THE BELL SYSTEM TECHNICAL JOURNAL

VOLUME XL MAY 1961 NUMBER 3

Functional Design of a Voice-Switched Speakerphone

By W. F. CLEMENCY and W. D. GOODALE, JR.

(Manuscript received January 16, 1961)

A new hands-free telephone, known as the 3A Speakerphone System, is described. It provides, by means of switched-gain techniques, almost complete freedom from distant-end talker echo and singing. The gain-switching action is virtually free of clipping or blocking — objectionable side effects that are often introduced with voice control of gain. The switching threshold is varied automatically in accordance with room noise to avoid blocking in the receive channel. Performance characteristics are shown, with particular emphasis being given to the parameters chosen to meet rather stringent performance objectives.

I. INTRODUCTION

The original 1A Speakerphone System[1] used amplification in both its receive and transmit channels to compensate for the acoustic loss that was introduced by placing telephone instruments at greater distances from users than is normal with a handset. Among the operational difficulties[2,3] with such a system are talker echo and singing, particularly under reverberant room conditions.

To avoid the limitations that are inherent in the simultaneous use of high gain in the two transmission channels, the voice-switched 3A Speakerphone System has been designed. It changes the gain in each of the two channels in accordance with the direction of the stronger speech signal. Voice-switching techniques have been used before in communication equipment, and have been found to possess limitations of their own. However, by application of certain design principles,[4] the new voice-switched speakerphone substantially eliminates talker echo and singing, and greatly reduces objectionable side effects of the type often encountered with voice control. Moreover, these advantages are realized for a wide range of conditions. This paper is intended to provide a functional description of the 3A Speakerphone System and to show its performance characteristics, with particular emphasis on the parameters chosen to meet rather stringent performance objectives.

II. TRANSMISSION DESIGN PROBLEM

From a transmission standpoint, the fundamental difference between a speakerphone and a handset telephone is the distance between the instruments and the user. The loss introduced in the receive and transmit channels with normal arrangements anticipated for the 3A Speakerphone amounts to some 20 db in each direction. The increased gain provided in each channel to compensate for this loss, and the acoustic coupling between the loudspeaker and the microphone, introduce four operational difficulties: (a) sustained feedback or singing, (b) room echoes returned to the distant talker, (c) increased levels of transmitted reverberant energy, and (d) higher transmitted room noise. While voice switching is effective in essentially eliminating the first two, it provides no relief from the latter two difficulties.[3]

The two problems to which voice switching applies as a solution are illustrated in Fig. 1, which shows acoustic and electric levels for the two types of telephones, when they are producing equivalent transmission levels.* For the marginal incoming volume level and hybrid balance selected as an illustration, it will be noted that, when the speakerphone is transmitting, Fig. 1(b), the acoustic sidetone is about 24 db above the speech pressure at the microphone. When it is receiving, Fig. 1(c), the return echo to the telephone line is about 29 db above the incoming speech level. These conditions, of course, result in singing. If the gains were reduced to a point just below singing, the return of talker echo to the connected telephone might still be objectionable on low-loss connections.

The figure suggests that, if 30 db or more of gain were interchanged in the two channels in response to signal flow, the singing problem would be eliminated and talker echo would be tolerable for the selected conditions.

* It is assumed that the appropriate sound and electric levels are measured with meters having dynamic characteristics similar to the VUmeter and used in the approved manner.

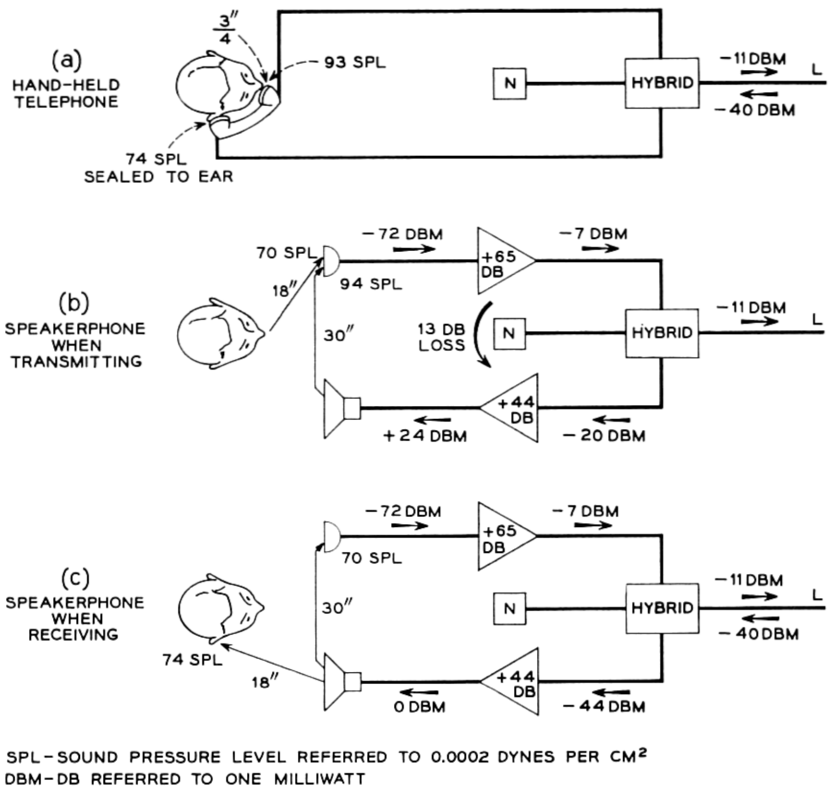

Fig. 1 — Comparison of transmission problem for handset and speakerphone

Actually, because the speakerphone's performance is more dependent on room acoustics and room noise than is that of a handset telephone, it is desirable to switch 35 to 40 db of gain between channels under certain conditions.

III. GENERAL DESCRIPTION

In the 3A Speakerphone, the interchange of gain between the receive and transmit channels is effected by control circuits operating on a linear differential basis; i.e., the channel having the stronger signal has the higher gain. To produce smooth gain changes without noticeable clipping of the speech syllables and without interference by the expected ranges of incoming line noise or room noise at the speakerphone, the control and the variable gain circuits are carefully designed on both a transient and a steady-state basis. The time factors of speech, conversational habits, noise, and room reverberation must all be considered in determining the appropriate circuit characteristics.

In order to reduce the objectionable effects of the gain changes, the 3A Speakerphone switches only the amount of gain that is required for stability at the loudspeaker volume desired by the listener. This means that on low-loss connections, for which the volume control setting is low, a small amount of gain is switched; on higher-loss connections requiring a higher volume control setting, a greater amount of gain is switched.

Automatic variation of the switching threshold of the control circuit in accordance with room noise at the speakerphone, to avoid blocking the receive channel, eliminates the need for any adjustment by the installer or user. Also, by proper selection of the time constants and the use of the linear differential control feature, there is no need of any adjustment for almost all conditions of room reverberation.

Before describing the voice-switched gain circuit in detail, a broad over-all picture of the circuit and its operation is presented in the schematic of Fig. 2. The transmit channel consists of the microphone M, amplifiers A_M and A_T, and the transmit variolosser TVL. The receive channel consists of the loudspeaker S, amplifier A_R, and receive variolosser RVL. The two channels are coupled in the usual way with a hybrid coil to the telephone line. The balance network N of the hybrid incorporates current-sensitive variable elements which utilize the dependence of the loop direct current on the distance from the central office to compensate partially for different line impedances.

Variolossers TVL and RVL are variable-gain devices regulated in a complementary manner by the direct current flowing through them. This direct current is obtained from the manual volume control VOL, and from the combination of control amplifier A_C, rectifier R_C, and timing circuit T_C. The input to A_C is the microphone signal taken from the output of amplifier A_M and modified by the control variolosser CVL. An increase in the direct current through CVL, produced by two circuits, increases its loss. The "switchguard" circuit, consisting of amplifier A_G, rectifier R_G, and timing circuit T_G, produces a direct current which increases the loss of CVL in response to the voltage across the loudspeaker, and thus guards against false switching of the received signal due to the microphone voltage resulting from the loudspeaker output. Another direct current is produced by the NOGAD* from any nearly constant microphone voltage due to the noise at the speakerphone location.

* NOGAD = noise-operated gain-adjusting device.

The operation of the voice-switching circuits can be explained by considering several conditions:

First, the quiescent condition of no local speech or noise at the microphone and no incoming signal from the line. No direct current flows from the control rectifier R_C through TVL and RVL, and no direct current flows through CVL from R_N and R_G. Consequently, the gain of the receive channel is high and that of the transmit channel is low. The speakerphone is thus prepared for loudspeaker reproduction of a line signal of any level.

Second, only a line signal is present and is being reproduced by the loudspeaker. The switchguard, as a result of the loudspeaker voltage, increases the loss of CVL, and thus limits the input to A_C resulting from the acoustic coupling between the microphone and the loudspeaker, so that no current is produced through TVL and RVL. The receive channel remains in its high-gain state, allowing uninterrupted loudspeaker reception of the incoming signal.

Third, local speech at the microphone with no room noise and no incoming signal. No direct current from R_G and R_N flows through CVL; the output of A_M is coupled with minimum loss in CVL to A_C, causing direct current from R_C to flow through TVL and RVL. The gain of the transmit channel is high and that of the receive channel is low, so that speech at the microphone is readily transmitted to the line.

Fourth, the same conditions as for the third case with the addition of room noise at the microphone. The loss of CVL is increased by direct current from the NOGAD, so that the noise signal at the input of A_C produces only a small current through TVL and RVL. Local speech, however, does not cause an increase of the direct current from the NOGAD, because it is designed to give very little response to fluctuating signals. The increased loss of CVL, due to the noise at the microphone in this case, can be overcome, however, by the higher level of speech signal normally produced under noisy conditions, and sufficient speech signals can reach the input to A_C so that the resulting direct current from R_C through TVL and RVL clears the transmit channel and blocks the receive channel.

Fifth, a line signal is present and room noise exists at the microphone. Both the NOGAD and the switchguard produce direct current through CVL, and its increased loss prevents receive blocking because of noise or acoustic coupling between loudspeaker and microphone.

Sixth, an incoming speech signal is present and local speech exists at the microphone. Direct current is supplied to CVL from the switchguard, and a signal voltage exists at the output of A_M due to speech at the microphone. The relative signal strengths are evaluated in CVL: if the line signal is great enough, the set will remain in receive; if the speech signal at the microphone is great enough, the set will transfer to the transmit state. Actually, because of the rapidly fluctuating levels characteristic of speech the set is continuously switching between the transmit and receive states, so that either party can quickly react to speech from the other party.

IV. TRANSMIT AND RECEIVE VARIOLOSSERS

The variolossers RVL and TVL in the receive and transmit channels act, in principle, like those used in the compandors of carrier systems.[5,6] In

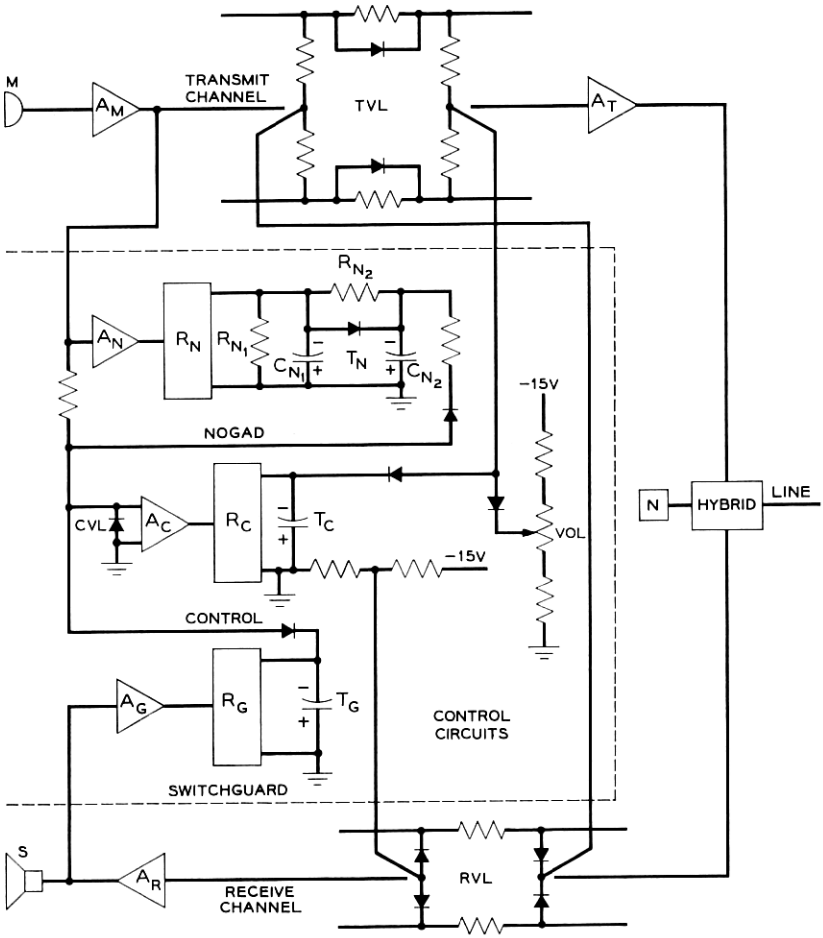

Fig. 2 — Schematic of 3A Speakerphone.

the present application, the variable impedance element is a diffused silicon varistor, the ac resistance of which varies inversely in a smooth and continuous manner as a function of the direct current through it. In TVL, as can be seen in Fig. 2, the varistors are part of the series path of the transmit channel, so that a decrease of their impedances by an increased flow of direct current increases the gain of the channel. In RVL the varistors are part of the shunt path, and a decrease of their impedances by an increased flow of direct current decreases the gain of the receive channel. The characteristics of the variolossers are shown in Fig. 3. For the range of direct current provided by the control rectifier R_C and the volume control VOL, the change of gain of TVL is 36 db and that of RVL is 42 db. Because the same control current flows through the tandem connection of the dc paths of the variolossers, their gains always change in a complementary manner.

Each variolosser is provided with a dc shunt path, not shown in Fig. 2, which serves to match the two characteristics.

As shown in Fig. 3, the sum of the transmit and receive gain changes is never greater than that for zero current. This avoids possible singing for

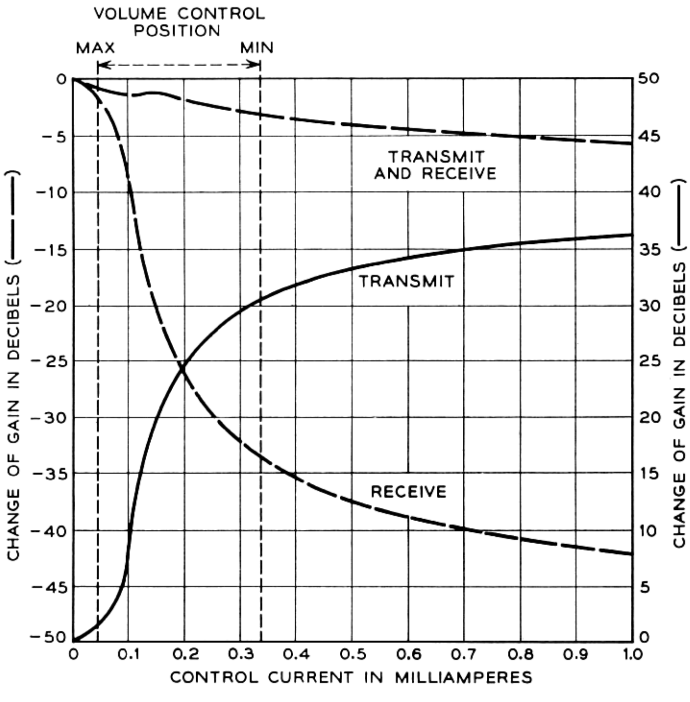

Fig. 3 — Variolosser characteristics.

any setting of the volume control and for the entire range of control current as it varies with speech level.

Because the varistors are nonlinear circuit elements, care must be taken to minimize distortion of speech signals in the variolossers. In the 3A Speakerphone, the signal levels have been adjusted so that the departure from linearity is small over the range of direct currents and signals encountered up to the overload points of the amplifiers in each channel. Furthermore, the variolossers are arranged as balanced circuits to keep control current variations from producing interference or "thump" in the speech channels.

In addition to the current from the control rectifier R_C, the loudspeaker volume control VOL supplies to the variolossers a quantity of direct current dependent on its manual setting. This, as can be seen in Fig. 3, varies the gain of the receive channel and also provides a bias current from which the control current increases in response to speech at the microphone. The resulting effect is that the amount of gain switched by the current from R_C is small for the lower volume control positions and becomes greater as the volume control is advanced, as portrayed in Fig. 4.

On the majority of calls, only a small amount of gain is switched, because low volume control settings give adequate loudspeaker output.

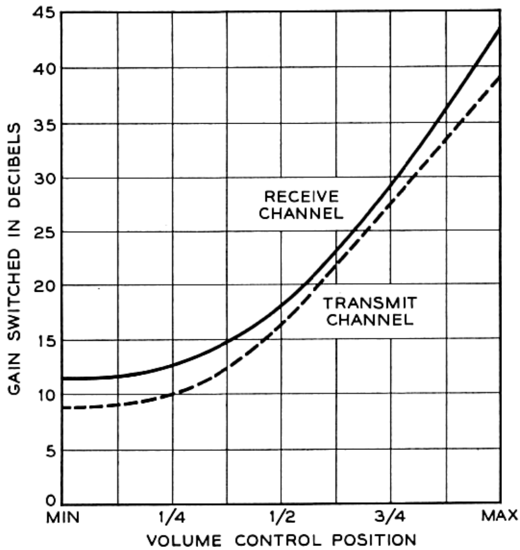

Fig. 4 — Gain switched for various volume control positions.

As a result, simultaneous speech is permitted with barely noticeable clipping, and the changes in the transmitted background noise, resulting from the gain switching, are not apparent at the connected telephone. However, sufficient loss is switched so that the talker echo retransmitted is not objectionable, even for moderately reverberant rooms. On the other hand, when the volume control is near its maximum position on high-loss connections, the resulting voice switching and noise background effects are more noticeable, but tend to be masked by noise at the distant end, because of the lower received signal there.

V. SWITCHING CONTROL CIRCUIT

Through mutual interactions, the control path, the switchguard path, and the NOGAD circuit control the state of TVL and RVL in response to signal flow in the speech channels. In the absence of both a received signal and ambient room noise, speech signals from the microphone pass through CVL with minimum loss, are further amplified by A_C, rectified by R_C, and impressed on timing circuit T_C. When the voltage developed by R_C is larger than the dc bias voltage in series with TVL and RVL, switching action occurs. Fig. 5 shows the steady-state transmit switching characteristic measured with a 1000-cycle signal from the microphone for various volume control positions, when the switchguard and NOGAD circuits are inactive. At the maximum volume setting, a gain change of 37 db occurs as the microphone voltage V_m varies from -97 to -81 dbv. At lower settings the gain change is reduced, but the point at which switching is completed remains the same.

When a loudspeaker voltage is present, the rectified output of the switchguard changes the loss of CVL in proportion to the loudspeaker voltage over a wide range. The switching action of the transmit channel with a loudspeaker voltage V_s of -20 dbv and the NOGAD inactive is shown in Fig. 6. In obtaining these data, the sidetone path is interrupted at the input of RVL and proper terminations attached. A receiving signal to produce V_s is connected at the input of RVL. Then a low microphone voltage is applied and slowly increased. At a critical value of V_m, in this case -63 dbv, the gain of the transmit channel abruptly increases. Simultaneously, the gain of the receive channel, as shown in Fig. 7, abruptly decreases.

This critical value of V_m and the value of V_s existing before the gain transfer define a receive to transmit, R TO T, transition point. In order for the set to return to the R state, V_m must be reduced considerably below the critical value, because of the removal of loss in CVL resulting from the

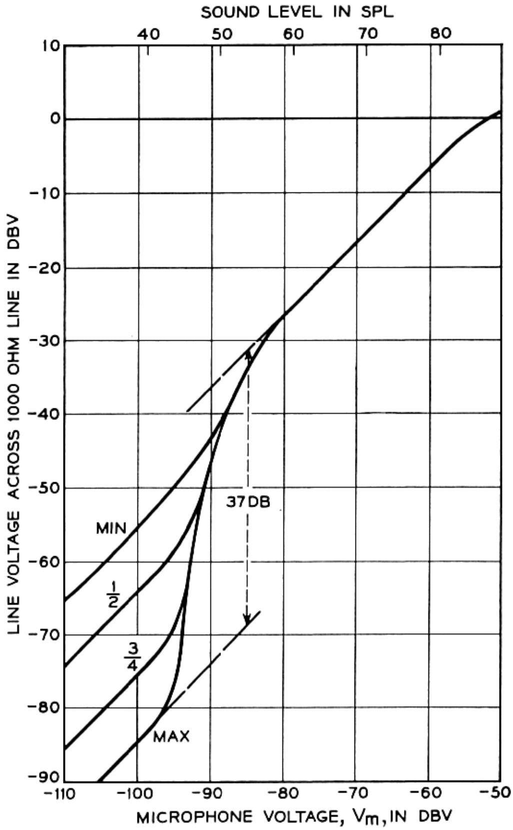

Fig. 5 — Transmit switching characteristic, with volume control position as parameter.

lower loudspeaker voltage at the input of the switchguard. At a second critical value of V_m, in this case -84 dbv, the set returns to the receive state. These values of V_m and V_s determine the T TO R transition point with $V_s = -20$ dbv. The hysteresis effect helps maintain one direction of transmission until a brief interval of low-level speech or a pause occurs in the signals passing through the channel having the higher gain, or until a deliberate attempt is made to interrupt by increasing the signal level in the other channel. The magnitude of this hysteresis effect depends upon the amount of gain switched, as determined by the volume control setting.

The R TO T transition curve of Fig. 8 shows the relationship between V_m and V_s at the transition points. Over a range of 30 db this relation-

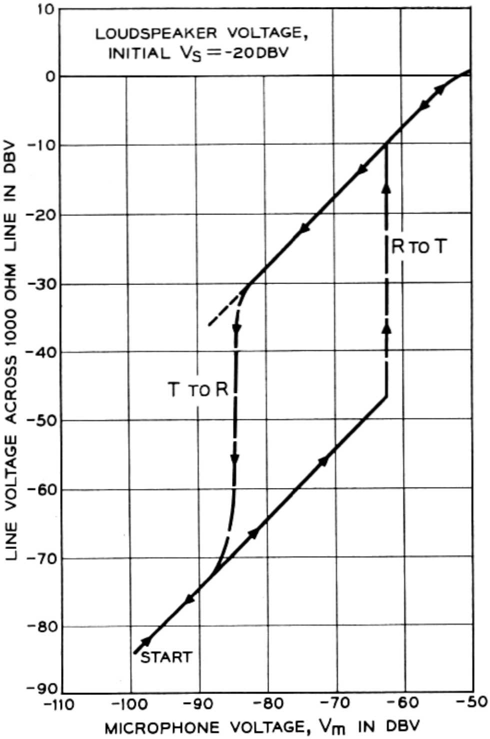

Fig. 6 — Transmit characteristic with receive signal V_s ; NOGAD inactive, maximum volume control setting, V_s applied first.

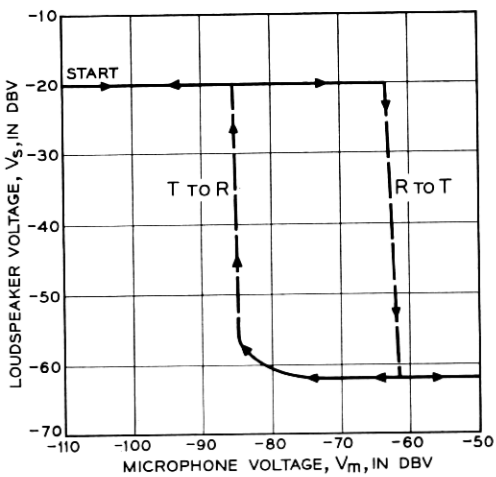

Fig. 7 — Receive characteristic with microphone signal V_m ; NOGAD inactive, maximum volume control setting, V_s applied first.

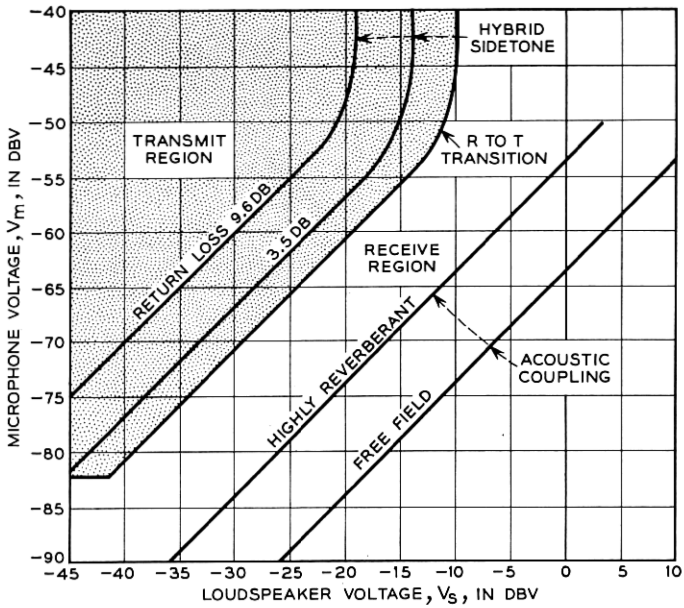

Fig. 8 — R TO T transition, with loudspeaker 3 feet from microphone on table.

ship is linear, and differential switching takes place. At its lower end the curve turns toward the horizontal, which means that V_m is constant for all values of V_s less than −42 dbv. For these low values of V_s, the switchguard does not produce enough current to increase the loss of CVL. At the upper end, overloading of amplifier A_M turns the R TO T transition curve toward the vertical. The significance of this is that, for steady loudspeaker voltages greater than −10 dbv, the set will not switch into the transmit state, regardless of the microphone input level. The electrical noise on the telephone line corresponding to this level, at maximum volume setting, is 60 dbrn,* which would only exist as a result of trouble conditions on a line. When receiving high-level speech, the loudspeaker voltage will exceed this value momentarily, but, due to the fluctuating nature of speech, high values are quickly followed by low values or pauses, during which speech at the microphone can switch the system into the transmit state.

Fig. 8 also shows the acoustic coupling and hybrid sidetone relationships. The two acoustic coupling lines depict the microphone voltage V_m produced by the loudspeaker voltage V_s for a highly reverberant (about

* dbrn = decibels above reference noise.

1 second reverberation time) and a free-field condition. The data represent the highest values observed over a 200-cps band centered at 1000 cps. Since only the microphone, the loudspeaker, and the air path are involved, this coupling is linear over the working range of voltages. The hybrid sidetone curves represent the loudspeaker voltage V_s produced by V_m as a result of hybrid coil unbalance, with the volume control at maximum and the set held in the receive state by opening the dc path of the variolossers. At high values of microphone voltage these sidetone curves are no longer linear and, because of overloading in amplifier A_M, turn towards the vertical in much the same fashion as does the transition curve. The sidetone curve for a return loss of 9.6 db represents a 2-to-1 unbalance between line and network impedance, while the return loss of 3.5 db represents an unbalance of 5 to 1.

For proper switching performance, the R TO T transition curve should lie above the acoustic coupling line but below the hybrid sidetone line.[4] If the transition curve touches or lies below the acoustic coupling line, receive blocking will occur. This means that the set will go into the transmit state instead of remaining in the receive state, because the microphone output resulting from the loudspeaker signal overpowers the switchguard. On the other hand, if the transition curve lies above the hybrid sidetone curve, transmit blocking will result; that is, the set will go into or remain in the receive state because the switchguard action of the sidetone loudspeaker voltage prevents switching into the transmit state.

In the 3A Speakerphone the margin between the transition curve and the acoustic coupling line has been made large, so that there is considerable freedom in positioning the loudspeaker and the microphone with respect to each other and with respect to walls and furniture. Less margin is provided between the transition curve and the hybrid sidetone curve, but conditions influencing this margin are better controlled. The return loss of 3.5 db, which still gives a 4 db margin, occurs only for an extreme condition, such as having three sets bridged on a loop. Aided by the increase in the total loss of TVL and RVL when transferring from the receive to the transmit state, as shown in Fig. 3, it is found that, for the transient signals of speech, hybrid sidetone will not cause transmit blocking with either the line terminals open or short-circuited.

VI. TRANSIENT SWITCHING PERFORMANCE

Both the speed with which the channel gains are switched in response to increases of speech levels and the duration of gain holdover after decreases of speech levels must be chosen to minimize clipping of the initial syllables, final syllables, and weaker syllables of a speech burst. In the 3A Speakerphone, a suddenly applied microphone voltage V_m of -70 dbv requires approximately 20 milliseconds to build up the line signal to within 6 db of its final value. This build-up time was selected as a compromise between slight initial clipping and too much sensitivity to impact room noise. The rate of gain change during a switching cycle is also important. By using a high rate of gain change when the gain is low and then decreasing this rate as the gain approaches its maximum, the "swishing" effect due to the rise and fall of the transmitted background noise is reduced.

The decay characteristic is shaped so that, after a microphone signal stops, the total time to revert to the receive state is approximately 300 milliseconds. In the first 115 milliseconds the transmit channel gain decreases about 5 db, and thereafter it falls at the more rapid rate of 0.17 db per millisecond. It has been found that a shorter decay time curtails the weak consonants at the ends of words and produces an undesirable expansion or "pumping" action on speech. On the other hand, a substantially longer decay time slows down the reply of the other party to an objectionable degree.

The time constant of the switchguard path, determined by τ_G in Fig. 2, is similar to that of the control path. The decay time of the switchguard is of particular interest because switching performance when receiving in a reverberant environment is dependent upon it. This decay time of the switchguard, which is the recovery time of the gain of the control path, occurs at an essentially constant rate of 0.2 db per millisecond. It will completely prevent any switching from R to T in rooms having reverberation times up to 0.5 second, while the speech sounds are decaying in the room. Actually, because of other factors, satisfactory operation is obtained in rooms with reverberation times approaching 1 second. These are: (a) speech sounds do not end abruptly and their decay time adds in part to the decay time of the circuit; (b) prolonged reverberation causes partial operation of the NOGAD circuit and delays the gain increase of the control path; (c) the transient signals of incoming speech do not build up the reverberant sound to its full steady-state level.

When one party replies quickly after the other has ceased talking, the over-all transfer time, determined primarily by the release actions of the control path and the switchguard path, is dependent on the relative levels of the signals in each channel. This transfer time becomes longer when the signal from the replying party is weak with respect to the signal from the party relinquishing control. On local calls, for which low volume control settings are adequate and less gain is switched, the receive replies can obtain control more quickly, and thus more rapid interchange is obtained, especially for the case of overlapping speech. The receive channel normally requires approximately 250 milliseconds to return to full gain under typical conditions. The maximum time is slightly over 300 milliseconds and the minimum is approximately 150 milliseconds for normal local speech levels at the microphone. In the transmit direction, the differential attack time is even more dependent upon relative signal levels, varying from 250 milliseconds when local speech levels are low with respect to the received loudspeaker speech levels to 20 milliseconds when the reverse conditions occur. A typical value can be considered to be 150 milliseconds.

VII. NOGAD CIRCUIT

To avoid noticeable clipping of the transmitted signal under quiet conditions, a sound level of approximately 45 db SPL initiates switching into the transmit state. This is less than the noise level produced by fans, air conditioners, and street and hall traffic at many locations where it is desired to use a speakerphone. Therefore, the possibility exists that the noise would block the receive channel. To prevent this, the noise-operated gain-adjusting device, NOGAD, passes a direct current, correlated with the noise, through CVL, and thus lowers the gain of the control circuit and raises the sound pressure needed to initiate switching. As people instinctively talk louder under noisy conditions, this introduces little or no switching degradation.

The NOGAD circuit recognizes the difference between the speech and the noise signals at the microphone on the basis that speech fluctuates more rapidly than most types of noise. The timing circuit T_N, shown in Fig. 2, consists of an *RC* circuit, R_{N_1}, C_{N_1}, with a time constant of about 1 millisecond, coupled to another circuit, R_{N_2}, C_{N_2}, having a build-up time of about 4 seconds and a short decay time, effected through the diode and R_{N_1} of about 100 milliseconds. Thus, when a signal consisting of both speech and noise exists at the microphone, the voltage across C_{N_2} is held closely to the nearly constant voltage component across C_{N_1} caused by the noise, and is little affected by the rapidly fluctuating voltages due to speech. The voltage across C_{N_2} then produces a direct current in CVL through a diode gate. This general method of discriminating between noise and speech is known in the echo suppressor art.[7]

With the circuit simply as described, an excessive time would be required for the set to adjust to room noise conditions after it is first turned on. To eliminate this delay, a precharging circuit, not shown in Fig. 2, of approximately 4 db. Some of this advantage undoubtedly accrues from binaural listening with the speakerphone. Fig. 11 shows the receive frequency response. It is characterized by a small peak at 450 cps, a sharp cut-off near 300 cps, and a more gentle fall off beyond 2200 cps. The sharp cut-off near 300 cps, assures good discrimination against noise induction from 60-cycle power and is not detrimental to articulation. The high-

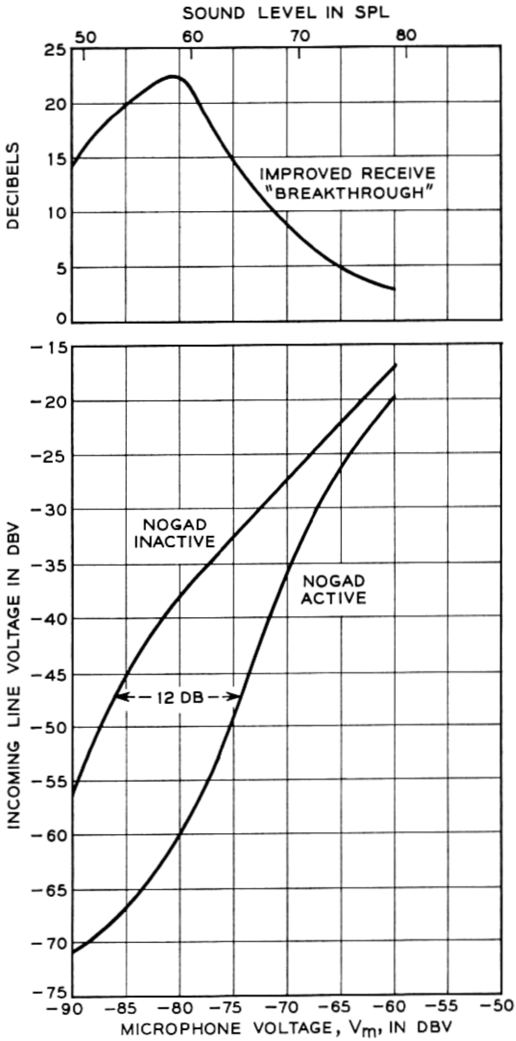

Fig. 10 — Effect of NOGAD on T TO R transition at maximum volume control setting.

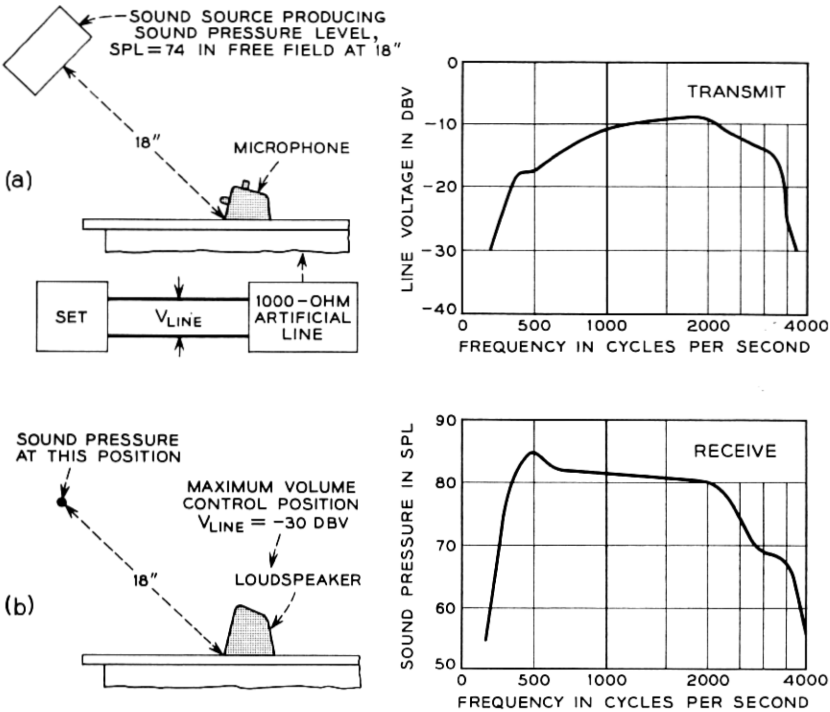

Fig. 11 — (a) Transmit and (b) receive frequency response.

frequency fall-off is caused by an acoustic interference between the direct sound wave and the wave reflected from the table top on which the loudspeaker rests. This loss is partially compensated for by sound diffraction and resonance effects occurring near the listener's ear.[8]

IX. CONCLUSIONS

The application of voice-switched gain in the 3A Speakerphone eliminates the talker echo and singing problems encountered in hands-free telephones having fixed gain, and provides very satisfactory transmission in the acoustic environment normal for homes and private offices. By careful design, the gain-switching action has been kept free of objectionable clipping or blocking, and is substantially unaffected by ordinary room noise. On most local telephone connections, the amount of gain which is switched is so small that the gain changes are barely noticeable. Finally, no adjustments are required for the noise and reverberation conditions most likely to be encountered.

X. ACKNOWLEDGMENTS

The authors wish to express their appreciation to D. Mitchell for many helpful discussions of fundamental problems, and to those other persons who participated in this development work.

REFERENCES

1. Clemency, W. F., Romanow, F. F., and Rose, A. F., The Bell System Speakerphone, A.I.E.E. Trans., Pt. I, **76**, 1957, p. 148.
2. Emling, J. W., General Aspects of Hands-Free Telephony, A.I.E.E. Trans., Pt. I, **76**, 1957, p. 201.
3. Gardner, M. B., A Study of Talking Distance and Related Parameters in Hands-Free Telephony, B.S.T.J., **39**, 1960, p. 1529.
4. Busala, A., Fundamental Considerations in the Design of a Voice-Switched Speakerphone, B.S.T.J., **39**, 1960, p. 265.
5. Stansel, F. F., N1 Carrier—Selection of Varistors for Use in Compandors, Bell Labs. Rec., **31**, 1953, p. 501.
6. Fracassi, R. D., The Compandor for N1 Carrier, Bell Labs. Rec., **31**, 1953, p. 452.
7. Wright, S. B., and Mitchell, D., U.S. Patents Nos. 1,814,017 and 1,814,018, July 14, 1931.
8. Wiener, F., and Ross, D. A., Pressure Distribution in the Auditory Canal, J. Acoust. Soc. Am., **18**, 1946, p. 401.