from *Electrical Communication*, Vo. X, No. 1, July 1931

The Single Side-band System Applied to Short Wavelengths

By A. H. REEVES

General Advantages of Single Side-band

THE theory of the "single side-band" principle of telephony was worked out by John R. Carson some years ago, and was first applied to carrier communication on wire lines, its chief advantage in this connection being that twice as many speech channels can be put into the same total band width as when using direct modulation with both side-bands. Since that date it has also been applied with considerable success to radio telephony on long wavelengths, in particular to the long wave New York-London commercial telephone circuit. It is, in fact, difficult to see how this circuit could be operated successfully without using the single side-band method, since, on long wavelengths, it would be a matter of some difficulty to obtain an antenna resonance curve passing both speech side-bands.

In radio work the halving of the band width given by the single side-band system assumed considerable importance as soon as the ether began to get crowded. An additional advantage for radio work is the increased efficiency due to suppressing the carrier and one side-band.

With the high power vacuum tubes now in use, at any rate of the water- or oil-cooled types, the factor limiting the transmitter output is the peak current or peak voltage on the plates of the tubes in the last stage of the transmitter. When this condition applies, it can easily be shown that, for a given peak power, an antenna gain in signal-to-noise ratio of at least 9 decibels is obtained by using the single side-band system, as compared with the use of transmitted carrier and two side-bands. A short analysis showing this result is given in Appendix I.

A further advantage, important in the case of a high powered transmitter, is the reduction of the power consumed when the carrier is suppressed. In the case of the low power modulation system, followed by stages of high frequency amplification, the last amplifier stage at least is usually of the class-3 type, i.e., the tubes in the last stage are biassed nearly to "cut-off." The anode currents in these tubes, therefore, are quite small in the absence of modulation, rising to peak values at intervals during speech, giving an average power consumption considerably less than that of the transmitted carrier method, where the anode currents do not change during modulation.

Difficulties of Application to Short-wave Work

The single side-band method has never yet been applied commercially to short wavelengths. The reason for this delay in what might seem to be an obvious application of the older art to the wavelengths now in use is probably twofold. In the first place, at any rate until recently, the saving of band width has not been a very important consideration on short wavelengths, as the total band width available at frequencies of the order of 10 to 20 megacycles is many times greater than that which was available on the long wavelengths.

The second reason has probably been the technical difficulty of obtaining local oscillators to re-supply the carrier frequency at the receiver in a sufficiently good state of synchronism with the suppressed carrier at the transmitter. If the re-supplied carrier frequency differs from the original carrier by more than about 20 cycles per second, noticeably bad quality—even when judged by the standards of "commercial" speech—inevitably results. At frequencies of the order of 60 kilocycles per second the synchronising problem presents no difficulty, necessitating merely a precision of one part in 3000. Any good local oscillator will give this stability of frequency quite easily over fairly long periods of time.

When the short-wave case is considered, however, it will be seen at once that the problem is much more difficult. On a wavelength of 15 metres the precision required amounts to one part in one million, a degree of frequency stability quite difficult to obtain on a "commercial" basis, even when using the most modern methods. It is, in fact, this synchronising problem which is the chief technical difficulty to be overcome in applying "single side-band" to short waves.

Advantages in Short-wave Work

Let us now consider some further advantages possessed by the single side-band method when used on short wavelengths which do not apply to the older long-wave case.

FADING

Short-wave fading may be divided roughly into two types:—

(1) Synchronous fading, in which all frequencies throughout the particular band width rise and fall simultaneously in amplitude.

(2) Selective fading, where the rise and fall in amplitude at the various frequencies in the range is more or less random.

With synchronous fading, without a local carrier oscillator, the extent of the resulting fade of the audio frequencies is double that when a local carrier oscillator is used, the fading being expressed in decibels. (This can easily be seen by referring to the expression for the audio voltage given in Appendix I.) This is a fairly important advantage, though not so great as it might at first appear. Modern practice in short-wave receiver design is such that the synchronous fading is already very largely taken care of by the use of some kind of automatic gain control. The effect of such a device, operated automatically by the carrier amplitude, is to change the gain of the receiver inversely with respect to the signal strength received, so that the resultant speech has a very nearly constant strength.

With an efficient automatic gain control practically the only noticeable effect of synchronous fading (except when it is very deep) is a change in the strength of the background noise. This noise fluctuation may, however, prove quite troublesome to a listener, so that the carrier suppressed system, which halves this noise fluctuation, has an advantage in this respect.

DISTORTION

It is, however, during selective fading that the suppressed carrier method gives the most noticeable improvement. By the older method the carrier frequency fades out almost completely for intervals of time sometimes approaching half a second, while the side-bands remain. The output of the speech detector, being intermodulation products of all the frequencies at the detector input, then shows second harmonics and other terms which may be considerably stronger than the fundamental speech frequencies. The resultant distortion is sometimes quite serious, this bad effect being most noticeable when some type of privacy system involving a frequency inversion is used.

This will be made clear by reference to the following example: Take the case of a simple inversion system in the range 300 to 3,000 cycles per second, in which 300 cycles is transmitted as 2,700, and vice versa. Consider a speech component at 1,900 cycles per second; after inversion we shall have 1,100 cycles, with which the transmitter will be modulated. In the receiving detector, when the carrier fades out selectively, the second harmonic at 2,200 cycles may be much stronger than the fundamental at 1,100; this resultant 2,200 cycles will be transformed into 800 cycles by the receiving inverter. We will have then at the receiving terminal a small fundamental frequency of 1,900 cycles, accompanied by a strong unwanted frequency at 800 cycles.

We must now take into account the fact that the ear seems to be disturbed much less by true harmonics or true intermodulation products of the original speech frequencies than by any other frequencies not having this relation. This effect is possibly due to the fact that the ear itself produces true intermodulation terms, and as the response of the ear is logarithmic, a few more harmonics do not have much effect. However this may be, experiment shows that any system producing undesired frequencies not having true harmonic relationship with the original speech, causes such noticeable distortion that a simple inversion system cannot be used successfully on short-wave radio links when there is serious selective fading.

It may be of interest to note here that to overcome this difficulty on the Madrid-Buenos Aires link, the International Telephone and Telegraph Corporation abandoned the simple inversion system, and is now using instead a displacement of the side-bands by 3,000 cycles away from the carrier. By this method the original amount of secrecy is maintained, but the harmonics and intermodulation terms produced in the receiving detector and elsewhere are no longer in the audible range. Displacing the side-bands in this way, while giving reasonable privacy with good quality, has the disadvantage of requiring double the usual band width.

The same remarks also apply to any secrecy scheme, in which frequencies are produced in the output of the receiving detector.

INCREASED SELECTIVITY

A further problem which is not solved by the single side-band method itself, but which is easily solved at the same time, is that of increasing the selectivity of the receiver approximately to the theoretical limit. In order to receive the side-bands at "commercial" speech quality on the double side-band system, it is necessary, on theoretical grounds, to transmit and receive a band width of slightly less than 6 kilocycles. To receive the whole of this band, and this band only, requires, of course, a good stability of the transmitted carrier frequency and equal stability, accompanied by very accurate tuning, of the receiver.

At present, the best commercial receiver is designed for a band width of about 8 kilocycles, a margin of 1 kilocycle on each side of the transmitted band being necessary to allow for slight frequency changes in the transmitter and slight changes of tuning at the receiver. There are many commercial short-wave receivers in use, moreover, of which the band width is nearer to 12 or 15 kilocycles. It is true that this increase of band width does not appreciably increase the receiver background noise level, as the only noise needing consideration, as mentioned above, is that due to the noise impulses beating with this carrier; and the impulses spaced more than 3 kilocycles from the carrier produce audio frequencies which can be cut out at a later stage by introducing a low pass filter. Reducing the band width to the theoretical limit, however, may greatly reduce interference due to other stations, and thus make possible closer spacing between adjacent channels.

INTERFERENCE WITH ADJACENT CHANNELS

Another advantage of the single side-band system is that the interference due to a side-band on receivers tuned to adjacent channels is less noticeable, in general, than the steady beat notes produced by neighbouring carriers.

EXTRA GAIN WHEN SELECTIVE FADING IS PRESENT

Lastly, as pointed out in Appendix I, during selective fading conditions a further gain in signal-to-noise ratio of up to 3 db. is frequently obtained.

A Commercial System

The object of this article is to describe one of the possible single side-band systems for use on short wavelengths, and to describe experiments carried out by the International Telephone and Telegraph Laboratories in conjunction with the Laboratories of Le Matériel Téléphonique, Paris, during the past year with this system, between the International Telephone and Telegraph Commercial Transmitting Station at Pozuelo del Rey, near Madrid, and an experimental receiving station at Trappes, near Paris. It is not claimed that the system to be described is by any means the only possible one, or indeed, that it is necessarily the best system for commercial telephony. Only the future can determine the relative merits of the various possible methods. It is claimed, however, that a commercial system has been worked out, and that the results so far obtained have been sufficiently successful for a fairly detailed description to be of interest.

A short general survey of the problem encountered will first be attempted, together with reasons for choosing, as a first experiment, the particular method detailed below. It is believed that in these experiments, synchronisation of the locally supplied carrier has, for the first time, been achieved automatically and for long periods of time. There are certain problems encountered in single side-band working which do not arise with the normal double side-band, transmitted carrier method, such as the reduction of unwanted frequencies, due to intermodulation in the transmitter to a sufficiently small value; these difficulties, however, are well known, and have already been studied during the single side-band work on long waves. For this reason, they will not be referred to in detail here. The fundamental new problem presented by short-wave single side-band is that of synchronisation, which, as already shown above, is not a difficult matter with long wavelengths, but with wavelengths of the order of 15 metres, presents a serious problem.

The Problem of Synchronisation

INDEPENDENT OSCILLATOR METHODS

There are two means of attacking the general synchronising problem. The first is to use at the transmitter and the receiver, independent oscillators of which the frequencies are so stable that when once they are adjusted to differ by only one part in a million, they will maintain this relation for comparatively long periods of time. This method, although by no means impossible, raises a somewhat difficult design and operating problem, even if the most modern means are employed.

It is true that by using the best modern quartz crystal oscillators, with accurate temperature control, this frequency stability has been obtained and even exceeded in the laboratory, but it is quite a different matter to maintain this high accuracy under the conditions of commercial operation. The same remarks apply to the other available frequency standards of high precision, *e.g.*, magneto-striction oscillators, tuning forks, oscillators automatically controlled by precision clocks, and the best purely electrical oscillators, accurately compensated for frequency changes. A solution is probably possible by all these methods, but it is not considered that such solutions would be convenient from a commercial standpoint.

PILOT METHODS

The second general method is by the use of some form of pilot signal transmitted in addition to the side-band and used automatically to synchronise the carrier supply oscillator at the receiver.

Let us consider in detail the possibilities of this method. Pilot signals may be divided into two classes:

(1) A pilot signal transmitted only during intervals when speech is absent.
(2) A continuously transmitted pilot signal.

The first type permits the use of the maximum energy supplied by the transmitter for the signal, but has the serious disadvantage that the resulting intervals during which there is no pilot may be longer than the minimum interval required to give the desired frequency control.

Any signal in the second category must fulfil two conditions:

(1) The peak voltage must be small compared with the peak voltage of the side-band; otherwise the available power in the side-band is appreciably reduced, and one of the chief advantages of the single side-band method begins to be lost.
(2) The presence of the pilot signal must not appreciably increase the band width required by the side-band.

We have, then, two classes of pilot signal fulfilling these conditions:

(*a*) A pilot signal at low frequency, by which is meant either a continuous wave modulated at low frequency, or two or more continuous waves beating together at a low frequency used as a synchronising signal.

In the term "low frequency" is included here the frequency given by a train of impulses. The resulting low frequency may be used in a number of ways, *e.g.*, an appropriate harmonic of the L.F. may be used as the required carrier, or the beat note may be used to drive a synchronous motor generator, the output frequency of which is suitably multiplied, and supplies the local carrier; but whatever method is used, the stages of frequency multiplication required make this system rather cumbersome for operating conditions, and furthermore, a small phase change in the low frequency gives some hundreds of cycles change in the high frequency. Also, a beat note dependent on the product of two or more transmitted continuous waves is more subject to synchronous fading than is a high frequency control signal.

(*b*) A pilot signal, consisting of one or more continuous waves of radio frequency, used directly to give or control the desired local carrier.

The continuous wave or waves may conveniently be at a frequency just outside the side-band, *e.g.*, 300 to 400 cycles on either side of the speech side-band. Or, if desired, a little of the original carrier may be transmitted, and used as the pilot signal. In any case, however, the peak power in the pilot signal should be at least

15 db. below the peak power of which the transmitter is capable. The utilisation of a signal of such weak power is not difficult from a signal-to-noise standpoint, owing to the fact that the signal is of one or more steady frequencies; circuits of very small band-widths may, therefore, be used to select them.

The band-width required for the side-band is about 3,000 cycles; but when choosing the band-width necessary for the pilot channel we have only to consider the maximum probable sudden fluctuation in transmitted frequency, *i.e.*, the fluctuation during the period in which the synchronizing circuit has not had time to take up this new stable position. A sudden fluctuation of 15 cycles per second is probably an outside limit, therefore, in the case of a reasonably good transmitter, so that the band-width required for the pilot signal is of the order of 30 cycles per second, *i.e.*, just one hundredth of the range required for the side-band.

As is pointed out in Appendix I, other things being equal, the energy in the background noise is proportional to the band-width, so that the ratio of signal-to-noise energies in the case of a pilot signal at a level 20 db. below the side-band will be just the same as the signal-to-noise ratio of the side-band itself. As the latter ratio must be, say, 10 to 15 db. for reasonably good speech, it is evident that the pilot will also be received under conditions in which the background noise does not prevent its utilisation.

Ways of Using a Pilot Signal

Some different possible ways of using such a pilot signal will now be considered. As is shown in the sketch of Fig. 1, the possible methods can again be divided into two classes:

(1) The selection, amplification and use of the pilot as the local carrier frequency, either directly in the case when some of the original carrier itself is transmitted, or by adding or subtracting the required audio frequency.

(2) The pilot may be used to synchronize a local oscillator.

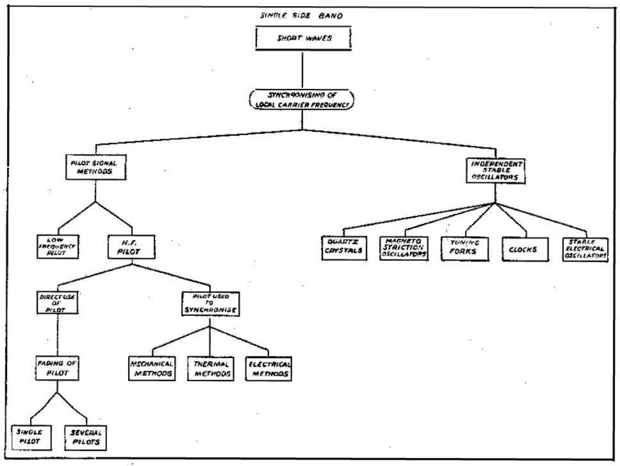
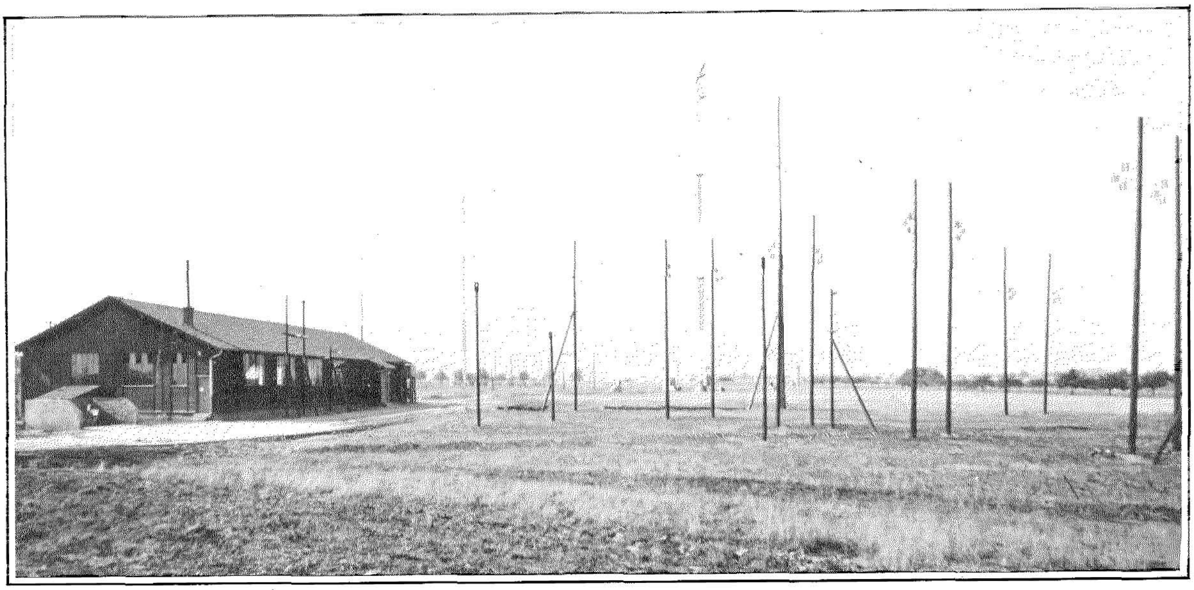


Figure 1

In considering the relative merits of these two methods, an inherent problem in short-wave work presents itself, viz., that of selective fading. Occasions will frequently arise when any one given frequency will fade out below the noise level, for periods of time sometimes approaching half a second in length. Therefore, if only one pilot is used, some form of time constant must be added, so that the locally supplied carrier will continue at the desired frequency during the fading periods. In practice this has been found to be one of the chief difficulties in the whole problem. When using the pilot signal directly, this time constant must take the form of a resonant circuit of very low damping—extremely low, in fact, as the resonant time constant required must be greater than half a second. Such a resonant circuit is, of course, very difficult to obtain, even when quartz crystals are used as the resonant element.

Even if such a resonant circuit were obtained, a large number would have to be used in parallel, with the resonance curves overlapping in order to cover the total probable fluctuation of the transmitted frequency. As this fluctuation must be assumed to be several hundred cycles in extent, and as the width of each resonance curve is less than half a cycle, it is clear that this solution becomes so cumbersome as to be impracticable.

Considerable improvement would, of course, be obtained by the use of two or more pilot signals, *e.g.*, two—one spaced at each end of the side-band. During selective fading these two pilots will only fade below the noise level simultaneously, at rare intervals, and for moments of very short duration. If arrangements are made, *e.g.*, by means of a relay, to use the greater of these two pilots to give the carrier, the time constant of the resonant circuit may



Radio Field Laboratory at Trappes. Antenna and Transmission Building.

be considerably reduced, but even then the solution is probably not a very practical one, at any rate whenever selective fading is present.

AUTOMATIC SYNCHRONISATION OF LOCAL OSCILLATOR TO THE PILOT

Let us now detail the second way of using a high frequency pilot signal, *i.e.*, by allowing it to synchronise automatically a local oscillator. In this case it is clear at once that the problem of the time constant is very much simplified. The possible methods logically divide themselves into three classes:

(1) The methods relying on mechanical means.
(2) Methods using thermal effects.
(3) Purely electrical methods.

Before going into further detail we will consider the general applicability of the three methods. In comparing their relative merits, the first question that immediately arises is the value of the time constant required. This depends on whether one or more pilot channels are used; if one only is employed, the time constant of the frequency control circuit must be such, at any rate during the fading out of the pilot, that the oscillator frequency does not change more than 20 cycles during the maximum period of time that the fade is likely to continue—in this case about half a second. On the other hand the time constant must not be so long that it does not follow the fluctuations in the received frequency. These latter changes are due to two sources: first, accidental changes in the transmitter itself, and secondly, frequency changes due to the transmission path.

The first type of change with any transmitter having a good master oscillator is always very gradual; in fact, a rate of change of a few hundred cycles in one hour is the maximum that is tolerated by present practice. Let us take 10 cycles per second per minute as the upper limit for this rate of change. At present there is little data concerning the second type of fluctuation; however, it is probably quite rare for this change, due to rapid alterations in the length of the transmission path to exceed 10 cycles per second about its normal value, and for this change to take place in a period of time less than 1 to 2 seconds. Let us take 10 cycles per second as the upper limit of the rate of change of frequency, due to the transmission path. Inasmuch as the total frequency change due to this latter fluctuation rarely, if ever, exceeds 10 cycles per second we see at once that this effect may be neglected altogether from the standpoint of "commercial" speech. We may take, therefore, the rate of change of 10 cycles per second per minute as the maximum in practice that the time constant of the frequency control circuit must follow. While considering the other limit, the constant must be such that the frequencies can never change by more than 20 cycles in half a second.

It will now be clear that between these two limits an apparatus can be designed satisfactorily by any of the three methods, *i.e.*, mechanical, thermal, or purely electrical circuits. Some mechanical method, such as rotating a vernier condenser on the oscillator, in accordance with the frequency of the received pilot signal, is perhaps the most obvious solution. When comparing such a scheme with the other two methods it is difficult to see its superiority; in fact it has a number of incidental disadvantages, such as the noise pick-up in the receiver from relay clicks, motor noises, etc. It is probable also that the proper maintenance of such equipment by ordinary operators would be less satisfactory than the maintenance of a system working on either of the other two methods.

Let us consider next the thermal methods. One convenient way of doing this would be to cause the pilot frequency to change the temperature control point of the thermostat of the local oscillator, which in this case could be of the quartz crystal type. The thermal time constant obtained in this way could be given a suitable value in accordance with the two limits fixed above. Such an oscillator would have the advantage of extreme immunity from all changes of frequency, except the one desired. In many respects, in fact, such a thermal system has a lot to recommend it.

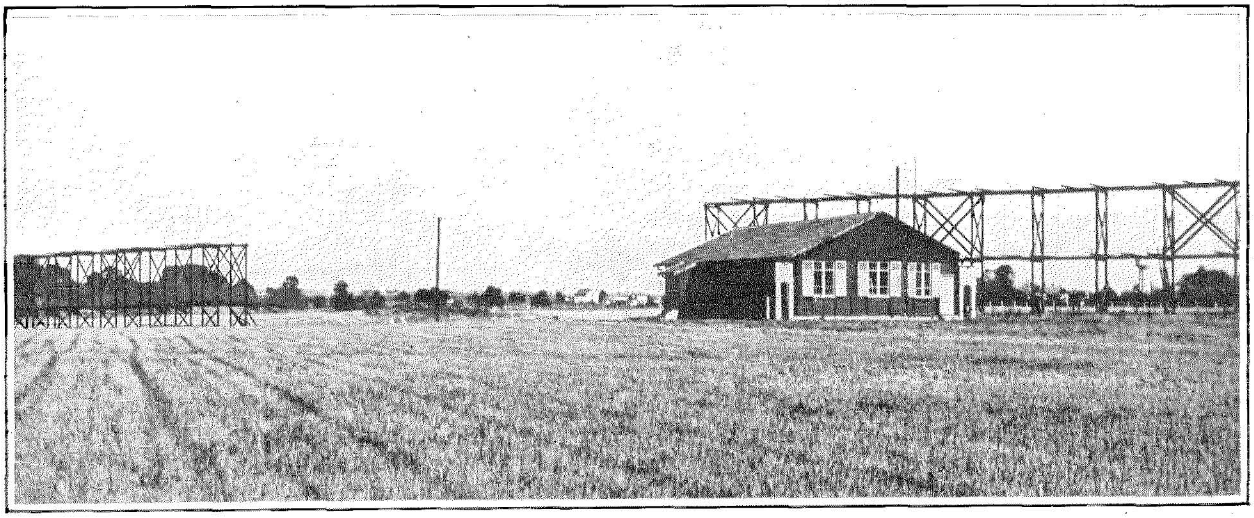
The frequency control can also be obtained fairly simply by purely electrical means, one of which, described in detail below, is the system used in the successful experiments which are now going to be discussed.

From the standpoint of "commercial" speech alone, the superiority of a purely electrical control method over the thermal system is doubtful. In the present case, the real reason that has governed the choice of system for preliminary experimentation is based on other considerations.

REASONS FOR CHOICE OF A PURELY ELECTRICAL METHOD

Considering the probable future development of radio during the next few years, it is almost certain that the single side-band method will be of utility, not only for commercial radio telephone links, but also for telegraphy, facsimile and television. It has been thought better, therefore, to start experiments on some system, the general methods of which will be applicable also at a later date to these other uses of radio.

In the case of these other applications it is not sufficient for the transmitted and received carriers to be synchronised with a precision of 20 cycles per second. This is true, of course, also, of high quality speech. For example, it is essential to avoid phase distortion, which can only be done by exact synchronisation of the frequencies and by keeping the phase variations between the two oscillators within narrow limits. This requirement at once rules out any method based on independent oscillators at each end of the link. It also practically rules out mechanical and



Radio Field Laboratory at Trappes. Antenna and Reception Building.

thermal methods of using a pilot channel, as will now be explained.

In short-wave transmission, even when the signal is received by one path only, the phases of the side-band components are frequently distorted backwards and forwards about their normal values by this transmission path, which is not of constant length. If, however, reception by more than one route is prevented, the result is simply a variable "delay" introduced between the transmitter and receiver, *i.e.*, the phase distortion of each component is directly proportional to its frequency. Taking the case of a side-band 20,000 cycles wide in the range of 20 megacycles, this means that the phase change will be constant within 0.1 per cent. for all the side-band components and the pilot, and may be sufficiently compensated for by an equal phase correction at all frequencies. To achieve this result, it is necessary for the phase of the re-supplied carrier to follow the phase variations of the pilot signal.

These fluctuations may sometimes be quite rapid; hence it is essential under these conditions for the synchronising time constant at the receiver to be quite short, *i.e.*, of the order of a small fraction of a second—a time constant value very difficult, if not impossible, to obtain by a mechanical or thermal means. By electrical methods, however, it is comparatively simple to make the value of the time constant very flexible. It is probably quite easy to have it short enough to follow all variations met with in practice.

Although the system used by the International Telephone and Telegraph Laboratories, and described below, is not one of exact synchronisation, it is a purely electrical method, which can be changed at a later date when required, in order to synchronise exactly. Another desirable feature is for the time constant to be variable when following exactly the fluctuations of phase and frequency. Due to the transmission path, the constant must be quite short, but on the other hand during the fades-out of the pilot signal the time constant must be long enough for the oscillator to maintain its frequency and phase within the required limits. Clearly this can only be carried out by means of a time constant, the value of which is varied automatically during these fades-out. This particular feature is only obtainable at all easily by purely electrical methods.

By a mechanical means, we could, of course, make the time constant longer or infinite during the fades, *e.g.*, by cutting off the mechanical control of the tuning condenser altogether by means of a clutch; but such devices operate relatively slowly, and in any case, as stated above, the time constant cannot easily be made short enough by such means when the level of the pilot signal is high. It is very difficult to conceive of a simple way to vary the time constant by thermal methods.

SOME PURELY ELECTRICAL METHODS

Let us next consider several possible purely electrical ways of changing the oscillator frequency or phase. Perhaps the most obvious solution is to use an iron core oscillator for changing the frequency by partially saturating the core. This method, however, is not suitable except at comparatively low frequencies.

A second possible means is to use the phenomenon possessed to some extent by any electrical oscillator, *i.e.*, a change of frequency with change of grid bias. By this method, however, it is difficult to obtain more than a very small frequency change. The third method, the one used in these experiments, is illustrated in Fig. 2.

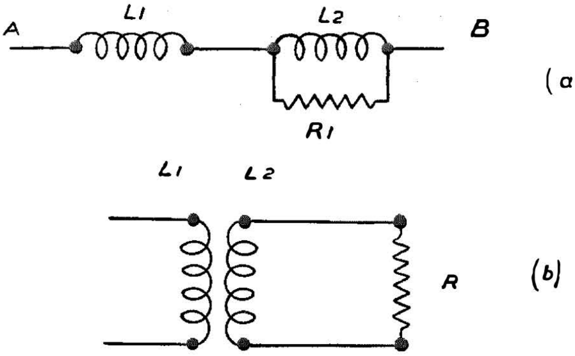


Figure 2

In 2 (*a*) a small inductance, shunted by a resistance, is connected in series with another inductance. It is clear by referring to Appendix II that the total effective inductance of the combination between points A and B depends on the value of the resistance. The same result applies, of course, to the circuit of 2 (*b*), except when the coupling between L-1 and L-2 is unity. When L-1 is the tuning inductance of a triode oscillator, the frequency of this oscillator will depend on the value of R. The resistance R may now be replaced by the plate-filament resistance of a second triode, in which case the oscillator frequency may, in general, be varied by changing the grid volts of the triode. It is this latter method that has been found the most convenient in practice in the experiments here described. A control triode of low plate resistance (about 2,000 ohms) is employed, and using a mean frequency of 500 kilocycles, a change of plus or minus 5% can easily be obtained.

The next question arising is how much frequency change is required. The answer to this depends, of course, on the degree of stability over fairly long periods of time (*e.g.*, 3 or 4 hours) of the transmitted frequency, and the average frequency of the receiving oscillator itself. If a quartz crystal be used to stabilise the first beating oscillator of the receiver (assumed to be of the superheterodyne type), and taking a reasonably good transmitter, experience has shown that a range of about plus or minus 3 kilocycles is sufficient to take into account the worst case. If the intermediate frequency in the receiver (or the first intermediate frequency in the case of more than two detectors) be taken as 500 kilocycles, this gives a necessary range for the oscillator variation of plus or minus 0.6 per cent. It is clear, then, that the control method explained above gives a range which more than amply fulfils the requirements.

DETAILS OF SYSTEMS ADOPTED

The frequency changing triode referred to above must now be controlled by the frequency of the pilot signal. The method of control, of course, depends on whether exact or approximate synchronisation is required. In these experiments the approximate case only has so far been attempted, and the circuit used for this is shown in its simplest form in Fig. 3. "A" represents the high frequency amplifier, first beating oscillator, first detector and first intermediate amplifier of the receiving set.

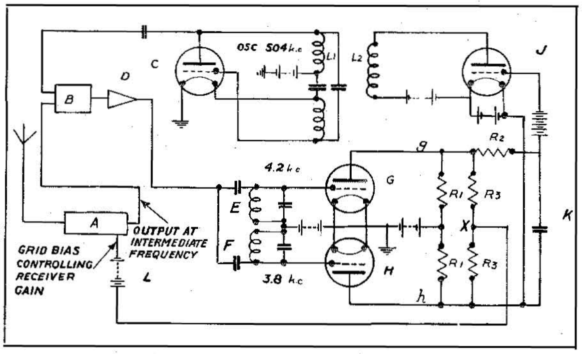


Figure 3

The resulting output at 500 kilocycles is applied to the rectifier B in combination with a little of the output of the synchronised oscillator C at which the average frequency is set (504 kilocycles); the output of rectifier B is passed through the low frequency amplifier D which has a resonance point at 4 kilocycles. The output of D is coupled loosely and equally to the two circuits E and F, tuning to 4.2 and 3.8 kilocycles respectively. The latter tuned circuits are connected as shown to the grids of the balanced rectifier system GH. The resonance curves of E and F are adjusted so that at exactly 4 kilocycles the losses due to the two circuits are equal and at a value of about 8 db. greater than the losses at the resonance points. If, then, the frequency of C is exactly 504 kilocycles, equal voltages will be applied to the grids of E and F, thus giving zero potential difference between the two plates of G and H.

When C differs from the incoming pilot by exactly 4 kilocycles, the potential difference between the plates of G and H is applied between the grid and filament of the control tube J, the plate coil of which is coupled to the resonant circuit of C. Assume now that for some reason the beat note between C and the pilot signal becomes slightly greater than 4 kilocycles; the plate current of rectifier G will now exceed that of H, thus increasing the negative grid bias on tube J, raising its plate resistance and hence lowering slightly the frequency of C—by this means tending to restore the beat note between C and the pilot to its original value of 4 kilocycles.

The essential principle of the synchronising action will now be evident; in practice it has been found easily possible by means of this circuit to cause the oscillator C to follow variations of the pilot signal amounting to plus or minus 5 kilocycles, the resulting beat note never differing by more than 20 cycles from the normal value of 4 kilocycles. Figure 3 also explains how the time constant is introduced into the frequency control circuit. The resistance R-2 is of the order of 5 megohms, and condenser K about 20 microfarads, giving a time constant of 100 seconds between impulsive voltage fluctuations across G H and the grid-filament of J.

Let us now consider another feature, shown in Fig. 3, which has been found essential to the successful operation of the system. It is, of course, impossible in practice for the characteristics of tubes G and H to be exactly similar; in other words, if the plate currents are exactly equal for one particular case in which the voltages across E and F are equal and at a certain level, the outputs of the rectifier tubes will not balance exactly when the voltage level on the grids is altered, although these two voltages may still be equal. Further, if exactly similar tube characteristics were obtainable, the exact balance condition between the two plate currents could never be used in practice. Some voltage difference is always necessary between *g* and *h* in order to give the desired frequency change on oscillator C.

If now the frequency remains unchanged, but the level of the pilot signal is changed, as occurs during fading, the potential difference between *g* and *h* will be changed correspondingly, with the resulting tendency to change the frequency of C; *i.e.*, the frequency of C will depend to some extent, not only on the frequency of the pilot signal, but also on its amplitude, a quite undesirable effect. To overcome this difficulty, due to the above two causes, an automatic gain control has been added. It is only necessary for a potential difference corresponding to the amplitude of the pilot signal (and not its frequency) to be used to control inversely the gain of the receiver in such a way that a very small change in this potential difference makes a considerable change in gain.

To achieve this result in the circuit of Fig. 3 the resistance R-3 is added, and the potential of the mid point X will be the average between that of *g* and that of *h*. When the beat is in the region of 4,000 cycles, therefore, at which point *g* and *h* change inversely and almost equally with frequency change, the voltage of X will depend almost entirely on the amplitude of the pilot and not on its frequency. X is connected to the battery L, which is used to counterbalance more or less the anode voltage of the rectifier tubes, thus providing a suitable controlling grid bias for one or more tubes of the receiver.

The gain control circuit is in fact very similar to that used to counteract the effects of fading in the normal double side-band type of receiver used by the I.T. & T. Corporation for short-wave work. By this means, when the pilot signal changes from a value only just above the noise level up to a level of 60 db. above this point, the 4 kilocycle volts applied to the grids E and F only change by about 10 per cent. The time constant of this automatic gain control circuit is made short enough to follow the most rapid type of fading met with in practice (neglecting, of course, the very rapid fading of more than twenty periods per second which sometimes occurs, but which is usually small in extent): a suitable value for this time constant has been found to be about one-twentieth of a second.

It is now evident that it is only when the pilot falls below the noise level that continuous control of the local oscillator is lacking. Taking the case when the pilot signal is reduced at the transmitter by 20 db. below the peak value of the side-band, and when conditions are such that the resulting signal-to-noise voltage ratio of the speech is of the order of 3 (a minimum value for a commercial circuit), the pilot at the receiver has rarely been found in practice to fall below the noise level for longer than one-tenth of a second, and as far as is known never longer than half a second. These figures, of course, are based only on results between Madrid and Paris covering a period of about five months and using a particular wave length (52 metres) with particular antenna systems at each end. It is thus difficult to say that the figures apply generally; this remains to be seen. The time constant of the frequency control circuit, made up of resistance R-2 and condenser K is sufficient to maintain the frequency of oscillator C well within the required limits during this worst case of fading lasting half a second.

Design of Complete Receiver

Having achieved sufficiently accurate synchronisation, the next question is the design of the receiver. It is clear that an ideal single side-band receiver should take account of the accurate frequency control available to reduce the band width passed by the speech circuits to the theoretical minimum of about 3 kilocycles. This can be done most conveniently by using the triple detection type of receiver, the first intermediate frequency being of the order of, say, 500 kilocycles, and the second intermediate frequency in the region of 50 kilocycles.

The first beating oscillator is controlled by a crystal, the second oscillator being controlled automatically by the pilot signal in such a way that the resultant second intermediate frequency never varies by more than 20 cycles. The band filters at 500 kilocycles have a margin of plus or minus about 4 kilocycles in addition to the width of the side-band plus pilot, thus giving a total band width of about 12 kilocycles; the final and high degree of selectivity is obtained by a band filter at the second intermediate frequency (50 kilocycles).

As the side-band frequencies at this stage have been properly stabilised, this second band filter can be accurately adjusted to cut off at exactly the points required to pass the side-band only. The receiver actually used in the tests described here is of such a design that the second intermediate frequency is, however, in the region of 20 kilocycles rather than 50, merely because 20 kilocycle apparatus was immediately available.

The next consideration in designing the complete receiver is the question of eliminating, so far as possible, the fading of the output speech. As has been remarked before, short wave fading may be divided roughly into two classes: synchronous and selective. During selective fading, in general, the average level throughout the whole side-band does not change considerably, and the gain of the receiver should, therefore, remain constant. When, on the other hand, synchronous fading appears, it is necessary to change the gain of the receiver inversely and, automatically to compensate for this.

Almost certainly the best method would be to use two or more pilot signals, *e.g.*, one placed on each side of the side-band. The average level of these two pilots will then give a good idea of the average level in the side-band itself, and, therefore, of the amount of synchronous fading present.

The frequency spacing of the two pilots may be so chosen that selective fading will not, in general, change their average very appreciably. If this latter average level were then used to control inversely the gain of the receiver used for the speech, it seems probable that the average speech output would remain substantially constant in most cases. This method, however, has not yet been tried out in the comparisons referred to. Another line of attack is based on the fact that, in general, selective fading is of short, and synchronous fading of longer, duration. If two separate automatic gain controls are used, one of very short time constant for the pilot channel only, as explained above, and the other of a time constant of the order of 30 seconds or more for the speech and the pilot channel, a fairly smooth speech level is found to result.

This type of smoothing is not perfect, however, as the assumption that the synchronous and selective types of fading have different time periods, is by no means always true. The outline of the complete circuit used is shown in Fig. 4. The constants are not, in every case, of the optimum values, but were determined by the apparatus that was immediately available.

The pilot signal used corresponds to a speech frequency of 3.47 kc. The reasons for this position of the pilot are two-fold:

(1) The maximum speech energy lies in the low frequency range below 500 cycles; it is, therefore, advisable to space the pilot as far as possible from these regions of maximum energy. In order to avoid an appreciable increase of band width, however, a distance of more than 400 or 500 cycles from one end of the side-band should not be exceeded.

(2) The audio output obtained in an ordinary receiver when listening to such a transmitter will give speech that is inverted and displaced by 400 cycles. Such an output is practically unintelligible, thus rendering the system private, although, of course, not secret, as an oscillating detector adjusted accurately at the correct frequency will always give straight speech. This latter objection, of course, also applies to any simple form of privacy system depending on inversion or displacement.

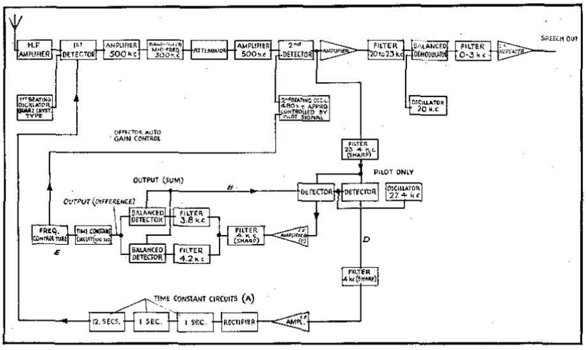


Figure 4

The pilot is normally at a level 20 db. below the overloading point of the transmitter. Owing to the very small band width used in the receiver for selecting the pilot, the resulting signal-to-noise ratio due to the pilot is about the same as that of the speech side-band, which is 10 or more decibels higher in level. It is thus easily sufficient to carry out any desired synchronising or gain-control processes. The necessary reduction of the peak value of the side-band to avoid overloading due to the presence of the pilot is very small (of the order of 0.1 db.) so that practically the full gain of the single side-band system is realised.

The automatic gain controls are similar to those already in commercial use in short-wave receivers developed by the Bell Laboratories and the International Telephone and Telegraph Corporation. The action consists simply of a rapid increase in grid bias voltage on a detector or amplifier in the earlier stage of the receiver whenever the D.C. output of the final detector to be controlled exceeds a predetermined limit. Such a change in grid bias reduces the gain of the tube in question until the final level is again at the correct value. In each case, between the output detector and the tube of which the gain is being controlled, there is inserted a time constant circuit of a value suitable to the particular function desired. The circuits A shown in Fig. 4 are built up in three sections in order to give a correct curve of output voltage against time, which is more suitable than can be obtained with a single section. The gains of the various detectors and amplifiers are adjusted so that the chief component of background noise in the final speech output is due to the grid circuit of the first high frequency amplifier; and the antenna is such that, except under unusual circumstances, the chief component of this first grid noise is caused by static. The same remark applies to the circuit of the pilot channel. The gains are such that, when no signal is being received, the static under conditions of minimum strength is beginning to affect the frequency control tube and also the two automatic gain controls. It is of no advantage to increase the gain beyond these limits, and it is evident also that this arrangement gives the best possible signal-to-noise ratio obtainable with the particular antenna used and under the given transmission conditions.

Synchronisation Limits

Even when receiving signals giving a final speech-to-noise ratio less than that required for a commercial circuit, but good enough for passing service messages, and even when the worst fading met with in practice is present, the pilot may still be kept 20 db. below the peak of the side-band, and the first intermediate frequency (500 kc. range) may change by about ± 5 kc. without changing the second intermediate frequency by more than 20 cycles. This limit clearly gives an ample margin. The transmitter frequency rarely changes by more than 1 kc. from its normal value, and a change of 2 kc. in the crystal beating oscillator is rare even when an accurate thermostat is not used.

The only other variable factor is the natural frequency of the 480 kc. oscillator when uncontrolled. This can easily be kept within one part in a thousand of its normal value, *i.e.*, 500 cycles. It is very desirable that the first beating oscillator be of a quartz crystal type, firstly to keep its slow frequency drift within the required limits, and secondly to avoid changes of frequency due to a change in signal strength when the latter is at a high level. On account, chiefly, of the action of the automatic gain control on the first detector, to which the beating oscillator is coupled, a change in frequency of the latter is, in general, caused by variations in the received signal, unless special care is taken to avoid this. The use of a crystal oscillator is not a serious objection, as a commercial receiver is not generally used for more than about three wavelengths.

Single Side-Band Transmitter

The block schematic of the transmitter used in these experiments is shown in Fig. 5. The problem was to evolve a transmitter design giving distortionless single side-band speech with sufficient stability but using, as far as possible, the existing double side-band equipment. The method of solving the problem may perhaps be of interest for illustrating the extreme flexibility of the low power modulation type of transmitter; it will be seen that the change-over from double to single side-band working can be carried out comparatively simply and rapidly.

The single side-band is obtained by modulation in three successive stages. The first two are at 19 and 250 kilocycles respectively, the modulators being balanced in each case, and the side-band being selected by appropriate filters. For the final modulation the existing harmonic generator of the transmitter was employed with the negative grid bias increased to a point such that, within certain limits, the voltage in the output harmonic was linearly proportional to the grid voltage. By this means, when the side-band output of the second balanced modulator is applied to the harmonic generator grid, this latter tube gives the second harmonic and acts as a modulator at the same time.

This third modulator is not balanced, and consists of a single tube, but a sufficient suppression of the carrier and the unwanted side-band is obtained by increasing the selectivity of the rest of the transmitter by sharply tuned circuits.

This type of transmitter circuit has been found to work quite well in practice.

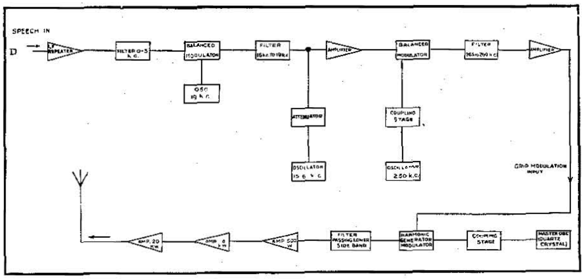


Figure 5

Tests Carried Out

During the period from April, 1930, to March, 1931, single side-band tests have been carried out by the International Telephone and Telegraph Corporation over the following three links:

(a) Buenos Aires to Madrid.
(b) Local tests at Madrid (Pozuelo to Griñon, 80 kilometres).
(c) Madrid-Paris.

As no single side-band transmitter had been installed at Buenos Aires, the tests on this link have consisted, so far, chiefly of experiments to check the automatic synchronising device, using the carrier of the ordinary double side-band transmitter as the pilot signal. The wavelength used was 15 metres.

During the second series of tests, viz., those between Pozuelo and Griñon, the complete system was tried out, using the provisional transmitter that has just been described. Although the distance was only 80 kilometres (between the Madrid transmitting and receiving stations of the Madrid-Buenos Aires link) the conditions were artificially made more or less similar to those prevailing on a commercial link by using directive antennae at both transmitter and receiver pointing in the wrong directions; by this means, the field strength and fading were found to be approximately those of the commercial Madrid-Buenos Aires link. For these local tests, a 30 metre wavelength was employed.

The final series of tests between Madrid and Paris, extending from November, 1930, to March, 1931, used the same transmitter as in the local Madrid experiments, but on wavelengths of the order of 30 to 52 metres. In each case the choice of wavelength was influenced solely by the transmission path conditions during the hours when experiments were possible. Owing to the exigencies of traffic, these hours were limited almost entirely to the period from 10 P.M. to 10 A.M.

Results of Tests

Buenos Aires to Madrid.—These tests, carried out with a receiver not yet in its final form, gave a maximum frequency difference between transmitted and re-supplied carrier frequencies of about 20 cycles per second, except during unusually heavy fading periods. No attempt was made to reduce the strength of the carrier below the normal value. These results were thought sufficiently promising, and showed that the system had good future prospects. On the other hand, the fact that adequate synchronisation could not quite always be obtained on account of severe fading, showed the necessity of some improved device to compensate for these severe fading periods.

Local Tests at Madrid.—Before the start of the second series of tests locally between Pozuelo and Griñon, the following two improvements were made:

(1) The first beating oscillator of the receiver was converted to the quartz crystal type as shown in Fig. 4.

(2) The automatic gain control B (Fig. 4) on the frequency control branch of the pilot circuit was improved by increasing the gain of the amplifier C so that any level of pilot signal between the measurement found in these local tests and the value only just above the noise would effect the frequency control tube equally. The output branch D of the pilot channel controlling the gain of the whole receiver through a comparatively long time constant had not yet been installed.

Test 1: Synchronism.—No long period stability tests of the synchronising device were attempted during these preliminary tests, but for periods of half an hour's duration a difference of 20 cycles per second between transmitter and receiver carriers was rarely exceeded.

Test 2: Speech Quality.—Except during the few occasions when the synchronism was not good enough, the speech quality was found to be quite satisfactory; in fact, during selective fading conditions, as had been foreseen, the quality was better than that given by the normal double side-band system.

Test 3: Improvement in Signal-to-noise Ratio.—This ratio was measured, both in the double and single side-band cases, at the final low frequency output by a standard volume indicator. As signal, a tone of 1,000 cycles was employed. Complete fades are, of course, frequently obtained in both the single and double side-band cases, so that exact measurement was difficult. An average value had to be estimated in each case, and the means of a number of observations taken. A further difficulty was the fact that, although the system was changed from double to single side-band, and vice-versa, as quickly as possible, the transmission conditions usually altered quite appreciably during the change-over periods. The results measured in this way showed a gain of 17 db. with the same peak power in the antenna in the two systems. This figure is considerably higher than the value expected theoretically: between 9 and 12 db. The cause of the discrepancy has not been satisfactorily explained.

Inasmuch as the type of fading is necessarily different when using the two systems, it is clear that the average value will not have exactly the same meaning in one case as in the other. It may be that the explanation lies here.

It is true that if the double side-band received carrier is only just above the noise level, the total noise received in this system will very appreciably depend also on the received band width, as explained in Appendix I.

As the band width of the receiver is about 20 kc. the maximum increase of the double side-band noise over that of the single side-band case would be

$$\sqrt{\frac{20}{6}} = 1.75 \text{ approximately}$$

i.e., about 4.5 db., giving a total gain from double to single side-band of about 16 db. In the tests referred to, however, the double side-band signal-to-noise ratio was never less than about 12 db., so that it seems that this latter possible explanation cannot apply.

Test 4: Speech Quality.—This was found again to be perfectly satisfactory, and better than during double side-band working.

MADRID-PARIS TESTS

Test 1: Synchronism.—In the absence of gain control D, Fig. 4, the synchronism was not always perfect, as the range of the fading (taking the minimum as the minimum value lasting more than half a second) frequently exceeded the control range of gain control B. To overcome this trouble, and at the same time to reduce the fading of the speech output, the separate control D was introduced, as shown in Fig. 4. By this means, the total range of the fading control on the signal was increased to a point such that the correct synchronising action never ceased for periods longer than about a quarter of a second.

The time constant E, Fig. 4, was quite sufficient to keep the oscillator frequency sufficiently constant during these short periods. At the beginning of March, 1931, during the stability tests on the complete system, covering ten consecutive nights, the re-supplied carrier frequency never differed from the suppressed carrier by more than 6 cycles per second, except for two or three moments when the transmitter was accidentally shut down—a result definitely better than had been hoped for.

Test 2: Speech Quality.—The received quality in Paris was found to be always as good as that obtained from the double side-band, and very definitely better during bad selective fading conditions.

Test 3: Signal-to-noise Ratio.—In these Madrid-Paris tests an improved type of signal was used. Instead of a single tone a noise-producing buzzer was employed giving audio frequency energy more or less evenly distributed throughout the whole speech spectrum. This type of signal was thought to be a much better imitation of speech from the point of view of fading than the single tone. The average result from a number of observations showed a gain of 12 db. when compared with the double side-band receiver with the automatic gain control off, and of 13 db. compared with the same receiver with its gain control on. This is evidently equivalent to increasing the power of the transmitter slightly more than 16 times. The peak powers in the antenna were carefully adjusted to be the same in the two systems. The curve showing one night's comparisons between the double and single side-band with the ratios plotted against time is shown in Fig. 6.

Test 4: Fading.—After the extra gain control D had been added the fading on the speech output was in general definitely less than the best that could be obtained with double side-band working with automatic gain control. Improvements showed that more particularly (as, in fact, all other advantages of single side-band working during bad selective fading) the usual automatic gain control on double side-band receivers is quite incapable of dealing properly with selective fading. When the carrier fades out the gain of the receiver is usually increased very considerably, giving a loud distorted speech output.

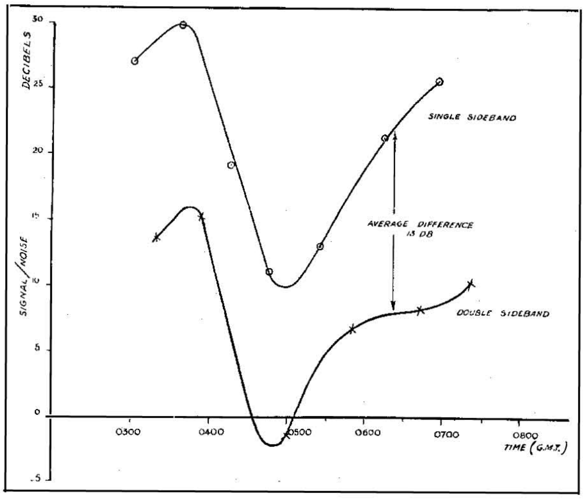


Figure 6

With the single side-band arrangements shown in Fig. 4, however, speech fading is very largely compensated for. The selective fading is usually found to be fairly rapid, so that the gain control D, working through a large time constant to the first detector of the receiver, has no time to operate. This is clearly the correct condition, as during selective fading the average power in the speech side-band is substantially constant. When synchronous fading occurs, however, it is usually of comparatively long duration, and thus operates control D, keeping the speech output fairly constant. The only way of making a really accurate comparison between two anti-fading devices is by a long series of observations over an extended period. From the results so far obtained, however, between Madrid and Paris it seems that the average fading, when using the present single side-band system, is about half that with the normal double side-band receiver, and during exceptionally bad conditions, the improvement is greater than this.

Test 5: Intelligibility—No accurate tests on the improvement of intelligibility were attempted, but on several nights rough comparisons were made by using short word lists, and changing over as rapidly as possible between the two systems. Improvement in the intelligibility was, of course, much more marked when the signal-to-noise ratio was low. One result which seems fairly clear, however, is that when the percentage intelligibility in double side-band is of the order of 20, the corresponding figure with single side-band working is usually between 70 and 80. In other words, a change from double to single side-band is capable of changing an uncommercial into a good commercial circuit.

Test 6: Stability of System.—As has been mentioned above, continuous tests, each lasting from two to three hours, were carried out on ten consecutive nights. Speech was used at regular intervals throughout all the tests, and the results showed a perfectly commercial circuit with good quality throughout the whole period, with no interruption whatever except during one or two short periods when the transmitter was shut down. It was actually unnecessary to touch the receiver at all throughout the tests, except for the preliminary tuning each night. Actually, however, when a frequency difference of 5 or 6 cycles was observed (usually once half-way through the night), the de-modulator oscillator was re-adjusted to restore the zero beat conditions.

Conclusion

The results obtained have definitely shown that one system of single side-band working, applicable commercially to short-wave links, has been evolved.

The author of this article wishes to acknowledge his indebtedness to Mr. C. W. Earp and Mr. H. T. Roberts, of the International Telephone and Telegraph Corporation, for their valuable assistance during the above experiments and the preparation of this paper.

APPENDIX I

Signal-to-Noise Ratio Improvement

In comparing the two systems let us assume equal peak voltages; and in the double side-band case, let the transmitter be modulated to 100 per cent.,

(1) DOUBLE SIDE-BAND CASE.

If the carrier is represented by

$$V \sin \omega t$$

then the peak voltage of each side-band is:

$$\frac{V}{2}$$

In the speech-producing detector of the receiver, let the audio voltage produced by beating each side-band separately with the carrier be represented by

$$\frac{KV^2}{2}$$

where K is, of course, a constant depending on the receiver, the transmitter, and the intervening circuit conditions. If now the phases of the side-bands with respect to the carrier are the same on arriving at the receiving detector as they were originally in the transmitter, the audio frequencies produced by each side-band separately with the carrier will be in phase in the detector output; the resultant is therefore

$$KV^2$$

Now let us consider the total background noise, the noise level being defined as that occurring when the incoming carrier is unmodulated. If the carrier energy in the speech detector is large compared with the noise energy picked up in the band width considered (this is always the case in "commercial" radio telephone circuits) then the only audio noise that need be considered in the output of the speech detector is that produced by noise impulses beating with the carrier. Thus the background noise voltage is:

$$KV\sqrt{\sum_{\omega-F}^{\omega+F} V^2}$$

where the noise input to the detector is

$$\sum_{\omega-F}^{\omega+F} V$$

and assuming in an ideal receiver that only frequencies from $(\omega-F)$ to $(\omega+F)$ are passed, where F is the highest audio frequency to be considered, and assuming also that the noise impulses arrive at random times. Hence the resulting signal-to-noise ratio R_1 is equal to:

$$\frac{V}{\sqrt{\sum_{\omega-F}^{\omega+F} V^2}}$$

(2) THE SINGLE SIDE-BAND CASE.

In order to give the same peak voltage at the transmitter as before, the tone which in the previous case gave 100 per cent. modulation must now be transformed into a single radio frequency of peak voltage 2V. Assume now that the peak voltage of the local oscillator re-supplying the carrier at the receiver is V_0; then the audio tone produced by the incoming side-band is $2KVV_0$ and the noise is:

$$KV_0\sqrt{\sum_{\omega-F}^{\omega+F} V^2}$$

Assuming that the receiver passes the upper side-band only in this case, then the signal-to-noise ratio R_2 will be:

$$\frac{2V}{\sqrt{\sum_{\omega}^{\omega+F} V^2}}$$

Assuming that the noise impulses are distributed evenly throughout the band width, then

$$\sum_{\omega-F}^{\omega+F} V^2$$

is double

$$\sum_{\omega}^{\omega-F} V^2 \text{ , so that } \frac{R_2}{R_1} = 2\sqrt{2}$$

giving thus a gain of 9 db.

In the case of short wavelengths, however, there is another fact to be considered. During fading, selective phase distortion frequency occurs in the transmission path, causing the phase of one side-band with respect to the carrier, in the double side-band case, to be more or less random as compared with the phase of the other side-band. In this case the average resultant of the two side-bands will give an audio voltage of

$$\frac{KV^2}{\sqrt{2}} \text{ rather than } KV^2.$$

R_1 is then:

$$\frac{V}{\sqrt{2\sum_{\omega-F}^{\omega+F} V^2}}$$

giving an improvement due to the single side-band system of 12 db. In practice a gain of 9 to 12 db is obtained, according to the transmission conditions.

APPENDIX II

Variable Reactance Circuit

Referring to Fig. 2 (*a*), we have the following expression for the impedance $(R+jX)$ between points A and B:—

$$R+jX = R_1\frac{\omega^2L_2^2}{R_1^2+\omega^2L_2^2} + j\left(L_1+L_2\frac{R_1^2}{R_1^2+\omega^2L_2^2}\right)$$