The Effects of Time Delay and Echoes on Telephone Conversations

By J. W. EMLING and D. MITCHELL

(Manuscript received August 9, 1963)

A brief history of the problem of echoes and delay in telephone connections is first given. Actual delays involved in typical circuits are shown. This is followed by a discussion of the effects of delay only on typical conversations. The sources of echo in typical telephone connections are then discussed, together with measures which have been taken to reduce echo by improving return loss. Methods of controlling the effects of delayed echo are then summarized; the summary includes a brief introduction to echo suppressors and some of the new problems they introduce.

I. INTRODUCTION

About 30 to 40 years ago, the effects of time delay and echo became of great concern to engineers planning transcontinental telephone systems. As a result, much effort was devoted to understanding these effects and finding ways to control them.[1] Fortunately, the development of high-speed transmission systems made the problem less severe than originally anticipated, and satisfactory ways were found to handle the amounts of delay that were then involved. As a result, interest in the effects of delay declined and little work has been carried out in this field during the last twenty years. But, as happens so often, the historical cycle is beginning to repeat. With the growth of intercontinental telephony and particularly the proposal of satellite systems involving one-way path lengths of 50,000 miles or more, the problems of delay have again come to the fore.

Because of this renewed interest, Bell Laboratories has resumed its studies of echo and delay. The basic problems, first considered 30 or more years ago, have been reviewed to see if improved solutions might result from the technology that has evolved in the intervening period. This work has been addressed not only to technical problems but also to new testing techniques for the evaluation of echo, delay and the control of these factors. Throughout the recent program, emphasis has been placed on the longer delays which are the unavoidable result of expanding the telephone network to provide a truly world-wide service.

This paper summarizes the earlier information on the effects of delay and echo, and outlines the fundamental problems they introduce. It thus serves as a guide to the literature of the past and also as an introduction to two companion papers dealing with the more recent studies in this area. The paper by Brady and Helder[2] traces the evolution of the echo suppressor, outlines the basic design problems involved and covers recent work aimed at applying modern technology to its design. The paper by Riesz and Klemmer[3] covers recent work on the subjective evaluation of delay and echo suppression in telephone communication.

II. HISTORICAL BACKGROUND

The problem of delay became acute during the 1920's and 1930's because voice-frequency circuits 500 miles and more in length were being set up for the first time in loaded cable. The propagation speed of these facilities was much slower than that of the open wire used previously, and the delay became sufficient to cause objectionable echoes. This initiated activity which led to the invention of the echo suppressor and also to intensive efforts to improve hybrid balance. Echo suppressors are capable of minimizing the most serious effects of echoes, but it was soon found that they also introduced other difficulties by interfering with the free two-way flow of speech. These impairments become more severe as the echo suppressor has to cope with longer delays. It was also found that delay alone introduces difficulties when it becomes large enough.

These facts naturally stimulated the development and use of the higher-speed carrier systems. Since the 1930's the use of these systems has grown rapidly, and at present practically all circuits of over 25 miles transmit at speeds of at least two-thirds that of light.

Table I shows the delays encountered on typical circuits of 30 years ago, on present-day circuits and on those which may be in service in the future through the use of satellites. This table clearly brings out the reasons for the intense interest in delay in the earlier period, the reduction of interest as high speed circuits were introduced, and the renewed current interest.

III. FUNDAMENTAL PROBLEMS

Problems due to echo and delay fall in three general categories. First are those due to the delay alone; second, those due to echo; and third,

TABLE I — AMOUNT OF DELAY IN LONG TELEPHONE CIRCUITS

Circuit	Approx. Length (miles)	Facility	Approx. Signal Velocity (mi/sec)	Approx. One-Way Delay* (msec)
(a) Actual Circuits in the 1930's				
New York to Chicago	900	VF on loaded cable (H-44)	20,000	13
New York to Dallas	1,850			93
(b) Considered in the 1930's but Not Established				
New York to San Francisco	3,200			160
(c) Present-Day Circuits				
New York to Hagerstown	260	VF on loaded cable (H-44)	20,000	13
New York to Chicago	900	K carrier	110,000	8
New York to San Francisco	3,200	TD-2	186,000	19
New York to London	4,100	submarine cable and others	120,000 for cable	35
Hawaii to London	10,000	submarine cable + 3000 mi of TD-2		80
(d) Circuits Involving Satellites				
New York to London	13,000† to 18,000	6,000-mile satellite + TD-2	186,000	70 to 97
	49,000	24-hour satellite + TD-2	186,000	265
Hawaii to London	29,000† to 38,000	two 6,000-mile satellites + 3,000 miles of TD-2	186,000	160 to 210
	98,000	two 24-hour satellites + 3,000 miles of TD-2	186,000	530

* Includes an allowance for terminals

† Length of path varies with satellite position.

those introduced by echo suppressors. This paper concentrates largely on the first two categories, leaving the last to the companion papers.

When two persons who are separated by a large distance wish to converse, there are two fundamental factors which must be taken into account. In their simplest form, these are illustrated in Fig. 1. This shows customer A at the left and customer B at the right connected by a four-wire circuit all the way from microphone to receiver (each line in the

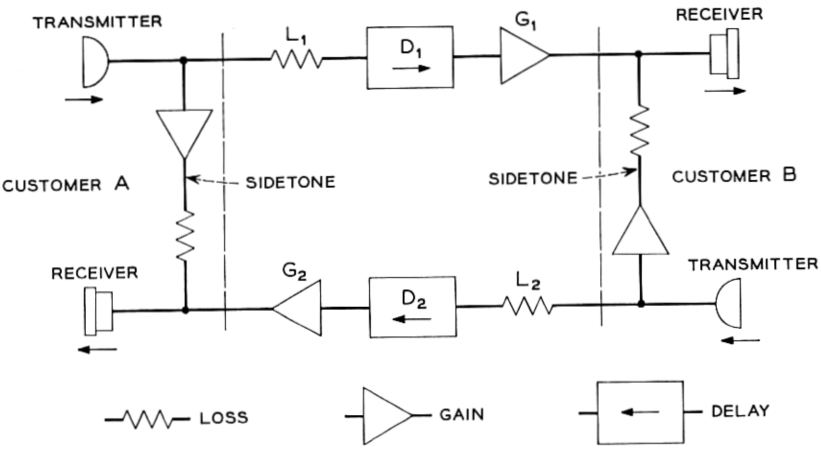

Fig. 1 — Four-wire toll circuit — no echo.

diagram actually represents two wires). There will be losses, indicated as L_1 and L_2, which must be compensated by gains G_1 and G_2 or the two will not be able to talk at all. In addition, there will be time delays between them, indicated by D_1 and D_2, for which no compensation is possible. In a practical case, the losses, gains, and delays are roughly equal in the two directions; they are not lumped as indicated in the figure, but are distributed over the entire circuit. For most long circuits, the net loss $(L - G)$ is made positive but close to zero.

Unlike losses that can be reduced as desired by amplification, delay cannot be reduced below an absolute minimum set by nature. This minimum equals the distance divided by the speed of light (186,000 mi/sec). As indicated previously, many of the older transmission facilities had speeds well below the speed of light, but current facilities achieve speeds closely approaching the theoretical maximum, and hence are minimum-delay facilities.

Delay changes the conversational process in a fundamental manner. We shall have more to say about this later, but for the moment it is only necessary to point out that when one talker stops talking he cannot hear the other reply until a time $D_1 + D_2$, or about $2D_1$, later.

For very long circuits, requiring the use of much gain, the separation of the two directions of transmission is economical, and it permits the use of high gains with a minimum of complication to avoid singing. On shorter circuits, however, it is far more economical to use "two-wire" circuits — that is, the same path for the two directions of transmission. Since by far the largest number of telephone circuits are short in length (i.e., customers' loops and local trunks), the aggregate economic benefit of using two-wire transmission on these circuits is very large. In addition to the greater economy of the facilities themselves, it is obviously more economical from the standpoint of switching (or interconnecting) circuits to use two wires instead of four.

Thus, the usual long distance call today involves both two- and four-wire circuits, much as shown in Fig. 2. The long four-wire toll circuits are converted to two-wire for transmission over short trunks between long distance and local offices, and over the customer's loop. At the end of the loop, the telephone reconverts to four-wire for connection to the customer's transmitter and receiver.

The conversion from a four-wire path to a two-wire path is accomplished by a hybrid coil. The effectiveness of the hybrid coil depends on the degree of match (or balance) between the impedances of the line and the network N. This balance is never perfect and therefore results in some of the incoming energy being returned as echo E_1 or E_2.

It has been found that echo becomes increasingly objectionable as the delay is increased. This is because close-in echoes tend to be masked by sidetone speech, but the masking effect decreases rapidly after speech ceases. Echo can be made less objectionable by increasing net loss, but loss obviously must not be increased beyond the point where talking is

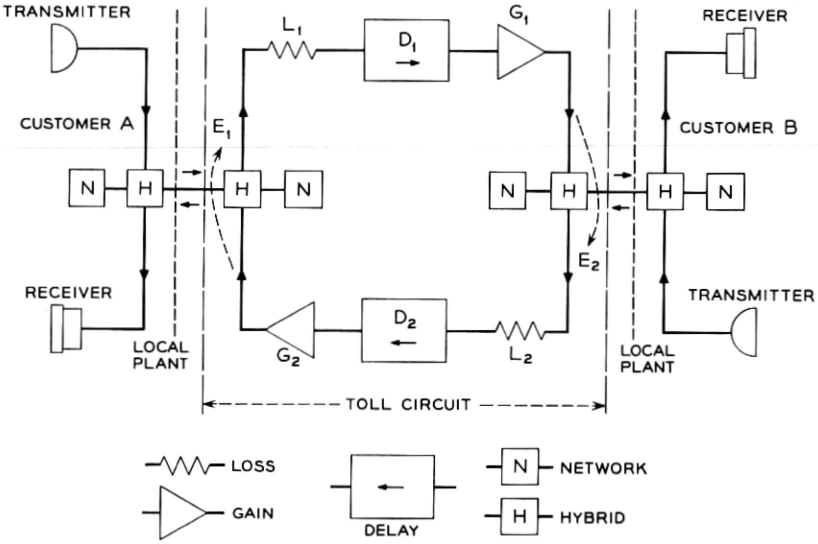

Fig. 2 — Four-wire toll circuit operated two-wire has echoes E_1 and E_2.

relatively comfortable. It was this situation which led to the invention of the echo suppressor. The fundamental point to be observed, however, is that increasing the delay results in two serious effects. The first is the increase in the response time and the second is increase in the severity of echo effects.

IV. EFFECTS OF DELAY ALONE

Delays of tenths of seconds are in the region of typical human reaction times and can be expected to have important effects on the structure of conversation even with four-wire circuits which are completely free from echo.

Fig. 3(a) illustrates a short portion of a typical conversation. The to-and-fro speech is shown for a case where there is zero delay in the telephone circuit. Line A shows the "talk spurts" TS from talker A at one end of the connection. A "talk spurt" is broadly defined as a portion of speech coming entirely from one talker.

In line B, the responses of talker B are shown. As is indicated, the talk spurts of each party vary considerably in length; and also the response times may vary. The "response time" RT, as shown, is defined as the interval from the time one party hears the other stop speaking until he himself starts speaking. RT may be negative if one party starts before the other stops.

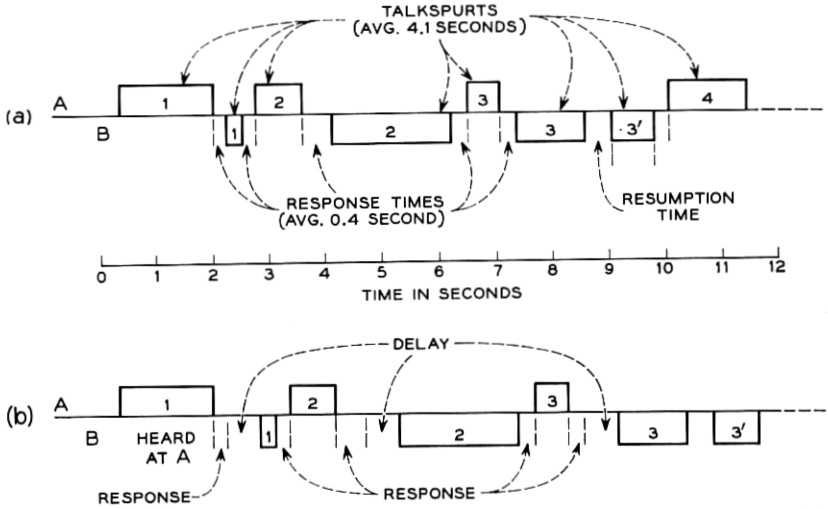

Fig. 3 — Typical time intervals in a two-way telephone conversation: (a) no circuit delay, (b) same conversation observed with added round-trip delay of 0.6 second.

Talk spurts, as defined, are not necessarily continuous. This is illustrated toward the right of line B by a break between B_3 and B_3' which provides an interval defined as "resumption time."*

In Table II some data are given showing typical TS and RT times as determined some years ago in observations made by A. C. Norwine and O. J. Murphy on actual telephone conversations.[4] These data were taken on slightly over 2800 telephone calls over a New York-to-Chicago private line used exclusively for Bell System business.

It is of interest that there are a rather large number of very short talk spurts. Thus the most frequent value (mode) is only 0.25 second. This probably represents the usual frequent monosyllabic replies. On the other hand, the median TS is two seconds and the average is a little over four. There are also occasional quite long talk spurts.

TABLE II—TALK SPURT (TS) AND RESPONSE TIME (RT) INTERVALS (IN SECONDS)

	Min.	Mode	Median	Mean	Max.
TS	0.09	0.25	2.0	4.14	143.8
RT	−3.95	0.24	0.32	0.41	5.04

For response times, the differences between mode, median and mean are less but there are a few long pauses and some negative times which represent double talking.

The time diagram in Fig. 3(b) illustrates one way in which large round-trip delays can affect conversation. This is a somewhat oversimplified model in that it assumes that the delay does not affect conversational structure. The added response time of 0.6 second is actually almost three times the mode of the normal RT and about twice the median value. Thus, if people did not change their conversational habits, this increase in response time would substantially add to the amount of time required to carry on a given amount of two-way conversation. The percentage increase in total time depends also on the length of typical talk spurts. Using the figures given in the preceding table, it can be estimated that the average conversation might be lengthened by from 6 to 12 per cent if the talkers continued to talk in just the same manner as before.

Experience with long delays, which is covered in the Riesz-Klemmer paper,[3] however, indicates that the effect of delay is considerably more complex than this simple model. Actually, calls tend to be shorter instead of longer when long delay is present. Apparently there are subtle

---

* Some workers in this field define B_3 and B_3' as separate talk spurts.

factors of annoyance and discomfort which cause the talkers to change their conversational structure and which make them wish to terminate the conversation sooner.

The simplified illustration, however, shows that even if talkers learned to adapt themselves to delay and discipline themselves to wait patiently for replies, there still would be degradation. Thus it seems inescapable that the addition of delays of one-half second or more will substantially reduce the true value per unit time of a toll connection. The very essence of most telephone calls is the ability to conduct a rapid exchange of ideas. This may not apply to some other types of communication, however, which involve relatively large amounts of one-way information.

The illustration is perhaps even more useful in pointing out the possibility that delay modifies the conversational pattern. As noted previously, even with zero delay some response times are negative. That is, there is a tendency to start talking before the other user has completed his talk spurt. With zero delay, this interruption is noted promptly, and the period of "double talking" is small. With long delays, however, the person who breaks in may talk for some time before the other is aware of the attempted interruption. In fact, it may be associated with a later, and quite different, part of the talk spurt than that which occasioned the break-in. If a response is long delayed, the original talker may resume his conversation before it is received, in which case the response to one piece of conversational material may appear as an interruption to a later, and possibly different, idea.

Thus, delay not only increases the tendency to double talk, but also increases the potential for confusion. As will be shown elsewhere,[3] these characteristics greatly complicate the suppression problem.

V. ECHOES AND THEIR EFFECTS

5.1 *Sources of Echo*

Any impedance mismatch in a transmission system will reflect energy back toward the source and be a potential cause of echo. If there is a mismatch at each end of a transmission line, the energy will be repeatedly reflected back and forth until dissipated by the line. In a strictly four-wire system, these reflections are ordinarily of no moment to the telephone user, since the amplifiers are one-way transmission devices and confine the echoes to small sections of the system with short transmission time and sufficient loss to rapidly damp out the reflections.

Echoes are likely to become important, however, as soon as two-wire transmission is employed, since the echo can now be returned with not much more attenuation than is encountered by the direct wave. Even though the major part of the system is four-wire, the effect of the two-wire portion needs to be considered since it usually controls the magnitude of the echoes.

As noted earlier, a commercial telephone connection always involves some two-wire transmission: a typical situation is illustrated in Fig. 2. Within the telephone station, the user's inherent four-wire transmission is converted to two-wire to fit into the exchange plant. At a toll office, it is reconverted to four-wire for transmission over long distances, and corresponding conversions are made at the far end. In practice, there may be other conversions, since four-wire transmission is sometimes used in the local plant, but is converted to two-wire for switching. Two-wire transmission is seldom employed in the toll plant, but some conversion from four-wire to two-wire for switching does occur. We need not consider these additional conversions for an understanding of the echo problem, but it should be appreciated that they complicate its solution.

Four-wire to two-wire conversions are made by means of a hybrid coil circuit. This is a form of bridge in which a network is provided to balance the line as shown in Fig. 4. If the impedance of the balancing network N is equal at all frequencies to the impedance of the two-wire line, the energy from the incoming four-wire branch is equally divided between the line and the network and the conversion is accomplished without echo. The "balance" in the hybrid circuit is then considered perfect. If there is any mismatch between the line and network impedances, the balance will be less than perfect and some of the energy will be transferred to the outgoing branch of the four-wire system.

Referring to Fig. 2, it will be noted that speech from customer A can

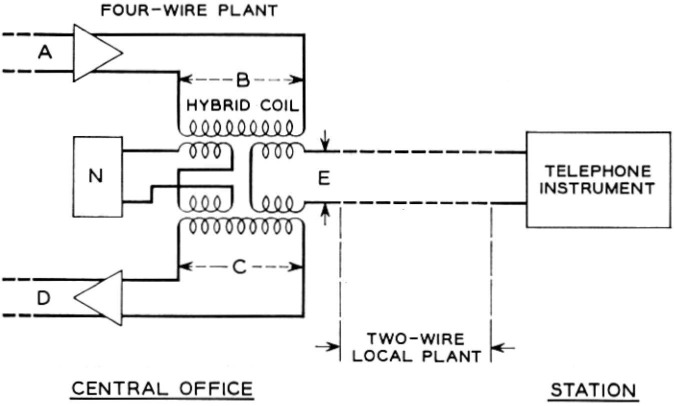

Fig. 4 — Occurrence of echo in two-wire local plant.

be transferred from one side of the four-wire circuit to the other at either his own telephone set, in which case it is called "sidetone," or at the distant, or B toll office. In this latter case, it is called "talker echo."

The principal difference between sidetone and talker echo is one of timing. Within the telephone station, transmission is essentially instantaneous and the sidetone appears concurrently with the talker's speech. The time for transmitting sounds to a distant toll office may, however, be many milliseconds. The talker echo may consequently be returned with a noticeable delay, in which case he will hear his own words reflected back much as if he were talking toward a cliff or other source of acoustical reflection.

Some of the talker echo can also be reflected at office A and back again toward the listener at the B office. This is referred to as "listener echo," since it is the distant listener who hears this some time following the arrival of direct speech. This process continues until the losses in the circulating path reduce the echo below audibility. This is illustrated in Fig. 5, reproduced from a paper[5] by G. M. Phillips.

It is customary to simplify discussions of echoes, as we have just done, by considering the transfer of energy as if it occurs at the hybrid coil in Fig. 4 (i.e., at the two to four-wire conversion point). In dealing with steady-state conditions, this is permissible. The magnitude of the steady-state echo is determined only by the relative impedances of the network and the line at the frequency of interest, and it can be computed if the steady-state impedance of the line is known at all frequencies. In problems involving transients, however, it is important to realize that the actual reflection occurs at any point or points in the two-wire line between the office and the station at which an impedance mismatch occurs. The time at which an echo occurs is therefore determined not only by the

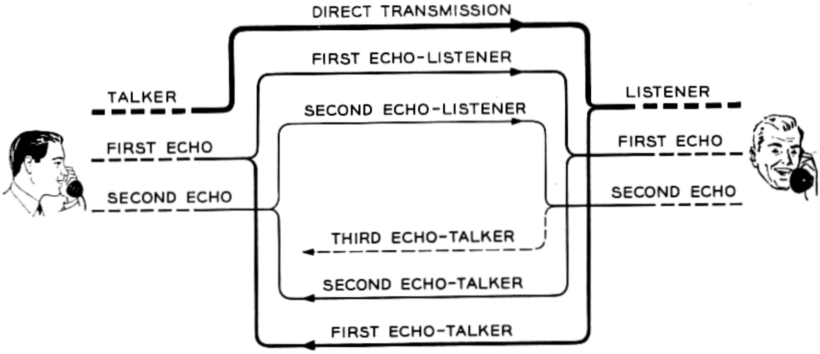

Fig. 5 — Telephone echo paths in a typical four-wire circuit.

transit time to the distant office but also by the transit time between the hybrid coil and the *point of reflection.*

Reflections do, in fact, occur at many places in a two-wire transmission line. Unless transformers have very large mutual inductance, they will usually cause some reflection at low frequencies. In voice-frequency circuits, variations in cable mutual capacitance and irregularities in the spacing of loading coils will introduce many small reflections. Thus in practice the echo tends to be a "smear" of many reflections.

However, by far the largest irregularity and associated reflection usually occur at the telephone station. Voice-frequency telephone lines will ordinarily have a characteristic impedance with a negative angle varying from a few degrees in the case of loaded lines to as much as 45 degrees with nonloaded lines. The magnitude of the impedance may be reasonably constant with frequency, for loaded lines, or decrease rapidly with increasing frequency for nonloaded. The telephone station impedance, on the other hand, is largely affected by the nature of the receiver. The angle is positive and the magnitude increases with frequency. It is evident that this is a very poor match with the line and a large source of echo.

Another important reflection may occur at the local central office. If the circuit (loop) from this office to the customer consists of loaded cable, it will provide a reasonably good match to the facility connecting the local office to the toll office. Where it is nonloaded cable, however, the match is poor because of the low magnitude and large negative angle of this facility.

The station and loop mismatch are problems of long standing, and there has been continuing effort devoted to their reduction. The problem is not so much technical as it is economic. The station and the loop connecting it to the central office are the most numerous elements in the plant (the Bell System has over 65,000,000 stations), and methods for solving the echo problem by improving their impedance characteristics would involve very large aggregate costs because of the large multiplying factor. Instead, it has proved more economical to reduce echo by measures applied to trunk circuits, since they are far less numerous.

The impedance of the two-wire local plant is a highly variable quantity. For any one loop, it varies greatly with frequency and also varies with the type of set and the length and type of facility. Thus, it is impractical to balance these impedances perfectly, and it is customary to select a network which provides the best compromise balance over the range of conditions commonly encountered. This network is 600 ohms in series with a capacitance of 2 mf in most toll offices, but 900 ohms with the same capacitance is used for local offices and some toll offices.

The most convenient way to express the amount of echo quantitatively is by using the concept of return loss, that is, the difference in db between the energy delivered to the circuit at a given point and the energy returned at this point as echo.* Obviously, the return loss of the local plant relative to the compromise network is a very complex matter, since it involves both frequency and facility. From the standpoint of echo, the frequencies between 500 and 2500 cps are of most importance, and it is customary to use a single value "average" over this band as a description. The effects of the various types and lengths of facility are expressed in terms of a statistical distribution of this "average" return loss or more particularly in terms of the mean and standard deviation of this distribution.

Typical values for return loss on telephone connections with compromise networks are a mean of 15 db with a standard deviation of 3 db at the toll office. At the user's local office, the corresponding figures are 11 and 3 db. The higher mean at the toll office is largely accounted for by the attenuation of the toll connecting trunks between the local and toll offices.

Return losses in the toll plant between one toll circuit and another are usually considerably higher than these values, and as noted earlier, return losses at the occasional points where toll circuits are switched on a two-wire basis are usually not controlling in the connection.

A long and continuing effort has been aimed at improving local return losses. So far, it has not proved economical to greatly reduce the unbalance at the telephone loop and station; however, impedance equalizers are used extensively on the toll office end of the exchange trunks giving access to the toll plant, and similar arrangements for other parts of the local plant are being developed. It appears that these and similar measures will achieve small but significant further improvements in local return loss over the next few years.

5.2 *Effects of Echo*

Not all echoes are harmful. For example, the telephone user likes some sidetone (which is talker echo without delay) because it gives him the impression that the circuit is "alive." Recent tests at Bell Laboratories[7] have indicated that the preferred sidetone volume is comparable to the volume he would like to receive from the distant talker. One reason for preferring this much sidetone is the high level at which the talker hears his own speech (largely via paths within the head) when he

---

* Appendix A discusses the return loss concept at somewhat greater length.

covers one ear with a telephone receiver. Obviously, his telephone sidetone must be noticeable compared to the head sidetone to obtain the desired impression of a live circuit.

When a telephone user is not talking, he does not have the masking due to his own voice, and he notices much weaker echoes. Delay which displaces talker echoes so that they can fall in silent intervals between speech sounds increases the chance that the talker will hear them. When the delays are short, the effects are small, since they are not much displaced from sidetone, and are manifested as sidetone having a hollow sound. With longer delays, the echoes clearly stand out, and, if the round trip delay is about one-quarter second and the returned echo is sufficiently loud (roughly 10 db above sidetone), it becomes so distracting that talking is virtually impossible. Even at very low levels, such long-delayed echoes are very annoying.

The annoyance of talker echo was recognized in the early days of long distance telephony, and tests to determine the tolerable echo were reported by Clark and Mathes[6] in 1925. In 1953, this work was repeated with current telephone circuits[5] with the results shown on Fig. 6. This shows the relation between round-trip delay and the minimum loss required to attenuate the echo sufficiently to provide a commercially tolerable condition in the judgment of an average listener. "Commercially tolerable" means that although an echo was discernible it was not loud enough to be objectionable. The curve represents the average tolerance to echo of all the listeners, but any individual may differ from the average by a considerable amount. About 68 per cent of customers have

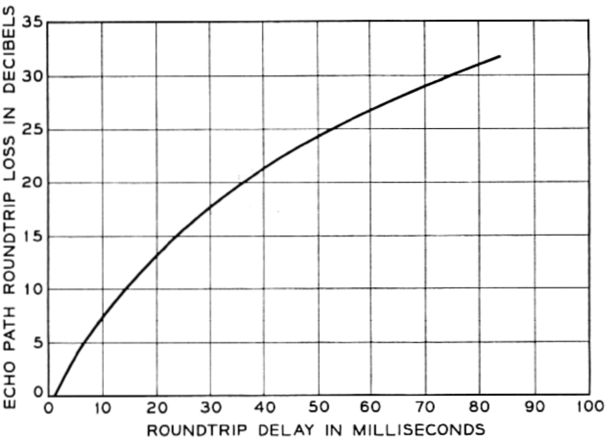

Fig. 6 — Average listener judgment of minimum echo attenuation necessary for just-tolerable talker echo conditions on modern telephone sets.

a tolerance to echo lying within ±5 db of this average curve. The remaining 32 per cent would be about evenly divided above and below but differing by more than 5 db from the average.

These tests were made with a typical loop connecting the talker to his local central office. The round-trip losses and delays shown are measured from the talker's central office to the point of reflection and return to the same central office. In a practical case, the loss in the echo path is made up of twice the net loss plus the far end return loss.

It might be thought that listener echo would be more troublesome than talker echo, since the user is then always in the "listening" condition, and hence does not have the masking effect of his own sidetone to reduce the effect of echo. In actuality, it has been found that where the round-trip delay is around one-quarter second or more, a given amount of talker echo is just about as annoying as the same amount of listener echo. Hence, talker echo is controlling on typical telephone circuits because the listener echo is always additionally attenuated by a second (near end) return loss.

5.3 *Controlling the Effects of Echo*

There are two fundamental ways in which echo can be controlled: (a) by reducing the delay and (b) by increasing the loss in the echo path, i.e., controlling echo magnitude. The use of carrier systems with their high propagation speeds has reduced the delay on a large part of the land line system to the point where echo is negligible. Since the speed is now close to that of light, little further reduction of delay is possible, and control of echo magnitude must be adopted where length is great enough to cause significant delay.

Control of echo magnitude has, over the years, taken many forms. Four-wire circuits and four-wire switching are used very largely in the long distance plant and have largely eliminated echoes except from the end links.* As noted earlier, it has so far not proved economical to reduce greatly the reflection at the telephone station.

Another way to reduce the effects of echo is to increase the loss between the talker and the point of major echo. It is to be noted that each db added will decrease the echo by 2 db. Attractive as this may be from an echo standpoint, it is obvious that it must not be pushed to the point where the direct transmission path is adversely affected.

---

* Even though the toll transmission paths are almost entirely four-wire, two-wire switching is still used extensively at smaller toll offices. By careful adjustment of the balancing networks and office wiring, the reflections from the four- to two-wire conversions are kept minor compared to the end link reflections.

The approaches just discussed are all simple, direct attacks on the problem. There is also a slightly more sophisticated approach, the echo suppressor, which is discussed in more detail in a companion paper.[2] This is a voice-operated switching device which reduces echo by introducing loss in the return transmission path of the four-wire system. Other sophisticated devices are conceivable. For example, since the return loss problem arises from the inability to obtain good return loss from a single balancing network facing a large variety of line impedances, it might be possible to make available a series of balancing networks from which the one giving the best balance for the particular circuit in use would be selected. This solution has been looked into quite extensively but, so far, the relatively small benefits which might result have not appeared to warrant the complications of implementation.

Another proposal frequently advanced has been the use of a self-balancing hybrid. Most of these schemes involve the measurement of current and voltage associated with the line impedance and the introduction of compensating currents or voltages in the balancing arm. F. B. Llewellyn has shown that this result is theoretically unattainable with any *linear networks*, i.e., with any arrangement for which the inserted compensating voltage or current is instantaneously, linearly related to the line value. This formerly unpublished document is reproduced in Appendix B with the author's permission.

In effect, this proof applies to a *single fixed* network which might of course be a complex device. However, it should be noted that this does not outlaw some more sophisticated approaches involving nonlinear networks or networks which are *adjusted* in accordance with some combination of the incoming signal and its echo. The latter class of device is sometimes called a "linear adaptive system." Although such arrangements may be technically achievable, they too have up to now appeared to be too complex and expensive to justify their use.

5.4 *Echo Control in the Bell System*

The use of four-wire circuits, four-wire switching and careful office balancing in two-wire offices has essentially eliminated echoes except from the end links. Also, some return loss improvement has been achieved through impedance correcting networks. Thus, for circuits with round-trip delays under about 45 milliseconds, it is possible to assure acceptable echo conditions on about 99 per cent of the calls by engineering the long distance circuits to have enough loss.[8]

When round-trip delay gets beyond about 45 milliseconds, however, the losses required would be unacceptable. Such circuits are operated at a low loss with echo suppressors. In general, these devices are used on most circuits over 1000 to 1500 miles in length. Under these circumstances, acceptable echo conditions are obtained on most calls, and the received volume of direct speech approaches the preferred value.

In principle, the echo suppressor is very simple and can be illustrated by the central echo suppressor invented by Clark and Mathes[6] in the early twenties. This device, shown schematically in Fig. 7, consists basically of two voice-operated switches. Speech from A traversing the upper four-wire path operates switch X_2, which disables the lower return path for the time necessary to prevent the return of echoes generated at the B end. Similarly, speech from B operates the switch X_1 and blocks echoes generated at the A end.

It is apparent that such a device may well interfere with normal conversational patterns. When the device is fully effective against echoes, it modifies the transmission system so that it is no longer a full duplex link capable of carrying intelligence in both directions simultaneously. Instead, it approaches a half duplex system which can be used in either one direction or the other but not in both at once. This will interfere to some extent with the free two-way flow of conversation, especially in those cases noted in Section IV where talk spurts overlap, most often caused by one talker trying to interrupt the other.

The echo suppressors now commonly in use are considerably more sophisticated than indicated by the simple model described above. Since these devices are discussed in detail in a companion paper, it suffices here to point out that they have been used with considerable success over a period of 30 or more years with the round-trip delays of 50–100 milliseconds currently encountered. We should also add that an echo

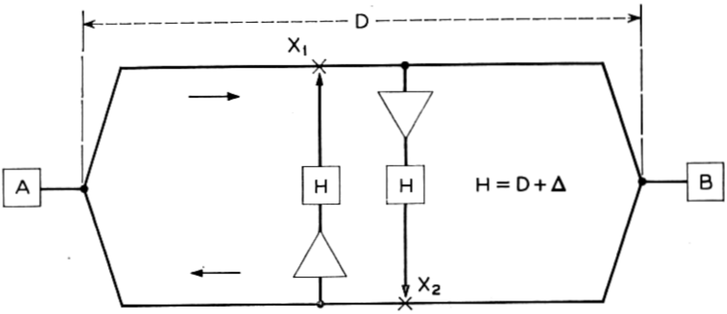

Fig. 7 — Central echo suppressor; *H* is slow release time of switch.

suppressor is essentially a compromise device. There is theoretically no great difficulty in suppressing echoes. The problem arises in accomplishing this while working with the wide variation in circuit characteristics, noise, and speech volume of the commercial telephone system. At the same time, there should be little interference with the rapid interchange of ideas which gives telephony so much value. Many of the requirements tend to be conflicting, and good design involves weighing the relative effects of a considerable number of subtle psychological factors.

Additional problems arise when several echo suppressors are used in tandem. Obviously the probability of interfering with normal conversational patterns increases with the number of suppressors used on a circuit. A particularly serious interference can occur where two or more complete suppressors are used with delay between them. In such a case it is possible to completely block both directions of transmission. This phenomenon is called "lock-out."

Lock-out can arise with the situation illustrated in Fig. 8. Let us assume that A has momentarily stopped talking and the transmission path is functioning in each direction. If B starts talking he will first operate switch 3 and at time *D* later he would also operate switch 1. However, if talker A resumes talking before time *D* his path is open to his suppressor and he will operate switch 2. Now both directions of transmission will be blocked and remain so until one party stops talking. Since neither one knows that the other is talking the "lock-out" may continue for a long period.

It should be noted that the important factor in producing lock-out is the delay *D* between X_2 and X_3. This delay provides a time storage which allows each talker to independently capture control of the suppressor nearest to him.

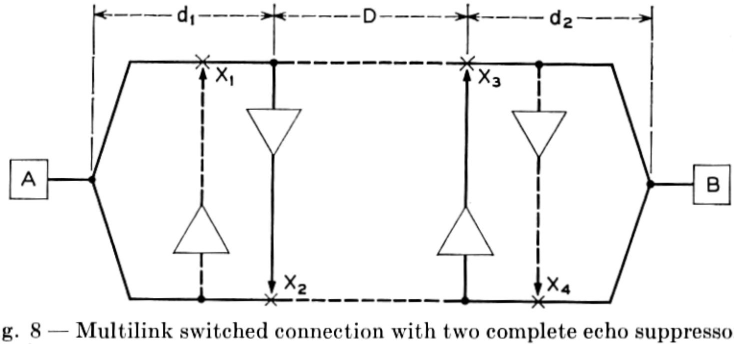

Fig. 8 — Multilink switched connection with two complete echo suppressors — lock-out can occur.

Lock-out was studied at length by A. W. Horton in the 1930's.[9] He showed how the probability of its occurrence can be calculated from a knowledge of the circuit and speech parameters.

Some protection against lockout can in theory be provided through echo suppressor design, but in the long run it may be more desirable to avoid the use of tandem suppressors by suitable switching arrangements.

VI. CONCLUSION

In resuming the work on echo and delay, we found ourselves with a solid background of fundamental knowledge but were faced with a number of new problems. Perhaps first in importance was the determination and evaluation of the effect of using delays far longer than those experienced hitherto. Second, and closely related, was the question of designing suppressors to control these larger delays, and particularly the application of modern techniques in the expectation that they would provide more effective devices than those designed for the shorter delays. These are the main subjects of the papers which follow.

APPENDIX A

Further Notes on Return Loss

While return loss, in principle, can be measured at any point in a circuit, it is most conveniently specified and measured at the four-wire, two-wire junction. At this point, both the incident and the reflected wave can be determined from measurements on the transmitting side of the hybrid using a signal source on the receiving side. The relative strength of the wave delivered to the two-wire line is determined by opening the network: that is, producing a complete unbalance. The relative strength of the echo is then determined by connecting the network and measuring again. The difference between these two measurements, both expressed in db, is the return loss. If the network exactly balances the impedance of the line, there will be no reflected wave, and the return loss will, of course, be infinite.

It should be noted that it is the difference between the measurements made with the network leg open and terminated in a network that expresses the return loss; not the total loss from receiving to the transmitting side. This latter, or transhybrid, loss exceeds the return loss by the losses inherent in the hybrid coil. Referring to Fig. 4, the incident energy at E will be seen to be lower than at B by the coil loss B to E (ideally 3 db and practically about 3.5 db). Similarly, the reflected energy at C is lower than that at E by the coil loss. Thus, the transhybrid loss B to C exceeds the return loss by twice the coil loss (about 7 db in practice). If amplifiers were introduced to just compensate for the coil losses, the difference in levels between A and D would be the return loss.

Return loss may also be expressed in terms of impedances. At the hybrid

$$\text{return loss} = 20 \log \left| \frac{Z_L + Z_N}{Z_L - Z_N} \right|$$

where Z_L and Z_N are the line and network impedance respectively. At any point in the two-wire circuit the return loss is also defined as above, where Z_L and Z_N are interpreted as the sink and source impedances.

Return loss is a function of frequency. From the standpoint of echo, the frequencies between 500 and 2500 cps are of most importance. Often a band of resistance noise covering this band is used as a source for measuring "average" return loss.

It will be appreciated that the reflections which give rise to echo can also cause oscillation or singing if the gain around the four-wire loop equals the loss at any frequency. Usually the high and low frequencies are controlling and return loss at the most critical frequency is referred to as a "singing" return loss to distinguish it from the loss at the frequencies important from an echo standpoint.

Return loss is often used to describe impedance characteristics and the uniformity of transmission lines. It is not practical to manufacture all cable pairs with precisely the same mutual capacitance; there will be some variation with length and from pair to pair. Similarly, it is not practical to introduce loading coils at precisely the theoretical spacing; some departures will be required to fit the geographical situation. It is customary to describe the effects of such variations in the impedance of the facility in terms of the "structural return loss." This is the return loss of an infinitely long (or a suitably terminated) practical cable compared to the impedance of an ideal cable with no irregularities. Typically, structural return losses run about 20–25 db for exchange plant circuits. Toll grade cable traversing areas which place few restrictions on the location of loading coils may have structural return losses of about 30–35 db. This is much higher than the 11–15 db typical of local return losses.

APPENDIX B

*Proof That a Self-Balancing Hybrid Cannot Be Constructed with Linear Networks**

By F. B. Llewellyn

In hybrid coils, the degree of balance that may be realized between the balancing network and the attached load impedance is one of the fundamental limitations in the design of telephone systems. From time to time, the proposal has been made that some means of producing an automatic self-balancing coil should be sought. Various schemes for accomplishing this have been suggested. Essentially most of them, in one way or another, involve the idea that it should be possible to insert vacuum tubes in such a way that they measure the current and voltage associated with the line impedance and introduce compensating values into the arm containing the balancing network. In this way, it was hoped to maintain the hybrid balance regardless of changes in the line impedance.

The following offers a proof that this result is theoretically unattainable with any linear network† whatever, regardless of whether it contains amplifiers, negative impedances, gyrators or any other linear elements, active or passive, that have been conceived in the past or may be thought of in the future. Of course, this does not rule out the possibility of obtaining operating improvements by certain *nonlinear* devices, and voice-operated echo suppressors provide a good example of such a nonlinear device.

As a starting point in the proof, Fig. B1 on the accompanying sketch shows the generalization of the configuration to be considered. The box contains the network elements of the hybrid itself while the four pairs of terminals represent, respectively, the attached line Z_L, the balancing network Z_N, the transmitting impedance Z_T, and the receiving impedance Z_R. Signal generators may exist in all of these except the balancing network Z_N. Consequently, the figure may be generalized even further, as shown in Fig. B2, by allowing Z_N to be included as part of the network within the box.

---

* This proof was prepared as a Bell Telephone Laboratories internal document dated October 13, 1950 and is published here with the author's permission.

† Note added by author in 1963: "With the exception of the line impedance Z_L all of the other impedances in the network are taken to be independent of time. Though implied in paragraph 1, this point has caused some confusion."

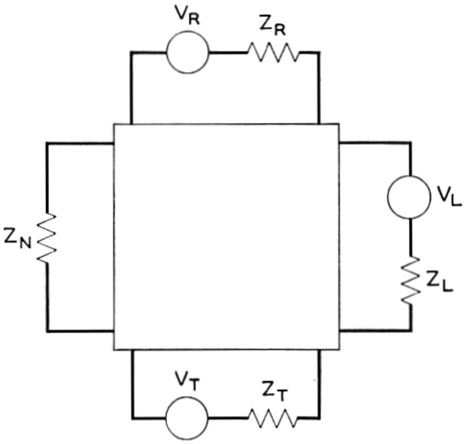

Fig. B1 — General prototype of hybrid terminations.

The conditions for an operative hybrid balance are three in number and may be set down as follows:

1. When the system is driven from V_R (that is, when $V_L = V_T = 0$), the current I_T is zero.

2. When the system is driven from V_R (that is, when $V_L = V_T = 0$), the current I_L is not zero.

3. When the system is driven from V_L (that is, when $V_R = V_T = 0$), the current I_T is not zero.

The requirement for self-balance is that these three balance conditions

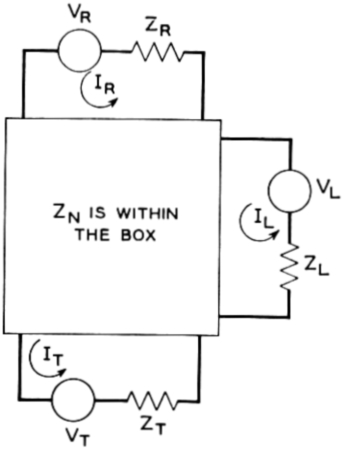

Fig. B2 — Equivalent of Fig. B1 for present analysis.

shall remain in force regardless of changes in the value of the line impedance Z_L.

Let us see what these conditions mean when applied to the general linear equations which describe the external behavior of the configuration in Fig. B2. These may be set down as follows:

$$\begin{array}{ccc|c} I_R & I_L & I_T & \\ \hline Z_{11} + Z_R & Z_{12} & Z_{13} & V_R \\ Z_{21} & Z_{22} + Z_L & Z_{23} & V_L \\ Z_{31} & Z_{32} & Z_{33} + Z_T & V_T \end{array} \tag{1}$$

where the Z's with double subscripts are general impedance constants representative of the self-impedances and transfer impedances of the network within the box. Thus they are independent of the terminating impedances Z_R, Z_L, and Z_T.

When the system is driven from V_R (that is, when $V_L = V_T = 0$), the current I_T becomes, from (1)

$$I_{T_R} = \frac{V_R}{\Delta} [Z_{21}Z_{32} - Z_{31}(Z_{22} + Z_L)] \tag{2}$$

where Δ is the determinant of the matrix in (1). According to the first of the three balance conditions, this current must be zero, which requires:

$$Z_{21}Z_{38} - Z_{31}(Z_{22} + Z_L) = 0. \tag{3}$$

If self-balance is to be imposed then (3) must remain satisfied regardless of changes in Z_L. This can be accomplished only if *both* of the following are true:

$$Z_{31} = 0 \tag{4}$$

$$Z_{21}Z_{32} = 0. \tag{5}$$

To repeat this in words, we have found that the condition for self-balance requires that Z_{31} should be zero, and, moreover, that *either* Z_{21} or Z_{32} should be zero.

Let us see what these requirements mean in terms of the second and third of the three balance conditions. To do this, we need the equation for I_L when $V_L = V_T = 0$ and the equation for I_T when $V_R = V_T = 0$. From (1) these are, respectively:

$$I_{L_R} = \frac{V_R}{\Delta} [Z_{31}Z_{23} - Z_{21}(Z_{33} + Z_T)] \tag{6}$$

$$I_{T_L} = \frac{V_L}{\Delta} [Z_{31}Z_{12} - Z_{32}(Z_{11} + Z_R)]. \tag{7}$$

In the event that (4) is satisfied, these become

$$I_{L_R} = -\frac{V_R}{\Delta}[Z_{21}(Z_{33} + Z_T)] \tag{8}$$

$$I_{T_L} = -\frac{V_L}{\Delta}[Z_{32}(Z_{11} + Z_R)]. \tag{9}$$

Then when (5) is satisfied in addition, we see that I_{L_R} is zero in the event that $Z_{21} = 0$ and I_{T_L} is zero in the event that $Z_{32} = 0$.

It therefore must be concluded that the self-balance condition cannot be satisfied without violating either the second or the third of the three general hybrid balance conditions. Consequently, the proof is complete that a self-balancing hybrid cannot be constructed with linear networks.

REFERENCES

1. Blackwell, O. B., Time Factor in Telephone Transmission, Trans. A.I.E.E., **51,** 1932, pp. 141–147; reprinted as Bell System Monograph 619.
2. Brady, P. T., and Helder, G. K., Echo Suppressor Design in Telephone Communications, B.S.T.J., this issue, p. 2893.
3. Riesz, R. R., and Klemmer, E. T., Subjective Evaluation of Delay and Echo Suppression in Telephone Communications, B.S.T.J., this issue, p. 2919.
4. Norwine, A. C., and Murphy, O. J., Characteristic Time Intervals in Telephonic Conversation, B.S.T.J., **17,** April, 1938, pp. 281–291; reprinted as Bell System Monograph 1074.
5. Phillips, G. M., Echo and Its Effects on the Telephone User, Bell Laboratories Record, **32,** August, 1954, pp. 281–284.
6. Clark, A. B., and Mathes, R. C., Echo Suppressors for Long Telephone Circuits, Bell Telephone Laboratories Monograph 165, presented at A.I.E.E., June, 1925.
7. Noll, A. M., Subjective Effects of Sidetone During Telephone Conversation, to be published.
8. Huntley, H. R., Transmission Design of Intertoll Telephone Trunks, B.S.T.J., **32,** September, 1953, p. 1019.
9. Horton, Arthur W. Jr., The Occurrence and Effect of Lockout Occasioned by Two Echo Suppressors, B.S.T.J., **17,** April, 1938, pp. 258–280; reprinted as Bell System Monograph 1073.