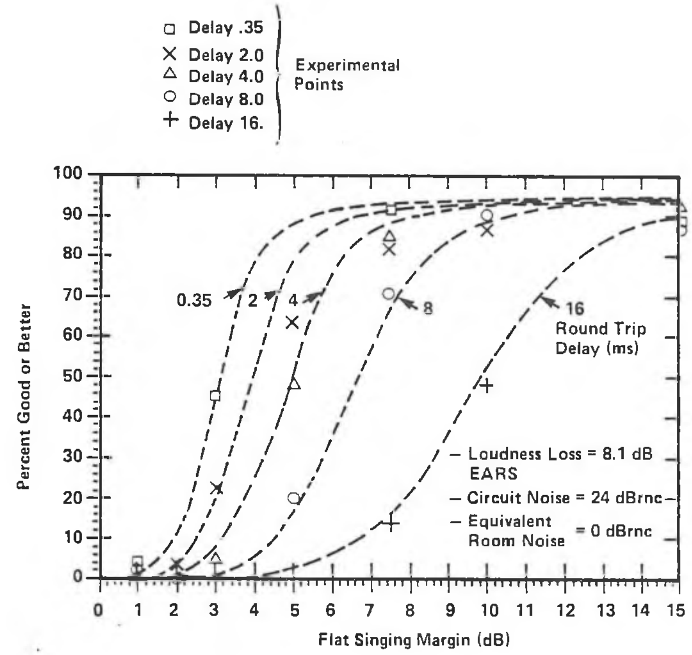


Figure 9 Loss/Noise – Listener Echo Opinion Model (% GOB)

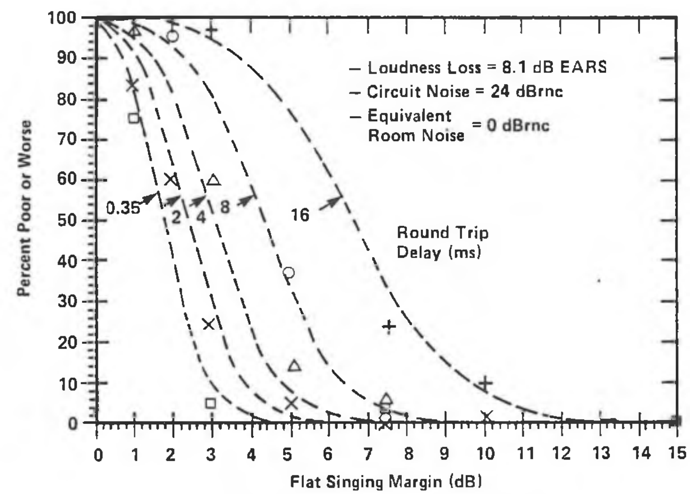


Figure 10 Loss/Noise – Listener Echo Opinion Model (% POW)

3.3 Test Methodology

Bell-Northern Research and other telecommunications organizations in other countries, have the capability of simulating virtually any telephone connection in their laboratories, for the purposes of both objective and subjective evaluation[12]. Figure 7 is one example of a connection which simulates the listener echo impairment of interest here.

The approach in Figure 7 is to play pre-recorded high quality telephone source speech into the simulation facility where various levels of echo, round trip delay and noise are imposed. The resulting impaired speech is then re-recorded and later played to and rated by test subjects; the ratings are based on the previously mentioned five-point scale and are made under controlled acoustical and electrical background noise conditions as shown in Figure 8.

3.4 Test Results

Figures 9 and 10 present the listener echo subjective test results, where on the ordinates the two grade-of-service indicators refer to the percent of subjects rating the connections good or better (% GOB), and poor or worse (%POW) respectively. In each figure the abscissa is the flat singing margin parameter which is the measure of uniform echo path loss around the four-wire path in Figure 7. (When this loss is 0 dB, the positive feedback nature of the echo path causes the circuit to oscillate or "sing"). Flat singing margin corresponds to a measure of echo attenuation. Thus a low singing margin implies high echo energy which from Figures 9 and 10 is rated very poorly; on the otherhand a high singing margin (high echo attenuation) corresponds to low echo energy which is rated much more favourably. For high enough singing margins the grade-of-service levels off at a satisfactory level governed by the background loudness loss and circuit noise conditions. The parameter in both Figures 9 and 10 is the round trip delay. Notice that grade-of-service deteriorates as the delay is increased.

Figures 9 and 10 provide the basis for answering certain network planning questions such as the one noted earlier pertaining to permissible remoting distances. For example, it is known that a high percentage of local telephone network connections (of the type illustrated earlier in Figure 6(a)) can provide singing margins of 8 to 10 dB or higher[13]. Under these circumstances Figures 9 and 10 indicate that listener echo will basically not be perceived as long as round trip delays are 4 ms or less. Thus with a knowledge of switching, processing and transmission delay performance of the connection, the above 4 ms delay figure can be translated directly into a permissible remoting distance. This indicates that distances of up to 80 km are possible between a host switch and a remote switching unit.

4.0 CONCLUSIONS

This paper has examined a number of electro-acoustic factors having a bearing on telephone network performance. Emphasis is placed on various listening factors such as continuous analog and speech-correlated digital noise, and listener echo induced frequency response effects.

In particular it is demonstrated that appropriate subjective testing methodologies and evaluations can answer planning and performance questions related to the contemporary mixed analog and digital evolutionary phase of the telephone network. For example in this paper the specific issues addressed were: tandeming of encoding/decoding processes, combining continuous analog and correlated digital noise, and determining local network remoting distances in a listener echo environment.

A thorough understanding of electro-acoustic factors affecting present day telephone speech quality will enable the network to evolve smoothly from the all analog to all digital stage of the future.

5.0 ACKNOWLEDGEMENTS

Some of the ideas included in the introductory part of this paper were suggested by P. Coverdale. The listener echo test results reported herein were contributed to by human factors consultations from J. Clegg, subjective testing support from N. Berry, as well as statistical analysis and grade-of-service modelling from M. Doughty.

6.0 REFERENCES

1. Gruber, J. and Williams, G., "Acoustic Factors Determining Telephone Speech Quality", (ABSTRACT), J.A.S.A. Supp. 1, Vol. 69, Spring 1981, pp. S76. Presentation to 101st A.S.A. Meeting, May 18-22, 1981, Ottawa, Canada.

2. Richards, D.L., "Telecommunication By Speech", Halsted Press, Wiley, 1973.

3. Canadian CCITT Contribution COM XII-No. 148-E, "Determination of Subjectively Equivalent Noise, and Threshold of Speech Correlated Noise", April 1979.

4. Daumer, W.R. and Cavanaugh, J.R., "A Subjective Comparison of Selected Digital Codecs for Speech", BSTJ, Vol. 57, No. 9, November 1978, pp. 3119-3165.

5. Chow, P., Meyer, F., Mills, F. and Pelletier, G., "The DE-4: A Fourth Generation Channel Bank", Telesis, Vol. 6, No. 3, June 1979, pp. 7-14.

6. CCITT Document AP VII-No. 53-E, Geneva 1980, Supplement No. 3 (AT&T) "Transmission Rating Models", pp. 17-44.

7. Telesis: Special Issue: DMS-100 Family of Digital Switches, Bell-Northern Research, Vol. 7, No. 4, 4th Quarter 1980.

8. CCITT Recommendation G.712, Orange Book Vol. III-2 Line Transmission, ITU Geneva 1977, pp. 415-423.

9. Marshman, G. and Mills, F., "The LD-1: A Second Generation Digital Line", Telesis Vol. 3, No. 6, March/April 1974, pp. 168-174.

10. Williams, G., Gruber, J.G. and Psimenatos, N., "Transmission Planning for Local Digital Networks", 2nd IEE Int. Conf. on Telecommunication Transmission, London England, March 17-20, 1981, pp. 205-208.

11. Neigh, J.L., "Transmission Planning for an Evolving Local Switched Digital Network", IEEE Trans. on Com., Vol. COM-27, No. 7, July 1979, pp. 1019-1024.

12. Canadian CCITT Contribution COM XII-No. 13-E, "A Preliminary Combined Loss/Noise - Listener Echo Opinion Model for a Uniform Echo Path Loss", April 1981.

13. Terry, J.B., Younge, D.R., and Matsunaga, R.T., "A Subscriber Line Interface for the DMS-100 Digital Switch", NTC Record, Vol. 2, Wash. D.C., Nov. 1979, pp. 28.3.1 - 28.3.6.