Ministry of Defence

INTERIM

Defence Standard

00-25 (PART 9) / Issue 1

30 April 1991

HUMAN FACTORS

FOR DESIGNERS OF EQUIPMENT

PART 9: VOICE COMMUNICATION.

AMENDMENTS ISSUED SINCE PUBLICATION

AMD NO	DATE OF ISSUE	TEXT AFFECTED	SIGNATURE & DATE

Revision Note

Historical Record

Arrangement of Defence Standard 00-25

The arrangement of the Parts comprising Def Stan 00-25 is shown below:

PART 1 - Introduction
PART 2 - Body Size
PART 3 - Body Strength and Stamina
PART 4 - Workplace Design
PART 5 - The Physical Environment: Stresses and Hazards
PART 6 - Vision and Lighting
PART 7 - Visual Displays
PART 8 - Auditory Information
PART 9 - Voice Communication
PART 10 - Controls
PART 11 - Design for Maintainability
PART 12 - Systems

Two or more Parts may apply to any one equipment and it is, therefore, essential that all Parts be read and used where appropriate.

HUMAN FACTORS FOR DESIGNERS OF EQUIPMENT

PART 9: VOICE COMMUNICATION

PREFACE

i This Part of this Defence Standard presents descriptive detail of voice communication for aiding designers, reflecting factors likely to affect equipment design such as frequency range, distortion and masking by noise.

ii This Part of this Defence Standard is published under the authority of the Human Factors Subcommittee of the Defence Engineering Equipment Standardization Committee (DEESC).

iii This Standard should be viewed as a permissive guideline, rather than as a mandatory piece of technological law. Where safety and health is concerned, particular attention is drawn to this Standard as a source of advice on safe working limits, stresses and hazards etc. Use of this Standard in no way absolves either the supplier or the user from statutory obligations relating to health and safety at any stage of manufacture or use.

iv Users of this Standard shall note that some material may be claimed to be subject to copyright in this or other countries. Copyright where known is acknowledged.

v This Standard has been devised for the use of the Crown and of its contractors in the execution of contracts for the Crown and, subject to the Unfair Contract Terms Act 1977, the Crown will not be liable in any way whatever (including but without limitation negligence on the part of the Crown its servants or agents) where the Standard is used for other purposes.

vi This Standard has been agreed by the authorities concerned with its use and shall be incorporated whenever relevant in all future designs, contracts, orders etc and whenever practicable by amendment to those already in existence. If any difficulty arises which prevents application of the Defence Standard, the Directorate of Standardization shall be informed so that a remedy may be sought.

vii Any enquiries regarding this Standard in relation to an invitation to tender or a contract in which it is invoked are to be addressed to the responsible technical or supervising authority named in the invitation to tender or contract.

viii This Defence Standard is being issued as an INTERIM Standard. It shall be applied to obtain information and experience of its application. This will then permit the submission of observations and comments from users using D Stan Form No 22 enclosed.

CONTENTS

HUMAN FACTORS FOR DESIGNERS OF EQUIPMENT

VOICE COMMUNICATION

Section One. General

0 Introduction

The primary topic of this part of the Standard is transmission of audible speech messages, from speaker to listener, via airborne sound. Consideration is also given to the case where the sound energy is temporarily converted into an electrical signal, for transmission (eg through a telephone line), then reconverted into mechanical vibrations, audible to the listener.

Various aspects of speech production and reception, pertinent to voice communication, are reviewed, but first the frequently met concept of signal-to-noise ratio will be explained. The signal-to-noise ration (S/N) is an important basic measure in communication, since it effectively determines the likelihood that a signal will be accurately received. It describes the intensity level of a speech sound (the signal), relative to the intensity level of the irrelevant background noise. The measurements of signal and noise are both made at the same location (such as a listener’s ear), and are both determined across the same range of frequencies. As an example, S/N for the speech between 2 people in an office may be about 1000:1. When the ratio is as large as this, then the communication will be successful.

Users will readily accept a system with large S/N, and the complex cognitive aspects of speech communication play little part in resultant high intelligibility. If, however, the ratio is poorer, then communication will be less successful. Only high frequency (ie common) words, in sentences of high redundancy, could be expected to be understood and the system would prove to be continual source of irritation to its users.

1 Scope

1.1 This part of the Standard discusses basic characteristics of speech, hearing and the mechanisms whereby people produce and perceive speech.

1.2 Factors affecting face-to-face and electronically assisted speech communication are discussed.

1.3 This document also describes how to predict, measure and assess speech intelligibility in a given noise environment.

1.4 This document applies solely to the case where talker and listener are communicating in the Standard atmosphere down to Sea level. Communication by divers breathing a helium rich atmosphere is specifically excluded.

2 Related Documents

2.1 The documents referred to in this Part of this Defence Standard are listed in annex B.

2.2 Reference in this Standard to any related documents means in any invitation to tender or contract the edition and all amendments current at the date of such tender or contract unless a specific edition is indicated.

2.3

DOCUMENT	SOURCE
British Standards American National Standards (ANSI) International Organization for Standardarization (ISO)	BSI Sales Department Linford Wood MILTON KEYNES M14 6LE Tel: 0908 221166
Defence Standards STANAGS	Ministry of Defence Directorate of Standardization Kentigern House 65 Brown Street GLASGOW G2 8EX Tel: 041-248 7890

3 Definitions

3.1 For the purpose of this Part of the Defence Standard the following definitions apply. Other definitions are supplemented in BS 4272: Part 3 Group 08.

3.2 Hertz (Hz). SI unit of frequency, indicating the number of cycles per sec (c/s).

3.3 Leq. The steady sound level which would produce the same energy over a stated time period as specified time - varying sound. Provided the "Leq" and long-term r.m.s are equivalent.

3.4 Octave

(a) The tone whose frequency is twice that of the given tone.

(b) The interval of an octave, together with the tones included in that interval.

(c) Three-tenths of a decade, a decade being a frequency interval of 10:1, see below.

3.5 One-third octave. A frequency interval whose upper frequency is 2 exp 1/3 times the lower frequency. Now in acoustics, third-octaves are, in fact, one–tenth decades in which the upper frequency is $10^{0.1}$ times the lower. Tenth decades are about 99.9% of a third-octave bandwidth, so the difference is negligible. The advantage is that tenth decades can take a more convenient series of centre frequencies.

3.6 Sinusoidal. An alternating quantity is said to be sinusoidal when its waveform, plotted to a linear time base, is a sine wave.

3.7 Intensity. The magnitude of a sound’s power or pressure, subjectively perceived as loudness.

3.8 Sound Pressure Level. The Sound Pressure Level of a sound, in decibels, is equal to 20 times the logarithm to the base 10 of the ratio of the r.m.s sound pressure (p) to the reference sound pressure (P_0).

$$L_P = 20 \text{ Log } (p/p_0) \text{ dB}$$

3.9 A-weighting. An electronic weighting network with a frequency response which approximates to that of a human ear by de-emphasizing low frequencies and extreme high frequencies. Sound levels measured with the A-weighting in the sound level meter are referred to as A-weighted sound levels and quoted in dB(A). A-weighted levels correlate well with subjective quantities such as loudness and annoyance. The risk of damage to the ear caused by loud noise is also better predicted from A-weighted rather than unweighed levels. The A-weighting response is defined in BS 5969 (IEC 651).

3.10 Threshold of audibility. The minimum level of sound which is just detectable under specified conditions. It should be noted that in this definition, and that of **3.11** there is no sharp dividing line between never perceiving a sound and always hearing it. There is a range of intensities over which the sound is judged to be present with a probability between zero and unity. O indicates "never heard", ie below threshold and 1 implies "always heard". A probability of 0.5 indicates that the listener detects the sound 50% of the time. Thus, a region of uncertainty exists and moreover, the region moves, depending upon the expectations of the listener. For these reasons, the threshold is best defined as that level of sound which is detected with a specified probability, in a specified situation. Ideally, the situation should be made "criterion free", so that listener expectancies cannot affect the result.

3.11 Absolute threshold. For a given individual, the threshold of audibility of a specified sound in the absence of other sounds.

3.12 Minimum audible field. At a specified frequency, the should pressure level of a pure tone corresponding to the modal value of the absolute threshold of normally-hearing young persons. Standard values are given in BS 3383: 1988 and ISO 226: 1987.

3.13 Masking. The process by which the threshold of audibility of one sound is raised by the presence of another (masking) sound. Also the amount by which the threshold of audibility of one sound is raised by the presence of another (masking) sound. Usually expressed in decibels.

3.14 Masked threshold. The threshold of audibility for a specified sound in the presence of another, masking, sound.

3.15 Presbycusis/Presbyacusis. The natural loss in hearing sensitivity, especially at high frequencies, that occurs with age.

3.16 Formant. A resonant peak in the spectrum of a voice. The position of the formants and their movements on the frequency scale are the important acoustic features which identify most speech sounds.

3.17 Phoneme. One of the basic speech sounds from which syllables and words are constructed. Each phoneme has a phonetic symbol in a phonetic alphabet. Not all phonemes are capable of being pronounced in isolation.

3.18 Intelligibility. A measure of the percentage of words or sentences correctly communicated from talker to listener under specified conditions.

3.19 Sidetone. The sound of a talker's own voice while talking. Sidetone can be heard through the normal air path from mouth to ear, by conduction through the head to the inner ear, or via a microphone and earphone or loudspeaker.

3.20 Syllable. A unit of speech consisting of a vowel sound and possibly one or more consonants.

3.21 Frequency. The number of repetitions (cycles) of a sound wave per second, subjectively perceived as pitch.

3.22 Resonance. The resulting vibration of a system when subjected to an applied force. The resonant frequency is that at which a given system will tend naturally to vibrate (resonate) with greatest amplitude.

3.23 Waveform. A graphical representation of measured sound amplitude over a period of time.

3.24 Spectrum. A graphical representation of measured sound amplitude across a range of frequencies.

3.25 r.m.s. Measure of any alternating waveform, the square root of the mean of the squares of continuous ordinates through one complete cycle.

Section Two. Hearing Characteristics

4 Auditory Communication Window

4.1 The limitations imposed by hearing are introduced first because they represent the broadest bound requiring consideration.

4.2 The effective lower bound of hearing. Referring to figure 1 the bottom curve shows the minimum audible field for a pure tone as a function of its frequency. The sensitivity of the healthy young ear increases as the frequency of the tone rises from 60 Hz to about 4,000 Hz, and thereafter decreases. At a frequency of 2,000 Hz the absolute threshold is approximately O dB, corresponding to an r.m.s sound pressure of 20 micro Pa. Below about 60 Hz and above about 10,000 Hz absolute threshold rises rapidly and imposes the lower and upper frequency limits of hearing.

4.2.1 The right-hand ordinate shows r.m.s sound pressure in Pascals (Pa), from which it can be seen that the young, unimpaired ear is extraordinarily sensitive in the region of 3,000 Hz. Indeed, if the ear were much more sensitive (approximately 25 dB) the noise caused by the collision of air molecules with the ear drum would be heard. It is rare that an environment is sufficiently silent that a sound at absolute threshold would be audible. Thus, in the frequency range 200-4,000 Hz it is rarely the absolute threshold of hearing that limits communication for a person with normal hearing.

4.2.2 Absolute threshold naturally varies from person to person. A threshold within about 10 dB of the minimum audible field at low frequencies and within 20 dB at higher frequencies would generally be regarded as falling within the normal spread for healthy ears.

4.3 The effective upper bound of hearing. When sound become very intense the ear can actually feel them and listeners report a tickling sensation shortly followed by an experience of pain. This transition, from hearing to feeling, occurs at about 120 dB SPL for tonal sounds, an r.m.s sound pressure of 20 Pa. This pressure is one million times the sound pressure at threshold.

NOTE: Where background noise levels are very high, personnel have adapted to this by the use of amplified signal levels at, and slightly above, 100 dB(A) for reasonably effective voice communication. This is partly a psychological adaption and partly a result of a temporary impairment of the hearing system. With prolonged and repeated exposures this temporary impairment can become permanent. Communication would undoubtedly be improved if background noise and signal levels could be reduced by 10 dB or more. In high noise levels it is important that amplified signals should never be higher than the minimum necessary to overcome that noise. Where background noise is low, amplified signals should not be permitted to contain peaks at 85 dB at the ear.

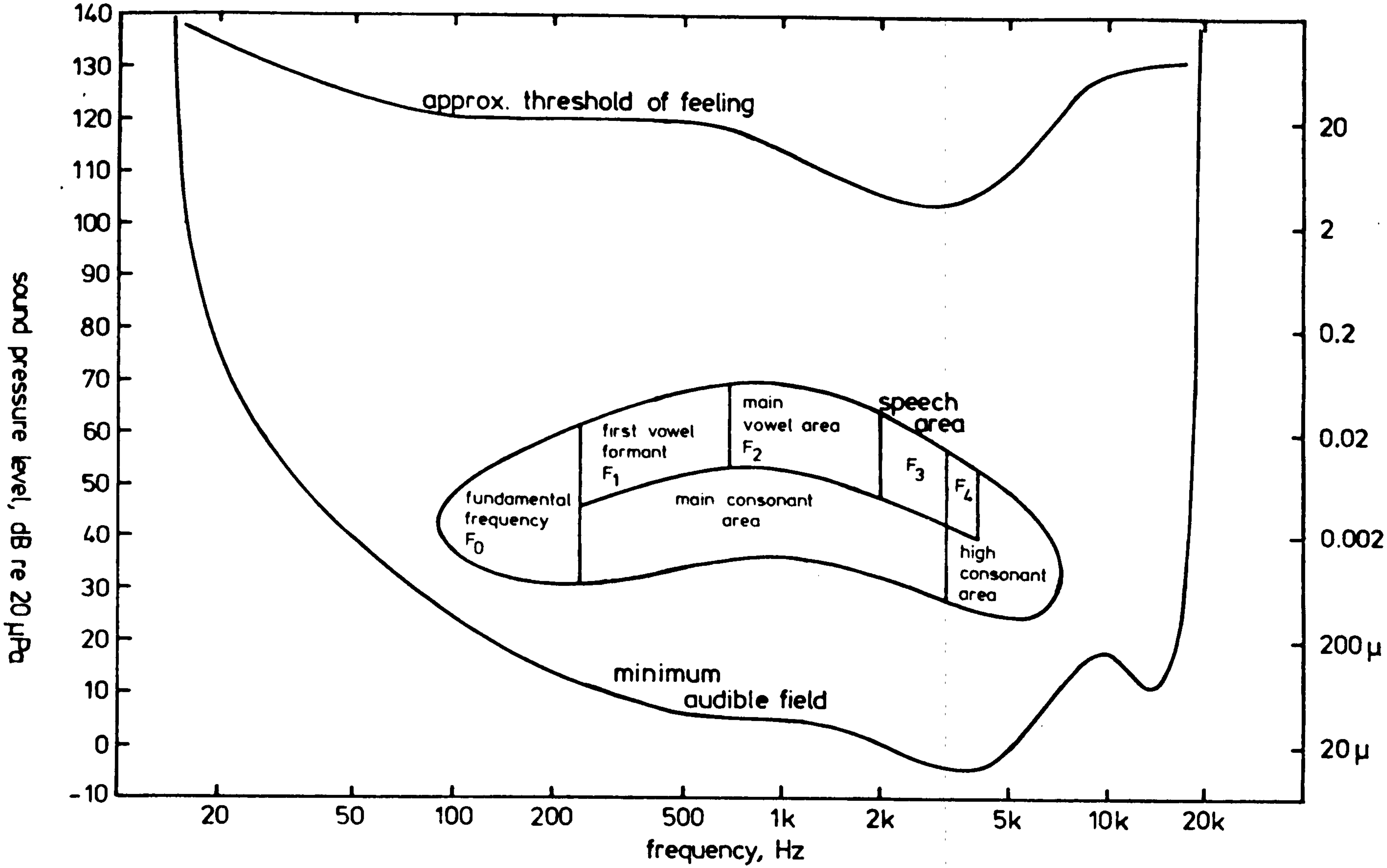

Fig 1 The range of human hearing, and the approximate ranges of various speech sounds when the talker is about 1m from the listener

4.3.1 Although physical pain does not actually result until 120 dB SPL is exceeded, the effective range of hearing is nearer 100 dB. The effective upper limit is more difficult to define than the effective lower limit because it is imposed by the degree of annoyance, discomfort and aural distortion that the listener will tolerate, all of these factors show large individual differences. It is clear that, as the level rises about 85 dB people will find sound irritating, and by 95 dB most people will take action to reduce the sound level, or remove themselves from the environment. Harmonic distortion is largely absent for levels less then 60 dB, but rises rapidly and is a significant problem in levels over 90 dB. The noise levels at which aural distortion becomes a problem are also the levels above which some form of hearing protection sound be used. Unfortunately some military environments have background noise levels so high that even with hearing protection the noise level at the ear may exceed 90 dB.

4.4 The limitations imposed by age and noise exposure. The hair cells which transduce mechanical energy into nerve impulses are nerve cells, and like all nerve cells are not regenerated when they die. For this and other reasons the threshold of hearing rises with age. Sensitivity to high frequencies is lost first - a phenomenon referred to as presbyacusis. The effective is not particularly noticeable for people under the age of 50; but it becomes progressively more evident above this age. Therefore frequencies above 5,000 Hz become somewhat unreliable for communication.

4.4.1 Some environments such as vehicles and engine rooms can have noise levels well above that which actually puts the ear at risk. Because of the ear structure, a noise-induced hearing loss appears typically as a loss of sensitivity in the region of 3,000-6,000 Hz. This loss is initially temporary and will recover, but with repeated and prolonged exposure can become permanent. Temporary hearing losses can take from minutes to hours to recover and affect communication even after the casual noise has stopped. A temporary hearing loss makes speech sound dull, muffled and indistinct particularly during lulls in the background noise. Hearing protection should be provided and worn to minimize such effects. The effects of hearing protection on speech communication are described in **7.8.**

4.5 The communication window. The limits of hearing so far discussed have established the spectrum region boundaries, inside which communication can effectively take place.

Section Three. Speech Characteristics

5 Speech Sounds, Their Spectra, Dynamics and Level

5.1 The action of speaking and the comprehension of natural speech although accomplished without a moment's thought, is a very complex process: words combine into phrases and clauses which link into sentences and sentence groups. The whole structure is, then, modulated to indicate to the listener how the subgroups are related, and how they are to be interpreted. The speech signal is a complicated time sequence of identifiable sounds, called 'speech elements' or phonemes. These are the basic elements of speech from the point of view of the phonetician. There is a large vocabulary of phonemes, from which any word or phrase in any language may be assembled. Any one language uses only a subset of these phonemes. In the English language there are about 40. There is a rough correspondence between the phonemes of speech, and the letters of the alphabet to enable the three important classes of speech sounds, and their primary physical characteristics to be described.

5.2 The production of speech sounds. Speech sounds occur when some portion of the vocal tract is constricted, and air forced through the constriction, produced turbulence and rapid change in air pressure which are perceived as sound. The method of constriction distinguishes the three classes of speech sounds as follows:

(a) Vowels.

(b) Voiced consonants.

(c) Unvoiced consonants.

5.2.1 Vowels. If lips and teeth are open and the airstream is not blocked off by the tongue so that the only constriction in the vocal tract is the vocal cords, then the result will be a vowel–sound. When the ligaments, or vocal cords, are pulled tight by muscle and at the same time air is forced through by action of the diaphragm, the vocal cords vibrate and produce a 'voiced' sound. The air is broken into a stream of regular, short puffs occurring at the rate of 75-150 pulses per second for men and 15-250 pulses per second for women. The spectrum of this pulse train is a set of harmonics of the basic pulse rate (or fundamental, F_0). and the amplitude of successive harmonics falls slowly as the frequency increases. A diagram of this type of spectrum is shown at (a) in figure 2. Cavities of throat, mouth and nose operate together to produce resonances which amplify or attenuate the individual harmonics. Vocal cavities typically produce two pronounced resonances in the spectrum, referred to as the first and second formants. The spectrum, at (a) in figure 2, is modified by the double resonance of the vocal tract to produce the vowel-like spectrum shown at (b) in figure 2, the first and second formants are marked F_1 and F_2 on the abscissa. Up to six formants can be identified in some vowel sounds, but the first two or three are most important.

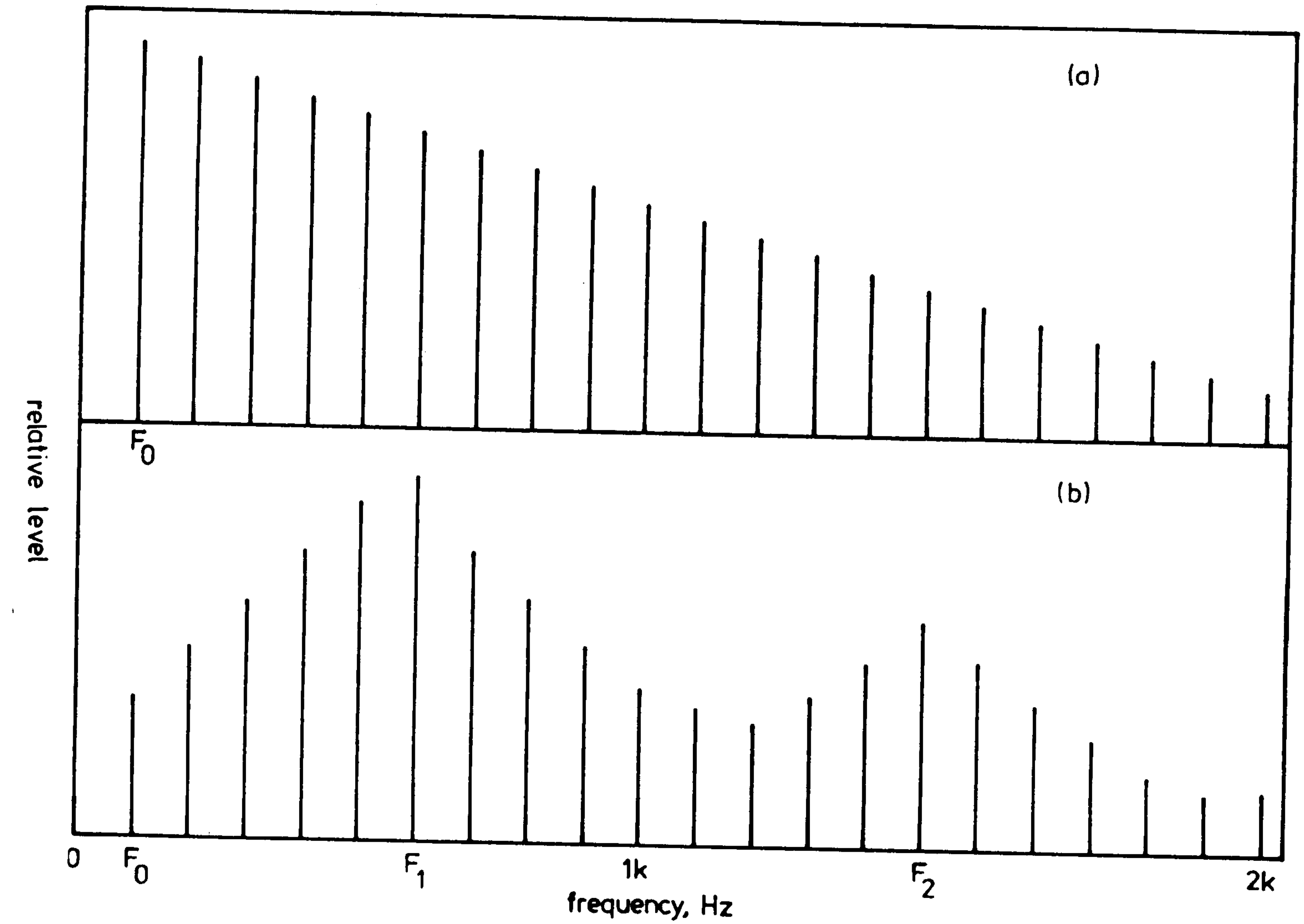

Fig 2 Spectrum of the pulse train produced by the vocal cords (schematic)

a) spectrum at the vocal cords
b) spectrum after modification by resonances in the vocal tract

5.2.1 (Contd)

The positions of the formants on the frequency scale are controlled by the shape of the mouth and positions of tongue and lips. The specific vowel is determined mainly by the frequencies of the first two formants relative to each other and to the fundamental frequency. The entire set of components is perceived as a unit, with a pitch corresponding to the fundamental of the harmonic series, even though there may be little or no energy at the fundamental itself.

5.2.2 Voiced consonants. Constriction of the vocal cords and the articulators (lips, teeth and tongue) can be combined, in which case the product is typically, a voiced consonant. Perhaps the best examples are the nasal sounds associated with the letters m and n where the vocal cords are vibrating, and air is exhaled through the nose rather than the mouth to give a different resonance. Other examples of voiced consonants include the plosives, such as b, d, g in which the mouth starts closed, and which are identified by the transition, lasting a few milliseconds, to the following vowel. The power of voiced consonants is greater than that of unvoiced consonants but less than that of vowel-sounds.

5.2.3 Unvoiced consonants. If the vocal cords are open, but air flow through the mouth tract is constricted, by the positioning of the lips, teeth, tongue or palate, the result is an unvoiced consonant. Examples include the sounds symbolized by s, sh, th, and by h when used at the front of a word. The constrictions cause turbulence in the airflow of the vocal tract which results in the production of a broadband noise. As in the case of vowels, the spectrum is modified by the resonance of vocal tract and so it has some tonal quality and specific frequency content. However, since the waveform is random and aperiodic, the sound is a noise rather than a set of harmonics as in the case of vowels.

5.3 The frequency range of speech sounds. Vowel sounds contain power at frequencies as low as 50 Hz, and the sibilants and fricatives, s and f, contain significant concentrations of power up to frequencies as high as 8,000 Hz. Figure 1 shows in a highly schematic form the frequency ranges and approximate levels of the speech sounds. The levels shown are for unamplified natural speech at about a metre form the talker. The different sections of the spectrum are not however, equally important for the comprehension of speech.

5.3.1 The distribution of speech levels across the spectrum. If the spectrum between 50 and 6,000 Hz is divided into consecutive third octaves, and the long-term r.m.s or average pressure within the different third–octave bands is measured for continuous speech, then the relative levels in the bands are roughly as shown in figure 3. The bulk of energy is in the three octave bands centred at 125, 250 and 500 Hz.

Even though the third-octave band around 5,000 Hz is 10 times the width of the band at 500 Hz, much less energy occurs in the higher band. Fully 84% of total speech energy occurs at frequencies below 1,000 Hz. The bulk of energy is in three octaves around 500 Hz.

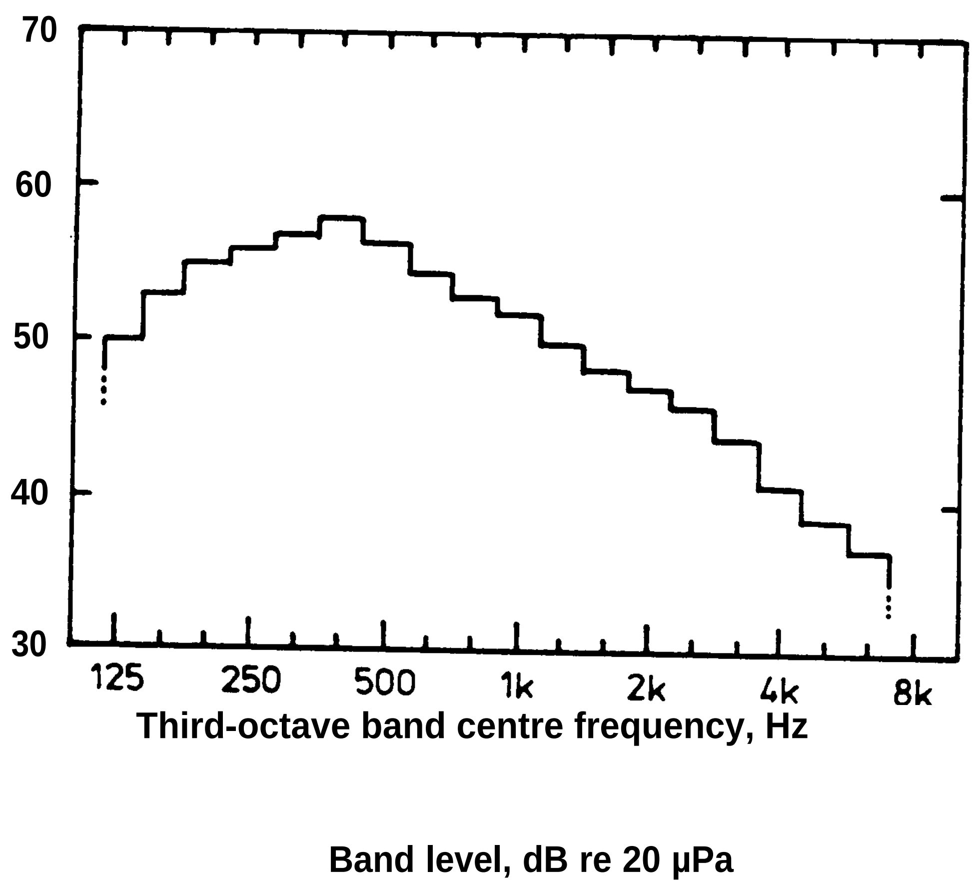

Fig 3 Third-octave band spectrum of male speech
(overall level is 65 dB)

5.3.2 The distribution of contributions to intelligibility. Different sections of the spectrum do not contribute to the comprehension of speech in proportion to the sound level occurring in the band. Sound at frequencies below 300 and above 6,000 Hz contributes almost nothing to the intelligibility of speech. If instead of plotting the sound level within each third-octave band we plot the relative contribution of that band to the intelligibility of speech we obtain the histogram of figure 4. The figure shows that it is the bands in the region 1,000 to 3,000 Hz that contribute most to the intelligibility of speech, even though only 16% of speech energy occurs about 1,000 Hz. The region have 1,000 Hz contains the low level, but all important, consonant sounds, as well as the main vowel formants.

5.3.3 Redundancy in speech sounds. A phoneme is usually identified by more than one acoustic feature. The features associated with each phoneme are not confined to one frequency region of the spectrum. Thus no single band is essential to the intelligibility of speech - if one band were filtered out there would often be alternative information in other bands for a listener to identify a phoneme. In other words, the relative contributions of each band to speech intelligibility as shown in figure 4 should not be regarded as adding up to 100% intelligibility.

5.4 The effective upper frequency limit for speech. A high fidelity communication channel such as a VHF FM public broadcast radio channel has an upper frequency limit of 15,000 Hz and provides good quality music and speech reproduction. Medium Wave (AM) broadcasts have an upper limit of 4,500 Hz - speech is usually highly intelligible, but does not sound natural. The audio bandwidth is restricted by international agreement, so that more stations can be crowded into the given radio waveband. A telephone has an upper frequency limit of typically 3,400 Hz- consonants which are distinguished from one another by their high-frequency content, such as 's' ad 'f' are often difficult to distinguish without the context and meaning of the word or sentence as a whole. The effective upper frequency limit for face to face speech does not usually arise. But for electronic channels, the effective upper limit for speech is a trade-off between fidelity, intelligibility and the bandwidth allowed for a cable or radio link. Generally 3,000-3,5000 Hz is a good, practical upper limit for speech channels where bandwidth must be restricted, without too much degradation of intelligibility. If there is no engineering reasons to restrict audio bandwidth then the upper limit for speech frequencies is best left as high as possible ie 15,000 Hz.

5.5 Lower frequency limit. Frequencies down to 80 Hz or lower must be transmitted by the communication system if all of the speech energy in the talker’s voice is to be heard by the listener. However, whereas most of the speech energy is below 1,000 Hz the bands above 1,000 Hz are more important for intelligibility. The lowest speech frequency transmitted by telephones and communications radios is usually set to 300 Hz. This lower limit removes more speech energy from the signal than does the upper limit of 3,400 Hz, but it does relatively less harm to speech intelligibility. There is probably no advantage, in terms of intelligibility, to having a lower limit below 300 Hz.

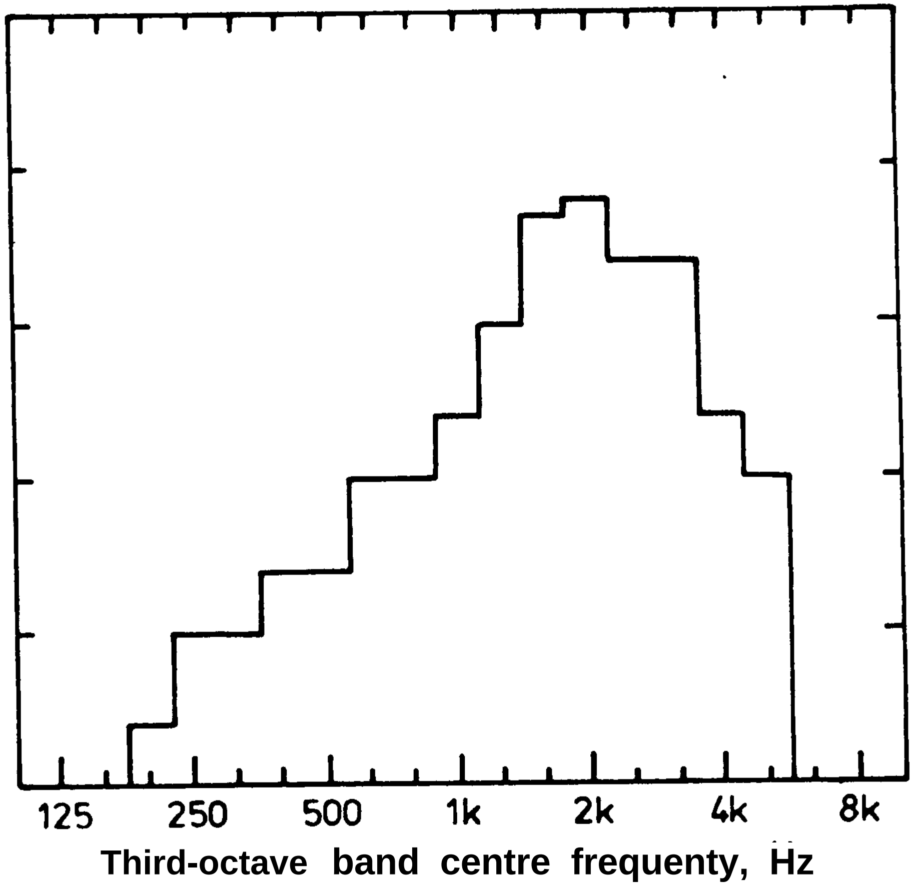

Fig 4 Relative contributions of third-octave bands
to speech intelligibility

5.5 (Contd)

When the communication system has a limited dynamic range, it can actually be advantageous to raise the lower frequency limit to 500 Hz or 600 Hz. By filtering out the higher-level low-frequency energy, which contributes little to intelligibility itself, the remaining higher frequencies, which are more important for intelligibility, can be further amplified without overloading the channel.

5.6 Speech dynamics. The r.m.s sound level measured over short periods such as 1/8th of a second, comparable to the duration of a phoneme, varies over a range of about 30 dB during natural continuous speech. The lower levels correspond to weaker consonants, the higher levels to the vowel sound. The 'speech peaks' measured in this way are typically 12 dB above the long-term r.m.s level of continuous speech, the 'speech minima' about 15-20 dB below. Instantaneous pressures of the speech waveform are typically 10 dB above the speech peaks, and about 22 dB above the long-term r.m.s. The dynamic range of a full speech signal, from instantaneous peaks to quietest significant sounds is therefore about 40 dB for a typical speaker. A trained speaker or actor with clear articulation can reduce his or her speech dynamic range to, perhaps, 20 dB. This range of speech levels is a necessary and natural part of speech.

5.7 Measuring and specifying speech levels. Speech levels may be measured or reported in different ways. Each way is useful and has its advantages under particular circumstances, but each method gives a different value and the method must be specified or the level is meaningless. An unqualified statement that 'the speech level was 70 dB' is of no value without further information. Speech levels may be sound levels in dB or dB(A) re 20 micro Pa at a specified position or distance from the talker’s lips, but may also be voltage levels in an electrical circuit in dBm or dBV. Usually this is clear from the context and/or the decibel reference level. Wherever speech levels or speech to noise ratios are quoted the method of measuring the speech and noise shall be specified.

5.7.1 Long-term root mean square level. The long-term r.m.s level is one of the most frequently quoted measures of speech, because it is repeatable and stable. It can be obtained using some third-octave or narrow-band spectrum analyzers, or more simply from an integrating-averaging sound level meter which conform to BS 6698 and IEC 804 and has an "Leq" setting. The quantity "Leq", the equivalent continuous sound level, is usually defined as the steady sound level which would produce the same energy over a stated time period as a specified time-varying sound (in this case speech). Provided the "Leq" period is the same as the measurement interval, the "Leq" and long-term r.m.s are equivalent. Speech levels are usually unweighed but often weighted. To avoid any confusion it should be stated which. A specialized instrument known as a speech voltmeter can also be used to give long–term voltage levels in a circuit, or, with a suitably calibrated microphone, long-term sound levels. In measuring an overall level or an octave or third–octave band spectrum, stable spectra approximating the long-term spectra are obtained for naturally punctuated, continuous speech with integration periods as short as one minute.

5.7.2 Sound level meter measurements. The conventional sound level meter is by far the most widely available, general purpose sound measuring instrument suitable for measuring speech. Many instruments described for example as "measuring amplifiers" have settings which conform to the standard specifications for sound level meters (BS 5969 and IEC 651). Most basic sound level meters offer an A-weighting and a linear (20 Hz-20,000 Hz) or C-weighting. The C-weighting does not affect any of the speech frequencies and speech levels in dB(C) can be regarded as unweighed. Two meter time constants are generally provided; these were originally designated "Fast" and "Slow" and are still commonly known as such, but more recent standards refer to Time-weighting "F" and Time-weighting "S" respectively. Either may be used in speech measurements provided it is stated which. Measurements made with a basic sound level meter are not long–term averaged: usually the peak levels for the words in continuous speech are averaged by eye, using the slow time constant to make this easier. This is not as precise or stable as long-term average, but generally adequate for many purposes.

NOTE: More sophisticated sound level meters are available which can measure the long–term r.m.s level of a signal and, in some cases, the statistical distribution of levels.

5.7.3 VU Meter. A VU (Volume Unit) Meter has a short rise and longer fall time and is intended to measure "programme peaks" in electronic circuits. Such meters are useful for maintaining a high signal level in a radio channel or tape recorder without overloading. The VU Meter does not register instantaneous waveform peaks, but peaks in a short-term average level. There is a standard specifying the characteristics of VU Meters (American Standard C16.5-1961); most VU Meters fitted to domestic cassette recorders do not conform to the Standard. The reference level, O VU (zero VU) is 1 dBm (1 milliwatt in a 600 ohm impedance). Other quasi-peak reading meters are available such as those defined in BS 5428, Part 9. Their characteristics differ from those of VU Meters and they are widely used by European and UK radio stations.

NOTE: The technique of reading a VU Meter is to ignore exception peaks and average the remaining typical peaks by eye. Readings of speech peaks on a VU Meter are typically about 12 dB above the long-term r.m.s level.

5.8 Typical speech levels. For face to face conversation or telephone speech in a typical office the long-term r.m.s level of male speech is, on average, about 65 dB at 1 metre. Female speech is about 2-3 dB less. Levels vary from talker to talker within a range of about 6-8 dB. Speech levels are also lower in quieter rooms and higher in noise, even though the speaker would generally consider himself or herself to be talking at a normal level. A maximum shout (for males) has a long-term r.m.s level of about 85-90 dB, but this cannot be sustained for long.

NOTE: The following are roughly equivalent: a long-term r.m.s speech level of 65 dB unweighed; average peaks on a sound level meter with fast time-weighting of 71 dB(C) or dB(unweighted); average peaks on a sound level meter with slow time constant of 68 dB(C). Values in dB(A) are typically 1-5 less than unweighed values or values in dB(C), but this varies from person to person and with the vocal effort.

5.9 Speech levels at different distances from the mouth. For distances beyond about 50 mm the mouth can be regarded as a point acoustic source where sound pressure is inversely proportional to the distance from the source (and sound intensity is inversely proportional to the square of the distance). Beyond this distance a doubling of the distance gives a 6 dB reduction in voice level: thus the level at 250 mm is 6 dB higher than at 500 mm and 12 dB higher than at 1m. Closer than 50 mm the mouth must be regarded as a complex multiple source.

Section Four. Speech Communication and its Assessment

6 Categories of Communication

Speech communication can be divided into three categories; (i) unamplified speech, normally face-to-face, (ii) normal speech where the speech waveform is transmitted via electronic channels, and (iii) vocoded speech and synthetic speech. The reverse situation, where the speech is Human, but the "listener" is a machine (ie Direct Voice Input) will not be considered, since this Standard deals only with human perception. In each case the most important aspect to consider is speech intelligibility. In the absence of any British Standard, references are made to International (ISO) and American (ANSI) standards.

7 Unamplified Speech

7.1 Unamplified speech is of importance in offices, workshops, test sites, in vehicles for crew and passengers, and many other situations. Many factors influence intelligibility, the background noise level and spectrum, reverberation, the hearing and clarity of speech of the communicators, whether communication is restricted in content and the number of words used, and whether the language used is the native tongue of the talkers and listeners. Usually the background noise in each situation is by far the most important factor to consider.

7.2 Noise and vocal effort. A talker will automatically adjust his vocal effort according to the background noise in which he is immersed to maintain a suitable speech to noise ratio. The speaker normally hears his own voice while speaking. This acoustic feedback is known as 'sidetone'. Sidetone is primarily carried from the talker's mouth to his ears through the air, but he also hears his how voice through the vibrations transmitted through the tissues of his head. The contribution of headborne sidetone to the speech level as perceived by the talker misleads the talker in his judgement of his speech sound level. Consequently a speaker in noise does not fully compensate for the influence of the background noise; each 10 dB rise in noise level is usually matched by about a 5 dB increase in the speech level.

7.3 Speech in a quiet to moderate noise environment. When a person is speaking at normal conversation level is a reasonably quiet room the long-term r.m.s speech level will be approximately 65 dB SPL at a distance of one metre. However, the natural level will vary considerably with background level and social conditions. When speaking confidentially in a quiet room a speaker may use a level as low as 40 dB, which is just above a whisper.

7.4 Speech in a moderate noise environment. In rooms with a moderate level of noise, such as computer rooms and offices, speech intelligibility is not limited by background noise, and people will still speak at a level of around 65 dB (long term r.m.s at a distance of one metre) by preference. The background does not prompt people to raise their voices until the A-weighted sound level reaches 50-55 dB(A). Above this level speakers tend to raise their voice by 5 dB for every 10 dB increase in the background noise level. If noise levels are not excessive, the whole question of the appropriate vocal effort can be left to the speaker and listener.

7.5 Speech in a relatively noisy environment. Eventually, as noise levels increase a speaker will become conscious of the need to speak in a very raised voice or need to shout to make themself understood by a listener as close as one metre. This should be taken as a warning that noise levels are likely to be 85-90 dB(A) or higher, and therefore potentially damaging to the ear. The noise should then be measured, so that an effective remedy may be implemented. Preferably, this will take the form of noise reduction at source, but an alternative would be the provision of personal hearing protection (see **7.7).** Beyond this the speaker and/or listener are likely to try and reduce the background noise level or move out of the environment if they have a choice. In extreme noise levels the speaker is limited by vocal effort and vocal ability. A loud shout has a level of about 85 dB at a distance of one metre. The maximum shout, which cannot be sustained for more than a few words, is about 90 dB.

7.6 Assessment of interference by noise. The most direct way of determining speech intelligibility is to carry out intelligibility tests with panels of listeners and talkers. This is expensive and time consuming, and is rarely employed for normal face-to-face communication. Speech interference by noise is more easily determined by estimation from physical measurements. The simplest of these is the A-weighted speech-to-noise ratio, although this can be a poor measure under some circumstances and better measures exist.

7.7 Speech Interference Level. The most widely used method of assessing a noise environment and rating its interference with speech is the Speech Interference Level. This exists in two slightly different forms. The Preferred-frequency Speech Interference Level (PSIL) is the arithmetic mean of the measured noise levels in the three octave bands centred on 500 Hz, 1,000 Hz and 2,000 Hz. The SIL as specified by the American National Standards Institute (ANSI S3.14-1977) and the International Organization for Standardization (ISO TR 3352-1974) is the arithmetic average of the measured noise levels in the four octave bands centred on 500 Hz, 1,000 Hz, 2,000 Hz and 4,000 Hz. These three or four bands cover the most important speech frequencies. Both versions of the SIL are good indicators of the ability of broadband noise to mask speech and have the advantage of being readily obtained with a portable sound level meter fitted with octave-band filters. Figure 5 shows the relationship between the ease of face-to-face communication, the distance between talker and listener and the SIL (ANSI/ISO version). The figure shows the maximum separation between speakers 'for just reliable communication' at different levels of vocal effort according to ANSI. The ISO is slightly more conservative, and quotes the distances should in Table A for satisfactory conversation. In addition to rating face-to-face communication the SIL can give some indication of the ease of telephone conversations: telephone use is usually satisfactory if the SIL is below 60, difficult if the SIL is between 60 and 75, and unsatisfactory if the SIL exceeds 75.

NOTE: The SIL is purely a measure of background noise. Speech is not measured but it is assumed that speakers adopt vocal efforts appropriate to the noise. This assumption is reasonable.

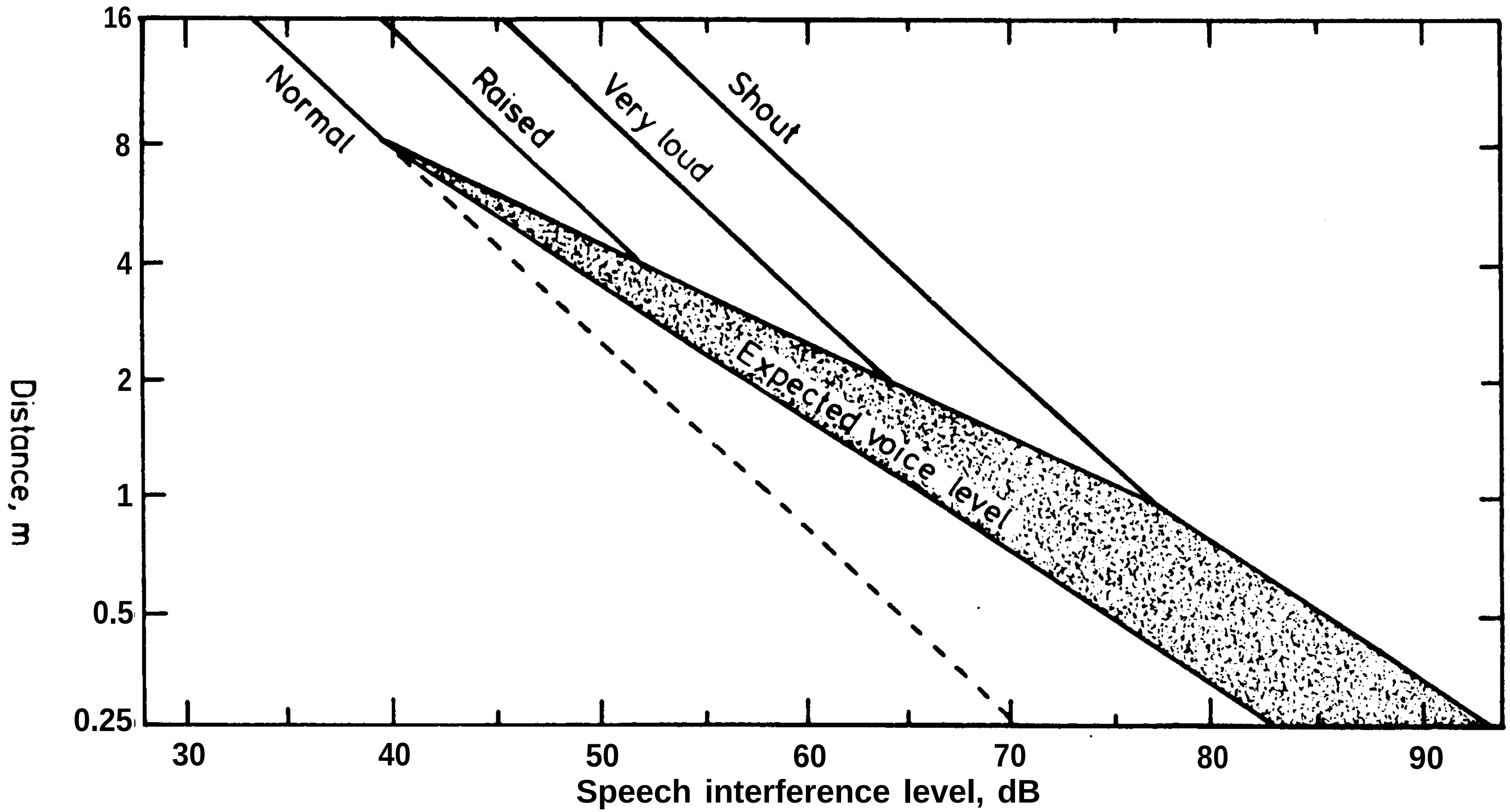

Fig 5 Talker to listener distances for 'just reliable' communication according to ANSI S3.14
(The region below each line shows the talker to listener distance and noise level combination for which just reliable face-to-face communication is possible. Each line is labelled to indicate the relative vocal effort)

Table A

Distances at Which Communication in Noise is Satisfactory

From ISO Technical Report 3352-1974

Speech Interference Level (four-band version) dB	Maximum distance at which normal conversation is considered to be satisfactorily intelligible. metres	Maximum distance at which conversation in raised voice is considered to be satisfactorily intelligible. metres
35	7.5	15.0
40	4.2	8.4
45	2.3	4.6
50	1.3	2.6
55	0.75	1.5
60	0.42	0.85
65	0.25	0.50
70	0.13	0.26

NOTE: For the purposes of ISO 335, intelligibility is regarded as satisfactory if the Articulation Index (see **8.3)** is 0.4 or greater.

7.8 Hearing protection and communication. Hearing protection has two effects of speech communication. Firstly, in noise levels about 85 dB(A) hearing protection will improve the reception of speech and warning sounds by persons with reasonable hearing. The protection reduces the speech and background noise equally at any frequency so that the same speech features will be heard whether or not the protection is worn. At the same time the combined speech and noise input to the ear is reduced from a high level and the distortion of the ear will also be reduced, resulting in a slight improvement in intelligibility. This improvement is not generally obtained if the listener has impaired hearing. Secondly, when the person talking wears hearing protection and the noise at his ear is reduced but the sound of his own voice is only slightly affected, since he hears his voice by headborne vibrations. Without conscious thought he will speak more quietly, since he can hear himself above the noise perfectly well. The reduction in voice level reduces the speech to noise ratio for the listener, and there is a net reduction in intelligibility. If protection is likely to be worn it is wise to assume that this will have an adverse effect on communication similar to increasing the SIL by 3-5. This should not be used as an argument against hearing protection; workers may be able to adjust, with training, to using higher vocal efforts.

8 Amplified or Electronically-Assisted Speech

8.1 Examples of amplified and electronically-assisted speech include public address in buildings or ships, telephones, radio links, intercoms (including those in aircraft and ground vehicles). The factors which influence the intelligibility of unamplified speech. In addition the bandwidth frequency response and distortions of the transmission channel must be considered. The persons communicating electronically may be in vastly different environments: communication may be easier in one direction than the other. As with face-to-face speech, noise is usually the most disruptive factor. The noise can be background noise from the speaker’s location transmitted along with the speech, noise generated in the communication channel itself, or background noise present at the listener’s location.

8.2 Assessment of speech intelligibility. As with face-to-face communication the most direct way of assessing the intelligibility of a communication channel is to carry out intelligibility testing with panels of talkers and listeners. This is sometimes done, and sometimes a specification may be written in terms of an intelligibility score in one of the standard tests. For communication systems subject mainly to noise and bandwidth limitations and some analogue distortions there are simpler and quicker alternatives in which intelligibility can be estimated from physical measurements. The simplest effective method is to measure the Articulation Index (AI), a more recent alternative is to measure the Speech Transmission Index (STI). These physical measures are useful for rating communication systems on a scale from excellent to poor, but are not sufficiently precise to assess marginal channels under marginal conditions, or to separate out two competing systems to determine which is better under marginal conditions. For these applications, which frequently occur with military communications, there is as yet no adequate substitute for intelligibility testing.

8.3 Articulation Index (AI). The Articulation Index is a fairly detailed method of predicting speech intelligibility and takes into account the measured spectrum of the interfering noise, the speech spectrum measured or presumed at the listener’s ears, and the relative importance of each frequency band to the intelligibility of speech. The noise and speech spectra are usually measured in third-octave bands. The full procedure and work sheets are laid out in American Standard ANSI S3.5-1971, which also outlines alternative procedures for measurement bandwidths other than third- octaves. Strictly, AI is only valid for male speech. AI also presupposes that the talker’s vocal level will be known whereas in practice it may vary according to any feedback from the listener.

8.4 Values of AI. The AI can have a value in the range from 0 to 1. An AI less than 0.3 is generally rated as unsatisfactory, between 0.3 and 0.5 as acceptable, between 0.5 and 0.7 as good, and over 0.7 as excellent. The term "acceptable" in this context should not be interpreted as "the level to aim for", since a poor AI will lead to increased workload and a slower communication rate.

8.5 Speech Transmission Index (STI). The STI is very similar in concept to the AI in that it determines signal-to-noise ratios in frequency bands and weights and combines these according to their contribution to intelligibility. It differs in that the signal to noise ratio in each band is determined from the modulation transfer function in that band. In principle, a band of noise is modulated so that its envelope is a sine wave. The modulated acoustic signal is used as the input to the communication system in place of the speaker. Any noise in the channel will reduce the depth of modulation at the output, as will any reverberation or signal dispersion. In practice the modulated bands re one octave wide and several rates of modulation are used with each band. The measured modulation transfer ratios are combined to give a single number, the STI, which can have a value from O to 1. The inventors of the STI claim that an STI of between O and 0.3 can be regarded as bad, between 0.3 and 0.45 as poor, 0.45 and 0.6 as fair, 0.6 and 0.75 as excellent. As a recent development the STI has not yet been exhaustively tested for military use.

8.6 Frequency distortion. A listener can tolerate large amounts of distortion of the speech spectrum before intelligibility is severely affected. When the transmitted frequency range is for any reason limited then the AI calculation will reflect whether the channel is satisfactory or not. The AI will also cope with irregularities in the frequency response of a channel provided there are no peaks or troughs with slopes of 18 dB per octave or more. If such peaks and troughs exist the AI will overestimate intelligibility by an unknown amount. (Similar restrictions are likely to apply to the STI).

8.7 Amplitude distortions: peak clipping. Two common forms of amplitude distortion are peak clipping and centre clipping. Peak clipping occurs when a circuit is overloaded and the waveform peaks are flattened off. In extreme cases the speech waveform would be reduced to a series of square-waves. Surprisingly, in quiet conditions where the speech is not accompanied by noise, peak clipping has little effect on intelligibility, s shown in figure 6. Even the most extreme clipping reduces the percentage of words correctly heard (see clause **10)** to not worse than 70%. The effect of peak clipping on speech quality becomes noticeable when clipping exceeds 10 dB and the speech sounds somewhat "edgy". More than 20 dB of peak clipping makes speech "course and grainy" and the quality would be thought "poor", although the intelligibility would remain reasonable.

8.8 Peak clipping with noise. The effect of peak clipping in the presence of noise depends upon where the noise enters the circuit. If the speech is clipped before the noise mixes with it there can be an improvement in intelligibility. An example is where the talker is in a quiet environment but the listener is in noise. The improvement occurs because the clipping most affects the high level vowel sounds while leaving the important low level consonants relatively undistorted. The level of the vowels and consonants are now similar so that the consonants' signal to noise ratio and therefore intelligibility can be improved for a given maximum signal level. The effect of this clipping on the AI can be calculated (see ANSI Standard S3.5). If the noise enters with the speech at the microphone so that noise and speech are clipped together there is generally no benefit from clipping, and perhaps some degradation.

8.9 Amplitude distortions: centre clipping. Centre clipping can arise in a faulty class B amplification circuit and is a form of crossover or switching distortion. Even small amounts of this can degrade intelligibility since in the main the lower level signals including the consonants are the most distorted. Figure 6 shows this graphically.

9 Vocoded or Synthetic Speech

Certain assumptions implicit in the AI and STI about the nature of the speech do not apply to vocoded speech, synthetic speech and possibly for speech transmitted by single sideband radio. These assumptions are that the important acoustic features which make speech understandable are distributed within a certain range of levels and within certain frequency bands which can be weighted according to the amount of speech information which they contain. At present the only way of measuring intelligibility is in intelligibility test with panels of talkers and listeners.

10 Intelligibility testing. Intelligibility testing is the most direct and obvious way of testing the usefulness of a speech channel, and in some cases the only way. In principle a set of prescribed messages is read by the talker and the number of errors made by the listener is counted. Such testing has to be carried out on a large scale under highly controlled conditions with practiced talkers and listeners in order to achieve a high statistical confidence in the results. Accordingly such tests are usually carried out at specialist laboratories. These tests are most useful for

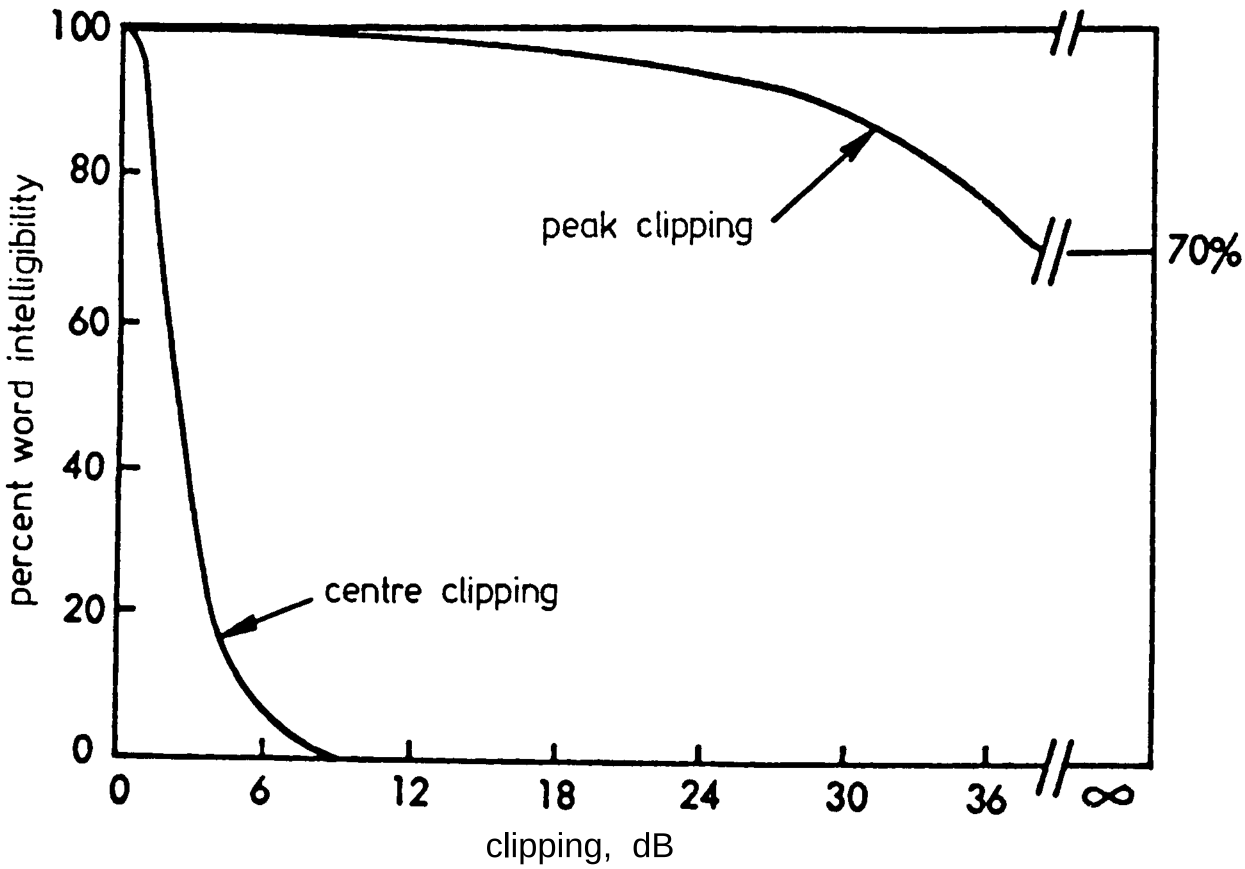

Fig 6 The effect of clipping on intelligibility under noise-free conditions

10 (Contd)

comparing communication systems directly under the same conditions on the same occasion. Comparing scores obtained on different occasions or from different laboratories is less precise, as there may be slight systematic differences. To overcome this difficulty, at least in part, it is good practice to include a known communication system in any series of tests as a reference standard. In all cases the value of the test results is only as good as the simulation of the conditions under which the system will be used (see **10.5).** Various types of test are available. These include (i) nonsense syllables, (ii) sentence tests, (iii) phonetically balanced word lists, (iv) the Modified Rhyme Test, and (v) the Diagnostic Rhyme Test. The latter three methods are the most frequently used.

10.1 Phonetically balanced word lists. Of various published sets of work lists the most frequently used is probably that given in ANSI S3.2-1960. This Standard gives twenty lists each of 50 monosyllabic words. Over a whole list the phonemes which make up the words are in proportion to their relative frequency of occurrence in normal everyday speech, hence the term 'phonetically balanced'. Various word lists are often combined and the orders of words randomized so that a listener cannot predict which word he will hear. The words are spoken in a carrier sentence; "Would you write . . . . . now". The results are given as the percentage of test words correctly identified.

10.2 Modified Rhyme Test (MRT). The spoken material in this test again consists of prepared lists of monosyllabic words. However, the listener has to select which word was spoken from a multiple choice answer sheet. Six possible responses are given to each spoken word, differing only in one consonant. These tests are useful for studying confusions between particular speech sounds.

10.3 Diagnostic Rhyme Test (DRT). Spoken material consists of 96 word pairs, the words in each pair differing only in the initial consonant. The pairs are always presented in the same order and the word spoken from each pair is determined by following one of 30 word lists, but as far as the listener is concerned, the word is random. Thus the listener may know that the first word is either Taunt or Daunt; the second either "Boot" or "Moot"; then "Cheat" or "Sheet", etc. Words are generally presented without a carried phrase at a rate of one word every 1.4 seconds. The listener has to select which of the two words was spoken, guessing if he is not sure. The result quoted for a DRT is not simply the percentage of words correctly identified, but is correct for guessing. A listener would be able to get 50% of words correct by random guesses. Fifty percent of words correct corresponds to a DRT Score of 0; 75% correct to a score of 50; and, of course, 100% correct to a score of 100.

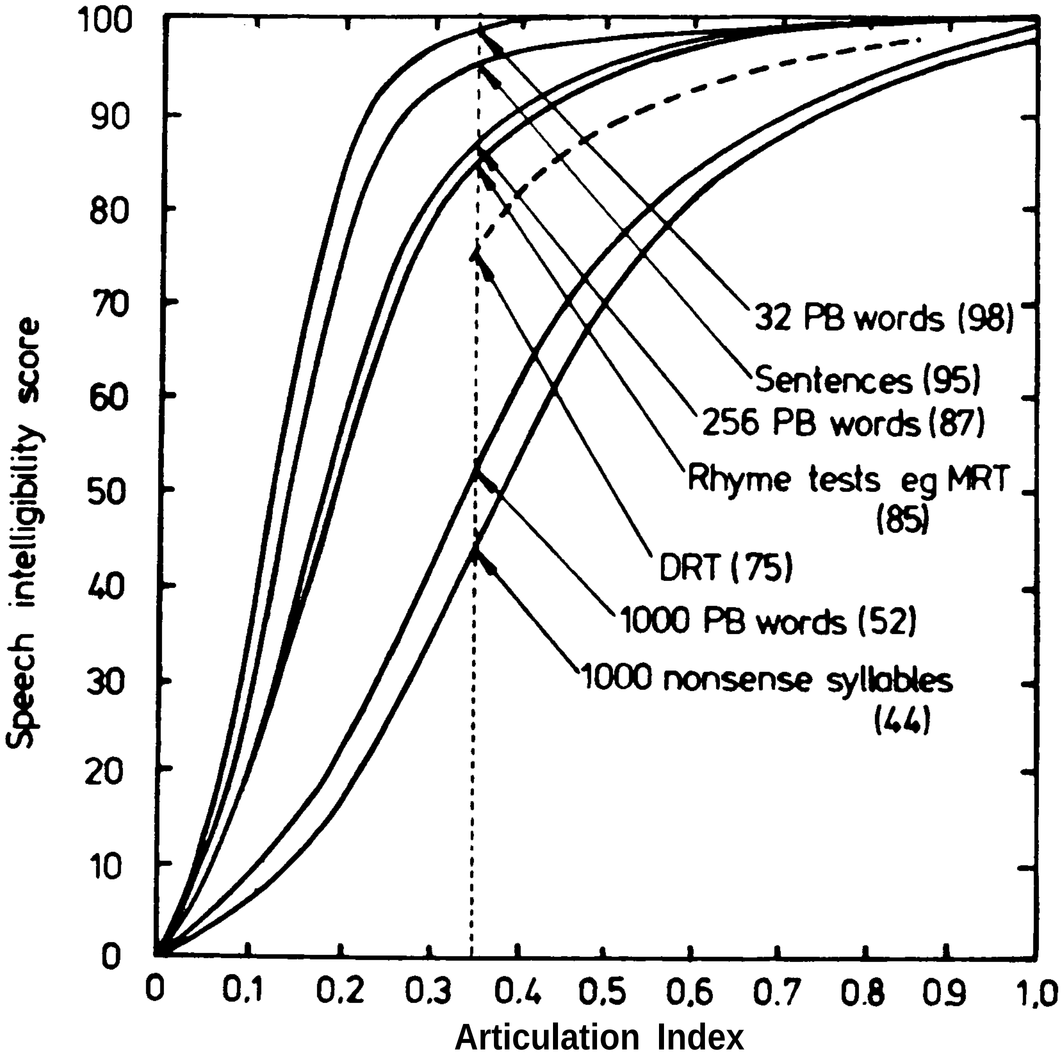

Fig 7 Approximate relation between AI and various measures of speech intelligibility (Based on ANSI S3.5 except for DRT curve, which is from Smith, CP 1986)

10.3.1 The words in each word pair are specifically chosen so that a particular 'attribute' of speech is present in one word but absent in the other. For example, "veal" and "feel" differ in that "veal" has a voiced and feel has an unvoiced initial consonant. The attributes tested are voicing, nasality, sustention, sibilation, graveness and compactness. Thus as well as obtaining an overall score for all words, a score for each of the attributes can be derived by analysing the responses to subsets of words. These attributed scores can be used to highlight specific problems in vocoders and diagnose where improvements may be made. The Diagnostic Rhyme Test is specified as the test for Linear Predictive Coders in NATO STANAG 4198 together with the minimum acceptable scores.

10.4 Comparison of intelligibility tests. The various intelligibility tests produce different quoted test results as shown in figure 7. The greatest differences can be ascribed to vocabulary size and to familiarity with, and hence predictability of, the test material. The figure also shows that the Articulation Index is a non-linear scale. (Note that this figure is approximate since particular distortions or noise may affect an intelligibility score but not the AI, or vice versa.) There is little practical difference between an AI of 1 and of 0.5. With an AI of 0.2 communication can only be satisfactory if the vocabulary used is small, connected or familiar, but if the penalty of misunderstanding one particular message is high and there is not time to repeat it, or spell out significant words in a phoetic alphabet then such a system would be unsatisfactory. An AI less than 0.2 represents an unworkable channel. Note that many military systems may be as poor as 0.3. In this region the slopes of the curves are very steep, the AI scale is quite compressed, and the AI does not effectively reflect important differences in intelligibility under poor conditions.

10.5 Realism in intelligibility testing. Whenever intelligibility testing is carried out the test conditions should be realistic if the results re to be of practical value. Tests on a vocoder for example might show a high intelligibility score when the speech input is free from noise, but the vocoder may fail completely if the long-term r.m.s speech to noise ratio at the input is only, say, 10 dB, and it might still work well if the speech to noise ratio is better than 15 dB. Such results re unpredictable. The input to a vocoder might on occasion be the output from another distorted communication link. Again results are not predictable without a proper simulation. Two microphones may be used as an input to a coder. Under quiet conditions no difference in intelligibility may be apparent. Under noisy conditions one of the microphones may have a better noise-cancelling effect and provide a less noisy input to the coder. Again care should be taken.

Section Five. Techniques and Hardware for Improved Communication

11 General. The environment of a talker and listener and the characteristics of a communication channel may be far from ideal, but careful consideration of hardware and signal conditioning techniques will often allow improved speech intelligibility, usually by improving the speech to noise ratio for the listener. Benefits such as improving operator performance, reducing noise exposure, risk of hearing disorders, mental and physical stress and fatigue, including annoyance and frustration factors are likely to be achieved.

11.1 Microphones. When the talker is in a noisy environment the speech to noise ratio at the microphone output can be optimized by placing the microphone very close to the mouth and by using a noise canceling microphone. Long-term r.m.s speech levels at 15-20 mm from the mouth can easily exceed 100 dB without raising the vocal effort unduly.

11.1.1 Noise-cancelling microphones. A noise-cancelling microphone used correctly close to the mouth gives good speech signals in all but the most extreme conditions. A noise cancelling microphone is open at the front and back so that sound can reach both sides of the diaphragm. The diaphragm moves in response to the instantaneous pressure differences between the front and rear ports, and the output voltage is proportional to the sound pressure gradient rather than the sound pressure. When the front of the microphone is very close to a sound source, say 10-15 mm, the pressure gradient is very steep, and the microphone output is high. For more distant sound sources, 0.5-l.Om away, and at low frequencies the pressure gradient is small, since sound entering by the front and back ports is more or less equal in pressure amplitude and phase, and exert opposing forces on the diaphragm.

11.1.2 Characteristics of noise-cancelling microphones. Unfortunately, it is only at low frequencies and long sound wavelengths, that the background noise pressures on the two sides of the diaphragm are sufficiently equal for useful cancellation to occur, and at frequencies above about 1000 Hz the microphone behaves like a normal pressure microphone. Below 1000 Hz a basic noise cancelling microphone will attenuate distant sound, the effect increasing as the sound frequency falls. About 3 dB attenuation can be expected at 500 Hz and 10 dB at 250 Hz. The attenuation continues to increase as the frequency reduces. Since interfering noise in many practical cases is predominantly of frequencies below 1000 Hz, for example in most helicopters, tanks or heavy machinery spaces, noise cancelling microphones have found widespread use. More complex microphone designs can achieve greater attenuations, but are rarely used for speech. The noise cancelling microphone has to be very close to the mouth, about 15 mm from the corner of the mouth, to be effective, and is usually mounted on a boom as part of a headset. Some modern noise-cancelling microphones will do considerably better than this, especially if mounted very close to the mouth as would appear to be the normal Army practice.

11.1.3 Wind induced noise in noise-cancelling microphones. Noise-cancelling microphones are inherently highly sensitive to wind noise generated by turbulent air flow over the microphone ports, far more so than normal pressure microphones. Because the microphone has to be used so close to the mouth, exhaled breath particularly with the plosive consonants can cause severe wind noise which will overload subsequent electronics. The overload gives severe distortion which makes the speech very "spitchy". A small polyurethane or similar foam wind-shield over the microphone is very effective in reducing wind noise to negligible levels, and should be an integral part of the microphone design. Another situation in which a wind shield would be particularly advantageous, for example, is that of a tank commander with his head outside the tank when travelling at speed.

11.1.4 Throat microphones. Throat microphones avoid picking up ambient noise by strapping a vibration transducer in contact with the larynx. This arrangement picks up the basic laryngeal speech excitation very well, but the much more important higher frequency components produced by the mouth, lips and teeth are largely missing. The intelligibility is so poor that the use of throat microphones in NOT recommended, a noise cancelling microphone is preferable in the majority of cases.

11.1.5 Mask microphones. In environments such as fast jet aircraft oxygen masks are worn and a mask microphone is built in. Usually the speech to noise ratio and intelligibility is adequate for a human listener, but may be marginal if the speech signal is fed to a vocoder.

11.2 High-pass filtering. High-pass filtering should be Mandatory, in situations where low frequencies predominate, as well as using a noise cancelling microphone. It is advantageous to including a high pass filter in the microphone preamplifier. A filter with a cut-off at 300 Hz passes most speech frequencies, but removes the low frequency noise which will either overload the subsequent circuits or cause the gain to be set low to avoid the overload. It is important that the filter is sufficiently early in the amplifying chain to avoid overload. Once overloading has occurred, the noise distortion components and intermodulation products with the speech will extend throughout the speech band and will be impossible to filter out.

11.3 Automatic Volume Control (AVC) or Audio Frequency Automatic Gain Control (AGC). If the system has a good signal to noise ratio, say 50 dB or more, it is possible to transmit the full dynamic range of speech, including inflexion, intonation vocal emphasis and variations from speaker to speaker. However, if the long-term speech to noise ratio is 20 dB or worse, and the speech at the talker's microphone is relatively free of noise, then AVC should be considered. AVC should be applied at a stage in the communication link before any noise mixes with the speech, but after the sidetone output to the speaker's headphones. The time constants of the AVC should neither be made too short, or distortion of the waveform would result. Nor should they be made too long, or occasional quieter words may be missed. Audio Frequency AGC should not be confused with Radio Frequency AGC fitted to radio receivers.

11.4 Sidetone manipulation. A person wearing a headset subconsciously monitors their vocal effort by listening to the sidetone fed from their microphone to their earphones. They try to maintain a reasonable level of sidetone. By using a lower level of sidetone the talker is encouraged to speak more loudly, thereby improving the speech to noise ratio for the listener. A high level of sidetone causes a speaker to speak more quietly, degrading the speech to noise ratio. Reducing sidetone to too low a level can be counter productive: the speaker will no longer make use of it to judge their voice level, it is quite disturbing not to hear one's own voice, and they may suspect that their equipment has malfunctioned. Setting an appropriate level of sidetone should be an important consideration in designing a communication system.

11.5 Peak clipping. Clauses **8.7** and **8.8** described how peak clipping can improve intelligibility of a relatively noise free speech signal from a microphone, in situations where high levels of noise mix with the speech before it reaches the listener. The clipping must occur before the noise mixes. Although peak clipping the audio waveform can improve intelligibility, a better approach is to use Radio Frequency clipping. The speech signal is transposed to a convenient high frequency before clippage, and is then transposed beck to the original audio frequency range. With AF clipping the distortion products are spread throughout the speech band. When the RF waveform is clipped, the distortion products do not overlap the transposed speech frequencies, and can be filtered out before the RF signal is transposed back. The final audio waveform has smoothly rounded peaks rather than flattened peaks. With both AF and RF clipping the smallest parts of the signal are undistorted but greatly amplified.

11.5.1 The effects of RF clipping on the perceived speech quality are
(i) the "edginess" and "graininess" of AF clipping are not present,
(ii) the effect of emphatic or careful enunciation is unchanged, and
(iii) the complete loss of envelope information with extreme degrees of clipping, say more than 20 dB, leaves the speech, although undistorted, very unnatural.

11.5.2 Using modern circuitry an RF clipper can be made as a small compact unit which can be used directly in, for example, a ship's pubic address system.

11.6 Frequency response shaping. With good noise cancelling microphones and given the mainly low frequency noise in many situations, a lifted response of bout 5 dB at 3000 Hz is beneficial to intelligibility.

11.7 Voice-operated switching (VOS). Often in helicopters of AFV crew members need to use both hands for main tasks and are unable to use a push-to–talk switch for the intercom. Therefore, all the crew may have their microphones switched on, picking up the cabin noise, though only one person will be talking. The noise level will be 6 dB higher if four microphones are picking up noise rather than one, and the speech to noise ratio from the person talking will be 6 dB poorer than necessary. Even when no-one is talking the unnecessary microphone noise adds to the crew's noise exposure. A voice operated switch detects the presence of speech at a microphone and switches the microphone through to the intercom automatically. VOs should not hang-up when overloaded.

11.8 Headphones. The choice of headphones is between a comfortable lightweight model and a heavier noise-excluding type. A headphone fitted with a boom microphone is referred to as a headset.

11.8.1 Lightweight headphones. Lightweight headphones and headsets are most acceptable when background noise levels re less than 70 dB(A). The amplifier driving the earphones shall be set so that the long–term r.m.s speech level cannot exceed 80-85 dB(A) at the listener's ear, to prevent the occurrence of sound levels high enough to damage hearing. Speech levels should be measured using an artificial ear conforming to BS 4669 (IEC 318) or with miniature microphones on real ears (see Rood, 1981). If the speech to noise ratio at the ear is not adequate the speech level should not be increased, but the noise level reduced by selecting a noise-excluding headphone. If the speech to noise ratio is not adequate with the noise excluding muff, it may then be necessary to increase speech levels, but the caution.

11.8.2 Noise-excluding headphones. Noise excluding headphones consist of earmuffs with earphone transducers ("receivers" or "telephones") built in. These muffs may be connected by a tensioning headband or may be mounted flexibly within a helmet, such as a flying helmet. Muffs or headphones can be worn with hard hats or helmets if the tensioning headband is worn behind the head, and a thin strap is passed over the head beneath the hat to prevent the muffs slipping downwards. However, as the muffs are supported, it is important that each muff fits around, rather than on, the outer ear and that a complete seal is obtained against the head around the entire circumference of the muff. If the seal to the head is not perfect the attenuation will be reduced: spectacle frames will reduce the attenuation by 5 dB approximately, and some hard hats or helmets will prevent the proper fit of muffs. Muffs worn over a NBC hood will give little attenuation. Modern foam seals on earmuffs are virtually as good as fluid-filled seals and less prone to damage. The required durability and flexibility of fit of earmuffs are defined in BS 6344: 1983, while the method of measuring the attenuation of external sound is specified in BS 5108: 1984 (1S0 4869). Although a good noise excluding headphone or earmuff can reduce noise at the ear by 30-35 dB at frequencies above 1,000 Hz for the average wearer, not all wearers will achieve an ideal fit. For this reason BS 5108 recommends that the "assumed protection"be quoted. This is the mean attenuation minus one standard deviation and is the attenuation likely to be achieved by 84% of wearers.

NOTE: Attenuation values obtained from tests made to American Standards are not comparable to values from tests to the British and International Standard since the procedures differ in important respects. Attenuations quoted from the ANSI Standard method are almost invariably higher and have a smaller standard deviation.

11.9 Active Noise Reduction (ANR). The attenuation of earmuffs is poorer in ambient noise of low frequency and may be as little s 10 dB at 250 Hz. Even this attenuation is useful. The noise attenuation of earmuffs may be improved at low frequencies using a technique known as Active Noise Reduction or ANR. This is illustrated schematically in figure 8. A miniature microphone placed inside each muff picks up the noise and the noise waveform is injected in antiphase into the earmuff through the communication earphone. The noise from the earphone partially cancels the noise inside the muff thereby reducing the noise level at the ear. The speech level from the earphone is also partially cancelled within the muff, but can be restored by increasing the signal gain before the feedback loop. About 15 dB of additional attenuation can be achieved at frequencies up to about 250 Hz, 5-6 dB at 500 Hz, and little to no extra at 1,000 Hz and above.

11.9 (Contd)

The passive attenuation is usually good at the frequencies where the active attenuation is poor and vice versa. ANR is coming into service in helicopters and armoured fighting vehicles. A specification for the performance of an ANR system is given in RAE Specification 2358.

11.10 Adaptive Noise Cancellation (ANC). Adaptive Noise Cancellation is a process whereby the noise can be subtracted out from a combined speech plus noise voltage waveform in an electronic communication line, such as a communication microphone output. ANC exists in two forms. In the first form a second microphone is positioned in the same noise field as the talker but where it will pick up little speech. A microprocessor digitally filters the noise signal from the second microphone and subtracts it from the speech plus noise signal, continuously adapting the filter characteristics to minimize the speech plus noise signal level. When the speech plus noise level is minimum, the maximum noise cancellation has been achieved. In the second form of ANC, the second microphone is dispensed with, and provided the noise is stable, the microprocessor can differentiate between the rapidly varying speech and slowly varying noise in the same signal. ANC is capable of reducing both tonal and broadband noise from speech in real time, but the hardware and software must be tailored to the specific application. ANC is at present still experimental.

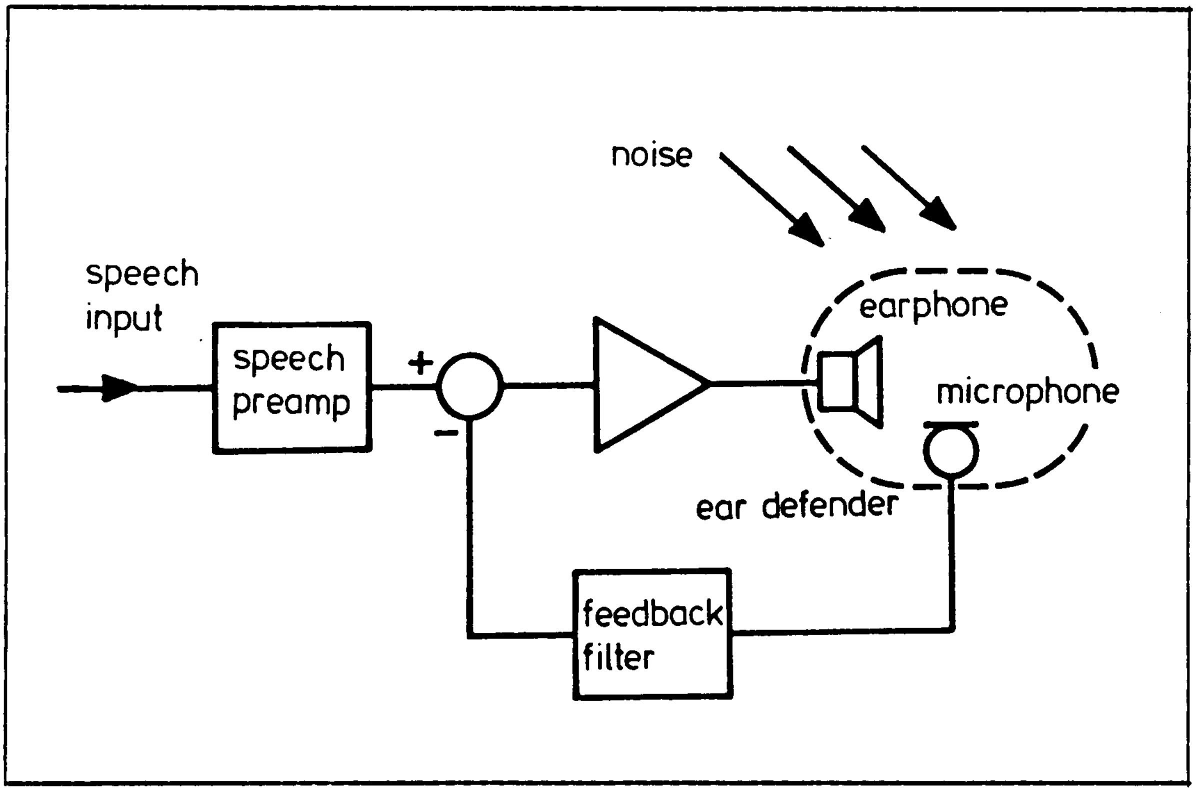

Fig 8 Schematic diagram of an Active Noise Reduction (ANR) system

(Based on Weeler and Halliday, 1981)

Section Six. Voice Procedure

12 Reasons for Voice Procedure

12.1 There is a wide range of pairs of people who may find themselves conversing on a speech link, and a corresponding range of reasons for difficulty in comprehension. Here are some examples:

(a) People who mumble, or who otherwise fail to articulate clearly.

(b) People trained in different disciplines or services.

(c) People who have marked regional accents.

(d) People of different nationalities, who may be using a common language which is native to neither.

12.2 There is also a wide range of conditions in which people may find themselves, which can impede comprehension. For example, the speech link may be technically bad for various reasons:

(a) The signal may be weak.

(b) The level of unwanted noise may be high.

(c) The electrical bandwidth may be restricted.

(d) Distortion of various kinds may be present.

12.3 One or both speakers may be under stress. They may be tired, or in a hurry, suffering physical discomfort, or engaged in battle or a complex work task. Under these conditions recently learned habits of speaking, or of expression, of the expected vocabulary and so on are liable to be forgotten and a speaker may fail to be understood.

13 Voice Procedure

13.1 For the above reasons groups of people who need to talk to each other have to be trained to use their speech link in standard ways. These ways should be as simple and obvious as possible. These standard ways are known as "voice procedure".

13.2 Principles of voice procedure. Different groups of people have unfortunately produced different voice procedures. There are however some general principles which should be maintained:

(a) Words should be simple and short; phrases, and complete messages should be short.

(b) Articulation should be clear, deliberate and unhurried, and a constant vocal effort should be maintained.

(c) The total vocabulary to be used should be kept as small as possible and should be specified: this is best achieved with a list of standard messages.

13.2 (Contd)

(d) Unusual words, codes or abbreviations can be spelled out using the standard phonetic alphabet, ie Alpha, Bravo, Charlie, Delta, Echo, Foxtrot, Golf, Hotel, India, Juliet, Kilo, Lima, Mike, November, Oscar, Papa, Quebec, Romeo, Sierra, Tango, Uniform, Victor, Whisky, X-ray, Yankee, Zulu.

(e) Numbers should be pronounced as wun, too, tree, for-wer, fife, six, sev-en, eight, niner, zero. Zero should always be used, not nought, oh, or nothing.

(f) '-teen' and '-ty' should be carefully distinguished, for example "thirteen, that is, wun tree .." or "thirty, that is, tree zero . ..". a decimal point should be indicated as, for example, "one decimal six fife".

(g) With poor reception the person receiving should acknowledge by repeating back salient points eg "roger, one decimal six fife" or request a repeat of the message, eg "say again?".

Index and/or Design Checklist

Not yet available . . .

Related Documents

British Standards

BS 2475: 1964	Octave and One-Third-Octave Band-Pass Filters (Equivalent to IEC 225)
BS 3383: 1988	Normal equal-loudness contours for pure tones under free field listening conditions (N)
BS 4669: 1971	Specification for an Artificial Ear of the Wide Band Type for the Calibration of Earphones Used in Audiometry (Equivalent to IEC 318)
BS 4727	Glossary of Electroacoustic, Power, Telecommunication. Electronics, Lighting and Colour Terms Part 3 Group 08 (1985) Acoustics and Electroacoustics Terminology (Equivalent to IEC 50 (801))
BS 5108: 1983	Measurement of the Attenuation of Hearing Protectors – Subjective Method (Identical to ISO 4869)
BS 5428	Method for Specifying and Measuring the Characteristics of Sound System Equipment Part 9 (1981): Programme Level Meters
BS 5775	Specification for Quantities, Units and Symbols Part 2 (1979): Periodic and Related Phenomena Part 7 (1979): Acoustics (Identical to ISO 31 Parts 2 and 7)
BS 5969: 1981	Specification for Sound Level Meters (Identical to IEC 651)
BS 6259: 1982	Code of Practice for Planning and Installation of Sound Systems
BS 6344	Industrial Hearing Protectors Part 1 (1989) Specification for Earmuffs
BS 6698: 1986	Specification for Integrating Averaging Sound Level Meters (Identical to IEC 804)

International Standards not yet Adopted as British Standards

ISO 1683-1983	Acoustics - Preferred Reference Quantities for Acoustic Levels
ISO TR3352-1974	Acoustics - Assessment of Noise with Respect to its Effect on the Intelligibility of Speech

NATO Standardization Agreements

STANAG 4198 13 February 1984. Parameters and coding characteristics that must be common to assure interoperability of 2400 BPS linear predictive encoded digital speech.

Foreign Standards Without British Equivalents

ANSI S3.2-1960	American National Standard Methods for Measurement of Monosyllabic Word Intelligibility
ANSI S3.5-1969	American National Standard Methods for the Calculation of the Articulation Index
ANSI S3.14-1977	American National Standard for Rating Noise with Respect to Speech Interference
ANSI S3.20-1973	American National Standard Psychoacoustical Terminology
ANSI C16.5-1954	American Standard Practice for Volume Measurements of Electrical Speech and Program Waves

Other Publications

Anderson, B W and Kalb, J T 1987;	English Verification of the STI Method for Estimating Speech Intelligibility in Noise. Journal of the Acoustical Society of America, 81, 1982-1985.
Fant, G 1973;	Speech Sounds and Features. MIT Press, Cambridge (Massachusetts), and London.
Borchgrevink, H M 1981;	Second Language Speech Comprehension in Noise - a Hazard to Aviation Safety. Paper 15 in AGARD-CP-311, Aural Communication in Noise.
Houtgast, T and Steeneken, H J M 1971;	Evaluation of Speech Transmission Channels by Using Artificial Signals. Acustica, 23, 355-367. (ref Speech Transmission Index)
Houtgast, T and Steeneken, H J M 1981;	STI - An Objective Measure for the Performance of Voice Communication Systems. Paper 17 in AGARD-CP-311, Aural Communications in Noise.
Kryter, K D 1985;	Noise and Man. Second Edition. Academic Press. (Chapter 4, Speech Communications in Noise)
House, A S, Williams, C E, Hecker, M H L and Kryter, K D 1965;	Articulation Testing Methods: Consonantal Differentiation with a Closed Response Set. Journal of the Acoustical Society of America, 37, 158-166. (ref Modified Rhyme Test)
Lower, M C, Flindell, and Wheller, P D, 1986,	The DRT Facility at RSRE; Information for Users. Report AC 575/1, ISVR, University of Southampton.
Miller, G A 1981;	Language and Speech. W H Freeman and Co, San Francisco
Moore, B C J 1982;	An introduction to the Psychology of Hearing. Second Edition. Academic Press. (Chapter 7, Speech Perception)
Richards, D L 1973;	Telecommunication by Speech. Butterworth and Co Ltd.
Rood, G M 1981;	Flying Helmet Attenuation, and the Measurement, with Particular Reference to the Mk4 Helmet. Paper 25 in AGARD-CP-311, Aural Communication in Noise. (ref Use of Miniature Microphones to Measure Sound at the Ear)
Royal Aerospace Establishment 1987;	Harrier GR5 Specification for Active Noise Reduction. RAE Specification No 2358.
Schmitt-Nielsen, A 1987;	Comments on the Use of Physical Measures to Assess Speech Intelligibility. Journal of the Acoustical Society of America, 81, 1985-1987.
Smith, C P 1986;	Comparison of the Effects of Broad-Band Noise on Speech Intelligibility and Voice Quality Ratings. Report RADC-TR-86-135, Rome Air Development Center, Air Force Systems Command, Griffiss Air Force Base, NY 13441-4700, USA.
Steeneken, H J M and Houtgast, T 1980;	A Physical Method for Measuring Speech Transmission Quality. Journal of the Acoustical Society of America, 67 318-326. (ref Speech Transmission Index)
Voiers, W D 1977;	Diagnostic Evaluation of Speech Intelligibility. Chapter 34, 374-387, in Hawley, M E (editor); Speech Intelligibility and Speaker Recognition. Dowden, Hutchinson and Ross, Inc, Stroudsburg, PA, USA. (ref Diagnostic Rhyme Test)

Other Publications

Wheeler, P D and Halliday, S G 1981; An Active Noise Reduction System for Aircrew Helmets. Paper 22 in AGARD-CP-311, Aural Communication in Noise

Widrow, B, et al 1975; Adaptive Noise Canceling: principles and applications. Proceedings of the IEEE, 63 , 1692-1716.

Minimum Performance Requirements

The following shall be satisfied, in any personnel location under any operating condition, where spoken communication is required.

1. Unaided voice communication. The Speech Interference Level, being the arithmetic average of the measured sound pressure levels in the four octave bands centred on 500 Hz, 1,000 Hz, 2,000 Hz and 4,000 Hz as defined in this Defence Standard, shall not exceed the levels shown in Table A of this Defence Standard. For this purpose, the use of a "normal conversation" shall be assumed where continuous communication over long periods is required, and the use of a "raised voice" shall be assumed only where brief and infrequent communication is required.

2. Voice communication using amplified or electronically-assisted speech. The following requirements shall be met between any two stations of a complete communication system, including where appropriate all other stations which are normally part of that system. Environmental factors including ambient noise shall be appropriate too the intended use of the system.

NOTE: It follows that a system which is adequate when only two stations are provided may prove inadequate when additional stations are provided; and that a system adequate in moderate levels of ambient noise is likely to be inadequate in more intense ambient noise.

a. Intelligibility.

(1) For those systems where a meaningful value of Articulation Index can be determined, the Articulation Index (AI) as defined in ANSI S3.5-1971 shall not be less than 0.5.

(2) For those systems where a meaningful value of Articulation Index cannot be determined, for instance systems using vocoded speech, intelligibility shall be determined using the Diagnostic Rhyme Test (DRT) (Voiers 1977). For those systems described in STANAG 4198, requirements of that STANAG shall be met. For other systems, the Diagnostic Rhyme Test score shall be 85% or greater, desirably 90% or greater.

NOTE: In some cases it may be appropriate to use other means of assessing speech quality, for instance Speech Transmission Index (STI) in place of AI, or Modified Rhyme Test (MRT) or realistic military orders in place of DRT. In such cases, it shall be shown that the use of such methods is appropriate, and that the results obtained are equivalent to, or exceed, the minimum requirements detailed above.

b. Maximum noise levels at the ear

(1) Where background noise is low, noise at the ear from speech and all other sources shall not exceed 85 dB(A), measured using the time weighting "F" defined in ISO 651.

(2) For any background noise level, during one complete day of normal operation, noise at the ear from speech and all other sources shall not exceed an 8-hour equivalent continuous noise level (L Eq (8)) of 85 dB(A), assuming typical expected use of the communications system. It is highly desirable that the long-term rms level at the ear should not exceed 85 dB(A) under any operating condition.

NOTE: This, together with requirements for intelligibility, will in general require that the background noise level at the ear, at times when speech or other information is not being transmitted, be substantially less than 85 dB(A). It follows that the use of noise-excluding headsets, possibly with an active noise reduction system, will often be required, and that microphones will need an effective noise-cancelling performance and should not be left permanently "live". Reduction of background noise, by engineering means or by re–location of personnel, should also be considered.

Index

Index (Continued)

Published by and obtainable from:
Ministry of Defence
Directorate of Standardization
Kentigern House
65 Brown Street
GLASGOW G2 8EX

Tel No: 041-224-2531/2532
Fax: 041-224-2503

The following Defence Standard file reference relates to the work on this Standard - D/D Stan/328/Ol/09.

Contract Requirements

When Defence Standards are incorporated into contracts users are responsible for their correct application and for complying with contract requirements.

Revision of Defence Standards

Defence Standards are revised when necessary by the issue either of amendments or of revised editions. It is important that users of Defence Standards should ascertain that they are in possession of the latest amendments or editions. Information on all Defence Standards is contained in Def Stan 00–00 (Part 3) Section 4, Index of Standards for Defence Procurement - Defence Standards Index published annually and supplemented periodically by Standards in Defence News. Any person who, when making use of a Defence Standard encounters an inaccuracy or ambiguity is requested to notify the Directorate of Standardization without delay in order that the matter may be investigated and appropriate action taken.

92/60028