An Interphone System for "Hands-Free" Operation in High Ambient Noise

E. L. TORICK AND R. G. ALLEN

Abstract—**A new interphone system has been developed providing a high degree of intelligibility and listener comfort in ambient levels up to 125 dB sound-pressure level. This paper discusses some of the novel electronic features employed, including 1) Automatic Gain Control (AGC) with recovery inhibition to prevent noise "pumping" during speech pauses, 2) Automatic Level Compensation (ALC) for automatic adjustment of listening level with changing ambient, and 3) highly reliable voice-actuated switching (VOX) at widely varying ambient levels.**

Introduction

OVER the past years, maximum noise levels produced by military vehicles have increased by several orders of magnitude. The present situation is such that direct voice communication in many circumstances has become impossible in armored vehicles or in the presence of jet and rotary wing aircraft, and, in addition, there has also evolved a real danger of hearing damage to unprotected ears. Unfortunately, the development of communications systems which provide adequate intelligibility has lagged far behind the development of increased mobile power.

In an effort to improve vehicular communications, CBS Laboratories recently embarked on a development program for the U. S. Army. The program encompassed vehicular noise studies, psychoacoustic testing, design of acoustic components, and the development of associated electronic circuitry for achieving reliable "hands-free" interphone operation. An overall experimental system incorporating some novel features was developed, accommodating three communications stations per vehicle. This paper describes the electronic aspects of the system.

Optimum Transmission in Noise

An armored, tracked vehicle is a highly resonant chamber closely coupled to many powerful noise sources. As a result, the sound pressure level within the vehicle is extremely high, often exceeding the Threshold of Feeling. Figure 1 is a typical noise spectrum for such a vehicle. This spectrum was measured in an XM577 vehicle traveling at 35 miles per hour. The overall sound pressure level was 121 dB sound-pressure level (*C* weighting). If we contrast this noise level with the average sound pressure level of 100 dB for loud speech at a distance of $\frac{1}{4}$ inch from the lips, we can understand the difficulty in obtaining oral communication.

The first step, then, in effecting communication in such an environment is to use a suitable transducer to extract speech signals from the noise. This difficult task is made possible by the use of a pressure gradient noise-cancelling microphone. The microphone used in the project is the M-93, an especially high-performance microphone recently developed for the Air Force by CBS Laboratories.[1] The microphone has a total noise discrimination in excess of 40 dB at 200 Hz and 10 dB at 3000 Hz.

The random noise response of the M-93 microphone is shown in Fig. 2. Here, the noise-cancelling performance is shown as the difference between the upper near-field response curve and the lower far-field response curve. By weighting the upper curve with a typical spectrum of loud male speech [2], and the lower curve with the vehicular noise response of Fig. 1, we can predict what sort of speech-to-noise ratio may be expected at the microphone output. This typically is 20 dB, as shown in Fig. 3.

Voice-Actuated Transmission

The first and most important step in providing "hands-free" operation in an interphone system is to eliminate the press-to-talk switch. Unless switching is accomplished by other types of body contact, two solutions are possible: "hot" or "open" mike techniques, or voice-actuated switching, commonly called VOX. The open-mike method was not chosen because of the high level of noise picked up by the microphone which, it was felt, would contribute to listening fatigue. Consequently a new system for VOX operation was developed.

In their most elementary form, speech recognition circuits for VOX systems have been primarily amplitude-controlled. When speech signals exceed a preset threshold, a circuit is closed; when levels drop below the threshold, the circuit is opened. This method is effective only to the extent that ambient noise levels remain constant at reasonably low levels and a good signal-to-noise ratio is maintained. However with the widely vary-

Manuscript received June 29, 1966. This program was sponsored by the U.S. Army Electronics Research and Development Laboratory, Fort Monmouth, N. J., under Contract DA 28-043-AMC-00257(E). This paper appeared in the *1966 IEEE Internat'l Convention Rec.*, pt. 2, pp. 148–156. Publication of this paper does not imply permission for public use of those circuit techniques which are proprietary developments of CBS Laboratories.

The authors are with CBS Laboratories, Stamford, Conn.

[1] Under Contract AF 30 (635) 18989.

Sony v. Jawbone
U.S. Patent No. 7,246,058
Sony Ex. 1006

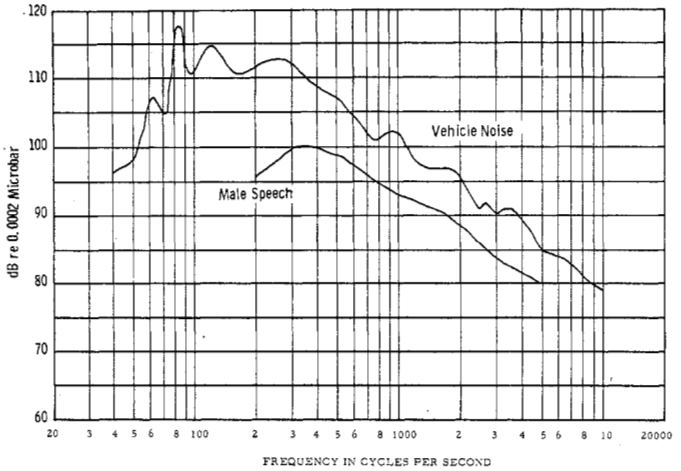

Fig. 1. One-third octave-band analysis of XM577 vehicle noise and male speech at $\frac{1}{4}$ inch.

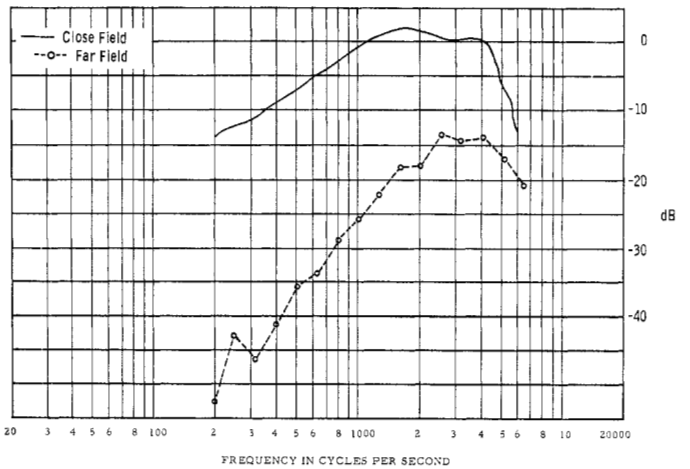

Fig. 2. M-93 random noise response.

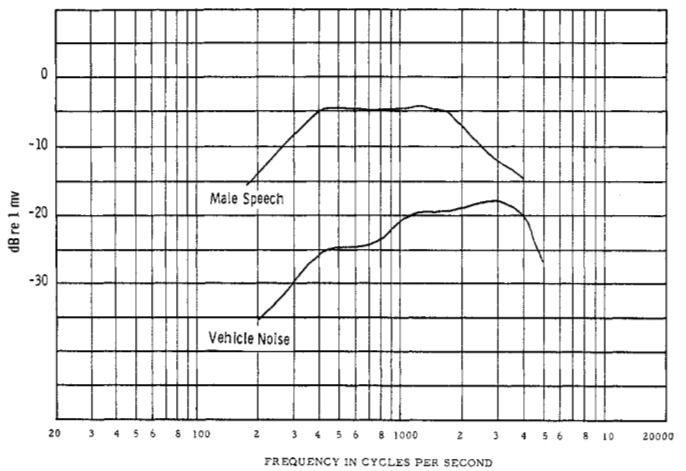

Fig. 3. Speech and noise at output of M-93 microphone.

ing noise levels experienced in track-laying vehicles, it is impossible to set a threshold level that is effective for all conditions. At one extreme, high noise will spuriously trigger the circuit; at the other extreme, low level speech will fail to initiate proper operation.

In this system, a novel solution to the problem has been devised. Two microphones are employed, one of which is used to sense the ambient noise. The noise-sensing microphone output is amplified and equalized to correspond to the far-field response of the noise-cancelling speech microphone. The outputs of both microphones are then continuously compared. When the rectified signals are equal, the switching circuit is inactive. However, when speech sounds are present at the appropriate microphone, the switching circuit energizes. Although switching occurs after a comparison of the two channels, the method is analogous to the presence of a virtual threshold which rises and falls with corresponding changes in noise levels. Investigation has shown that because of the highly reverberant nature of the vehicle's interior, the ambient noise is highly uniform in amplitude and spectrum. Consequently, a single, centrally located, noise-sensing microphone is used, feeding control units at all crew members' stations.

Figure 4 is a block diagram of a crewman's amplifier and control unit, containing VOX circuitry and automatic gain control (AGC). The basic variable gain device is a variolosser circuit, which has been previously used in two of CBS Laboratories' highly successful products developed for the commercial broadcasting industry.[2] The variolosser is essentially a diode shunt, the impedance of which is a function of the dc control current employed. The operation of the VOX circuit is as follows. The speech signal from the noise-cancelling microphone is amplified, peak-rectified, and fed to one input arm of a voltage comparator. The other input arm is fed with the rectified noise reference signal mentioned earlier. The output of the comparator feeds a Schmitt Trigger which is adjusted to fire at the syllabic rate of speech—approximately 150 milliseconds. In

[2] AUDIMAX®, Automatic Level Control, and VOLUMAX®, Automatic Peak Controller.

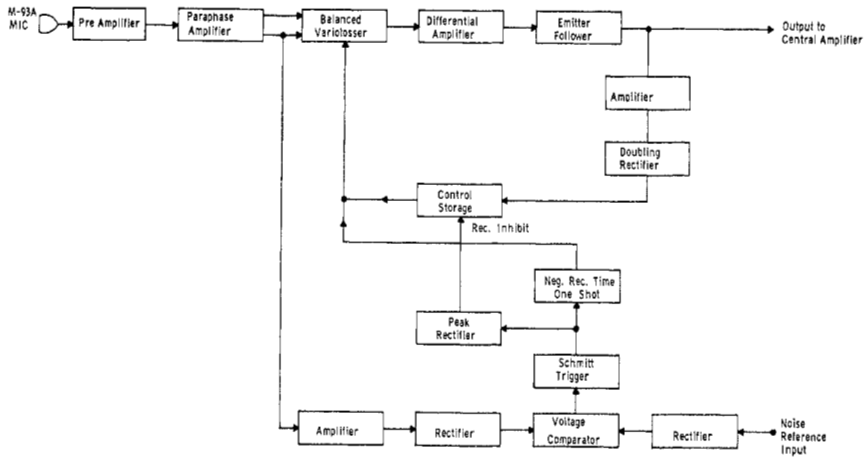

Fig. 4. Crew member's amplifier and control unit.

order to assure an uninterrupted VOX pedestal, the Schmitt Trigger feeds a circuit called a negative recovery time one-shot.[2] Conventional one-shots suffer from a finite recovery time. That is, after the circuit has been triggered to its quasistable state and returned to its stable state, a finite time must elapse before the circuit can be retriggered. This negative recovery time one-shot, however, will trigger to its quasistable state with every input pulse, and if already in that state when a trigger pulse arrives, it will remain "on" for an additional quasistable time period. The quasistable period of this circuit is two seconds, and its operation in the system is as follows. The circuit turns on, with a rise time of 2 milliseconds, thereby enabling the variolosser, when talking commences, and remains on continuously for the duration of the message. Two seconds after the last syllable it slowly returns to the squelched condition with a decay time of approximately 0.5 second. The two-second interval has been found desirable to cover brief pauses between words and sentences. When "VOX" is off, the microphone gain is reduced by about 25 dB.

Automatic Gain Control

Automatic gain control (AGC) is used to ensure a standard speech level at the output of each crewman's unit. In its basic form, the AGC amplifier utilizes the traditional feedback circuit. The signal derived from the amplifier output is rectified and the resulting control signal is "stored" in a memory capacitor. The voltage across this memory capacitor is the control voltage for the AGC loop, which is closed through an emitter follower and the variolosser. Consistent with good audio practice for control of average levels, the attack time of the AGC amplifier is 10 milliseconds. The circuit provides compression within ± 2 dB over a 20-dB range of input levels.

A distinguishing feature of this AGC amplifier is its unique method of operation in the recovery mode. Ordinary compression devices normally use a recovery time of approximately one second. This is a compromise value reflecting the designer's desire to use a faster recovery in order to maintain higher average levels, but having to settle for less in order to prevent intolerable increases in transmitted background noise during speech pauses. The solution to this problem is to inhibit recovery action during pauses, and to permit a rapid recovery when speech is present. This approach has proved to be unusually effective. It makes possible the use of a recovery time of 150 milliseconds (equivalent to the natural syllabic rate of speech) with a normal compression of 10 dB.

The Recovery Inhibitor works as follows. The output of the Schmitt Trigger of the VOX circuit is peak rectified and used to inhibit recovery of the memory capacitor. To understand this, one must realize that the Schmitt Trigger "off" level is $+24$ volts and when triggered "on," the output drops to approximately 1 volt. Consequently, during speech, the output of the peak rectifier is a low positive voltage of 1 to 2 volts. This applies a negligible voltage to the inhibit gate, and the voltage across the memory capacitor varies in the normal AGC mode. During a pause in speech, however, the peak rectifier output is $+24$ volts, which inhibits discharge of the memory capacitor. Thus, the control current, and therefore the gain, remains constant, preventing an undesirable increase in the background level. If speech stops long enough for the circuit to resquelch, the charge is still retained on the memory capacitor for several additional seconds, so that if the VOX fires within this time, it will not go to maximum gain but rather to the gain level used for the previous transmission.

Initial Consonant Loss

A characteristic of voice-actuated systems is the occasional loss of certain initial consonants. The reason for this loss is that many consonants have much smaller energy content than vowels. The level difference can easily be as much as 20 dB. Consequently, while a sys-

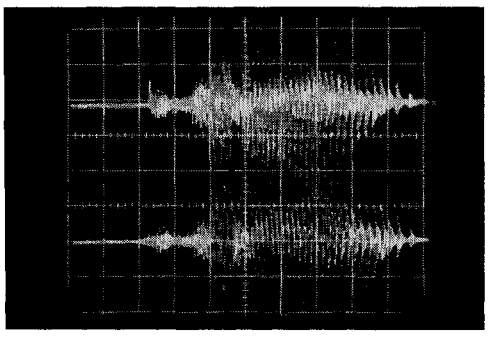

Fig. 5. Speech oscillogram—"sa."

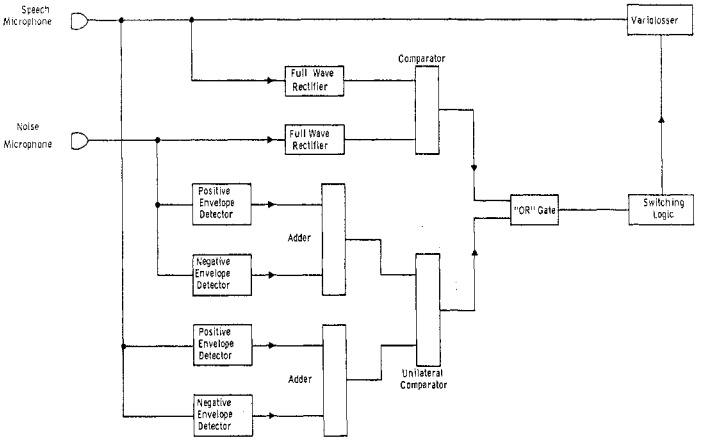

Fig. 6. Voice actuated switch with asymmetry detection.

tem may have an average speech-to-noise ratio of 20 dB, the ratio for the initial phoneme is 0 dB and the circuit cannot switch.

Figure 5 graphically illustrates an extreme example of this effect. We see here a pair of oscillograms of the phoneme pair "sa." The difference in level between "s" and "a" is 20 dB. The lower trace is the original signal; the upper trace is the signal after VOX switching. We can see that 100 milliseconds of the "s" is lost. It is unlikely that this effect can be completely eliminated, because of the limited speech-to-noise ratio. However, because of the two-second delay in returning to the squelched mode, the effect can occur only at the beginning of a sentence. To compensate for this minor effect, the use of an expendable word proves completely effective.

Additional Test for Speech Recognition

In the system described, amplitude difference constitutes an adequate means for identification of speech on the microphone line. Other techniques suggest themselves to improve the comparison reliability if needed. These techniques make use of peculiarities of speech waveforms. Speech waves contain periodic, sharply defined, glottal spikes, which cause a peak factor considerably higher than the rms amplitude. It is a well-known fact that the instantaneous peak amplitudes of speech waveforms are commonly 6 to 16 dB higher than the average or rms value. On the other hand, vehicular noise is more continuous and the peak vs. average values more nearly alike. Use of a very short attack time (less than 1 ms) in the peak rectifier for the speech signal, and a longer attack time (approximately 10 ms) for the reference noise signal will enable the difference in peak factor to aid the comparison.

An additional test can also be applied to improve switching reliability, based on the distinctive characteristic of human speech to produce asymmetrical waveforms. While not all speech sounds exhibit this quality, asymmetry often occurs in sharply spoken exclamations as at the onset of transmissions in high noise. Whereas speech waveforms may be highly nonsymmetrical, vehicular noise does not usually exhibit this characteristic. Therefore, by comparing the symmetry of the electrical signals in the speech and noise channels, it is possible to obtain additional information for the decision-making process.

Figure 6 shows a basic method for detecting and utilizing the effect of asymmetry in conjunction with the two-microphone VOX circuit. Positive and negative envelope detectors operate on both speech (plus uncancelled noise) and reference noise channels. The outputs of each pair of detectors are added algebraically and fed to a comparator circuit. Since the outputs of the adder may be negative or positive depending on the polarity of asymmetry of the original speech sound (proper phasing of the microphones is assumed), this comparator must be a unilaterally acting device. Thus,

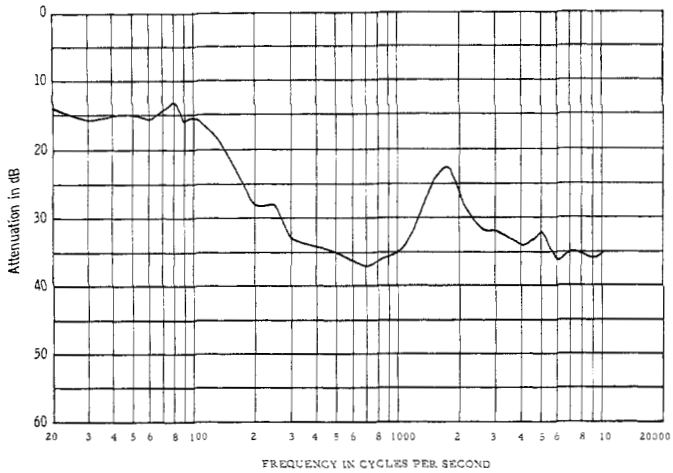

Fig. 7. Attenuation characteristic of circumaural ear defender.

when the asymmetry of the speech-noise channel is greater than that of the noise channel, the comparator produces an output of constant polarity regardless of the polarity of asymmetry. The output of this unilateral comparator is then fed to an "OR" gate in conjunction with the original VOX comparator. When either comparator produces an output, the electronic switch is energized.

The test for asymmetry cannot be used alone. When a high speech-to-noise ratio exists, this feature is not needed. However, when speech levels are close to the noise levels, the asymmetry test provides another means for recognizing the presence of speech. Furthermore, as described here, the circuit cannot be confused by asymmetrical noise even in the unlikely event of such noise occurring. The asymmetry test has not been tried in the vehicular interphone system because it was considered unnecessary in the existing sound-pressure-level ambients. However, for communication in higher noise levels it offers the means for direct extrapolation of the present approach.

Reception in Noise

In addition to extracting a maximally useful signal from the noise ambient, the intercom system must be able to shield the ears of the vehicle personnel from the noise and deliver an adequate level of speech signal to them. Most available ear protectors are effective at higher frequencies, but are inadequate at the low frequencies where most of the sound energy in a tracked vehicle is concentrated. CBS Laboratories solved this problem in a companion program[3] by the development of a well-sealed ear protection cup to which is coupled a larger added volume of approximately 200 cm^3. This added volume, which is filled with sound absorbent material, acts as a Helmholtz resonator at low frequencies and attenuates them very effectively. The extra volume is coupled to the ear defender through an acoustic low-pass filter so that voice frequencies are not coupled into it. Standard H143/AIC earphone transducers are used and are rigidly mounted in the ear defenders. Figure 7 is a graph showing the attenuation of the ear defender.

[3] Under Contract DA-28-043-AMC-00282(E).

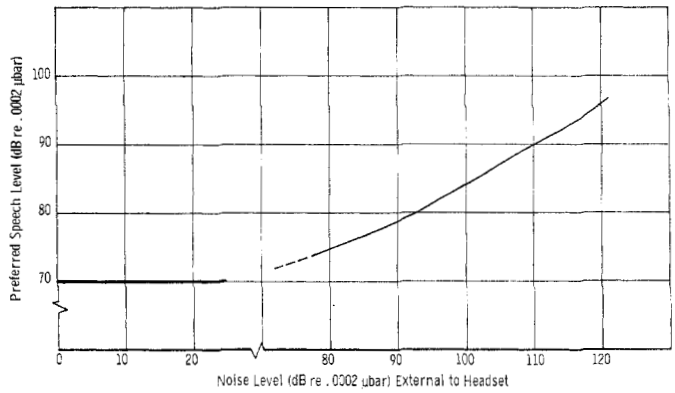

Fig. 8. Preferred speech level as function of ambient noise level.

The average sound pressure level in the armored vehicles varies from quite low levels when the engine is off, to 95 dB SPL when idling, and up to 125 dB at top speed. The typical range is 50 dB and, as might be expected, the sound power in the headphones necessary to achieve satisfactory communication varies considerably. To determine listener preference over this range of noise, a panel of listeners, wearing the previously described ear defenders, was placed in a simulated tracked-vehicle noise environment. The noise sound pressure level was varied and the subjects were asked to adjust the listening level of a transmitted message for maximum comfort and intelligibility. Figure 8 demonstrates the results obtained. It is noteworthy that the slope of the curve is approximately 0.5. That is, the average panel member preferred a 1 dB increase in program level for every 2 dB increase in ambient noise level.

Automatic Level Compensator

To complete the goal of "hands-free" operation, an Automatic Level Compensator (ALC) [3] was developed for automatically adjusting the listening volume level in response to the ambient noise level as indicated by the listener preference curve. Figure 9 is a block schematic of the central interphone amplifier unit which performs this function. The same microphone which is used to sense the ambient noise level for the VOX circuitry controls the ALC. In order to provide the desired control function over a 50 dB range, a nonlinear amplifier is used for the noise reference signal. The nonlinearity is achieved by the use of a diode compression network in the feedback path. The control element of the ALC circuit is a Raysistor which is placed in a feedback path in the output stage of the amplifier. It has been placed here, rather than as a shunt or series element in an earlier stage, in order to minimize amplifier noise and crossover distortion.

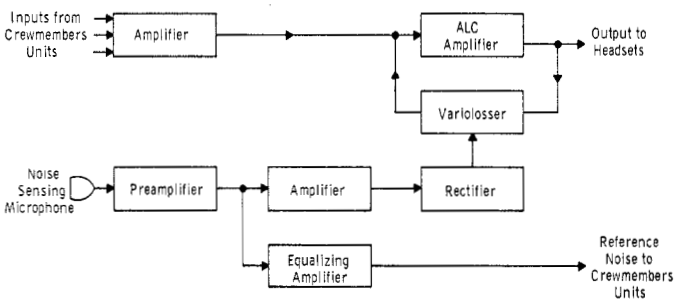

Fig. 9. Block schematic of central amplifier unit.

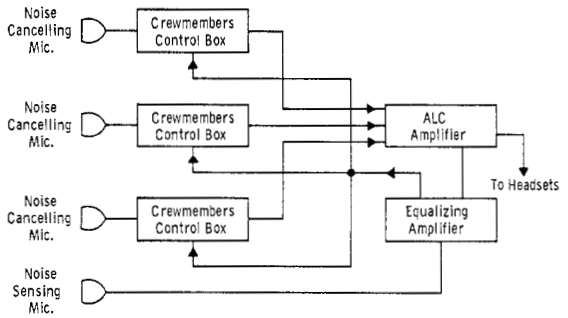

Fig. 10. Simplified system block diagram.

Summary

In summary, the simplified diagram of Fig. 10 offers a view of the total system. Three crewmen's stations have been provided. Each one of these consists of a helmet-mounted headphone and ear defender set, a noise-cancelling speech microphone, and a preamplifier and control unit including VOX, AGC and AGC Recovery Inhibition circuitry. The single interphone amplifier houses the shock-mounted noise-sensing microphone feeding each of the crewman's units and the Automatic Level Compensator which distributes audio to all headsets. No mention has been made of the various switching functions used for radio and telephone transmission since these are functionally similar to those used in the earlier Intercommunication Set AN/VIC-1 [4].

The system described here has been successfully demonstrated in the laboratory and in operational vehicular tests. As of this writing, a more extensive field test is currently being conducted by military personnel.

Acknowledgment

The authors wish to acknowledge the assistance and guidance of C. O'Malley, Contract Monitor for the government, and the efforts of their colleagues at CBS Laboratories—B. B. Bauer, Vice President for Acoustics and Magnetics, A. L. DiMattia and A. J. Rosenheck, who were responsible for acoustic investigation; and A. Kaiser and M. A. Stern who collaborated in the electronic design.

Bibliography

[1] H. Fletcher, *Speech and Hearing in Communication*. New York: Van Nostrand, 1953.
[2] J. Dixon, *Electronic Equipment Engineering*, vol. 11, p. 37, June 1963.
[3] I. Pollack, *J. Acoust. Soc. Am.*, pp. 1324–1327, vol. 29, 1957.
[4] MIL-A-55129 and MIL-C-55126.