SUMMARY TECHNICAL REPORT OF DIVISION 17, NDRC

VOLUME 3

TRANSMISSION AND RECEPTION OF SOUNDS UNDER COMBAT CONDITIONS

OFFICE OF SCIENTIFIC RESEARCH AND DEVELOPMENT
VANNEVAR BUSH, DIRECTOR

NATIONAL DEFENSE RESEARCH COMMITTEE
JAMES B. CONANT, CHAIRMAN

DIVISION 17
GEORGE R. HARRISON, CHIEF

WASHINGTON, D. C., 1946

CONTENTS

Chapter 9

THE INTERPHONE

In many ways the interphone stands as the prototype for speech communication systems. The microphone, amplifier, and headset represent the indispensable elements: two electro-acoustic transducers joined by an intermediate transmitting network. The problems which arise are typical of the problems encountered in any communication system: intelligible transmission of speech under a variety of operating conditions, exclusion of noise at all points in the system, and the maximum possible comfort and convenience for personnel. Thus the following discussion of aircraft interphones exemplifies concepts and working principles which apply to a wide range of other equipment.

The aircraft interphone in use at the end of World War II represented the culmination of a series of modifications of the original equipment. Even before the entry of the United States into World War II, persistent reports of the unsatisfactory performance of aircraft interphones were received. A program of testing in the field and in the laboratory was begun, and improvements soon became available in microphones, headsets, and amplifiers. Changes were adopted, communication was improved, and at the close of hostilities every major component of the original interphone had been replaced.

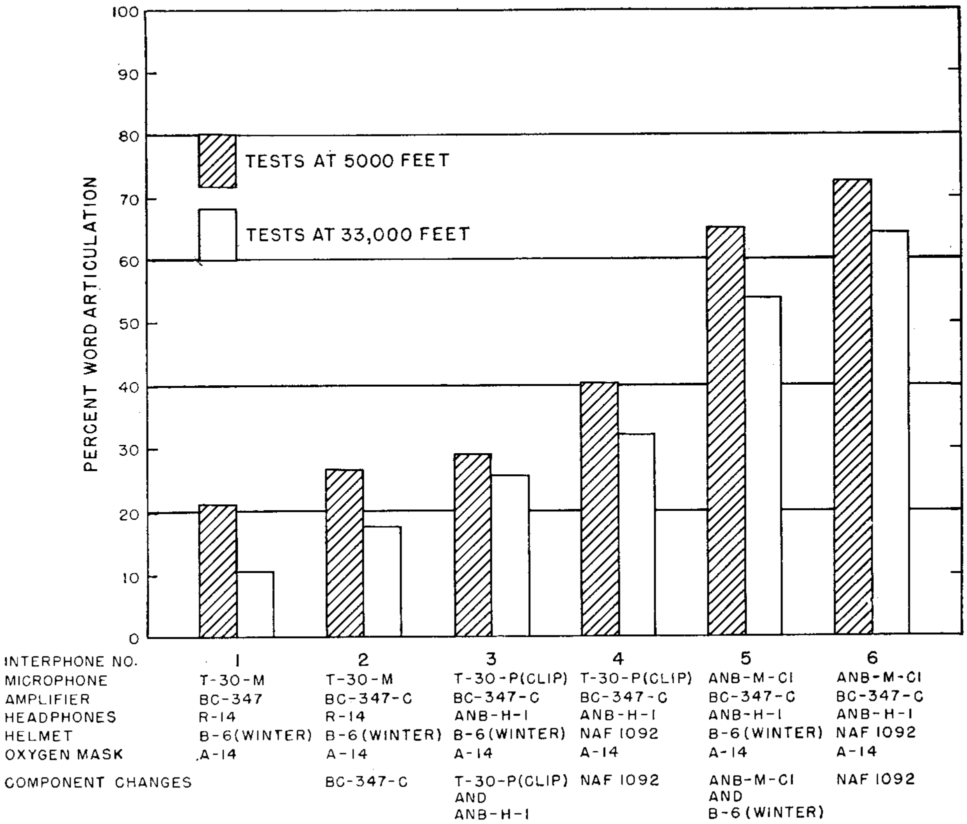


Figure 1. Bar graph illustrating the improvement in interphone performance as old components were discarded and replaced by newer ones.

The story of these improvements is graphically told in Figure 1. This chart shows the averaged results of a series of articulation tests on six aircraft interphones. The tests were run with trained Service personnel at low altitude (hatched) and at high altitude (unhatched) during flights in a B-17F bomber at Eglin Field, Florida.[5] Interphone No. 1, at the left of the chart, represents approximately the performance of the original interphone. As the old components were discarded and replaced by newer ones, interphone performance improved steadily. The final version gave 60 to 70 per cent correct word articulation under operating conditions which had limited the earlier equipment to 10 or 20 per cent, and further improvements were still under development.

The value of these improvements can be measured in terms of greater military coordination. The personnel of a combat vehicle must communicate with one another to operate smoothly as a unit, and they must maintain contact with distant points of command to operate as an effective component of a larger team. The interphone is the standard system for communication between members of the crew, and the effectiveness of the vehicle depends in large measure upon the effectiveness of the interphone.

9.1 EVALUATION OF PERFORMANCE IN NOISE

One of the important contributions of the Psycho-Acoustic and Electro-Acoustic Laboratories in the early years of World War II was an insistence that the greatest single hazard to effective military communication is the presence of noise, and that interphone performance can be properly assessed only under the acoustic stress of ambient noise. Satisfactory performance by an interphone system in quiet surroundings does not insure satisfactory performance in noise. The evaluation of performance in noise became, therefore, a problem of major importance.

In its simplest form, the interphone operates as shown in Figure 2. When a member of the crew wants to talk to another station, he depresses his push-to-talk switch. The output voltage from his microphone is then amplified by the interphone amplifier and heard in the earphones at all stations. In such a system there are three major ways in which noise can reach the listener's ear (see Figure 3).

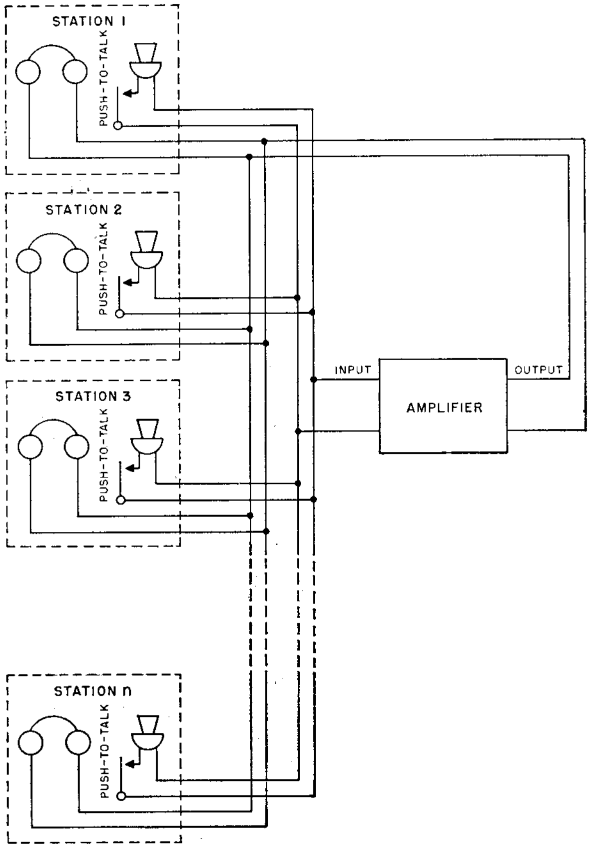


FIGURE 2. Interphone system (schematic).

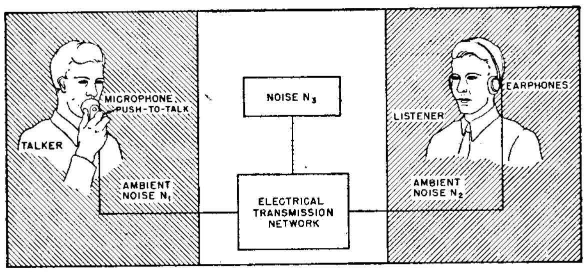


FIGURE 3. Noise interference in a voice communication system.

1. The microphone picks up a certain amount of the ambient noise N_1 around the talker. This noise reaches the listener via the interphone system. The noise spectrum reaching the listener may differ considerably from the spectrum of the ambient noise N_1. The noise spectrum acting upon the microphone is a function of the ambient sound field, of the type and shape of microphone, and of the position of the microphone relative to the mouth and face of the

talker. The spectrum is further modified before reaching the listener by the overall transmission characteristic of the interphone.

2. Noise may enter the listener's ear directly by leakage around the earphone and its cushion. This contribution depends on the noise spectrum N_2 at the listener's location and on the noise-excluding properties of the earphone cushion. These properties are a function of the type of earphone and cushion, and of the particular way in which they fit the listener's ear, and are independent of the rest of the system.

3. Noise may also enter the system by way of the electric transmission system itself. This source of noise, usually negligible for interphones, becomes important in other communication systems, e.g., a radio link. The noise spectrum at the listener's ear due to this cause is, in part, a function of the transmission characteristics of the system. Noise entering the system when the push-to-talk switch at other stations is accidentally pressed and noise picked up by the earphones acting as microphones at other stations can be lumped, if desired, with the effects of the noise N_3.

It is clear that an appraisal of the performance of interphones and their components must include not only their properties as electroacoustic transducers for the speech signal but also the degree to which unwanted sound is excluded.

9.1.1 Parameters Controlling Performance

The capacity of an interphone system[a] to transmit speech intelligibly is measured by a count of the number of discrete speech units recognized by the listener. The percentage of units correctly perceived is called the articulation score (see Chapter 5). The problem, therefore, is to correlate the efficiency of the interphone system, as measured by an articulation score, with its physical and psychophysical parameters.

Two theories relating the intelligibility of speech to the parameters of the interphone have been proposed for linear systems by the Bell Telephone Laboratories. According to one theory,[1, 9] the ability of the listener to perceive and recognize speech sounds depends *primarily* upon the relative intensities of the speech and noise spectra at the listener's ear, evaluated over most of the audible frequency range.

The spectrum level [SL] (see Section 4.1) of the received speech is a function of many variables, the most important of which are:

1. Voice level of the talker.
2. Orthotelephonic gain of the interphone (see Section 7.2.1).
3. Coupling between mouth and microphone.
4. Coupling between ear and earphone.
5. Speech material and composition.
6. Enunciation and training of the talker.

The spectrum level of the noise at the listener's ear depends on:

1. The ambient and electrical noise present.
2. The noise-excluding properties of the circuit, including microphone and earphone.
3. The efficiency and electrical properties of the circuit.

From this information and from certain assumptions and empirical relations derived from the fundamental properties of speech and hearing, an estimate can be made of the articulation scores to be expected. The estimates are most accurate when the spectrum of the noise at the listener's ear extends over a wide range of frequencies, the noise levels are moderate, and the response characteristic of the system is free from abrupt changes with frequency.

The correct calculation of scores on an absolute basis is difficult and of little practical meaning in view of the many variables involved. Comparison of measured and computed articulation scores on a relative basis, however, has yielded some useful results. A very desirable feature, from the design engineer's point of view, is that estimates can be made of the effects on the performance of the system of variations of circuit constants and elements. Predictions of performance based on this theory are considered in detail in Section 9.1.5.

The second theory[10] assumes that the differential sensitivity of the ear to frequency and level (see Section 3.5) is the controlling factor in the successful interpretation of speech sounds. Empirical functions were obtained at

[a] Talker and listener must be considered integral parts of a voice communication system.

the Bell Telephone Laboratories relating the speech spectrum, the differential sensitivity of the ear, the masking of sounds, and the articulation score. No use is made of this theory in the work reported here.

The important contribution of the two available theories is the demonstration that, in the simplest cases, the overall response of the system and its noise-excluding properties are of fundamental importance in determining performance. By assuming "average normal" talkers and listeners, standardized and uniform speech material, "constant" voice level, and negligible interaction between mouth and microphone, many of the troublesome variables can be effectively eliminated. However, this simplification, while useful for theoretical analysis, does not extend to many practical conditions of use. Hence, in evaluating interphone performance, articulation tests under simulated conditions of use should be performed wherever feasible.

9.1.2 Overall Response of the System

In the preceding section it was stated that systems can generally be rank-ordered as to performance under certain simplified conditions if the levels of received speech and noise at the listener's ear are known as a function of frequency. In order to obtain this information, it is necessary to consider the overall response of the system.

Meaningful comparisons of the *overall frequency response* of two systems can be most readily obtained by defining a suitable reference system. The orthotelephonic reference system (see Section 7.2.1) consists of a "normal" talker and listener[b] facing each other at a distance of 1 m in a quiet, nonreverberant room. The *orthotelephonic gain* [OG] at a given frequency is then expressed as the ratio of the level produced at the listener's ear by the system under test to that at the listener's ear when speech is transmitted from talker to listener over the 1-m path in free air. It is assumed, of course, that the talker speaks into the system under test in the same way he talks over the orthotelephonic reference system. By definition, then, speech heard over a system having an OG of 0 db at all frequencies will sound the same as speech heard over an air path 1 m in length.[c]

In order to obtain a quantitative measure of OG, it is necessary first to decide how the level at the listener's ear is to be measured. One method involves two measurements of the sound pressure at the listener's eardrum: first when he listens to the system under test and second when he listens to the talker over a direct air path 1 m in length. The ratio of these two pressures, expressed in decibels, is a direct measure of the orthotelephonic gain of the system.[11]

In practice, however, this measurement would be very difficult to perform. An alternative procedure, and one which yields the same results, involves the subjective loudness produced by a given sound source expressed in terms of the *equivalent free-field sound pressure*. The equivalent free-field pressure is obtained by requiring an observer located in a nonreverberant, quiet room to equate the loudness produced by a sound source some distance in front of him with that produced by the earphone. The equivalent free-field pressure is then defined as the free-field pressure, measured at the listener's location, which produces the same loudness as does the earphone under test. Specifically, by free-field sound pressure is meant here the pressure measured by a "point" microphone located at the position where the listener later places his head when making the loudness balance. This location is taken as the mid-point of a line joining the listener's ears.

The concept of equivalent free-field pressure, introduced by the Bell Telephone Laboratories, has the advantage that measurement of the sound pressure is more easily performed in the free field than at the listener's eardrum. Although there is some evidence to the contrary

[b] This is in contrast to methods using artificial voices and ears. The chief merit of using the latter lies in the simplicity achieved (see Chapter 10).

[c] If the experiment is actually performed in which a listener compares what he hears over a system with 0-db OG with what he hears over an air path of 1 m, certain differences, for example, in apparent localization may be experienced. Such differences are disregarded here.

(see Sections 3.2 and 10.3), it will be assumed here that for a given frequency and observer there is an essentially constant relation between the pressure at the eardrum and the corresponding equivalent free-field pressure for various types of earphones and pressure levels (see Section 3.1).

Consequently, the orthotelephonic gain of a system (in decibels) can be defined for any given frequency as follows:

$$\mathrm{OG} = 20 \log \frac{p_x}{p_0}. \tag{1}$$

This equation can be expressed in words. The OG of a system X is the ratio, in decibels, of the equivalent free-field pressure p_x, produced by system X, to the free-field pressure p_0, generated 1 m from the talker in a quiet, nonreverberant room. Since the talker must speak the same way[d] into both system X and the orthotelephonic system, the OG of the system is independent of the intensity and spectrum of the talker's voice.

If the OG of a system is known as a function of frequency, the speech spectrum heard by the listener is easily obtained. First, the spectrum of the talker's voice is determined in terms of free-field pressure p_0, at 1 m in an anechoic (echo-free) chamber. Average speech spectra of this type were given in Figures 1 and 2 in Chapter 4. (They are easily reducible to a distance of 1 m.) Second, this speech spectrum is added (in terms of decibels) to the OG of the system X. The result is the speech spectrum heard by the listener over system X, expressed in terms of the equivalent free-field sound pressure p_x.

9.1.3 Synthesis of the Overall Response

In practice, it is convenient to synthesize the overall response of the system in three steps. The response of the three basic components of the interphone system—the microphone, the earphone, and the network connecting these two electro-acoustic transducers—is determined separately. These three response measurements, if properly made, can be added to obtain OG.

This procedure may best be illustrated by reconsidering equation (1) above. This equation can be rewritten in terms of three factors, as follows:

$$\begin{aligned}\mathrm{OG} &= 20 \log \frac{p_x}{p_0} = 20 \log \left(\frac{e}{p_0}\right)\left(\frac{E}{e}\right)\left(\frac{p_x}{E}\right), \\ &= 20 \log \frac{e}{p_0} + 20 \log \frac{E}{e} + 20 \log \frac{p_x}{E}. \end{aligned} \tag{2}$$

The three terms of equation (2) can be identified with the response characteristics of the three elements of the system in the following way.

Microphone. The term 20 log e/p_0 is the ratio, in decibels, of the voltage developed by the microphone, held in its normal relation to the talker, to the free-field sound pressure measured 1 m away from the talker with the microphone removed. An audio-spectrometer (see Chapter 4) is used to measure e and p_0 in suitable contiguous frequency bands. The measurement of p_0 is made under orthotelephonic reference conditions. The quantity 20 log e/p_0 may be referred to as the *real-voice response* of the microphone.

Amplifier. The term 20 log E/e is the ratio, in decibels, of the voltage E developed across the listener's headset to the voltage e developed by the microphone. This term is simply the electrical response of the network connecting the two electro-acoustic transducers.

Earphone. The term 20 log p_x/E is the ratio of the equivalent free-field sound pressure p_x to the voltage E measured across the terminals of the earphone. More specifically, p_x is the free-field pressure necessary to produce the same loudness sensation as that produced when a voltage E is applied to the earphone. In making the loudness-balance judgment the listener is located in a quiet, nonreverberant room with the sound source in front of him. It is not necessary to use a speech signal. The quantity 20 log p_x/E may be referred to as the *real-ear response* of the earphone.

Addition of these three terms gives the orthotelephonic gain for the system at any given frequency.

[d] Difficulties in keeping the voice "constant" are encountered if marked interaction exists between the talker and the microphone he uses, as, for example, when an oxygen mask is used.

9.1.4 Response of the System to Noise

The level and spectrum of the noise at the listener's ear is the sum of the contributions from the noise transmitted by the microphone, the noise reaching the listener's ear through lack of insulation of the earphone cushion, and any contributions from noise due to other causes (see Figure 3). For noise of the type found in combat vehicles (see Chapter 2), the resultant noise spectrum at the listener's ear is obtained by summation (on a power basis) of the various contributions.

The noise entering the system through the microphone can be measured readily by using an audio-spectrometer (see Section 10.1.2). A "talker" (who, however, does not talk) holds or wears the microphone under test in a diffuse sound field corresponding to N_1 (see Figure 3), while the voltage developed by the microphone due to noise pickup is measured. By addition of the amplifier response and real-ear response, the equivalent free-field spectrum level of the noise due to N_1 is obtained.

To obtain the noise spectrum leaking through the earphone cushion, the listener might be immersed in the diffuse noise field N_2 wearing the earphones under test in normal conditions of use, and a measurement made of the noise spectrum at the eardrum would yield the desired results. A much simpler method determines the attenuation which the headset provides for pure tones (see Section 10.3.2). The attenuation, in decibels, is then subtracted from the noise spectrum N_2, and the spectrum reaching the listener's ear through the cushion is thus expressed approximately in terms of equivalent free-field sound pressures.

Still another method consists in measuring the masked threshold of the listener for pure tones when noise is leaking through the cushion. This is accomplished by energizing the earphone with single frequencies. This method is, in many ways, the most fundamental one, in that it gives a direct measure of the reduced auditory area available for communication. It takes into account such effects as masking by adjacent noise components (spread of masking). It can be used to test the noise exclusion properties of earphones (see Section 10.3.2), but is an even more valuable laboratory method if applied to the *total* noise spectrum at the listener's ear due to N_1, N_2, and N_3.

The contribution due to electrical noise N_3 is evaluated by an analysis of the voltage appearing across the earphone terminals. Addition of this spectrum to the real-ear response of the earphone yields the desired result.

9.1.5 Prediction of Performance

Now that we have considered in detail the procedures for determining the speech and noise spectra at the listener's ear, we can proceed to demonstrate how this information is used to predict the performance of an interphone system. The following computational procedures were devised originally by the Bell Telephone Laboratories[1a] and later refined in some respects by the Electro-Acoustic Laboratory.[14a]

It is the fundamental assumption of the theory that the contribution to intelligibility by any given band of speech frequencies may be calculated independent of the contributions made by other bands. These contributions can be expressed by means of an empirical function, the *articulation index* [AI], which has additive properties.[c] The total contribution to AI over a total range of frequencies is the sum of the contributions to AI of the component bands contained in that range, essentially independent of the manner of subdivision.

The sum of the contributions of all the bands of frequencies passed by the interphone can be written as follows:

$$\mathrm{AI} = \sum_{p=1}^{p=n} (\Delta\mathrm{AI})_p, \tag{3}$$

where $(\Delta\mathrm{AI})_p$ is the contribution to the articulation index of the pth band of speech frequencies in the total range under consideration. Clearly, $(\Delta\mathrm{AI})_p$ is a function of both the frequency limits of the band and of the speech and noise levels in that band.

[c] It is also assumed that there is a unique relation between articulation score and AI for a given speech material and crew.

The problem, of course, is to design the interphone in such a way as to maximize AI, i.e., to maximize the contribution from each of the p bands of frequencies. It is plausible to assume that this condition will obtain for low levels of noise and high levels of speech, and $(\Delta AI)_p^{max}$ will then depend primarily upon frequency.

We therefore put

$$(\Delta AI)_p = (\Delta AI)_p^{max} W_p. \tag{4}$$

The weighting factor W_p can be interpreted as the fractional part of $(\Delta AI)_p^{max}$ which is available when the speech and noise levels are unfavorable. Thus W_p is primarily a function of the speech-to-noise ratio in the pth band, and it is convenient to assume that it is independent of frequency. Hence,

$$AI = \sum_{p=1}^{p=n} (\Delta AI)_p^{max} W_p. \tag{5}$$

It is necessary, therefore, to evaluate $(\Delta AI)_p^{max}$ experimentally as a function of frequency and W_p as a function of the signal-to-noise ratio.

Relation between $(\Delta AI)_p^{max}$ and Frequency

This relation has been evaluated at the Bell Telephone Laboratories for nonsense syllables

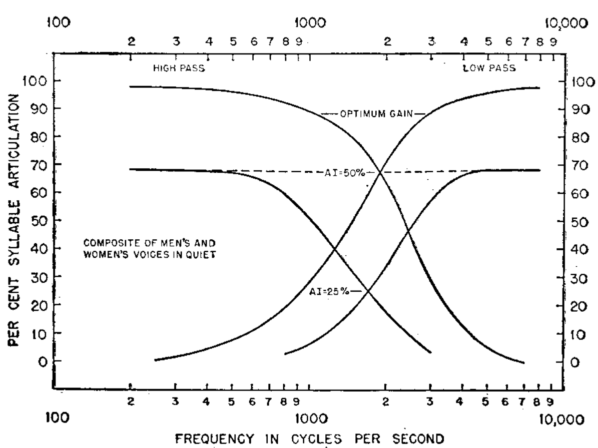


FIGURE 4. Articulation scores vs cutoff frequency of high-pass and low-pass filters inserted in a wide-band system. (BTL)

with both men and women talking. Articulation tests were made in quiet on a high-fidelity system into which low-pass and high-pass filters could be inserted. The results of these tests are presented in Figure 17 of Chapter 7.

If the optimal scores for each filter condition are plotted as a function of filter-cutoff frequency, the upper curves in Figure 4 result. It is evident from these curves that as much of the total intelligibility is carried by frequencies below 1,900 c as is carried by the frequencies above 1,900 c. Furthermore, the first point on the additive AI scale, the 50 per cent point, is thus determined. It appears that 50 per cent AI corresponds to an articulation score of 68 per cent. Another point can be obtained by returning to the orthotelephonic gain of the wide-band system which gives an articulation score of 68 per cent. From Figure 17, Chapter 7, we see that the wide-band system gives 68 per cent articulation with an orthotelephonic gain of about −31 db. If the articulation scores are now plotted as a function of filter-cutoff frequencies for this gain setting, the lower curves

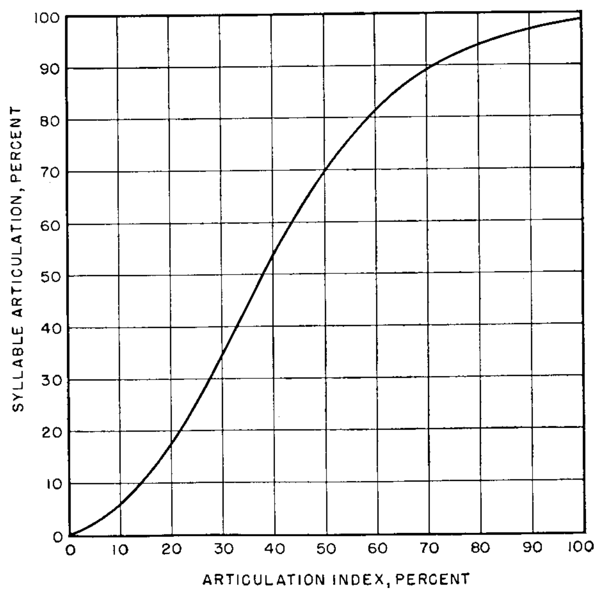


FIGURE 5. Approximate relation between syllable articulation and articulation index. (BTL)

in Figure 4 result. These functions cross at a point which defines an AI of 25 per cent.

By a similar fractionating procedure, the complete function relating AI to syllable articulation score can be plotted. The function has the general shape shown in Figure 5. For different speech material and crews, other curves would have been obtained. By means of this function, the ordinate in Figure 4 can be converted directly into AI. This has been done in Figure 6, where the cumulative contribution to the articulation index is shown as a function of the upper cutoff frequency. The slope of this function shows the relative importance of the various frequencies in their contribution to intelligibility. The similarity between this function (see Figure 6), the pitch scale (see Figure 7, Chapter 3), and the function which relates the position of maximum agitation on the basilar membrane to the frequency of the acoustic stimulus has been pointed out (see Section 3.3.1).

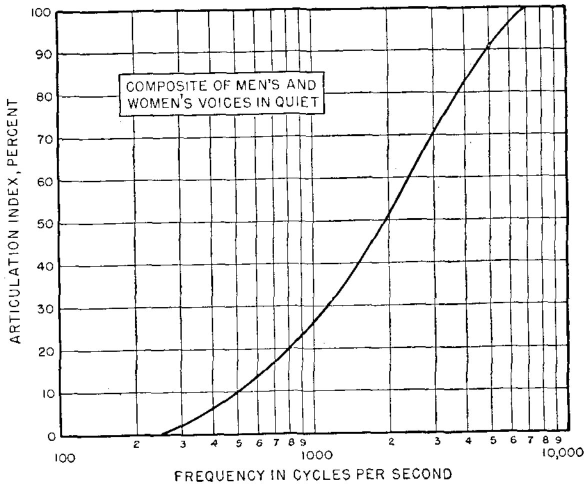


FIGURE 6. Relating articulation index and cutoff frequency of a low-pass filter inserted in a wide-band system. (BTL)

With this information, the range of important speech frequencies can be divided into a number of bands which contribute equally to the articulation index. Computations are normally made on the basis of 20 bands, each contributing 5 per cent.

RELATION BETWEEN W_p AND THE LEVELS OF SPEECH AND NOISE

As the simplest case, consider first the computation of W_p for quiet conditions. If the entire frequency range of speech is divided into 20 bands, each contributing 5 per cent to AI, then $(\Delta AI)_p$ can be plotted between 0 and 5 per cent as a function of the orthotelephonic gain for each of the 20 bands. When this is done, it is found that W_p depends principally upon the sensation level of the speech in the band, i.e., upon the amount by which the speech level exceeds the threshold of hearing in the band.

These results can be carried over directly to the case where noise is present if it is assumed that we are dealing with the masked threshold instead of the quiet threshold of hearing. Now we have previously seen (see Section 3.4.1) that the masking of pure tones by noise is directly proportional to the effective level Z of the band of noise frequencies (see Figure 10, Chapter 3). Hence W_p can be expressed directly in terms of the speech-to-noise ratio in each band. The spectrum level of the noise B in decibels is subtracted from the spectrum level of the speech B_s in decibels. Because the temporal distribution of amplitudes differs for speech and noise, a constant number of decibels is added to the speech-to-noise ratio $(B_s - B)$ in order to fit the experimental evidence.[1b]

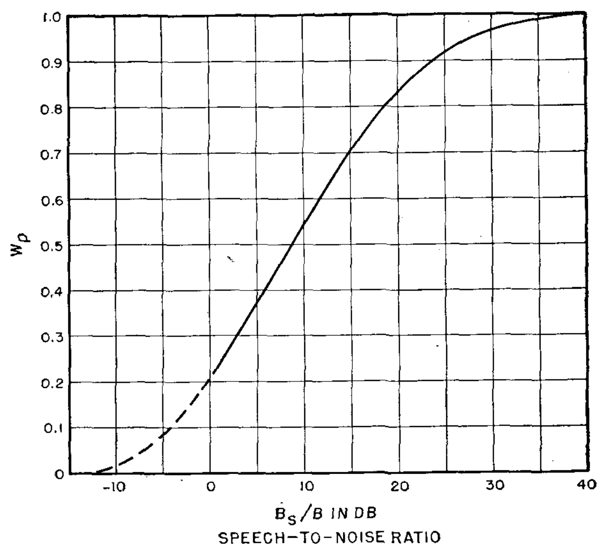


FIGURE 7. Weighting factor vs speech-to-noise ratio. (BTL)

Figure 7 shows approximately how the weighting factor W_p depends upon the speech-to-noise ratio $B_s - B$. If the signal-to-noise ratios have been determined for each of the 20 bands of speech frequencies, the corresponding values for W_p can be read from Figure 7. It can be seen that the speech-to-noise ratio must be of the order of 20 to 25 db before a band of speech frequencies makes its maximum contribution to the articulation index.

In making calculations, a work sheet like the one shown in Figure 8 may be helpful. The frequency scale is divided into 20 bands, each having a maximum potential contribution to AI of 5 per cent. The weighting factor can then be obtained from Figure 7 for each band. Addition of all contributions to AI according to equation (5) determines the total articulation index. If it is desired, once AI is known, the syllable articulation score can be predicted from Figure 5. In this manner, an approximation of the interphone's performance is obtained solely on the basis of the speech-to-noise ratio at the listener's ear.[f]

Figure 9 shows a comparison of theory and experiment. The system under test was a band-pass system of variable bandwidth. The noise spectrum at the listener's ear is shown in Figure 10, expressed in terms of the equivalent free-field spectrum level. While, in some instances, the agreement between the experiment and the calculations is not especially close, the theory seems to provide a workable estimate of the relative performance of the systems.

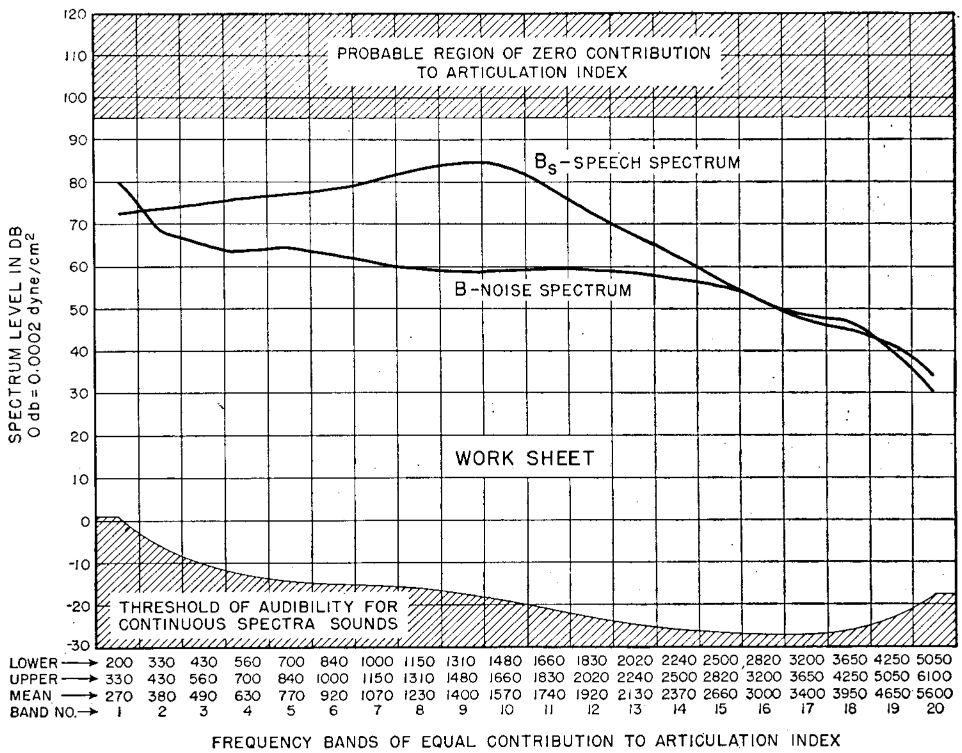


FIGURE 8. Work sheet for evaluating the articulation index of a communication system with noise present.

9.2 INTERPHONE COMPONENTS: THE MICROPHONE

The task of delivering a signal at a suitable speech-to-noise ratio over a wide band of frequencies is the joint responsibility of the microphone, the amplifier, and the headset. Failure on the part of any one of these components means failure of the entire interphone.

The first link in the chain is, of course, the microphone. Several different types of microphones were used in World War II: hand-held microphones, noise-canceling microphones, throat microphones, microphones mounted in enclosures. Each type was represented by a variety of models, and it is necessary, therefore, to limit the following discussion to those micro-

---

[f] Refinements have been added to the theory, taking into account such secondary variables as interband masking, self-masking, and nonlinear distortion of certain types.

phones which were actually used on a large scale and to the general factors governing the performance of each of the generic types.

9.2.1 Hand-Held Microphones

The two hand-held microphones most widely used in aircraft were the Army's T-17 and the Navy's T-38 (RS-38-A). Photographs of these

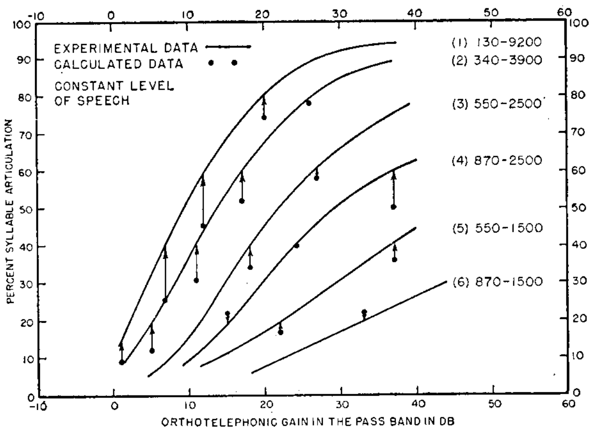


FIGURE 9. Measured and computed gain functions for band-pass systems in noise.

two microphones are shown in Figures 11 and 12. Both the T-17 and the T-38 are carbon-button instruments.

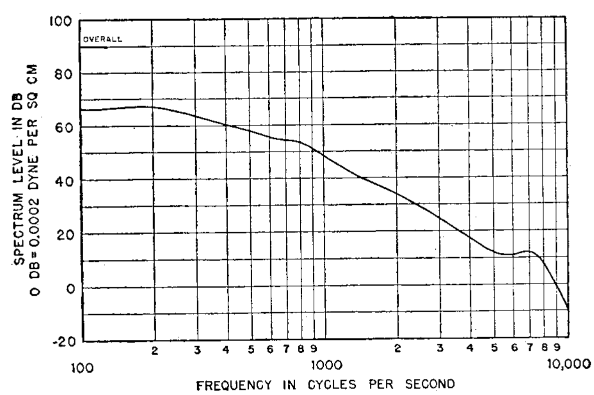


FIGURE 10. Noise spectrum at the listener's ear in articulation tests described by Figure 9.

Carbon-type microphones were adopted almost exclusively by the Armed Forces of the United States, despite the fact that all carbon microphones are essentially nonlinear devices. Furthermore, their frequency-response characteristic depends upon the level of the sound pressure for which the response is obtained. They require a separate source of current, their performance is likely to vary with change of position, and they produce a certain amount of "burning noise" due to the passage of current through the carbon granules. In addition, there is some deterioration of the carbon granules with age. Their extensive use is due primarily to their very high sensitivity, far exceeding the sensitivity of magnetic and dynamic microphones. In spite of their limitations, the T-17 and T-38 microphones proved to be reliable instruments for communication.

Hand-held microphones as a class have certain disadvantages which limit their usefulness. They do not exclude ambient noise as well as the other types of microphones, and they require the continuous use of one hand during communication.

Measurements of the speech-to-noise ratio were made for these and other microphones. Eleven sets of highly discriminative band-pass filters, having cutoff frequencies indicated on the figures, were used for the analyses. A talker held the microphone close to his mouth in an ambient noise field similar to that found in two-engined aircraft without sound treatment. The talker remained silent while the noise level from the microphone was measured for each of the pass bands. The noise was then turned off, and the analysis repeated as the talker spoke the sentence, "Joe took father's shoebench out; she was waiting at my lawn." The results of these measurements are presented in Figures 13 and 14. The speech-to-noise ratio in any band is found by counting the number of decibels between the levels of the two signals.

Measurements of this type are influenced by a variety of variables and therefore should never be regarded as absolute. Their value lies in the fact that they provide a basis for comparing different microphones all tested under identical conditions. Comparison of the T-17 and T-38 shows that the speech-to-noise ratios are nearly the same at all frequencies. This agrees with the usual finding that the two microphones yield approximately equivalent articulation scores when tested under the same conditions.

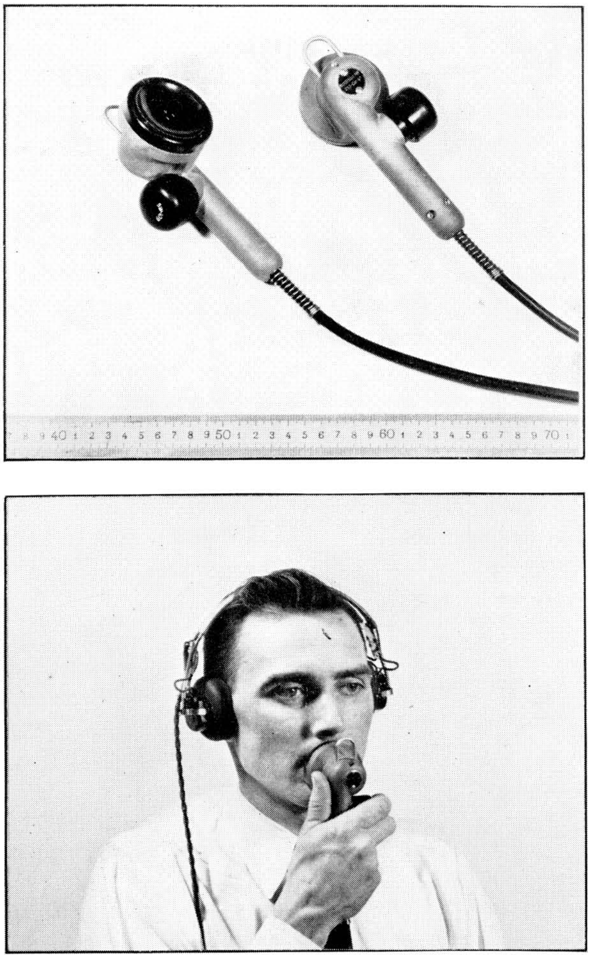


FIGURE 11. Hand-held carbon microphone T-17.

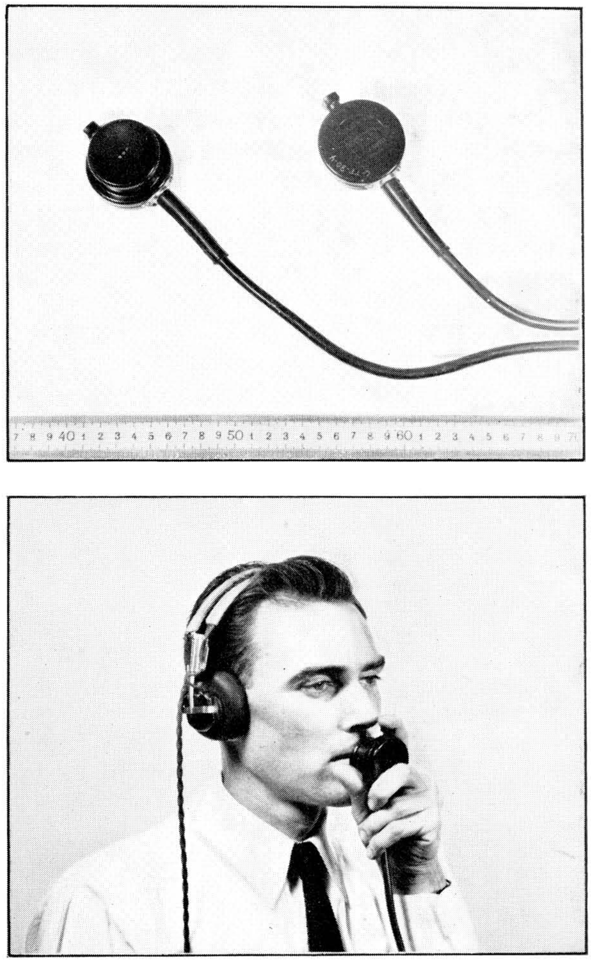


FIGURE 12. Hand-held carbon microphone T-38.

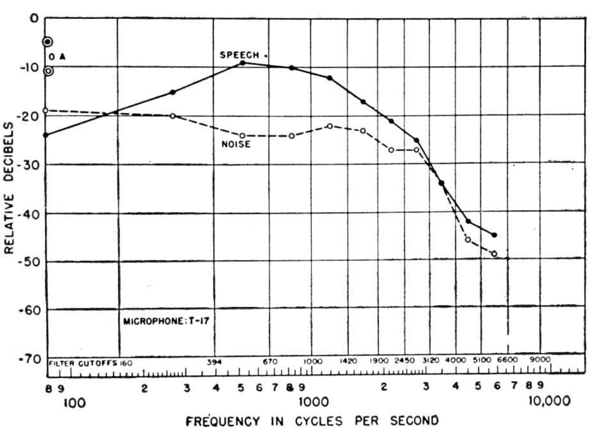


FIGURE 13. Speech and noise levels obtained for different frequency bands with the T-17 microphone.

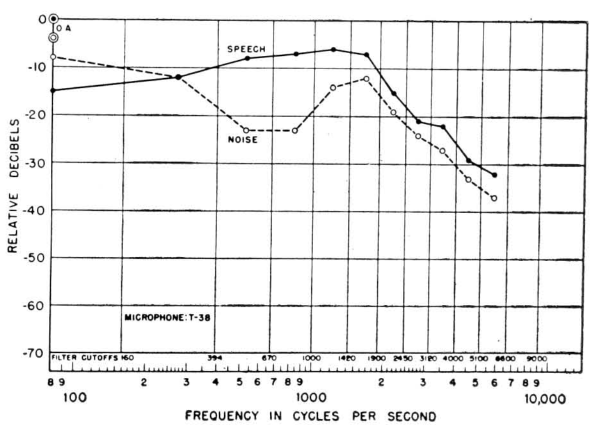


FIGURE 14. Speech and noise levels obtained for different frequency bands with the T-38 microphone.

9.2.2 Noise-Canceling Microphones

Noise-canceling, or differential, microphones have both sides of the diaphragm exposed to the impinging sound waves and are, therefore, sensitive to the pressure gradient in the sound field. When the differential microphone is close to and oriented toward the talker's lips, it is activated primarily by the talker's voice and not by the noise field.

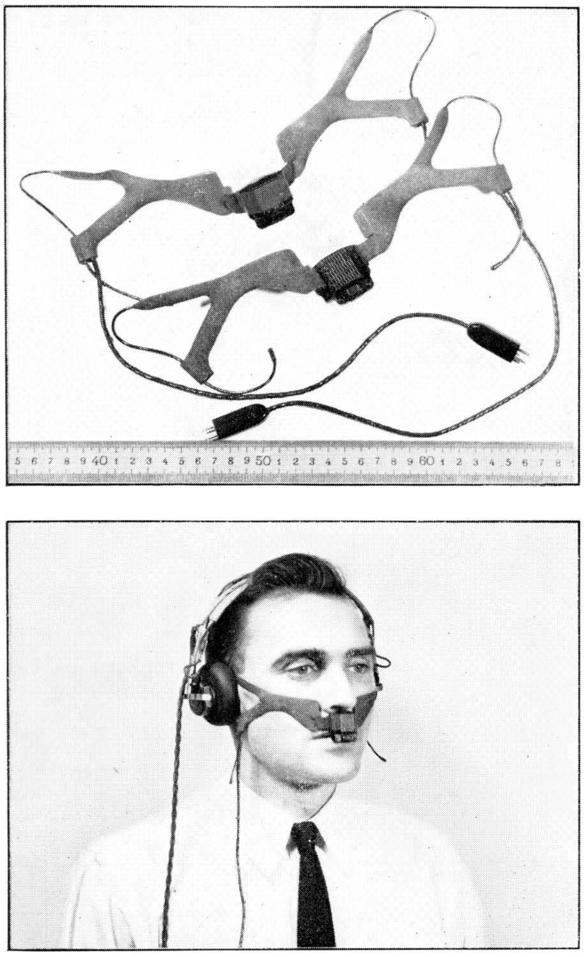


FIGURE 15. Noise-canceling T-45 microphone.

The noise-canceling microphone, T-45, used by the Army Ground Forces, is shown in Figure 15. A later improved version, M-5/UR, is shown in Figure 16. Because of the manner of suspension in front of the talker's lips, these were sometimes called "moustache" or "lip" microphones. This suspension leaves the hands free (if a suitable switch is provided), and thus has one great advantage over the hand-held microphone.

Speech and noise measurements for noise-canceling microphones, comparable to those presented for hand-held microphones, are shown in Figures 17 and 18. It can be seen from these figures that the speech-to-noise ratio is most favorable at frequencies below about 2,000 c. In this region the ratio is somewhat better than that of the hand-held microphone, although neither microphone performs very poorly here. Unfortunately, the noise-canceling microphone possesses an inherent limitation in its inability to discriminate against high-frequency noise, and in high ambient noise this

FIGURE 16. Noise-canceling M-5/UR microphone.

becomes a serious consideration.[13] It is usually in the region above 3,000 c that noise levels actually exceed speech levels, and in this region the noise canceler is just another microphone.

Figures 17 and 18 also include measurements of "wind noise." The talker wore the microphone in a 30-mph wind and adjusted the position of his head until the maximum level of noise was obtained. He then held this position, and the noise due to the wind was measured in each frequency band. The results show that the T-45 is more susceptible to wind noise than the M-5/UR, and this was confirmed by the results of articulation tests. In both cases,

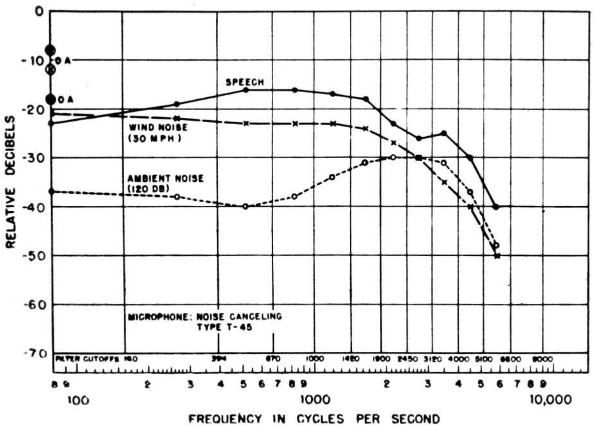


FIGURE 17. Speech and noise levels obtained for different frequency bands with the T-45 microphone.

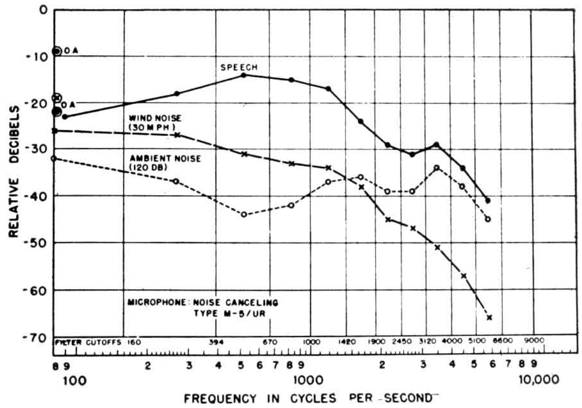
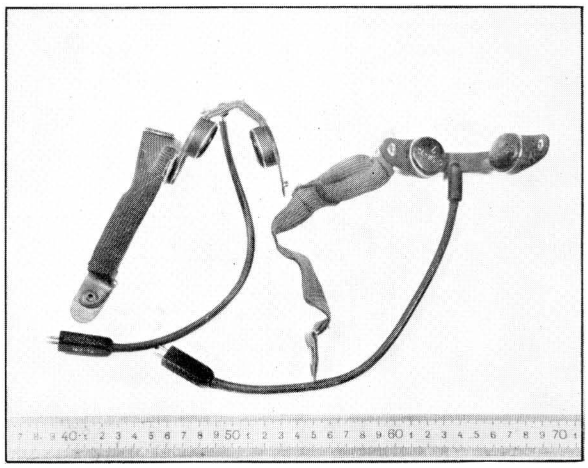


FIGURE 18. Speech and noise levels obtained for different frequency bands with the M-5/UR microphone.

however, the noise-canceling microphones generate a higher overall noise level in this wind than they do when used in an ambient noise of 120 db.

While the noise-canceling microphones possess limitations which cannot be overcome, they proved to be the most satisfactory devices available in moderate or low noise levels or in instances where acoustic feedback was a critical problem. It was a real misfortune that the sensitivity of aviators' lips, or their fondness for moustaches, and the relative lack of sensitivity of their necks should have delayed the substitution of this microphone for the ubiquitous throat microphone.



9.2.3 Throat Microphones

A device used widely by the U. S. Army Air Forces at the beginning of World War II was the throat microphone. In this assembly the microphone is strapped to the throat directly above the larynx. Such an arrangement possessed the advantages of apparently low noise pickup and free use of hands, and it would probably have been a very effective instrument but for the fact that the speech signal available at the larynx is intrinsically unintelligible. A photograph of the standard throat microphone, T-30, is shown in Figure 19.

Articulation tests showed persistently that the "mushy" speech picked up by a throat microphone provides a poor means of com-

FIGURE 19. Throat microphone T-30.

munication. That this inferiority was not due to poor design of the American version is illustrated by the articulation results shown in Figures 20 and 21. These figures present the

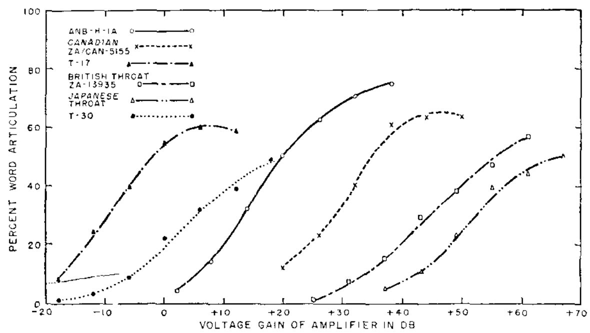


FIGURE 20. Articulation as a function of the voltage amplification applied to the outputs of six *different microphones*.

results of articulation tests run in 120 db of airplane noise with three hand-held and three throat microphones.[11] The Japanese and the British (ZA 13935) throat microphones are both magnetic instruments; the American T-30 is of the carbon type. The hand-held microphone was represented by the American T-17. The measurements included the Canadian dynamic ZA/CAN 5155 in a noise shield, and a special instrument consisting of the dynamic earphone ANB-H-1A used as a microphone. Figure 20 shows the articulation of each of these microphones plotted as a function of the voltage gain of the interphone amplifier. The two carbon microphones, T-17 and T-30, are clearly the most sensitive and require the least amplification, while the two magnetic instruments, the British and Japanese throat microphones, are the least sensitive. If these functions are now replotted according to the level of the speech received at the listener's ears, they group as shown in Figure 21. The three hand-held microphones are clearly superior to the three throat microphones.

This inferiority of the throat microphone was repeatedly demonstrated by tests in the laboratory and in the field, in spite of the fact that the overall noise levels seemed favorable. Figure 22 shows the speech and noise levels comparable to those presented above. It will be noted that the speech level falls rapidly above 1,500 c and this correlates with the "mushy," consonantless quality of the signal. In this instance, therefore, the usual advantage of a positive speech-to-noise ratio is of little value because the speech itself is so unnatural.

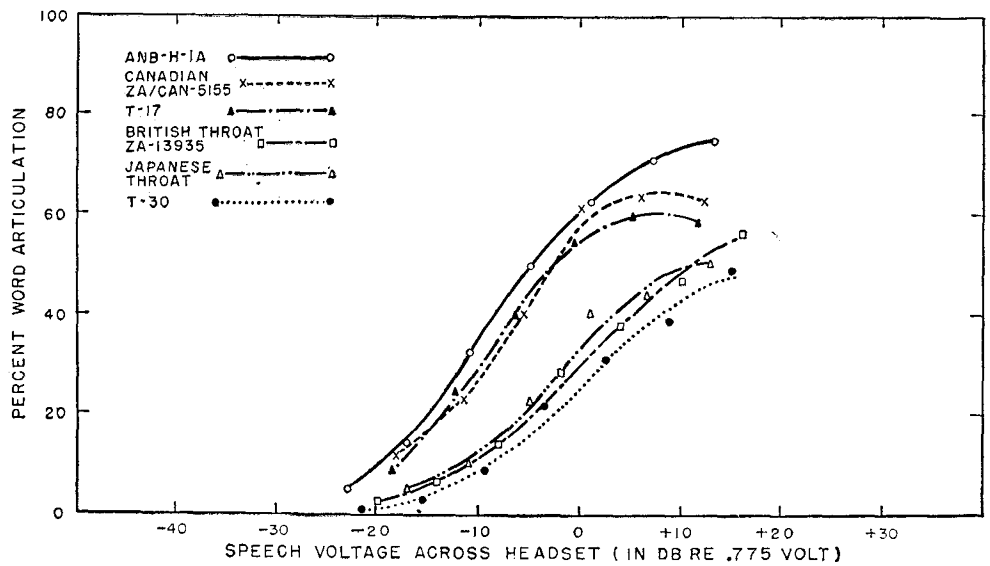


FIGURE 21. Word articulation provided by six microphones when the amplification used with each microphone was adjusted to produce approximately the same speech level at the listeners' ears. Each point is based on 200 words.

The second merit of the throat microphone, the freedom of the hands, also proved illusory in practice. The position of the microphone was critical and required frequent adjustment, and talkers who were frustrated by the poor performance of the instrument would often use

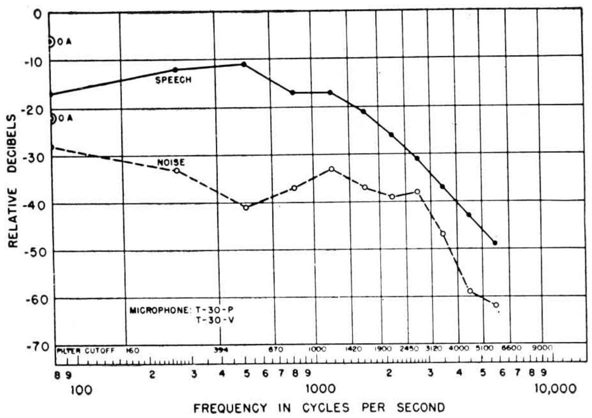


FIGURE 22. Speech and noise levels obtained for different frequency bands with the T-30 microphone.

one hand to press the microphone against their throats. All in all, the results with the throat microphone were not what had been hoped, and its belated replacement by other types of microphones was one of the major improvements made in the interphone system.

9.2.4 Microphones Mounted in an Enclosure

Mounting a small microphone inside an oxygen mask or a noise shield combines the advantages of noise exclusion with free use of the hands. The effect of such an arrangement, however, is to reinforce the low frequencies, and the speech acquires an unfortunate, "booming" character. One of the fundamental requirements, therefore, is that the mask microphone should have a poor low-frequency response in order to make the response of the mask and microphone combination approximately uniform. The fidelity and level of speech transmitted over these devices depends to a large extent upon the acoustic properties of the enclosure, and the interaction between the response characteristics of the mask and the microphone poses a very difficult problem for analysis. In most cases an empirical evaluation is necessary.

The mask microphone adopted by both the Army and Navy was a carbon instrument, ANB-M-C1. A magnetic mask microphone, MC-253-A, was tested extensively and was adopted for use by the British. A photograph of these two microphones is shown in Figure 23.

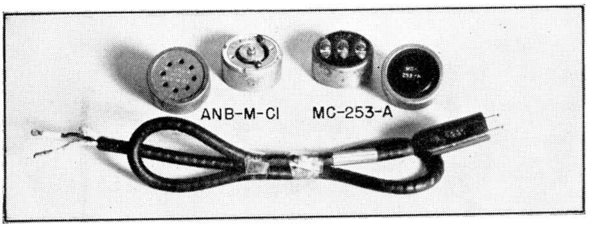


FIGURE 23. Mask microphones ANB-M-C1 (carbon) and MC-253-A (magnetic).

As in the case of the throat microphone, the interaction between the mask microphone and the voice precludes the simplifying theoretical assumptions made in Section 9.1.5. The articulation efficiency which can be realized with a properly designed mask microphone is considerably better than that obtained with throat microphones. When oxygen masks must be used, therefore, it is recommended that the microphone be placed inside the mask rather than on the throat. A photograph of the standard oxygen mask, A-14, used by both the Army and Navy, is shown in Figure 24. The microphone is mounted in the mask just above the outlet valve.

There are also some advantages to be gained at low altitudes by mounting the microphone in an enclosure. Noise shields were developed which provided a good speech-to-noise ratio, eliminated wind noise, and left the hands free. A photograph of the ANB-M-C1 microphone mounted in the Harvard D-17 noise shield is shown in Figure 25. Speech, ambient noise, and wind noise measurements are shown in Figure 26. The speech-to-noise ratio is excellent at all frequencies, and the effect of a 30-mph wind is much less serious than in the case of the noise-canceling microphones (see Figures 17 and 18). A noise shield was also developed by the Bell Telephone Laboratories which covered the nose and the mouth and which performed even better than the Harvard D-17 shield.

Neither of these shields was adopted by the Services, however.

An interesting comparison of the four types of microphones is given in Figure 27.[7] Articulation tests were run in ambient noise similar to that encountered in the cabin of a two-engined aircraft without sound treatment. Speech levels were held constant throughout all the tests. Five of the six microphones were nearly equivalent; the throat microphone T-30 was strikingly inferior. Best results were obtained with the ANB-M-C1 in a noise shield, but the superiority over the noise-canceling microphones was small. The hand-held T-17 ranks fifth in this group.

From such comparisons it appears that the noise shield has much to recommend it. Noise shields are not, however, as comfortable as some of the other classes of microphones and, in certain situations, they might restrict the useful field of vision. Also, they would have to be issued individually, whereas the noise-canceling microphone does not. These considerations probably contributed to the adoption of noise-canceling microphones rather than noise shields for use at low altitudes.

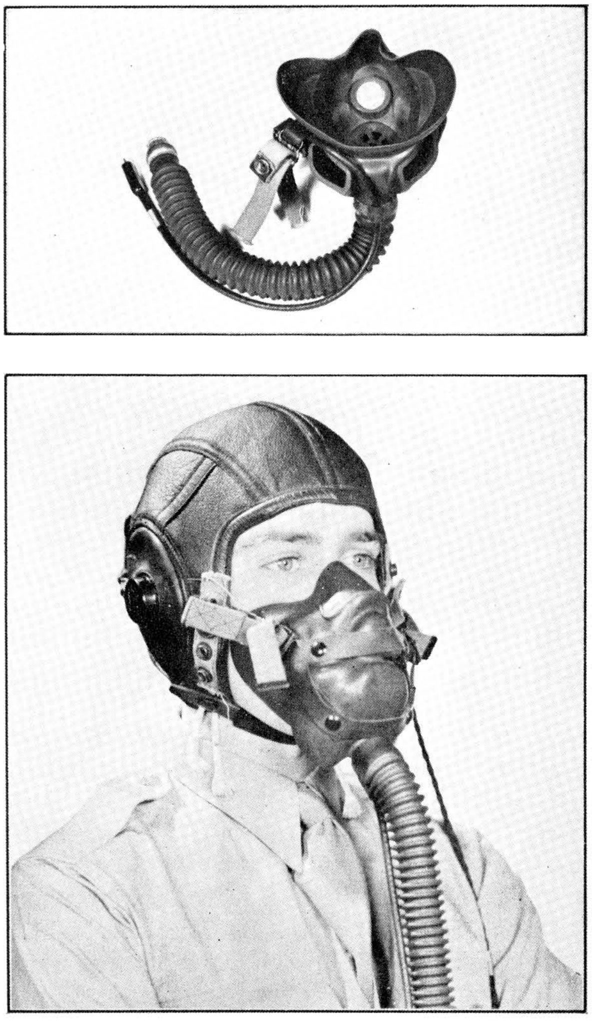


FIGURE 24. Oxygen mask A-14.

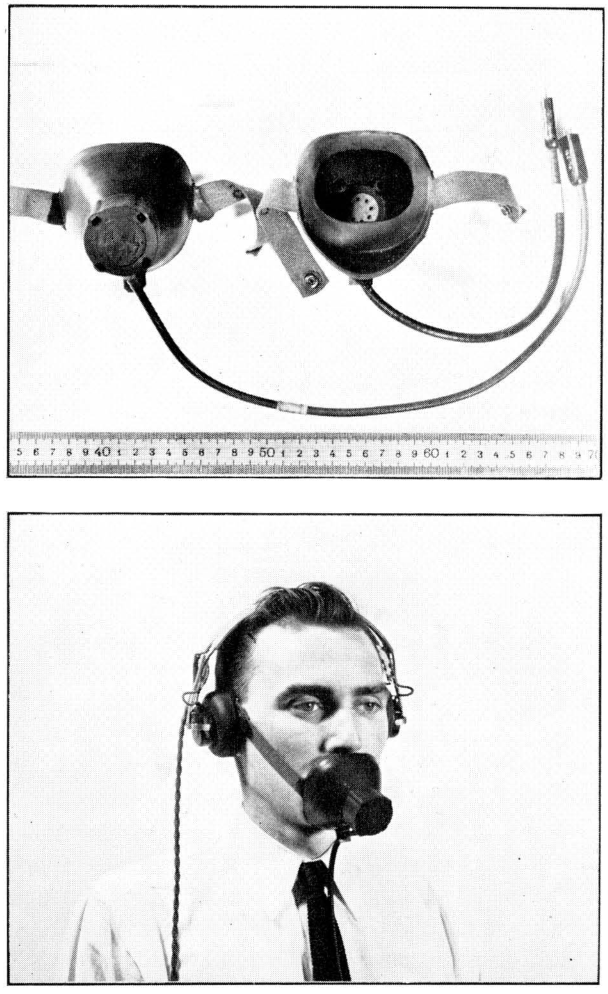


FIGURE 25. Noise shield D-17.

9.3 INTERPHONE COMPONENTS: THE AMPLIFIER

The second link in the chain from talker to listener is the interphone amplifier. The requirements for a satisfactory amplifier are straightforward and can be easily met by careful engineering: uniform response, sufficient voltage gain, and a power output adequate for the number of headsets used. The problem of

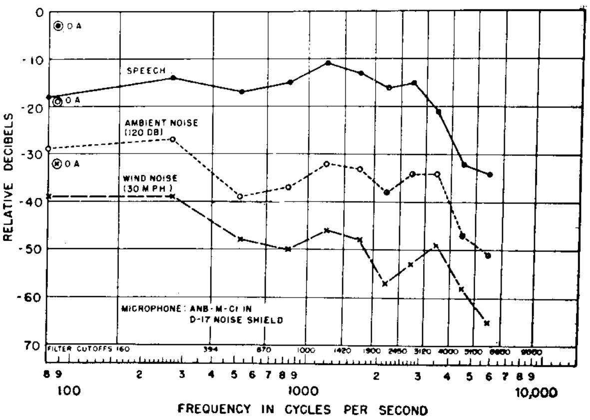


FIGURE 26. Speech and noise levels obtained for different frequency bands with the ANB-M-C1 in the D-17 noise shield.

noise exclusion, so serious for microphones and headsets, is not a factor in the design of the amplifier.

The interphone amplifier BC-212, with which the AAF began World War II, satisfied none of the fundamental requirements for satisfactory performance, and it was soon replaced by a more adequate unit, BC-347C. This amplifier possessed a satisfactory frequency response and was fairly adequate for communication at low altitudes. At high altitudes (see Section 9.5) there is a marked decrease in headphone sensitivity and in the voice level which a talker is able to maintain. Extensive flight tests at Eglin Field[6] showed that at altitudes of 35,000 ft neither the voltage nor the power-output capacity of the amplifier BC-347C were adequate. As a result of these tests, design recommendations were formulated. The frequency response at sea level should be essentially flat between 200 and 4,000 c, with an overall voltage gain of approximately 18 db. At 35,000 ft the overall gain should be increased to approximately 30 db, and the amplifier should be capable of delivering a peak power of at least 200 milliwatts per headset, preferably 400.

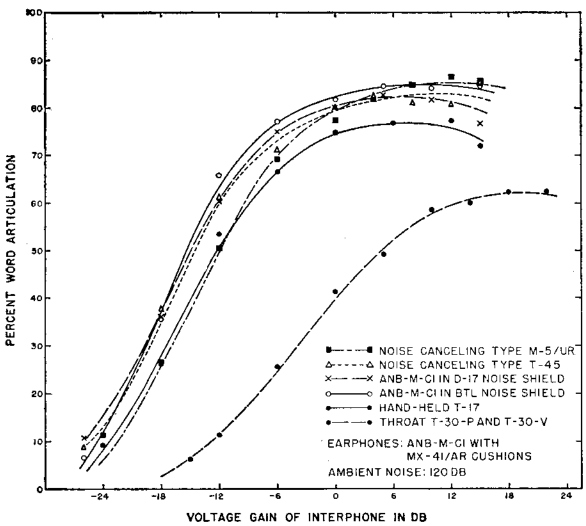


FIGURE 27. Word articulation as a function of voltage gain for six carbon microphones.

Following these recommendations, the interphone amplifier AM-26/AIC was adopted by the Army Air Forces. This amplifier possesses a power-output capability of approximately 5 watts as compared to approximately 0.75 watt available from the amplifier BC-347C which it replaced. In addition, the AM-26/AIC provides increasing voltage gain with altitude.

The improvements in the amplifier, therefore, were all changes in the direction of more uniform frequency response and more adequate output power. Once the minimum requirements had been clearly demonstrated by experimental tests, most of the difficulties were readily overcome.

9.4 INTERPHONE COMPONENTS: THE HEADSET

The headset consists of three components: the earphones, the earphone sockets, and the supporting headband, neckband, or aviation helmet.[g] Its performance depends in part upon the functioning of each of the components separately and in part upon the way in which they interact upon each other.

The basic function of the earphone is to convert electric into acoustic power. Its effectiveness depends primarily upon the characteristics of the earphone itself, upon the dimensions and

[g] For present purposes the aviation helmet is considered a special type of headband.

volume of the enclosed space, headband pressure, and cushion leaks. The earphone socket serves the dual role of coupling the earphone to the ear and of excluding unwanted noise. Here again, the socket's performance is modified by secondary factors, such as the size and weight of the earphone and the headband pressure. From such considerations it is clear that the functioning of each component depends in part upon the combination that is used.

The headset which provides a satisfactory speech-to-noise ratio must have good insulation and uniform frequency response. The headsets with which we began World War II were inferior on both counts. The Army Air Forces headset, HS-23 (see Figure 28), combined highly resonant earphones, R-14, with comfortable, sponge rubber sockets which provided little insulation against frequencies below 1,000 c. The Navy headset, NAF-38610 (see Figure 29), also used resonant earphones, CTE-49015 (TH-37), with poorly insulating sockets, CMO-49104 (TC-66), and a headband which did not supply sufficient pressure to seat the sockets properly.

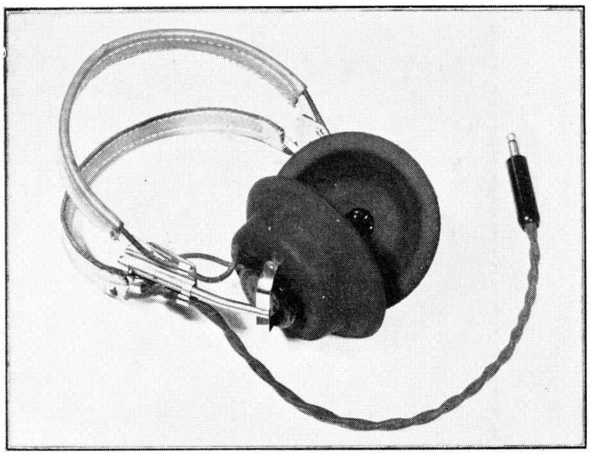


FIGURE 28. Headset HS-23.

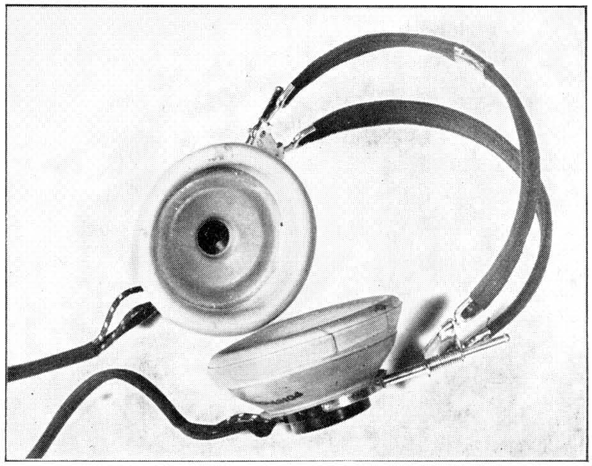


FIGURE 29. Headset NAF-38610.

A series of studies on the necessary characteristics for headsets,[2] together with the active cooperation of leading manufacturers, led to the development of magnetic and dynamic earphones with uniform response up to 4,000 c. Earphone sockets designed by the Bell Telephone Laboratories and by the Psycho-Acoustic Laboratory provided greater insulation at all frequencies. Such improvements made it possible for the AAF to adopt the magnetic earphone, ANB-H-1, and the Bell Laboratories' socket, MX-41/AR, as components in their later headset HS-33 (see Figure 30). Although these sockets provided an excellent seal, they were not so comfortable as the Harvard Socket, M-301. The Harvard design, a circumaural type of "doughnut" socket, was optional for headset HS-33 but was not generally issued because it was feared that the lower sensitivity of the headset could not be made up in additional power. Furthermore, the chamois cushion did not withstand heat, dirt, and humidity well, a consideration in tropical and ground force applications. The Harvard design was adopted, however, for use in aircraft helmets.

The Navy adopted a dynamic earphone, ANB-H-1A, designed to meet the same specifications as the Army's magnetic ANB-H-1. This dynamic earphone, when combined with an improved headband and cushions of the doughnut type, made up the H-4/AR headset (see Figure 31).

Early in the war the Army Ground Forces adopted headset HS-30, shown in Figure 32. The response of the earphones R-30 fell off badly above 2,500 c, and the headband and tip were not entirely satisfactory. When properly adjusted, however, the semi-insert tip furnished good protection against ambient noise. The semi-insert tip was retained, somewhat less effectively, in the later headset H-16/U (see Figure 33). The semi-insert coupler in this case is used in addition to a circumaural socket. The poor frequency response of the R-30 unit was

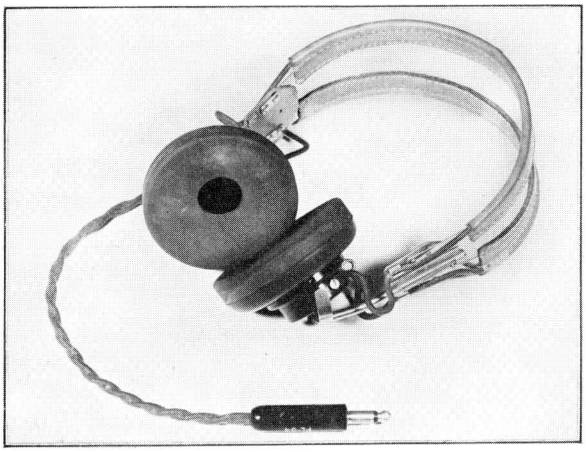


FIGURE 30. Headset HS-33.

retained, but the headband was greatly improved. The H-16/U has by far the most complicated construction of any of the headsets considered herein. Altogether, the poor frequency response of the R-30, the serious

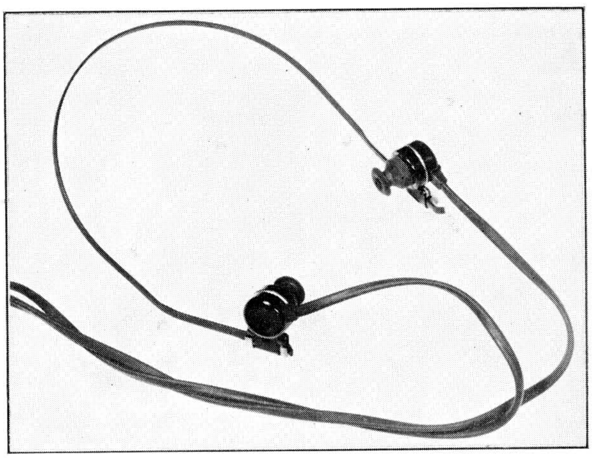


FIGURE 32. Headset HS-30.

discomfort of the HS-30, and the precarious seal provided by the H-16/U make these headsets the least satisfactory of any types now in use.

In general, circumaural sockets are more comfortable, but they provide a larger volume and less efficient coupling between earphone and eardrum than supra-aural cushions. The volume driven by an earphone in the supra-aural MX-41/AR socket is about 5 to 6 cu cm;

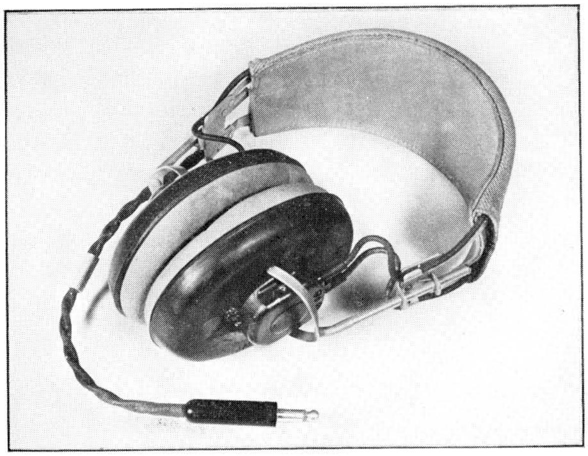


FIGURE 31. Headset H-4/AR.

with the Harvard circumaural sockets it is about 22 cu cm. The effect of the larger volume is essentially a loss in sensitivity of 5 to 10 db, depending upon the type of earphone. A more recent circumaural socket of Harvard design

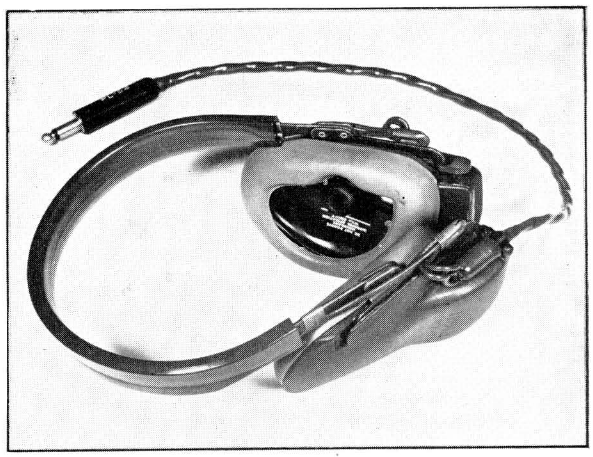


FIGURE 33. Headset H-16/U.

fits closely around the ear and thus provides the advantage of comfort along with a relatively small enclosed volume (about 11 cu cm). This socket was adopted by Canada and is shown in Figure 34.

Comfort is likewise an important factor in determining headband pressure. Lowering the pressure increases the comfort, but decreases the acoustic efficiency of the headset by allowing leakage around the socket. Leaks permit ambient noise to reach the listener's ears, and also produce a low-frequency resonance between 100 and 1,000 c. Data are available[12] which indicate the pressure necessary for each type of socket.

Articulation tests with dynamic (ANB-H-1A) and magnetic (ANB-H-1) earphones in three types of sockets[3] showed that the decreased sensitivity of sockets enclosing a large volume is directly reflected in the articulation scores. If the volume is large, more gain is needed in the interphone amplifier. The same tests also demonstrated that the dynamic earphones are slightly superior to the magnetic, regardless of which earphone sockets are used.

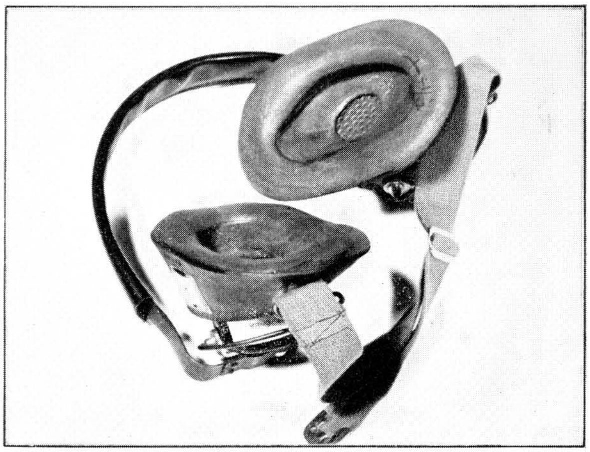


FIGURE 34. Canadian headset ZA/CAN 1637 with Harvard design 8-C earphone socket.

9.5 THE INTERPHONE AT ALTITUDE

As our fighting planes went to higher and higher altitudes, new problems and new complications were added to the already difficult task of communication in high ambient noise. Frequency-response characteristics change, sensitivity decreases, and talkers find it arduous or impossible to maintain an adequate voice level. Failure of interphone communication at altitude was a regular occurrence in the early days of high-altitude flying.

The various ills which afflict the interphone at altitude can be conveniently summarized in the term *altitude decrement* [AD]. The altitude decrement of an interphone is the sum of the decrements of its components, including the effects exhibited by listener and talker. In one way or another the altitude decrement can be traced back to the effects of reduced ambient pressure. Temperature changes from 20 C at 5,000 ft to − 20 C at 35,000 ft are also encountered,[8a] but tests have shown that changes due to temperature are of less importance to interphone performance than are pressure changes. Altitude decrement can thus be defined as the change in performance due to the reduction of the ambient pressure.

The effect of altitude on the human voice was described in Section 4.1.3 for the case where the talker does not wear an oxygen mask. Figure 14 in Chapter 4 shows the average altitude decrement for the three vowels ū, a, and ē, which is about 10 db between 400 and 4,000 c.

A more practical problem, however, is to determine the talker's behavior when he is wearing an oxygen mask. Experiments were undertaken by the Electro-Acoustic Laboratory[4] with a crew of trained men in an altitude chamber. These talkers spoke with "half effort" (see Section 4.1.3) at sea level and at altitude, and an audio-spectrometer was used to analyze the speech sounds. Several types of oxygen masks were used with a magnetic mask microphone (MC-253-A). The magnetic instrument avoids the complication of nonlinearity inherent in carbon microphones. The altitude decrement for the microphone (see Section 10.5) was then subtracted from the total decrement of voice-mask-microphone to obtain the decrement of the voice-mask combination.

The results of this investigation showed that the altitude decrement of the voice-mask combination is largely independent of the type of oxygen mask. The decrement is different, however, for different speech sounds. As a general rule the average altitude decrement of a given voice-mask combination at altitude

A_2 relative to altitude A_1 is given approximately by the expression,

$$AD = 10 \log \frac{\text{air density at } A_1}{\text{air density at } A_2}.$$

Figure 35 shows the average altitude decrement at 35,000 ft with a number of talkers

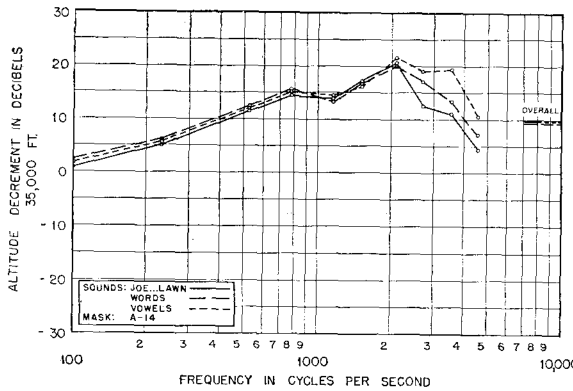


FIGURE 35. Average altitude decrement, experienced by talkers using oxygen masks, for various speech sounds at 35,000 ft. (The decrement of the microphone is not included.)

for different speech materials. Comparison with a similar curve obtained without an oxygen mask (see Figure 14, Chapter 4) reveals that the effect of the mask is to boost the levels at low frequencies. Between 400 and 4,000 c, however, the presence of the mask increases the altitude decrement.

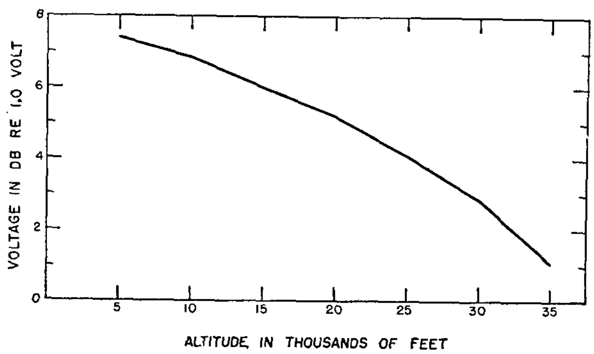


FIGURE 36. Peak instantaneous voltage at microphone terminals during articulation tests conducted in flight. (ANB-M-C1 microphone in A-14 oxygen mask.)

It should be stressed that these data were obtained with the talkers maintaining the same subjective effort at altitude as at sea level. In actual practice the talker monitors his voice in his own earphones, and when his voice level drops at high altitudes he will exert additional effort to maintain an adequate level. Thus his altitude decrement will be somewhat smaller than indicated above. For engineering purposes, the "half-effort" decrement can be regarded as including a desirable safety factor.

The decrements which are actually observed[8b] under flight conditions, where the talker monitors his voice level by his own side-tone, are indicated in Figure 36. These measurements, made in a B-17 bomber at Eglin Field, represent the decrement of the voice-mask-microphone combination as a function of altitude. Eleven announcers with ANB-M-C1 microphones in A-14 oxygen masks were used. The altitude decrement in this case amounts to about 6 db at 35,000 ft.

The variability in speech level from word to word for a given talker is approximately the same at altitude as it is at sea level. Differences between talkers are slightly greater at altitude than at sea level. The range of signal levels which the interphone amplifier is required to handle is, therefore, about the same at all altitudes; it is only the mean speech level which changes.

The altitude decrement of the voice-mask-microphone combination does not represent the total decrement for the interphone. The sensitivity of the earphones also changes at high altitudes. As the atmospheric pressure is reduced, the compliance of the air volumes increases, but the impedance of the diaphragm remains the same. The effect of this change depends upon the design of the particular earphone considered, but in general the frequency response is altered and the sensitivity is decreased (see Section 10.5).

In order to evaluate the effect of these combined decrements upon articulation efficiency under actual conditions of flight, an extensive testing program was undertaken at Eglin Field, Florida,[5, 6, 8] with the cooperation of the Aircraft Radio Field Laboratory and of the Army Air Forces Proving Ground Command. A representative testing crew of AAF men was selected and trained. During flights in B-17F bombers, the test crew was stationed as indicated in

Figure 37. Two experimenters participated in all the flights.

Typical results from articulation tests conducted under these conditions are presented in Figure 38.[h] The data were obtained with the ANB-H-1 earphones and an experimental interphone amplifier which permitted adjustment of the voltage gain in the system. In Figure 38

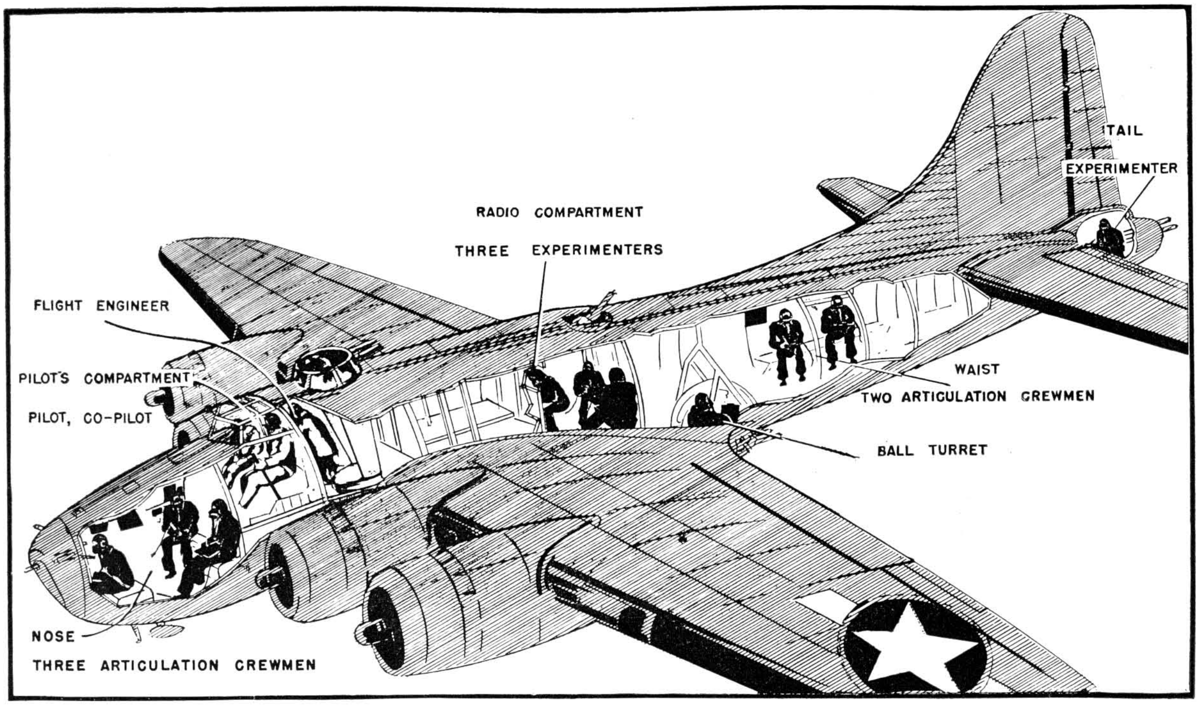


FIGURE 37. Crew positions in B-17F.

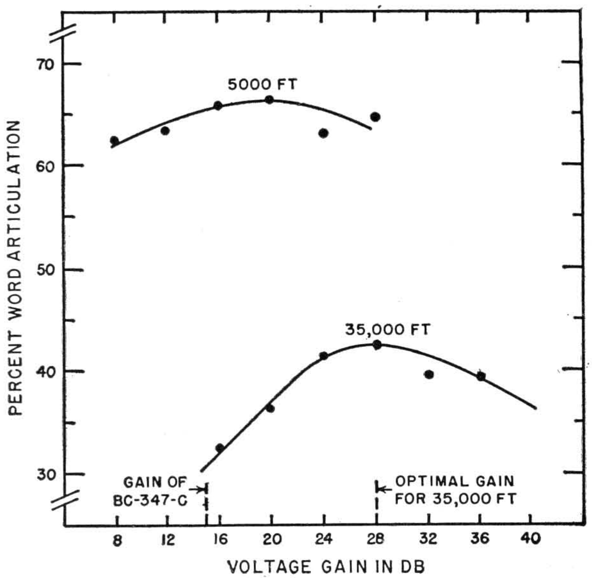


FIGURE 38. Functions showing the relation between word articulation and voltage gain at 5,000 and 35,000 ft.

the articulation of the system is plotted as a function of the voltage gain of the interphone amplifier, with altitude as the parameter. From these functions it is apparent that, by boosting the amplifier gain at high altitude, about one-third to one-half of the altitude decrement can be recovered.

On the basis of these and similar tests conducted both at Eglin Field and at the Psycho-Acoustic Laboratory, recommendations were made for interphone amplifiers (see Section 9.3) to give the optimal performance at altitudes. It is obvious, of course, that the advantage from increasing the voltage gain of the amplifier is counteracted by overload distortion if the power output is limited. The tests showed

[h] The difference in performance between 5,000 and 35,000 ft shown in Figure 38 is somewhat larger than the difference indicated for comparable equipment in Figure 1 of this chapter. The larger difference in Figure 38 resulted from the use of more difficult test words in this series of tests.

that the peak power per head set should be at least 200 milliwatts, preferably 400 milliwatts. The improvements in articulation which result from increased gain and power constitute very real advantages to bomber personnel to whom the interphone represents the key to teamwork necessary for the successful completion of a mission.