A METHOD FOR DETERMINING INTERNAL NOISE CRITERIA BASED ON PRACTICAL SPEECH COMMUNICATION APPLIED TO HELICOPTERS*

Harry Sternfeld, Jr., and Linda Bukowski Doyle
Boeing Vertol Company

SUMMARY

The objective of this study was to provide information regarding the relationship between the internal noise environment of helicopters and the ability of personnel to understand commands and instructions. A test program was conducted to relate speech intelligibility to a standard measurement called Articulation Index. An acoustical simulator was used to provide noise environments typical of Army helicopters. Speech material ("Command" sentences and Phonetically Balanced Word Lists) were presented at several voice levels in each helicopter environment. Recommended helicopter internal noise criteria, based on speech communication, were derived and the effectiveness of hearing protection devices were evaluated. Similar limits to satisfy other types of criteria can be developed using the methods presented in this paper.

INTRODUCTION

An investigation being conducted by the Army's Working Group on Aircraft Noise indicates that speech communication requirements, as well as hearing damage risk criteria, are important factors in determining interior noise level criteria for helicopters and that the speech communication requirements are especially important in troop carrying helicopters where misunderstood instructions in combat situations may lead to serious consequences.

Speech communication in the current generation of Army helicopters is known to be inadequate but little information has been quantified in order to assist in developing adequate criteria. Although several studies have been conducted in the general area of speech communication in office-type environments, no predictive method had been validated as being completely adequate for all types of noise environments and all types of speech content. In the helicopter both of these factors are somewhat unique. The ambient noise is generally of higher level and more dominated by both low frequency noise (from rotors) and pure tones (from transmissions and turbines) than most environments previously tested; while the speech tends to be simple sentences of limited content, also quite different in structure than the material used in more generalized speech testing.

* *This study was conducted under contract with the U.S. Army Aviation Systems Command, Contract DAAJ01-74-C-1054.*

In order to provide the needed information the Army (AVSCOM) has sponsored the study which is described in this paper to assist in establishing the relationship between the internal noise environment of helicopters and the ability of troops flying in helicopters to understand commands and instructions. Information of this type, which was not previously available, can be used as a basis for reevaluating and, if necessary, revising Specification MIL-A-8806-A, "Acoustical Noise Level in Aircraft, General Specification For."

The program, therefore, applied standard speech communication test procedures as specified in Reference 1, but added noise environments and speech material typical of those encountered in troop-carrying helicopter operations in order to assess Army requirements. The same techniques can also be applied to develop criteria for application to other purposes such as passenger carrying civil helicopters.

TEST PROGRAM

Speech Intelligibility Testing

Any test involved with speech communication is called speech intelligibility testing. Two measurements which frequently come up in tests regarding speech intelligibility are Articulation Index and Speech Intelligibility Scores. Intelligibility refers to those units of speech material which are complete and meaningful words, phrases, or sentences. Test subjects are asked to listen to noise and speech simultaneously; they then record, in writing, what has been spoken. The subjects are scored on the percentage of speech recorded correctly. These scores are called their Speech Intelligibility Scores.

Articulation Index, developed by French and Steinberg (Reference 2), is a rather powerful tool for measuring speech intelligibility. The Articulation Index (AI) is a weighted number representing, for a given set of speech and noise conditions, the effective proportion of the normal speech intelligibility. AI is computed from acoustical measurements (or estimates) of the speech spectrum and of the effective masking spectrum. The detailed method for calculating Articulation Index may be found in Reference 1 which also gives suggestions for refinements to AI to take into account such things as interruption in the noise, the reverberation in the listening situation, the vocal effort used by the talker and face to face talking.

Figure 1, from Reference 1, shows the correlation between the speech intelligibility scores and articulation index in certain noise and speech environments. By use of this figure AI can be used to predict speech intelligibility.

In intelligibility testing there are four main categories of test materials (i.e., speech material) used most often (Reference 3):

(1) nonsense syllables
(2) monosyllabic words
(3) spondiac words
(4) sentences

(1) Nonsense syllables are superior to words or sentences as test items when it is desired to determine accurately the effectivenss of a device in transmitting particular speech sounds. Two practical disadvantages are the speaker must pronounce the speech sounds very precisely and the test subjects must record the sound they hear in phonetic symbols.

(2) A standard monosyllabic list contains words so chosen that all speech sounds are approximately according to their frequency of occurrence in normal speech; hence, they are termed "Phonetically Balanced (PB)." There are 20 such lists of 50 words each, the spread of difficulty being approximately the same in each list and each list having nearly the same average difficulty. Rare and unfamiliar words have been avoided as much as possible.

Two variations on the monosyllabic word lists are the rhyme test and the modified rhyme test which are tests of phonemic differentiation. The rhyme test vocabulary consists of 50 monosyllabic word sets of five rhyming words each; the modified rhyme test consists of 50 monosyllabic word sets of six rhyming words each (Reference 4 and 5). Within a given set, the rhyming words differ in a consonantal spelling. The subjects are scored on whether they supplied the correct consonant. Reference 5 contains more information on the differences between the rhyme and modified rhyme tests.

(3) Spondiac lists contain words of homogenous audibility, i.e., lists in which each individual word is as difficult as each other word. Such words are most easily selected from those which have the syllables spoken with equal stress on each syllable; e.g., railroad, horseshoe, airplane. They are especially useful in tests whose design is to establish accurately the amplification or power level at the threshold of hearing.

(4) There are different forms of sentence lists. In one form, the test subject is required to respond to questions or commands by an appropriate word or phrase. A sentence would then be either "right" or "wrong," depending on whether or not it was clear that the subject understood the meaning.

In another type of test, the listener is required to record in writing the sentence that is read to him. In sentences of this type, there are "key words" that are to be marked as right or wrong. An effort is made to avoid clichés, proverbs, and other stereotyped constructions, as well as the too frequent use of any one word.

Noise Simulation

One of the obvious problems in conducting a speech intelligibility program involving helicopter noise is that of providing enough different helicopter interior environments (and systematic variations in environment), which are stable and repeatable enough to permit the required testing. Prior to this program, Boeing Vertol had developed a useful capability for synthesis of internal noise. This simulator (Figure 2) consists of a full-scale mockup of a 15-passenger helicopter cabin and cockpit. Loudspeakers and horns are mounted at various locations on and around the fuselage in order to provide an independent sound source for each noise component. Each speaker receives its signal from a separate amplifier which, in turn, is input from a separate track of a 14-track tape playback. There are three fundamental types of signals which are employed in the simulation, broadband noise, continuous pure tones and pulsed harmonic sets.

The broadband noise, such as engine and boundary layer, are produced by random noise generators and then shaped through a one-third octave band filter set to achieve the desired spectrum.

Pure tone components, such as transmission gears, occur in harmonic sets. Initially an oscillator is used to generate each harmonic and that signal, at its required level, is put on a separate track of a magnetic tape. This tape is then dubbed so that the desired harmonic combinations are recorded on a single track of the final tape.

The pulsed harmonic combinations are used for main and tail rotors. In this case each harmonic is stored in the memory of a digital averager until the entire required combination is in memory. The averager, which had been especially modified, was then instructed to read the stored signal out at a rate corresponding to the blade passage period. The key to making this system work lies in the use of on-line real-time data analysis of signals sensed by microphones at desired locations in the aircraft as the signals are generated. In this manner the input is adjusted so that the desired output is obtained, thereby automatically accounting for electronic component frequency characteristics and for "room" acoustics.

In any finite size enclosure with directive sources, the sound field, especially of pure tone components, will not be uniform. Anyone who is familiar with helicopters knows that substantial differences in transmission noise level can be experienced by merely moving one's head. Therefore any meaningful sound pressure level data must be obtained by space averaging both during simulation development and final analysis. To achieve this, occupancy was limited to four test subjects sitting in a rather tight square. During simulation development a microphone on a rotating turntable was used. Each noise component was provided from the appropriate loudspeaker and its level adjusted so that its average level during rotation reached the desired value.

A preliminary review was conducted of available data on internal noise in current Army helicopters. This compilation resulted in the scatter shown in Figure 3. Also shown is the level defined by specification MIL-A-8806-A indicating that most procurement, to date, has deviated from established requirements.

Since the main concern of this study is directed at current Army troop-carrying helicopters, this constraint limits the field to two models: the

Bell UH-1 (Huey) and the Boeing Vertol CH-47 (Chinook) series. Accordingly, the following were selected as representative for the test program:

1. CH-47C at 22 680 kg (50 000 lb) gross weight (Current Army configuration).

2. CH-47A at 14 969 kg (33 000 lb) gross weight (Original Army Chinook configuration which contained more acoustical treatment in the aft cabin than the CH-47C).

3. A "Paper Design" CH-47 which is a predicted spectrum of a Chinook designed to meet MIL-A-8806-A. This will be referred to as CH-47 'MIL Spec.'

4. UH-1H at 4309 kg (9500 lb) gross weight.

5. UH-1H with doors open which is a configuration often used to permit rapid egress of troops when approaching landing zones.

In order to develop the required simulations, the aircraft noise must be studied in detail in order to identify individual components by means of narrow band spectra shown in Figures 4 and 5. Analyses of this type of data from several locations in each aircraft was done to develop the spectra which were the basis for preparation of the test samples. A typical simulation consisting of two broadband and thirty harmonic components is presented in Figure 6.

Speech Material

This program, after much consideration of the available speech materials, used both standard Phonetically Balanced (PB) word lists and non-standard "Military Jargon" sentence lists. The 1000 PB word list is from Reference 6. The purpose of the PB testing is to provide continuity between this program and the results of other speech communication studies.

Since this program is concerned with determining the effect of helicopter noise on communication with troops, the "Military Jargon" sentence lists were developed as samples designed to reflect somewhat typical commands and questions which might occur in an assault helicopter. Some typical sentences were:

Hold it - move now.
Advance to the northeast sector.
Out the front door.

Sentences of this type are more pertinent to actual Army usage than the more theoretical PB word lists. There were five different military sentence lists, each one composed of twenty sentences. Each of the five lists was randomly scrambled once; hence, there was a total of ten lists (two hundred sentences) of the military type employed in the program. The two hundred sentences contained a total of 300 key words on which scoring was based. It will be noted that the military sentences on the whole are much shorter than the standard sentences of Reference 3 of the type whose test results are shown in Figure 1. Also, they are limited to fewer words; whereas the standard sentences avoid cliches and stereotyped expressions, the military sentences are composed mainly of these.

The PB word lists and "Military Jargon" sentence lists were tape recorded for playback through a loudspeaker located so that the sound came from the front of the cabin. The individual selected as the speaker had no strong regional accent and had clear diction. As the data was recorded the peak levels were monitored on a graphic level recorder to assist in maintaining a constant voice level. A raised voice was used rather than a conversational tone to produce the type of compressed range associated with higher voice levels. A calibration tone was included on each tape for the purpose of providing a constant reference for setting the desired playback volume levels. The use of this procedure ensured repeatability of data between test sessions.

Test Procedures

The subjects in this program were twelve males who were members of various engineering staffs of the Boeing Vertol Company. Potential candidates were screened by audiograms, conducted by the Boeing Vertol Medical Department.

Prior to the start of the test program, the twelve test subjects were divided into three groups, consisting of four subjects each (the subjects were allowed to form their own groups of four). At a pre-test session, the subjects were given their instructions, they familiarized themselves with the military sentences and a practice session was held. Since testing with phonetically balanced words is a much more rigorous procedure which requires training to a level where other subjects score at least 90% in a quiet environment, these practice sessions were held at each test session scheduled for military jargon testing. By the time PB word testing all subjects had exceeded the minimum requirements.

Each subject was equipped with writing utensils, a lap board and answer sheets. During a particular test (a test is defined as one aircraft level and one speech level), the helicopter noise was continuous; one word and/or sentence at a time was played with a finite time interval for each subject to write what he heard or what he thought he heard. Minor rests of about 1 minute were given at approximately 5-minute intervals with longer rest periods at every 15 minutes.

In order to ensure that each test would produce a meaningful range of results, a pretest was performed for each aircraft noise environment. In this pretest the voice playback level was varied until, in the opinion of the Test Director, only a few of the messages could be understood. The level was then readjusted until it was judged that most of the messages could be understood. At each of these levels the value of the 1000 Hz reference tone was measured. These calibration levels and one additional level which was the average of the above two were then used to set the voice levels prior to each test.

In addition to the full aircraft spectra discussed above, additional variations were obtained by completely eliminating individual components by disconnecting the appropriate tape track from the simulation system during playback. This not only provided more test environments but was also used to evaluate the relative importance of the various noise sources in affecting speech communication.

Additional testing was conducted in one single rotor and one tandem rotor configuration to evaluate the effect of hearing protective devices.

RESULTS

Data Analysis

The phonetically balanced word lists were simply scored as correct or incorrect for each individual word. Phonetic spelling and/or misspellings were counted as correct providing the word was recognizable. Failure to fill in any word was counted as incorrect. In the sentence lists, failure to correctly recognize any key word was scored as not understanding the sentence, hence there was usually more than one chance to miss a sentence.

The principles of calculating A.I., which are fully described in Reference 1, require measurement of both the ambient noise and speech levels at the observer's location. In order to define the aircraft noise levels, one-third octave band measurements were made at the left and right ear position of each subject for each separate environment tested. The left and right ear spectra were arithmetically averaged to obtain a single spectrum for each seat location for each test.

Reference 1 recommends the use of long-term rms spectra for speech measurement. Accordingly four sections of the PB tape were played through the speaker system and analyzed in a manner which provided one-third octave band spectra integrated over 32 seconds for each sample. Averaging these samples thus produced a 128 second rms spectrum. Since different playback levels were used for each test, a curve of absolute level against calibration tone level was made in order to assign amplitude values to the long term rms spectrum shape.

Articulation Index (A.I.) was calculated as defined in Reference 1. Of the three methods (full octave, one-third octave and 20 band) described in that reference, the one-third octave band method was used due to compatibility with available analyzing equipment. Speech peaks were determined by adding 12 dB to the measured long-term rms values of the speech actually used in preference to use of the idealized voice of Reference 1.

Speech Communication

The fundamental results of this program are shown in Figure 7 for all the combinations of aircraft noise condition and speech levels tested. Since we are concerned with evaluating the general situation in the cabin of military helicopters and not the characteristics of specific locations, the data for each of the four test locations has been combined to give a single value representative of the aircraft/speech combination tested. Each PB data point of Figure 8 is based on 1200 and each Military Jargon data point on 720 separate evaluations.

Comparison of these results with the more general data of Reference 1 is illustrated in Figure 8 and indicates that these latter criteria would be too conservative if applied to the helicopter. Since this difference appears to be true of both the sentences and PB words it does not seem that the type of speech material used is responsible, therefore the difference probably is due to the particular characteristic of helicopter cabin noise. Although the scope of this program cannot rigorously define the reason, a possible explanation may lie in the fact that in the speech interference range the helicopter noise is often predominated by pure tones generated by dynamic components. In this case the value of a one-third octave band of noise may be set by only a very narrow portion of the bandwidth, leaving the rest of the band available for much better speech communication than would be indicated by an Articulation Index calculated on the basis of the one-third octave band (or even the twenty band method). This implies that the use of an equivalent band level concept (by adjusting the measured band level by $10 \log \frac{b_t}{B}$ (where b_t is the bandwidth of the tone and B the bandwidth of the one-third octave band in question) might be applicable. Such a procedure would have to be applied exercising great judgment in cases where a particular level is set by several tones, or a combination of tones and broadband noise. A more practical approach is to assume the data developed in this program as being more correct than the Reference 1 curves for application to helicopter internal noise.

In evaluating speech communication in a given aircraft, there are essentially three parameters which must be considered:

(1) The voice level which may be used.

(2) The distance over which communication must take place.

(3) The reliability of understanding which must be achieved.

In order to perform the required evaluations, it was necessary to assume some voice levels and spectra. For the greatest general applicability, the spectrum shape used was the ideal voice spectrum of Reference 1 as opposed to the specific spectrum of the talker used in this program. The rms overall sound pressure levels corresponding to the descriptions were obtained using a simple test in which three males (who had helicopter experience) used voice levels which were felt to be typical of those employed in military helicopters. The levels were measured using a one-third octave band analyzer with a graphic level recorder set such that the response was an rms level corresponding to a standard sound level meter set for "slow" response. (This corresponds to a pen decay rate of about 40 dB/sec at a writing speed of 80 mm/sec or a 1-second time constant.)

For discussion purposes, three voice levels were evaluated:

"Loud Shout" (90 dB rms at speech peak value at 1 meter) was a level which could be used to issue command sentences to troops and was not unlike those employed in addressing ground troops during military drill.

"Short Duration Shout" (100 dB rms at 1 meter) was a level which could be sustained for up to about ten continuous words without rest.

"Maximum Effort" (110 dB rms at 1 meter) was a level which could only be sustained for a few words and resulted in some vocal strain; this level might be expected in emergency situations.

Having defined the voice spectra as just described, the Articulation Index can be calculated for each aircraft by using the appropriate internal noise data and the method of Reference 1. The "Military Jargon" curve of Figure 7, which was developed by this study, can then be used to convert the abstract Articulation Index to a more applicable speech intelligibility percent.

In order to assess the adequacy of speech recognition in Army helicopters, it is necessary to establish requirements which will permit troops to perform their mission. Unfortunately, no such standard has been established. Lacking any published guidelines, two interim criteria are suggested.

A minimum requirement should be based on communication essential to safety. For example, when using maximum vocal effort (110 dB rms at 1 meter), it should be possible to achieve 50% intelligibility which would at least attract the attention of a person anywhere in the cabin even if he cannot accurately understand the message content.

Although maintenance of a minimum communication in emergency communication in emergency situations is of first priority, this does not in itself ensure a comprehension which will permit troops to adequately follow instructions necessary to successful completion of their required missions. A reasonable suggestion might be based on 80% communication using a "Loud Shouting" level (90 dB rms at a distance of 1 meter). This would ensure that a troop commander, using the type of voice which might be used to instruct a platoon on the drill field, could achieve good simultaneous communication with personnel located within a radius of 1 meter of the speaker.

If it is desired to improve the speech communication reliability by increasing A.I. in an aircraft in an efficient manner, the reductions should be made in the frequency bands which contribute most directly to the A.I. Figure 9 illustrates the relative A.I. weighting factors normalized to their maximum value and suggest that noise reduction outside the frequency range of 100 Hz to 5 kHz will have relatively little payoff. Coupled with the above is the fact that it generally requires less weight to attenuate high frequency noise than low frequency. Pure mass attenuation, for example, provides 5-dB increased attenuation per doubling of frequency for the same surface density, while materials such as fiberglass greatly exceed that rate at the higher frequencies.

The procedure for establishing an internal noise level to meet a given communication requirement in an efficient manner is illustrated in Figure 10. The A.I. is calculated as described in Reference 1. The score of the aircraft noise limit required to obtain a desired value of the A.I. is found by trial and error solution.

At frequencies below 500 Hz, the aircraft noise levels need not be dictated by speech communication requirements but rather by hearing damage risk criteria.

Application of the above procedures to the suggested conditions results in the helicopter noise criteria curve of Figures 11 and 12. Other criteria based on other assumed requirements can be constructed in a similar manner.

Hearing Protection

The noise levels in helicopters not only affect communication but may, if high enough, cause temporary threshold shift, hearing damage, or at least discomfort. For these reasons, the use of protective devices may be recommended. It is important, however, to know if these devices have any adverse effect on understanding of spoken commands. To investigate this, testing was repeated in the CH-47C and UH-1H (doors open) configurations with the following protective devices:

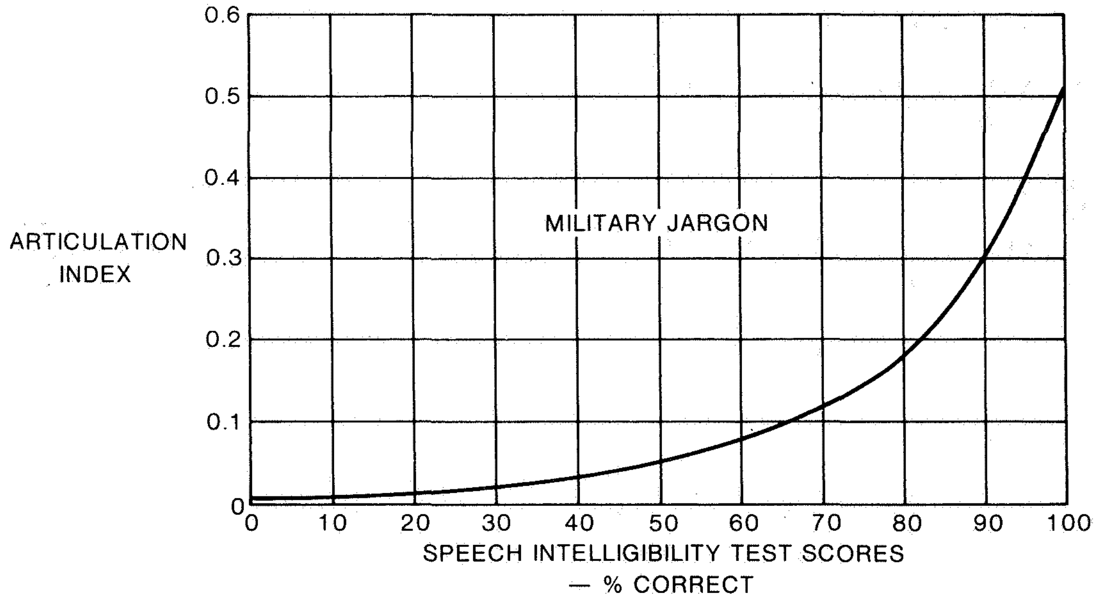
1) Army SPH-4 flight crew helmets.
2) Disposable ear plug (a spongy material which was compressed, inserted in the ears and then allowed to expand).
3) A non-disposable fitted ear plug (size for each person based on ear canal measurement).

The results which are shown in Figure 13 reveal that no adverse effect on speech communication can be attributed to the use of ear plugs and, in fact, they even enhance understanding. In the case of the SPH-4 helmet when evaluated in the UH-1H, a significant deterioration occurred. Although no measurements of noise inside the helmets were made, spontaneous comments from several of the test subjects indicated that with the helmet on, the tail rotor noise was extremely annoying. It is therefore suspected that the SPH-4 helmet may have a resonance in the tail rotor frequency range and may be amplifying one or more harmonics of UH-1 tail rotor noise.

The disposable ear plugs tested slightly better than the fitted ones, probably due to being uncritical with regard to fit. They were also judged to be more comfortable.

CONCLUDING REMARKS

In order to assure adequate speech communication in military helicopters, internal noise standards should provide for noise levels which will result in Articulation Index(es) in accordance with the following sketch (which is replotted from Figure 7):



In order to relate the Articulation Index to a sound pressure level spectrum, it will be necessary to further define

(a) the voice level to be required

(b) the distance over which communication is required

(c) the required reliability of communication

Additional investigation of Army requirements is needed to establish these parameters.

Ear plugs can be used to protect hearing with little or no degradation in speech communication.

The procedures developed during this program can be used to develop internal noise criteria for other applications such as civil transports.

REFERENCES

1. Anonymous: METHODS FOR THE CALCULATION OF THE ARTICULATION INDEX, ANSI S3.5-1969, American National Standards Institute, New York.

2. Kryter, Karl D.: THE EFFECTS OF NOISE ON MAN, Academic Press, New York, 1970.

3. Beranek, Leo L.: ACOUSTIC MEASUREMENTS, John Wiley and Sons, Inc., New York, 1962.

4. Fairbanks, Grant: TEST OF PHONEMIC DIFFERENTIATION: THE RHYME TEST, Journal of the Acoustical Society of America, Vol. 30, No. 7, July 1958, pp. 596-600.

5. Kryter, Karl D., and Whitman, Edward C.: SOME COMPARISONS BETWEEN RHYME AND PB-WORD INTELLIGIBILITY TESTS, Letters to the Editor, Journal of the Acoustical Society of America, Vol. 37, No. 6, June 1965, p. 1146.

6. Anonymous: METHOD FOR MEASUREMENT OF MONOSYLLABIC WORD INTELLIGIBILITY, ANSI S3.2-1960 (Revised 1971), American National Standards Institute, New York.

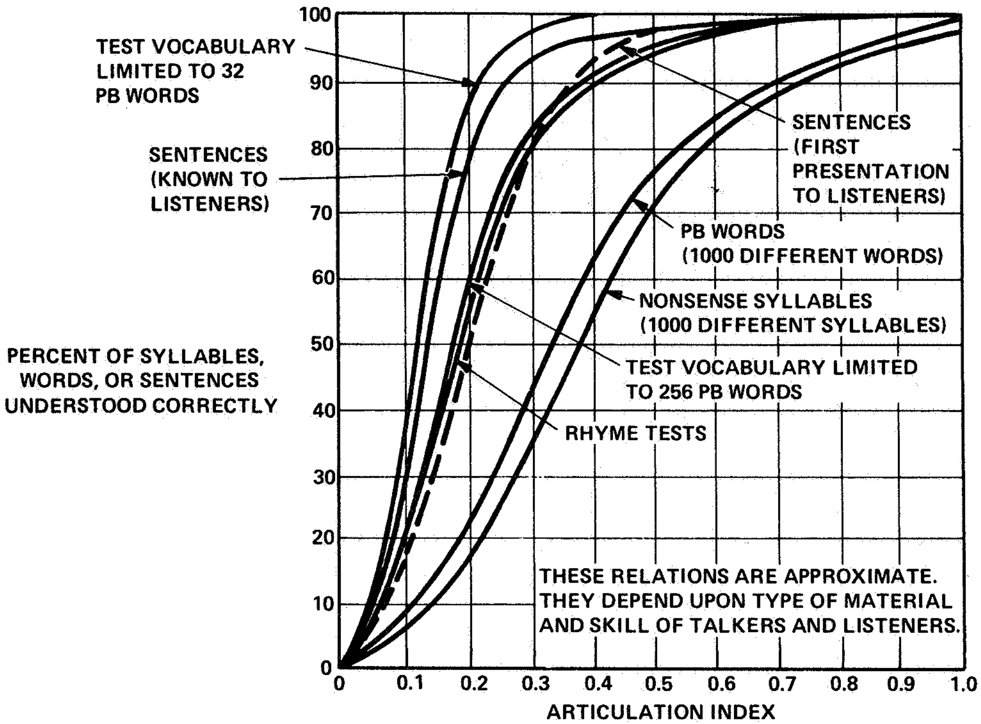


Figure 1.- Relationship between articulation index and speech intelligibility (ref. 1).

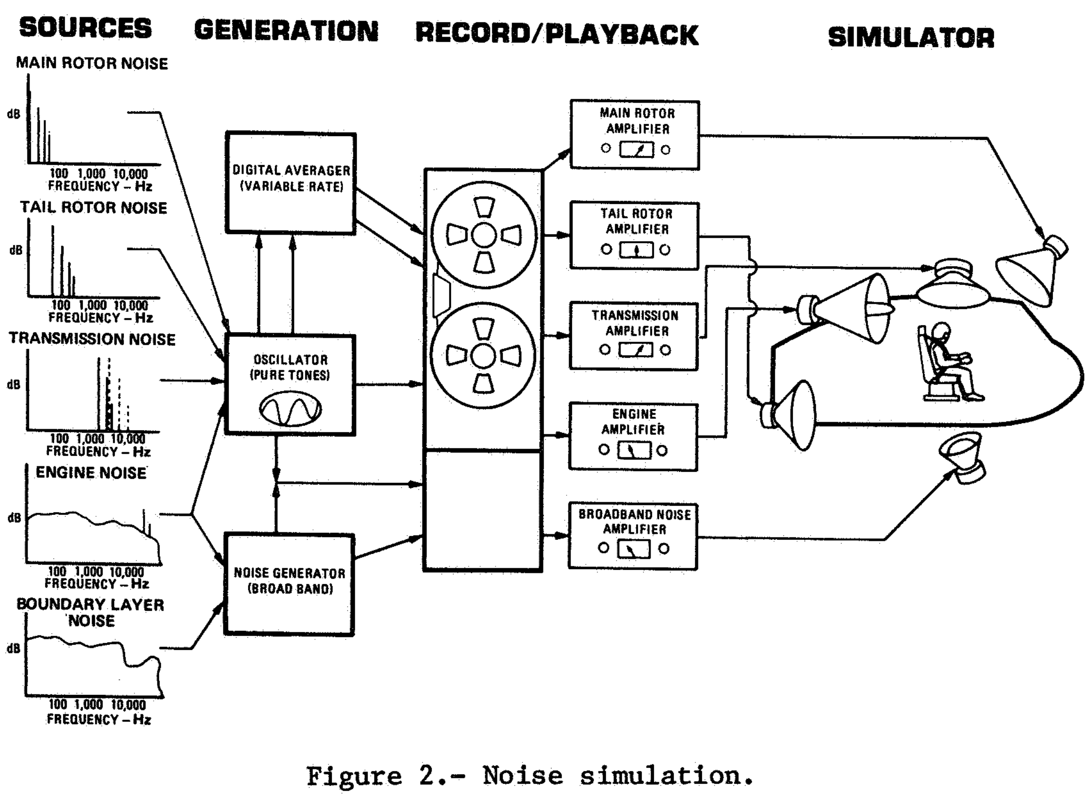
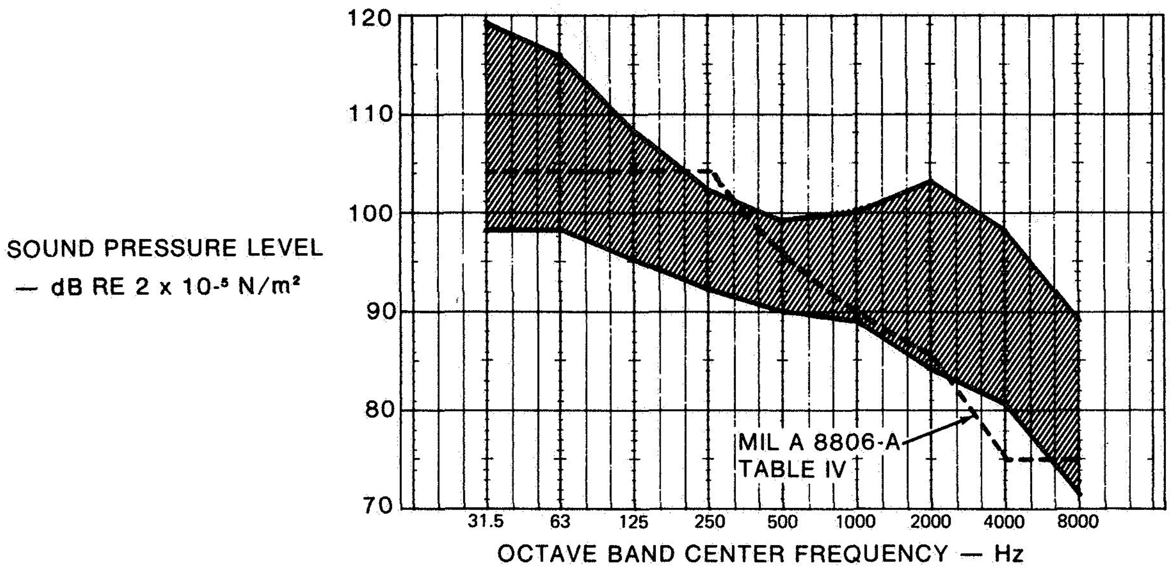


Figure 2.- Noise simulation.



(DATA INCLUDES CH-47C, CH-47A, UH-1H, AND OH-58)

Figure 3.- Scatter of center cabin noise data in army helicopters.

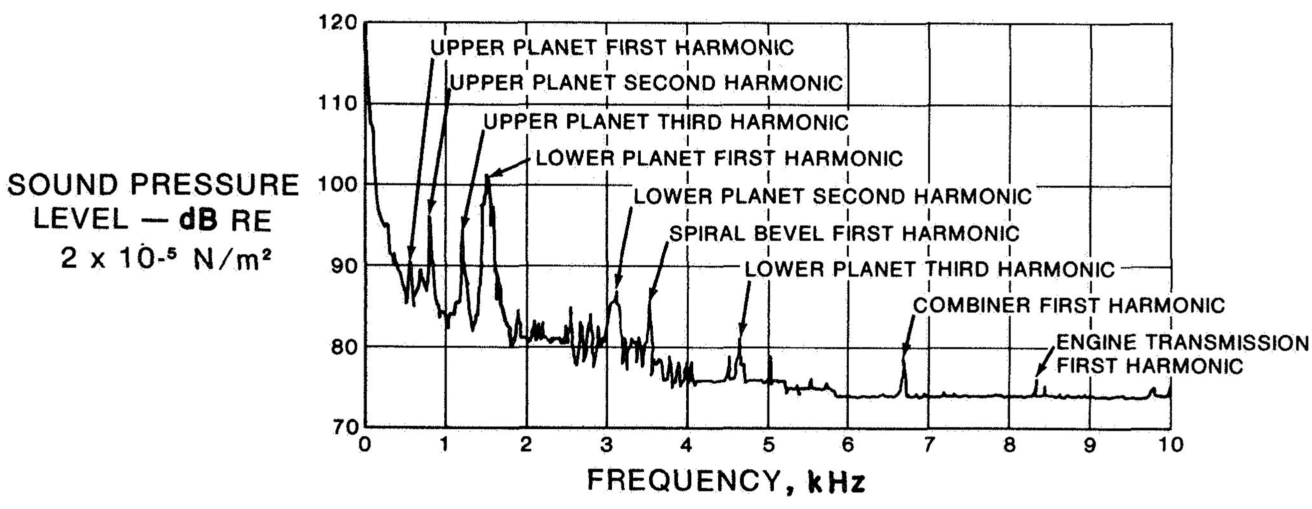


Figure 4.- Noise spectrum of CH-47C interior at station 320 (20 Hz bandwidth).

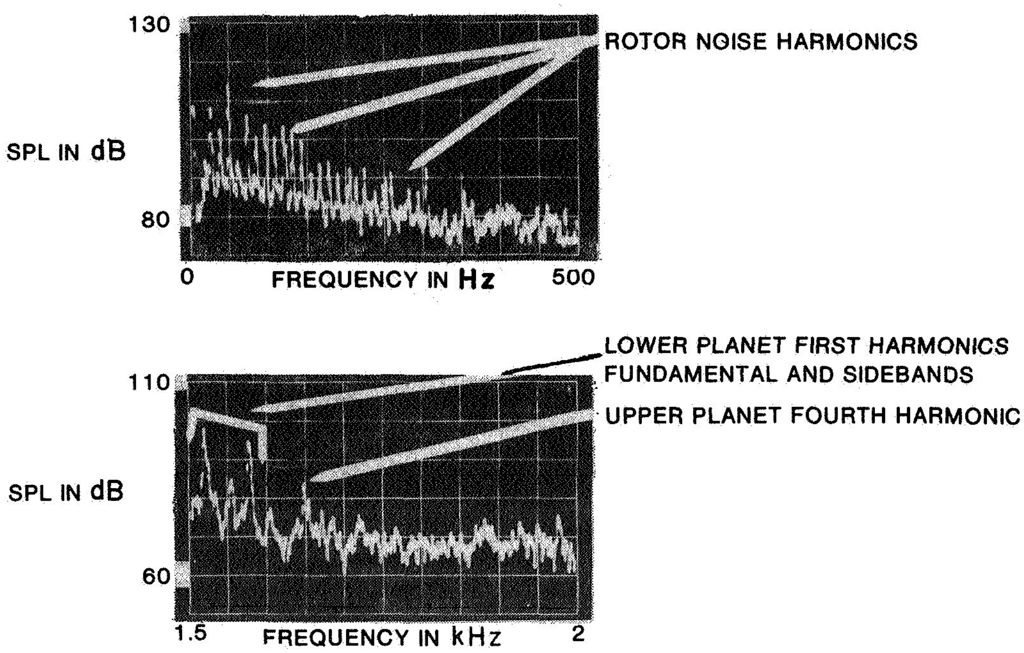


Figure 5.- High resolution noise spectra, CH-47C cabin.

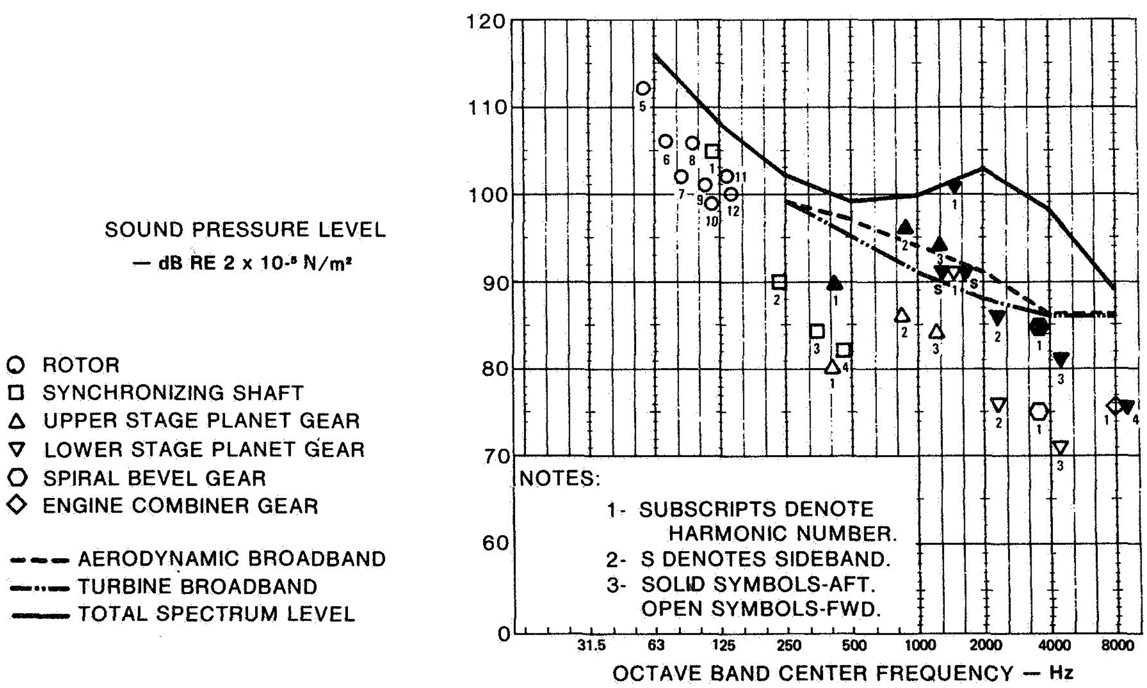


Figure 6.- Noise components of the CH-47C.

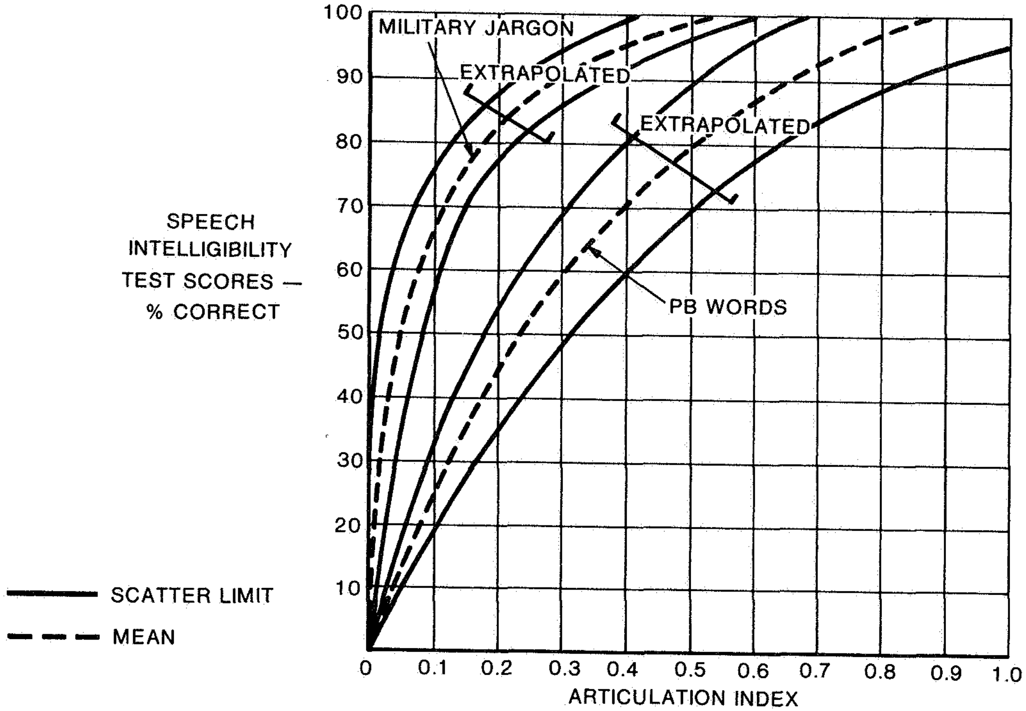


Figure 7.- Test results - relation between speech intelligibility and articulation index in helicopters.

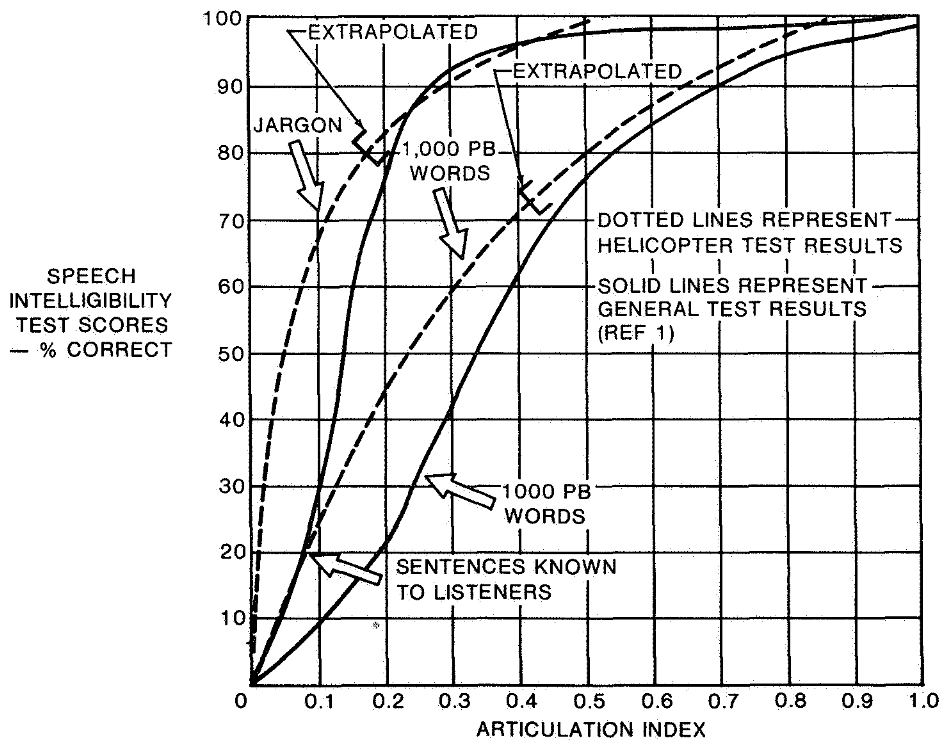


Figure 8.- Comparison of helicopter and general test results.

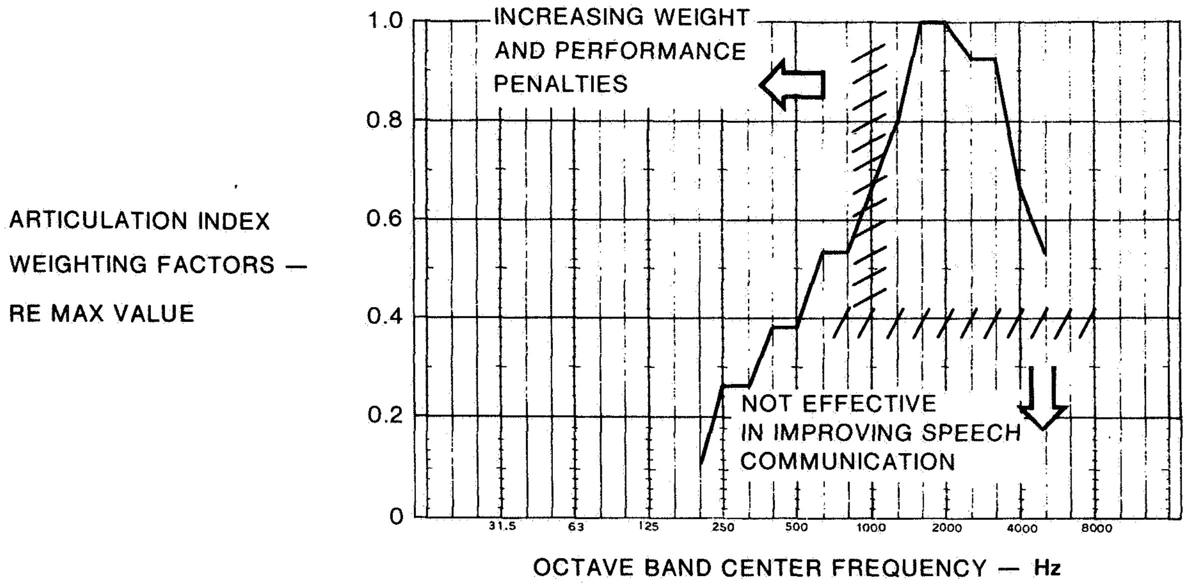


Figure 9.- Articulation index weighting factors.

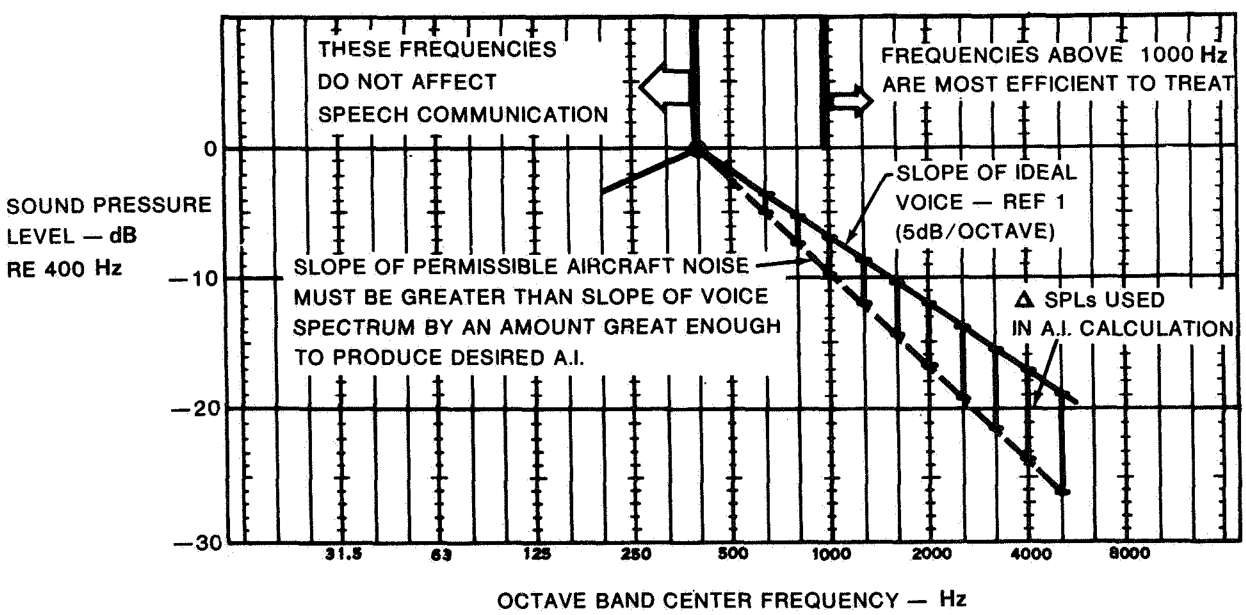


Figure 10.- Principles involved in construction of an aircraft noise criterion which results in minimum weight penalty.

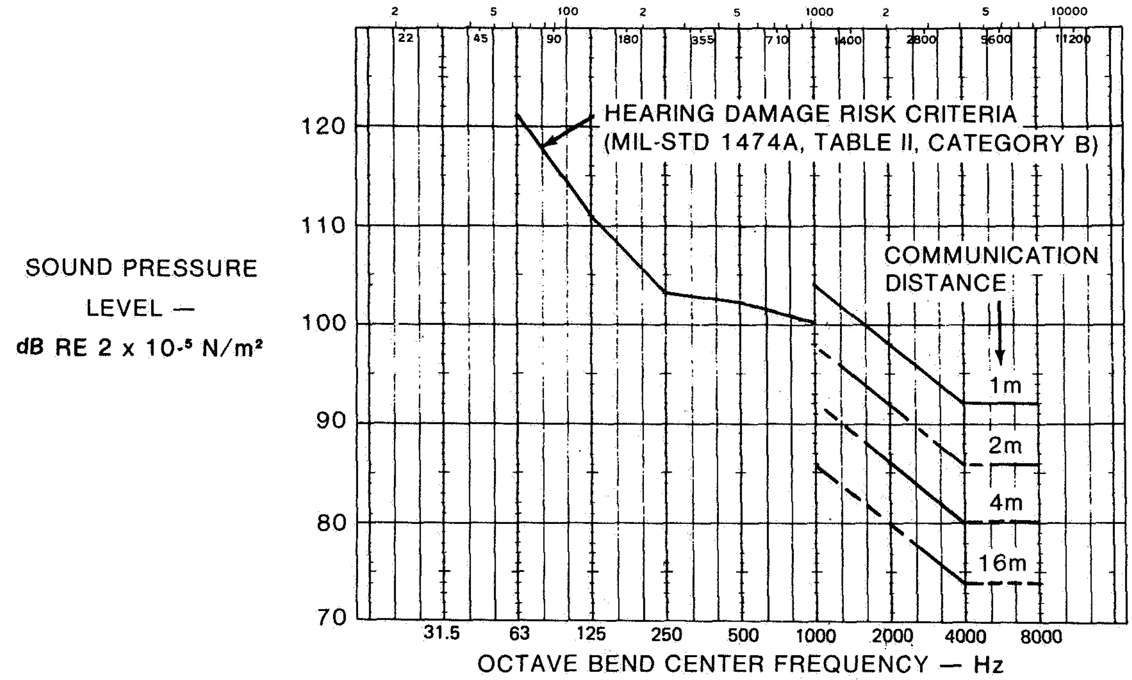


Figure 11.- Helicopter interior noise levels for required emergency commands, 50% speech intelligibility using a maximum effort voice, 110 dB rms at 1 meter.

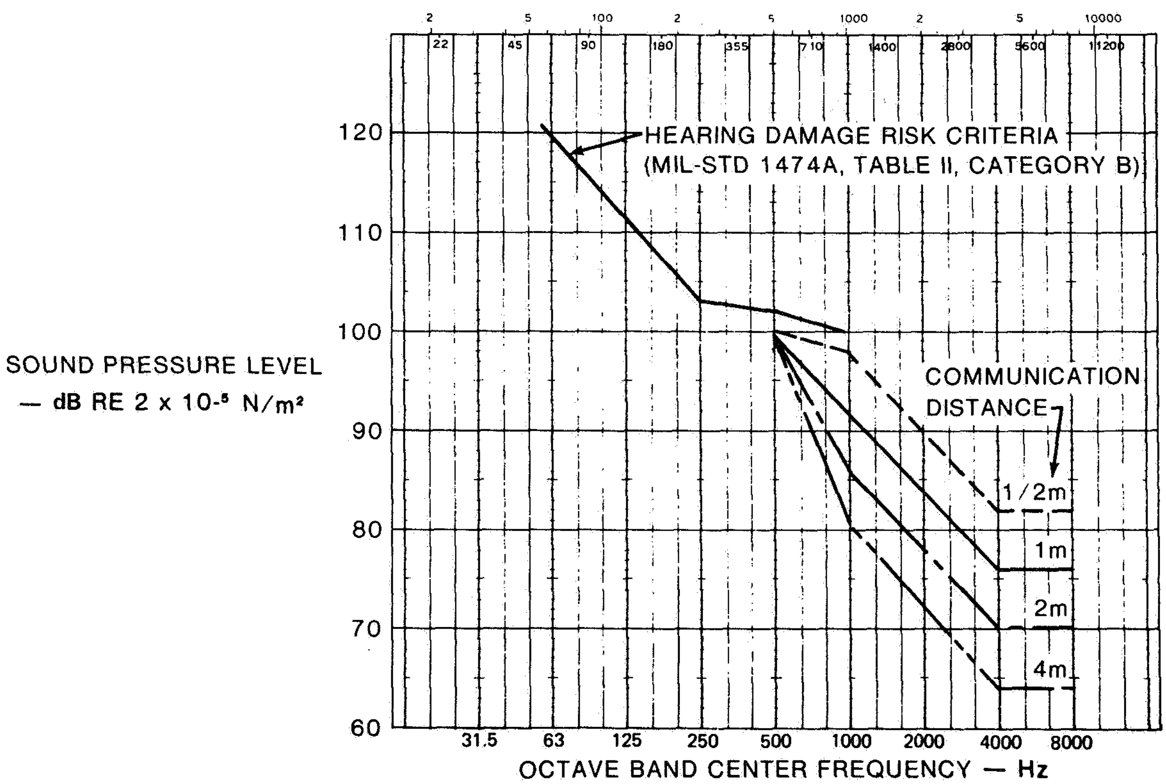


Figure 12.- Helicopter interior noise levels for troop instruction, 80% speech intelligibility using a loud shouting voice, 100 dB rms at 1 meter.

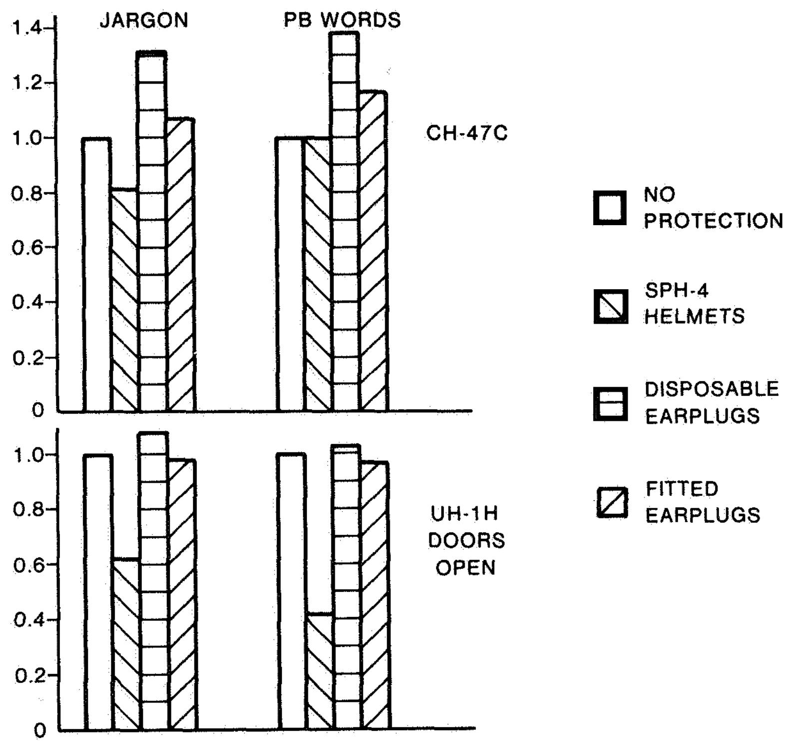


Figure 13.- Effect of hearing protection on speech intelligibility.