Manuscript no. 413/II

Guide lines for auditory warning systems on civil aircraft

Roy D.Patterson *)
MRC Applied Psychology Unit,
15 Chaucer Rd, Cambridge, England

*) This document was prepared for the Airworthiness Division of the Civil Aviation Authority while the author was on leave as a visiting research scientist at the Institute for Perception Research, Eindhoven, The Netherlands.

SUMMARY

The purpose of this paper is to develop a set of guidelines for the design and/or evaluation of the auditory warning systems used on the flight-decks of commercial aircraft. The principles that govern the design of auditory warnings are introduced in the first section of this paper by comparing the problems inherent in existing warning systems with a prototype of an advanced auditory warning. The guidelines are then developed in four separate sections concerned with (i) the overall sound level, (ii) the temporal characteristics, (iii) the spectral characteristics, and (iv) the ergonomics of auditory warnings. The use of voice warnings is considered briefly in a fifth and final section.

The resulting guidelines are presented, for subsequent reference, as a group in a separate appendix at the very end of the document.

CONTENTS Page

Page

APPENDIX B The Learning, Retention, and Confusion of Auditory Warnings

APPENDIX C Summary of Auditory Warning Guidelines

Instituut voor Perceptie Onderzoek
Den Dolech 2 – Eindhoven

RP/ne 82/01
02.02.1982

Manuscript no. 413/II

Guidelines for auditory warning systems on civil aircraft

Roy D. Patterson

0 INTRODUCTION

The purpose of the auditory warning system on the flight-deck of a commercial aircraft is to alert the flight-crew to dangerous conditions, to potentially dangerous conditions, and to the arrival of information on visual displays. All of the current warning systems perform the alerting function with exceptional reliability; furthermore, in the vast majority of cases, the information specifying the type of problem is successfully communicated. But the existing systems achieve their success at considerable cost, in that they flood the flight-deck with very loud, strident sounds. This has two unfortunate side effects: First, it makes the auditory warning systems unpopular with flight-crew. Second, and perhaps more important, many of the existing warnings disrupt thought and prevent crew communication, which at a critical moment makes an already difficult situation worse. The side effects are not unavoidable and the primary purpose of this paper is to explain how they can be minimised or even eliminated without reducing the reliability of the system.

Some of the more recent aircraft have as many as sixteen auditory warnings and alerts on the flight-deck. The experience of flight-crews, and some recent laboratory research, indicate that it requires an inordinate amount of training and retraining to maintain perfect identification of all of the members of such a set.

This has led to the suggestion that voice warnings should be introduced to reduce the number of auditory warnings required on the' flight-deck. Thus, although the primary concern of this paper is the design of auditory warnings, the appropriate number of warnings and the role of voice warnings are also considered in so far as the design of the warnings and the total system interact.

To understand the development of the new auditory warnings described in this paper, it is important to know the structure envisaged for the next generation of warning systems. The most relevant discussion of future systems appears in ARINC CHARACTERISTIC 726 which presents a consensus of the views of flight-crew, carriers, and manufacturers concerning the improvements that need to be incorporated. The system described in that document employs six to eight individual warning sounds, and a somewhat larger number of voice warnings, to present information at three priority levels on the flight-deck. The top priority is ‘emergency’; these situations require the immediate action of the flight-crew. There would be four to six of these immediate-action warnings, each composed of a unique warning sound followed by a voice warning that provides backup and, perhaps, a little more specific information. The second priority is ‘abnormal condition’ and it requires the flight-crew’s immediate awareness. There might be as many as 10 of these alerts. This entire group of warnings would be signalled by one particular warning sound (sometimes referred to as an ‘attenson’) and it would be followed by a voice warning specific to the condition that initiated the warning. The third level of priority is ‘advisory’ and it requires the crew’s awareness but not necessarily action. This group would be signalled by a soft warning sound, or attenson, but there would be no verbal message. The attenson simply indicates that the flight-crew should check the visual displays associated with operational or aircraft systems as soon as they have a convenient opportunity. In case of conflict the higher priority warning takes precedence.

In the current paper, it is assumed that a multi-level priority system will replace the existing system, and that the aim is basically

(a) to develop a set of four to six compatible sounds to serve as immediate-action warnings,

(b) to develop two or three attensons for lower priority alerts, and

(c) to integrate voice warnings into the auditory warning system.

0.1 Structure of the Current Paper

In the remainder of this section, the principles that govern the design of auditory warnings are introduced by first outlining the problems associated with the existing situation, and then describing a prototype of an advanced, immediate-action warning. In the main body of the paper, four groups of design principles are presented in separate sections and the guidelines that summarise the principles for practical purposes are developed. The role of voice warnings is discussed in the fifth, and last, section of the main paper. Since the paper is primarily intended to support the design and evaluation of auditory warning systems, the scientific justification for the guidelines is kept to a minimum in the main paper and the essential details are presented in two separate appendices. The conclusions are presented as a list of guidelines in a third and final appendix at the very end of the document.

0.2 The Existing Situation

The major deficiencies of existing warning systems will already be apparent to most readers.

0.2.1 Overall level

Most of the warnings are too loud. Some of the warnings are so loud that they not only disrupt thought but can actually prevent crew communication altogether. The background noise on the flight-deck of a modern commercial jet is relatively low, and it certainly does not justify the existing warning levels. The first topic of this paper, then, is the appropriate level for flight-deck warnings.

0.2.2 Temporal characteristics

The onsets and offsets of warning sounds are typically too abrupt. Although flight-crew do not seem to show overt startle reactions, the sharp onsets and high levels of some warnings are entirely sufficient to evoke startle reactions in the population at large, and they almost undoubtedly cause a temporary disruption of cognitive function in the flight-crew. Since no warning ever requires an instantaneous response there is no need to use abrupt onsets and risk startle reactions.

The temporal patterns of existing warnings are not sufficiently distinctive; most warnings have either no temporal pattern, that is they are continuous, or they have simple alternation patterns, as for example, the interrupted horn. Sounds with distinctive temporal patterns are less likely to be confused, and on the flight-deck it is important to use every means of making the warning sounds distinctive.

The total on-time of the warnings is far too high. The purpose of a warning is to draw the crew's attention and convey a small amount of information. There is no need to blanket the flight-deck with sound to accomplish this task. Short bursts of sound separated by 5 to 10 second silent intervals can present the information just as effectively while still leaving space for crew communication.

The temporal characteristics of auditory warnings are the topic of the second section of this paper.

0.2.3 Spectral characteristics

The spectral content of existing warnings is, in general, quite good; they typically contain many components spread throughout the spectrum, some harmonic and some inharmonic. As a result, they are distinctive sounds and spectrally-based confusions seem very unlikely.

In the more recent warning systems, where there are a large number of warnings, some of the lower priority warnings appear more urgent than higher priority warnings and this should be rectified. In addition, given the spectrum of the background noise on the flight-deck, the power of the low-frequency components of the warning sounds should be increased relative to the mid-frequency components, if the warnings are to sound the same during different stages of flight. The procedures for choosing a spectrum are the topic of the third section of the paper.

0.2.4 Ergonomics

The ergonomics of existing warning systems is deplorable. Most of the problems derive from the fact that the early warning systems had to be very simple in order to be reliable enough and small enough for the flight-deck. With the advent of micro-electronics, however, small, reliable systems that include ergonomic considerations are possible and should be implemented as soon as possible.

Ergonomic problems are perhaps most easily understood from the human point of view. To the flight-crew existing warning systems must seem rude and selfish; they burst onto the flight-deck with shouts of ‘emergency’ disrupting and preventing other activity until they are cancelled. Furthermore, they are totally lacking in a sense of perspective. When a warning occurs it is usually either a false warning or the direct result of a standard flight procedure, as when the overspeed warning sounds at the transition from level-flight to descent. Even when a true warning occurs, it almost always indicates a potential problem rather than a sudden emergency. But the existing warnings have only one mode; if they are on they are in the full emergency mode. The louder warnings are very disruptive, and so, when a loud, procedurally induced warning is impending, one of the flight-crew may be detailed to sit with a finger poised over the cancel button so that the warning can be terminated the instant it sounds. And when a loud false warning or a loud true warning occurs the crew's attention is directed, not to the real problem, but rather to the problem of finding the warning cancel button. These and other ergonomic problems are the topic of the fourth section of the paper.

0.2.5 Voice warnings

Although currently there are very few voice warnings, and the speech quality of those that do exist is not particularly good, it seems likely that they will improve rapidly and become an important part of flight-deck warning systems in the not too distant future. Consequently, their role is considered in a fifth and last section of this paper, although the principles of synthetic speech production and the criteria for choosing a speech system are not considered.

Voice warnings have the decided advantage that they are easy to learn and difficult to forget or confuse. They take a relatively long time to present their information, however, and there is already a lot of speech on the flight-deck, so voice warnings should probably be used to support rather than replace warning sounds in the case of immediate-action warnings.

0.3 A Prototype of an Advanced, Immediate-Action Warning

A prototype of an advanced, immediate-action warning will be introduced at this point to ensure that the purpose and direction of the paper are apparent from the beginning. It should be noted that the prototype was constructed to illustrate the design principles described in this paper; as yet it has not been tested on the flight-deck.

The temporal structure of a warning that might be used to designate the condition 'undercarriage unsafe' is presented in the diagrams of Figs. 0.1 and 0.2. They show, respectively, the basic pulse patterns used to present the warning, and the time course of the complete warning. In Fig. 0.1 each rounded hump represents a pulse of sound about one tenth of a second in duration. The waveform within the pulse is unique to a particular warning; it carries the spectral information of the warning sound and is never altered. A burst of six pulses defines the warning sound. The basic grouping of four, clustered pulses followed by two, irregularly-spaced pulses provides the rhythm of the sound which, combined with the spectral characteristics stored in the waveform, gives the sound its distinctiveness. The four rows of Fig. 0.1 show how the spacing and intensity of the pulses can be varied within the burst to vary the impression of urgency and avoid abrupt onsets. Each trapezium in Fig. 0.2 represents one play of the warning sound, that is one row of Fig. 0.1; the number in the trapezium indicates the row. The duration of a row in Fig. 0.2 is about 20 times that of Fig. 0.1, or about 32 seconds. The rectangle represents a voice warning and the heights of the rectangles and trapeziums indicate the relative intensities of the various components. When conditions dictate, the warning is initiated and proceeds as follows:

0.3.1 Overall level,

The warning comes on at a moderate level, well above the minimum required to draw the crew's attention, but well below the level where it would be aversive. This would make procedurally-induced occurrences of the warning much more acceptable. The warning is repeated and then a voice warning presents the same information and perhaps some minimal elaboration, again at a moderate level. The method for determining the levels is presented in Section 1.

At this point the warning is turned down automatically to a level that is still clearly audible but which can be overridden by a person speaking loudly. The warning stays in the background at the lower level for a reasonable length of time, depending primarily on the urgency of the condition, but if the fault is not corrected the auditory warning and the voice warning both return (the second row of Fig. 0.2) at the maximum of the appropriate range for warnings – a level still considerably below the louder existing warnings. The warning then returns to the background level as before, and the whole pattern is repeated should it prove necessary.

0.3.2 Temporal characteristics

The requirements that auditory warnings be arresting but not startling might, at first, appear to conflict. However, since the pilots are not required to make instantaneous responses to the warnings (in fact, instantaneous responses are discouraged) an arresting but not startling warning can be produced by bringing the warning on at a comparatively low level and increasing the level of successive pulses quickly as shown in the second and successive rows of Fig. 0.1. This amplitude envelope gives the impression that an object is moving towards you rapidly and then receding slowly, and this apparent motion draws your attention. At the same time, since the first pulse comes on at a moderate level, the warning does not cause a startle reaction. The basic pulse is similarly given a rounded top rather than an abrupt onset or offset to reduce the risk of a startle reaction.

As mentioned earlier, the grouping of four regular and two irregular pulses gives the warning sound a distinctive, slightly syncopated rhythm. The version of the pattern shown in the first row of Fig. 0.1 does not, however, sound particularly urgent when played at a moderate level and at the rate indicated. More urgent versions of the same pattern are obtained by compressing the first four pulses in time as shown in the last two rows of the figure. So long as the warning is composed of a group of four regular beats followed by two irregular but fixed beats, and the waveform within the pulse is not changed, it will sound like the same warning. In Fig. 0.2, the ‘3’ in the first pair of trapeziums indicates that the warning is initiated in the version that gives the impression of moderate urgency. After the voice warning it is changed to version ‘1’, which sounds less urgent when played at a lower level because the first group of pulses is well spaced and they are all the same level. But if the fault condition is not corrected the warning returns in the most urgent form ‘4’ which, combined with the maximum level, commands attention.

The total on-time of a warning can be quite small and still be entirely sufficient to present the required warning information. In the course of one 32–second cycle of the main sequence of the current example (a row of Fig. 0.2), the warning’s information is presented seven times, once verbally and six times acoustically. But the time taken to present the information is less than one quarter of the total, and of this, the verbal warning takes up almost half. Thus, in the first 32 seconds there is ample space for crew communication, and since the flight-crew would be familiar with the temporal structure of the warning, they would probably find that beyond the voice warning they could communicate without difficulty, while still being aware of the warning in the background. This stands in marked contrast to the existing situation where the warning must be cancelled before communication is possible – a procedure that increases the risk in a real emergency that the warning will be forgotten after cancellation as the crew attend to other aspects of the emergency.

0.3.3 Spectral characteristics

There are already many warnings with spectra that convey the impression of urgency well.

Where possible, to provide continuity between existing and future warning systems, the basic pulse of a new warning would be taken from the sound assigned to the same function in the system it replaces. This can be accomplished by digitizing a sample of the original sound, selecting an appropriate subsection, and rounding the onset and offset with a cosine gating function. In this way the original association of sound and function (for example, a horn sound with configuration faults) can be preserved while at the same time implementing the other improvements.

Although there is no need for new urgent sounds, there is a need for several less urgent, but attention-demanding sounds – the so-called ‘attensons’. They are required to indicate the arrival of a low-priority alert or other information on a visual display. These can be generated using strictly harmonic spectra that sound like musical chords.

0.3.4 Ergonomics

From the point of view of the flight-crew, the ergonomic problem is one of making the warning system more civilised. Some of the ergonomic improvements of the prototype have already been introduced: The measures implemented to reduce the risk of startle reactions (moderate initial levels, round-topped pulses and low-level initial pulses) will assist in making the system seem less rude. The automatic reduction of the warning level after the voice warning, and the reduction of the on-time, will help make the warning seem less selfish. The sense of perspective is improved by not starting with the loudest, most urgent version of the warning sound. Thus, the system should seem altogether more civil.

There is, however, a second motivation for the ergonomic improvements and that is safety. If the aversive character of the system can be changed, it may be possible to convince the crew not to cancel warnings as soon as they occur, thereby cancelling the protection the warnings provide. If the warning is not too aversive initially, and moves to a background level in a reasonable amount of time, the crew might feel that, in the case of procedurally-induced warnings, they could wait for the natural correction of the fault condition to extinguish the warning.

In the event of an unexpected warning, the crew would know that they have some time to determine whether there is a real fault before the warning sound changes to the most urgent form, and so again they might feel that they could leave the system on for the sake of safety. The ergonomics could be improved further, in this situation, by providing a response panel on which the crew could indicate to the system that they had heard the warning and correctly identified the problem area. If correct, the crew response would cause the warning to remain in the background mode for a longer period of time before changing to the most urgent form. If incorrect, the response would cause the system to change to the most urgent form of the warning immediately, thereby providing a useful check on the direction of the crew's attention.

The temporal and spectral characteristics of existing warnings are such that if two warnings are required simultaneously, the combined sound might prove confusing and cause the recognition of one or both warnings to be delayed. It has been suggested that an order of priority could be applied to the warnings in future systems, with the higher priority warning interrupting and suppressing the lower. In some cases, however, the order of priority is difficult to determine. Furthermore, the suppression of a fairly important warning might actually make it more difficult to analyse the overall problem correctly. A warning system comprised of advanced warnings like the prototype could signal two immediate-action faults virtually simultaneously with a minimum risk of confusion, because the bursts of sound carrying the information could be interleaved. It would require a small amount of coordination on the part of the warning system, but since the on-time of the sound bursts occupies less than one quarter of the total elapsed time, it would be feasible. This ability to signal two emergency conditions simultaneously would markedly reduce the problem of establishing a strict order of priorities for the immediate-action warnings. In addition, it would reduce the probability of having to suppress an immediate-action warning to a very low level.

0.3.5 Voice warnings

In the immediate-action warning, the role of the voice warning is to present one, highly redundant, repetition of the warning's information to eliminate the possibility of confusion; in this way the major advantages of speech are incorporated into the system. The voice warning is not repeated when the warning switches to background mode because it would be intrusive and increase the total on-time unnecessarily. It is also difficult to produce a background-level voice warning because of the large dynamic range required for speech. The vowels of speech are often 30 dB more intense than the consonants, and so if a voice warning were attenuated to produce a background version with the correct vowel level the consonants would be near or below masked threshold.

In the immediate-action warnings, the voice-warning component should probably have a ‘key-word’ format, rather than a ‘full-phrase’ format to keep the sound duration as brief as possible. The comprehension of key-word messages takes as long as a complete phrase with the same information, and the phrase is more redundant. But the keyword format helps minimise the total on-time and this is an advantage for immediate-action warnings. At the second level of priority, ‘abnormal conditions’, there will be more warnings and they will not be paired with individual sounds, and in this case, full-phrase format will almost undoubtedly be preferable.

1. THE OVERALL LEVEL FOR FLIGHT-DECK WARNINGS

1.0.1 The existing situation

The noise levels on the flight-decks of civilian jet aircraft are relatively low and do not appear to justify the sound levels of many auditory warnings. The literature does not reveal any substantial research to determine the appropriate level for flight-deck warnings, and there is a widely held suspicion that the levels of many existing warnings were set simply by making them as loud as possible to ensure that they would command the flight-crew's attention. This seems, at first, a reasonable approach in that, if the appropriate level is not known, it must be better to set the level overly high, even if it makes the warnings aversive to the crew. The difficulty with this approach is that one can be too conservative; in some cases the warnings are so loud that when they occur the flight-crew's attention is focused not on the cause of the problem but rather on the intensity of the sound and the search for the cancellation button. Furthermore, some of the warnings are loud enough to make verbal communication difficult or even impossible.

Although there has been little research on the correct level for flight-deck warnings, there have recently been three surveys of flight-crew opinion concerning aircraft warning systems. A structured-interview technique has been used to obtain opinions concerning the levels of the individual warnings on the crew member's particular aircraft (Ref. 1); no warnings were rated 'too soft' while many, including the most important ones, were frequently rated 'too loud'. A survey of warning-system philosophies (Ref. 2), includes a catalogue of common crew complaints from which the author concludes 'loud sounds tend to incapacitate'. Finally there is an IATA study (Ref. 3) in which 46 airlines, operating between them virtually all of the standard aircraft, were surveyed, and on the topic of warning intensity the conclusions are: 'Most aural alerts, as currently designed, are too loud', and subsequently, 'Most aural alerts are so loud that normal crew co-ordination cannot be carried on'.

There would, then, appear to be a clear requirement for a procedure to determine the appropriate level for flight-deck warnings. The procedure presented in this section has three basic steps: The background noise on the flight-deck is used to determine auditory threshold as a function of frequency on the flight-deck. Then, this general threshold curve is used to establish the appropriate range of levels for flight-deck warnings. Finally, individual warnings are adjusted so that their dominant components fall in the appropriate range.

1.1 The Range of Appropriate Levels for Flight-Deck Warnings

In this subsection it is argued

(a) that warnings should be 15 dB or more above masked threshold to ensure that they will be noticed, and

(b) that warning sounds should not be more than 30 dB above threshold or they will be aversive, and may disrupt verbal communication.

1.1.1 The lower limit – threshold + 15 dB

The audibility of a signal is determined by the background noise in the environment where it occurs. Thus, the problem of establishing the minimum level for an auditory warning reduces to one of determining the threshold imposed by the flight-deck noise, and specifying the minimum warning level in terms of that threshold. Because of the inherent variability of noise, and the variability in auditory processing, threshold does not occur at one precise signal level; rather the probability of detecting the signal rises from very low to very high as signal level increases over a range of about 20 dB.

The function relating the probability of signal detection to signal level is referred to as the psychometric function and it is important for two reasons: First, threshold is defined by this function. The measurement of the psychometric function and the precise definition of threshold are presented in the first part of Appendix A; a method for calculating threshold for flight-deck noise is presented later in this section. At this point it is sufficient to note that threshold is well defined and can be predicted with considerable accuracy. The second importance of the psychometric function is that it shows how the audibility of a signal increases with signal power in the region above threshold. Briefly, when a signal is 10 dB above threshold it is easy to hear and when it is 15 dB above threshold it is difficult to miss. Data on auditory frequency discrimination, the perception of loudness, and localisation all lead to the same conclusion; namely, once a signal is 15 dB above threshold the effect of the background noise on the signal is negligible. These results are reviewed in Appendix A along with two studies (Ref. 4, 5) showing that this generalisation extends to warning sounds like those used on the flight-deck.

As a guideline, then, the lower limit of the range appropriate for auditory warnings is 15 dB above the threshold imposed by the background noise.

1.1.2 The upper limit – threshold + 25 dB

Although the flight-decks of commercial aircraft are quiet by comparison with military aircraft, it still requires fairly loud warnings to reach the minimum of the appropriate range. As a result, the upper bound of the region is primarily determined by the desire to avoid annoyance and disruption. Annoyance is a complex psychological variable and shows considerable dependence on context, but it is clear that sounds in excess of 90 dB(A) annoy virtually everyone, and the pilot surveys (Ref. 1, 2 and 3) confirm that flight-deck warnings are no exception. Loud sounds also disrupt thought and communication, all of which indicates that the upper bound of the appropriate range for warnings should be no higher than safety considerations dictate.

The relationship between warning level and safety is considered in the first section of Appendix A. Evidence from studies involving existing warnings (Ref. 4 and 5) shows that increasing the warning level above that required to make it perfectly audible does not improve either the detection or recognition of the warning. Furthermore, loud continuous warnings tend to hold the crew's attention beyond the point where the problem has been identified, and they disrupt communication at a vital time. Thus, the overall safety level is probably improved by avoiding excessive sound levels.

As a guideline, then, since the minimum of the appropriate range for flight-deck warnings is already fairly high, the maximum should be kept to about 10 dB above the minimum; that is, threshold + 25 dB.

1.2 The Prediction of Masked Threshold

Since the range of appropriate warning levels is defined in terms of masked threshold, the problem of checking the level of an existing warning, or predicting the correct level for a new warning, reduces to one of determining threshold for the warning on the flight-deck. It might be possible to actually measure threshold for the warnings on some of the larger flight-decks, but it would be a time-consuming and costly procedure, and it would not solve the problem of the designer who wants to predict the required level in advance of the warning's construction. Fortunately, the theory of auditory masking has progressed to the point where it is possible to predict threshold when the background is a stationary noise like that on the flight-deck. The model is outlined in this subsection and the procedure for predicting threshold on the flight-deck is illustrated in the next subsection; the details are presented in the second part of Appendix A.

1.2.1 The power-spectrum model of auditory masking

The peripheral auditory system begins the processing of incoming sound with a fairly detailed frequency analysis, and it is in essence this initial analysis that determines whether one sound will mask another. The auditory system is largely insensitive to the phase of individual frequency components, particularly when the masker is a noise, and auditory warnings are long with respect to the integration time of the ear. As a result, a simple power-spectrum model can provide quite an accurate representation of the frequency analysis; indeed, to a first approximation, a noise will mask a signal whenever the spectra of the stimuli show that the noise power exceeds the signal power throughout the spectrum. The accuracy of the model depends on the analysing filter chosen to produce the spectra of the stimuli. If the filter smears the spectra of the incoming sounds to the same degree as the auditory system, the prediction can be highly accurate. The model is described in the second section of Appendix A.

Briefly, it is assumed that an observer trying to detect a signal centres an auditory filter at a local peak of the signal spectrum and listens for the signal through that filter. If the power of the signal at threshold is P_s, the long-term power spectrum of the noise is N(f), and the auditory filter shape is W(f), then the general equation for the power-spectrum model of masking is

$$P_s = K\int_{-\infty}^{\infty} N(f)W(f)\,df \qquad (1.1)$$

In words, the power of the signal at threshold is some constant, K, times the integral of the noise spectrum times the filter function. It is a 'power-spectrum model' because the stimuli are represented by their long-term power spectra. Thus, the filter shape is just a weighting function that imposes the limitations of the auditory system on the spectrum of the incoming stimulus.

1.2.2 Auditory filter shape

The filter shape can be measured experimentally; the method and the effects of centre frequency, stimulus level and age are reviewed in Appendix A, section 2. The filter is typical of a resonant, physical system: It has a well defined passband with an equivalent rectangular bandwidth, BW_{ER}, that is roughly 15% of the centre frequency. The skirts of the filter bounding the passband fall at a rate of just over 100 dB per octave. About 40 dB down from the top, the slope of the skirts becomes much shallower. The filter is close to symmetric on a linear frequency scale. A good approximation to the attenuation characteristic, or shape, of the filter is provided by a rounded-exponential function of the form

$$W(g) = (1 - r)(1 + pg)\, e^{-pg} + r \qquad (1.2)$$

where g is the normalised distance in frequency from the centre of the filter, f_c, ($g = |f - f_c|/f_c$). The parameter p determines the width of the passband of the filter. The function is a pair of back-to-back exponentials (e^{-pg}) with the peak rounded off by the term (1 + pg) and the dynamic range of the exponential limited by a floor, r. The term (1 - r) simply ensures that there is neither loss nor gain at the centre frequency.

1.2.3 The calculation of masked threshold

The filter shape can be substituted into the general masking equation, Eq. 1.1, to provide an expression for calculating threshold from an arbitrary- noise spectrum. The proportionality constant, K, can be assumed to have a value of 1.0 for practical purposes. Thus, the general expression for threshold is

$$P_s = f_c \int_0^{0.8} N(g) \left[(1 - r)(1 + pg)\, e^{-pg} + r \right] dg \qquad (1.3)$$

The constant is required to convert from the normalised frequency domain to physical power. Since the limit on the dynamic range is implemented by means of a constant, r, the integration is restricted in frequency to 0.8. The filter function and its integral are presented graphically in Appendix A, along with a discussion of their relationship and a more comprehensive discussion of threshold calculation. This expression can be used to predict threshold on any aircraft where the total noise power does not exceed about 95 dBA, that is, on helicopters as well as fixed-wing aircraft with either turbo-prop or jet propulsion. (Above 95 dBA the auditory filter broadens and a correction must be included.)

On the flight-deck of modern jet aircraft the noise spectra are fairly smooth. In this special case, the noise spectrum can be approximated by a constant, NL_c; the auditory filter can be approximated by its equivalent rectangular bandwidth, BW_{ER}; and the masking equation (Eq. 1.3) reduces to a very simple form:

$$P_s = BW_{ER}.NL_c \qquad (1.4)$$

where BW_{ER} is in Hz and NL_c is in (dynes/cm^2)/Hz. Typically, both the noise level and the signal—power—at—threshold are expressed in dB SPL; that is in tenths of log-units, where the reference level is 0.0002 dynes/cm^2. Thus a more convenient form of Eq. 1.4 is

$$10 \log P_s = 10 \log BW_{ER} + 10 \log NL_c \qquad (1.5)$$

where 10 log P_s and 10 log NL_c are in dB SPL. BW_{ER} is approximately $0.15f_c$, and it is the width that a rectangular filter with unit height must have to yield the same total area as the auditory filter. Provided the noise spectrum does not fall more than 6 dB across the equivalent rectangular filter, the average noise level in dB SPL can be approximated by the value at f_c

The procedure for calculating threshold as a function of frequency is illustrated in Fig. 1.0. The spectrum of the flight-deck noise is the solid line with dots. Two auditory filters centred at 1.0 and 4.0 kHz are shown along with their rectangular equivalents. The appropriate noise level for calculating threshold at 1.0 kHz is 50 dB SPL and the same value is appropriate at 4.0 kHz. Thus threshold at 1.0 and 4.0 kHz would be

$$10 \log P_s(1.0) = 10 \log(0.15x1000) + 50 = 71.8 \text{ dB SPL, and} \qquad (1.6a)$$

$$10 \log P_s(4.0) = 10 \log(0.15x4000) + 50 = 77.8 \text{ dB SPL.} \qquad (1.6b)$$

The values are plotted as dots in open circles in Fig. 1.0; threshold is greater at 4.0 kHz simply because the filter is wider in absolute terms at the higher frequency. The value of threshold at multiples of 0.5 kHz is shown by the line of dots, and they could be joined to provide a threshold curve for this noise spectrum.

1.3 The Evaluation of Existing Warning Levels

In this subsection the power-spectrum model of masking is used to predict threshold on the flight-deck of a Boeing 727 and a BAC 1–11, and the appropriate range for warning levels is derived from the threshold functions. Then, the levels of the configuration horns and firebells on the two aircraft are evaluated (Ref. 6).

1.3.1 The appropriate range for warnings on the Boeing 727 and the BAC 1–11.

The background noise was recorded on the flight-decks of a BAC 1–11 and a Boeing 727 during five phases of flight: takeoff, steady-climb, level-flight, descent, and approach. In the mid-frequency range, between 0.5 and 5.0 kHz, where the majority of the power in the warning sounds occurs, the level-flight phase of flight produced the loudest noise on both aircraft, followed by the descent and steady-climb phases respectively. On takeoff there were thumps when the aircraft crossed runway cracks and a steady low-frequency rumble, but neither of these noises is important because neither would mask the auditory warnings which are of long duration in comparison to the bumps and high in frequency in comparison to the rumble. The spectra of the steady-climb, level-flight, and descent noise on the flight-deck of the BAC 1–11 are shown in Fig. 1.1. The abscissa is frequency in kHz and the ordinate is power in dB SPL (the range of human hearing is roughly 0 to 120 dB SPL in the region of 2.0 kHz). The noise on the flight-decks of these tail-engine jets is produced primarily by the turbulence in the boundary layer of the air flowing over the nose of the aircraft. Thus, all the spectra fall gently from left to right, and the higher-speed, level-flight spectrum exceeds the lower-speed, climb and descent spectra. On this flight there was a long, slow descent delaying for landing clearance and so the spectrum is relatively low. The situation is similar on the 727, Fig. 1.2, except that engine noise was apparent in the 4.0–kHz region of the climb spectrum when full power was being used. On this flight there was a steep descent with engine power on, and so the spectrum of the descent noise is somewhat higher. Since the noise levels are greatest in level-flight, and since this phase commonly constitutes a large proportion of flight time, the level-flight spectra were used to generate the flight-deck threshold curves for these two aircraft.

Since the spectra are fairly smooth the threshold curve was calculated using the procedure described in Section 1.2.3. Threshold values were calculated at multiples of 0.1 kHz and the K value was reduced to 0.5 in this case because the warnings are essentially continuous. The threshold curves follow the noise spectrum quite closely but they do diverge from the noise spectra gradually as frequency and BW_{ER} increase.

It was argued in section 1.1 that the appropriate level for a warning is 15 to 25 dB above masked threshold. This region is shown by the shaded areas above the threshold curves in Figs. 1.1 and 1.2. The boundaries of the region have been approximated by straight lines for convenience. The shaded area, then, shows the region in which the major frequency components of the warning sounds should fall.

1.3.2 The warning horn and firebell on the Boeing 727

The two loudest warnings on the flight-deck of a 727 are the takeoff and undercarriage warnings, which are intermittent, and continuous horns. The spectra of the two warnings are shown in Fig. 1.3 as interrupted and solid vertical lines for the intermittent and continuous horns respectively: they show that the sound is composed of a set of harmonics of a fundamental close to 0.6 kHz. The warnings are produced by the same physical horn and that is why their spectra are similar. These horns are very loud indeed: the three primary components of the continuous horn in the region 2.4 to 3.7 kHz are 105, 104, and 106 dB SPL, fully 10, 16, and 23 dB above the <u>maximum</u> of the appropriate-level range. The corresponding components of the intermittent horn are 99, 108, and 106 dB SPL, or 3, 20, and 23 dB, over the maximum of the appropriate-level range. To put it another way, these warnings have about 100 times the acoustic power required to exceed the <u>maximum</u> of the appropriate range. Thus it would appear that the device could be attenuated by about 20 dB without reducing its effectiveness. The importance of this 20–dB reduction is that it would bring the warning levels down from the 105–dB range where they are very aversive, to the 85–dB range where they would be much more acceptable. At first glance this might appear to bring the 2.4–kHz component of the intermittent horn rather close to the minimum of the acceptable range; however, the minimum line provides an overly conservative estimate in this case because the intermittent horn is a take-off warning and, the noise level on take-off is considerably below the noise level in level-flight. And it was the latter that was used to establish the minimum of the acceptable range.

The vertical lines in Fig. 1.4 show the spectral components of the firebell on the 727. The overall level of this warning is much more appropriate than that of the horn, the component at 4.1 kHz being only 9 dB above the appropriate range for level-flight noise. It should probably not be turned down because in the climb when full throttle is being used the sound of the engines introduces a broad hump in the 4.0 kHz region of the spectrum which would elevate the appropriate range by about 8 dB in this region for the duration of the full power condition.

1.3.3 The warning horn and firebell on the BAC 1–11

The two loudest warnings on the flight-deck of the BAC 1–11 are the take-off and undercarriage warnings. As on the Boeing 727 they are intermittent and continuous horns produced by the same physical source. Their spectra are shown by the interrupted and solid vertical lines in Fig. 1.5. Both of the warnings are very loud, over 100 dB SPL. The intermittent horn is slightly louder and, as before, the components of interest are in the region 2.5 to 4.5 kHz. The absolute intensity of the horn is lower on this aircraft; however, the appropriate-level range is also a little lower because the background noise is lower on this flight-deck. The largest component in the continuous-horn spectrum is 102 dB SPL at 3.3 kHz, fully 15 dB above the maximum of the appropriate range. The two largest components of the intermittent horn are 105 and 96 dB SPL at 3.2 and 3.7 kHz respectively, and they are 17 and 16 dB above the maximum of the appropriate range. Thus it would appear that this warning could be attenuated 15 dB. The background noise in level-flight is considerably higher than in the climb or on approach and so these warnings will still sound too loud in these phases. But they should probably not be turned down more than 15 dB because they could be required in level-flight or descent. As with the Boeing 727, however, it is this initial reduction of 15 dB that is most important because it would bring the warnings down from over 100 dB, where they are very aversive, to under 90 dB where they will be much more tolerable.

The spectrum of the firebell is shown by the vertical lines in Fig. 1.6; the most prominent component at 3.8 kHz is just under 100 dB SPL. It is fully 20 dB above the maximum of the appropriate range. The level-flight spectrum is the highest throughout the frequency range on this aircraft – the engine noise is not particularly noticeable in the climb. Thus it might be possible to turn the firebell down a full 20 dB on this aircraft. However, since the detectability of this warning rests very heavily on one single component it might be better to turn the warning down 15 rather than 20 dB.

2. THE TEMPORAL CHARACTERISTICS OF FLIGHT-DECK WARNINGS

Once the levels of the louder flight-deck warnings are reduced, the most obvious problem with the existing warnings is their temporal characteristics. The transients of the warning sounds are too abrupt; the temporal patterns are too similar; and the ratio of on-time to silent interval is far too high. The prototype warning described in the introduction offers an economical means of alleviating several of these temporal problems; one carefully tailored pulse of sound is used to build a warning with a distinctive temporal pattern and a low on/off ratio. In this section, guidelines for the temporal characteristics of the pulse, and the combination of pulses into a distinctive pattern, are outlined.,

Whereas it is feasible to modify the level of existing flight-deck warnings without replacing the sounders, any improvement in the temporal characteristics of the sounds will require the introduction of micro-computers and thus a completely new system. The technology exists, however, and is currently being used to support navigation on the flight-deck. It will undoubtedly form the basis of any new warning system, and so micro-electronic systems are assumed to be available in the discussion that follows.

2.1 The Basic Sound Pulse

The fine structure of the waveform that defines a particular sound is not heard, in the sense that the rapid fluctuation in air pressure is not perceived as a rise and fall in loudness. Rather it is the envelope of the waveform that carries what is thought of as the temporal information of a sound, and it is the parameters of the envelope that are the main topic of this subsection – chiefly the rise and fall times, and the overall duration of the basic sound pulse.

2.1.1 The rise and fall times of the pulse

Some of the existing flight-deck warnings go from off to full on at a level over 100 dB SPL in under one one-hundredth of a second (10 ms). There are several factors that prompt the use of fairly fast rise times. However, the discussion that follows shows that onsets like those currently used are never justified and rise times in the range 20 to 30 ms would be preferable.

In the natural environment a rapid rise to a high sound level is characteristic of a catastrophic event in the listener's immediate surroundings. The natural response to such an event is an involuntary reflex in which the muscles are tensed in preparation for a blow or a quick response. But quick motor responses are not particularly useful on the flight-deck because large civil aircraft cannot change configuration quickly. Furthermore, since instantaneous responses often prove incorrect, they are specifically discouraged on the flight-deck and in pilot training. Thus the abrupt onsets of current warnings are not justified by a requirement for fast motor responses. Changes in sound level are useful for drawing a listener's attention, and the greater the rate of change the more demanding the sound. But the change does not need to cause a startle reaction in order to capture attention promptly and surely, and so abrupt onsets are not justified on this basis either.

In order to minimise the on-time of the complete warning, it is important to keep the duration of the basic pulse, and thus the duration of the rise time, as brief as possible. In addition, if sound pulses are fairly short and not too close together, they do not necessarily disrupt the communication they interrupt because the brain and the auditory system together can use the redundancy of speech and language to fill in the parts of the communication where the pulses occur. But, as before, neither of these requirements necessitates the use of sharp onsets.

When the onset of a sound rises faster than 10 dB/ms, the sound seems to come on instantaneously, and if the final level is over 100 dB SPL there is a significant chance of its startling an unwary listener. When the onset rises slower than 1 dB/ms the listener can actually hear the sound level rising, and if the final level is under 90 dB SPL, the chance of a startle reaction is minimal. In spectral terms, when the level of a sound is changed, each line component in the short-term power spectrum broadens momentarily. When a sound is gated abruptly, the dispersed energy dominates the perception and it is heard as a transient, varying from a gentle click to a loud bang, depending primarily on the overall energy.

The steady-state level of auditory warnings has to be fairly high and so the rate of rise in the region above threshold should not be greater than about 1 dB/ms. The final level recommended in the previous section was 20 to 30 dB above threshold, which leads to a suggested rise time of 20 to 30 ms for that portion of the ramp where the sound level exceeds threshold. This rate of rise is ample to command attention, and a duration of 30 ms does not add an undue amount to the total pulse duration.

The duration of the offset of the pulse is determined by the same factors that defined the onset duration and in the same way. If the onset ends more than about 100 ms before the offset begins, it could probably be a little shorter than the onset without increasing the risk of a startle response. However, shortening the offset a little would not reduce the total pulse duration significantly, and it is often convenient to let the offset function be the mirror image of the onset function.

Thus, as a guideline, the optimum duration for onsets and offsets is 20 – 30 ms in the region above threshold.

2.1.2 The form of the onset and offset

The shape of the gating function is not particularly important. The onset function should be concave down, or at worst linear, and the system should be critically damped so that there is no overshoot of the steady-state level. But within these broad constraints the details are not essential because they tend to manifest themselves as changes in the side lobes of the spectral-smearing function, and on the flight-deck all but the main peaks in the spectrum are masked by the background noise (provided the rise time and fall times are greater than 20 ms).

An example of a good gating function for the flight-deck is the quarter-sine function with a frequency of about 10 Hz. The sine wave rises from zero to full range over 25 ms, and since it has a continuously decreasing slope the rate of rise is lower at the higher levels. The second quarter of the same function provides a good offset function. It has the opposite curvature to the offset imposed by most physical systems, but as with the quarter-sine onset, it has the advantage of using lower rates of change at the higher levels.

In a digital system it is quite easy to generate and apply these gating functions to digitized sound samples to produce the desired envelope for the pulse.

2.1.3 Pulse duration

As a minimum the duration of the basic pulse should be at least twice the integration time of the auditory system so that the internal level of the sound is at its maximum long enough to ensure detection of the warning in the flight-deck noise. At the same time, the pulse should not be longer than necessary for three reasons: it increases the on-time of the warning; it reduces the intelligibility of the speech it interrupts; and it reduces the diversity of temporal patterns that can be used in the warning set.

The integration time of the auditory system is, unfortunately, not very well defined. It is clear that up to 50 ms detection of a tone in noise improves 3 dB per doubling of signal duration, and that beyond 300 ms increasing signal duration has little effect. But attempts to derive the integration time using an energy detection model, much as the bandwidth of the auditory filter was derived using a power-spectrum model, have been hindered by the listener's uncertainty concerning the onset time of the signal, and the problem of multiple observations when the signal is long with respect to the integration time. The more recent research leads to the conclusion that the integration time is under 50 ms, a conclusion in agreement with the observation that listeners can follow a ramp in the envelope of a sound when the rate of change is less than about 1 dB/ms. Thus we can expect a pulse duration of 100 ms or more to be sufficient to support reliable detection of warnings when the level of the warnings is calculated according to the power-spectrum model outlined in the previous section.

With regard to the maximum pulse duration, a series of experiments has been performed in which speech or a pulsating sinusoid was presented in conjunction with a pulsating masker. They show that if the masker pulses are 150 ms or less, and the gaps between the masker pulses are 150 ms or more, the auditory system is inclined to fill in the masker interruptions and the listener perceives a continuous signal. This suggests that the signal pulses should not be longer than 150 ms. Furthermore, the advantages of a low on/off ratio and the requirement for highly distinctive temporal patterns both argue for the use of short basic pulses.

Thus, as a guideline, the pulse duration should be 100–150 ms. A pulse with a steady-state portion 100 ms in duration, with rise and fall times 25 ms in duration, and with quarter-sine gating functions, would have an effective duration of about 125 ms, and be near the optimum in terms of its envelope parameters.

2.2 The Pattern of Pulses

The spacing of pulses in flight-deck warnings is important for three reasons:

1) The pattern of pulses can affect the probability of warning confusion.

2) Pulse spacing affects the level of interruption imposed by the warning.

3) Pulse spacing affects the perceived urgency of the warning sound.

The latter pair of factors are related and will be considered first since they follow directly from the previous discussion.

2.2.1 The effect of pulse spacing on disruption and urgency

It was noted in the previous section that when a warning is composed of short pulses and the pulses are spaced by more than 150 ms, the masking and disruption caused by the warning are considerably reduced. If the pulses are rounded and more than 300 ms apart, and if they are presented at a moderate level, the sound will not be perceived as urgent. This, then, is the type of pulse pattern appropriate for the attensons required to signal low-priority alerts and the arrival of information on flight-deck displays. The traditional hostess call and the information ding-dong used in airport terminals are both of this type.

As the pulse spacing decreases below 300 ms the sound becomes progressively more urgent and progressively more disruptive. The prototype warning presented in the introduction illustrates how the effects of pulse spacing may be used to produce ergonomic warning sounds. The warning begins with a set of four pulses spaced 50 ms apart; as a result, the warning interrupts communication for half a second, drawing the flight-crew’s attention, and conveying a sense of moderate urgency when played at a moderate level. The remaining pair of pulses are widely spaced and so cause virtually no disruption. After the voice warning, the pulse spacing of the warning sound is increased to avoid disruption and reduce the sense of urgency, so that when the warning is played at a low level it provides a background warning with the same basic character as the original. Finally, if the warning does not draw a response in a reasonable length of time, it is reiterated with no gaps between the pulses. When played at a high level this sound disrupts communication and is decidedly urgent.

The level of perceived urgency depends on the spectral characteristics of the sound pulse as well as pulse spacing, and so the precise gap values have to be established empirically for each new warning. The principles illustrated by the prototype warning are, however, generally applicable, and the specific values in the example provide a guideline for starting values.

2.2.2 Pulse patterns and distinctiveness

Although we now describe the auditory warnings on the flight-deck as a warning system, in point of fact these ‘systems’ are sets of warnings that have grown to their present size, a warning at a time, over a number of years. It is clear that when a new warning was added to an existing set, care was taken to ensure that the new sound was different from the other members of the set. But it seems that this judgement was based largely on the spectral content of the sounds, for although there are a wide variety of spectra, there is little variety in pulse pattern or pulse envelope. Most of the warnings have either no temporal pattern, that is they are continuous, or they have a simple alternation pattern with a repetition rate of 2 – 5 Hz.

When the total number of warnings was small there was no real need to use temporal dimensions to ensure that the warnings were distinctive. But as the number of warnings has grown, the lack of variety in temporal patterns has become more important. One of the pilot surveys concerning warning systems (Ref. 1) revealed that the warnings are not only intrusive but also confusing. It was difficult to determine from the survey data whether the pilots meant that the warnings cause confusion on the flight-deck by disrupting thought and communication, or that the warnings are themselves confusing, or both. As a result, a study was performed to determine whether flight-deck warnings are intrinsically difficult to learn and remember (Ref. 7). Groups of naive listeners were taught to recognise a set of ten auditory warnings drawn from the flight-decks of a variety of current civil aircraft. The results show that the first four to six warnings are acquired quickly; thereafter, the rate of acquisition slows markedly but the listeners do continue to learn and all but one listener acquired the complete set in under an hour. A summary of the experiment is presented in Appendix B, and the learning and retention data are discussed in Section 4 as they pertain to the question of how many warning sounds should be used on the flight-deck.

The relevance of the study for the temporal characteristics of warning sounds derives from the confusion analysis that the authors applied to the errors made during learning. The listeners’ responses were pooled and a table was made showing the distribution of responses made to each warning sound. Correct responses occur along the negative diagonal of such a table, and concentrations of responses off the negative diagonal (recurring errors) point to the important confusions. The significance of the confusions can be assessed statistically, providing the responses are scaled to reflect the listeners', response biases. The analysis is summarised in Appendix B. There were five significant confusions; they are marked by vertical lines in Table B1 which presents the warning sounds with a breakdown of their temporal and spectral characteristics. A brief listening test, in which members of a confusable pair were played alternately, immediately revealed that warnings with the same pulse-repetition-rate were likely to be confused even when there were gross spectral differences between the warning sounds. For example, two warnings with pulse durations of 260 and 300 ms and a common pulse-repetition-rate of 2.5 Hz were reliably confused, despite the fact that one had a spectrum dominated by two, fixed, line components and the other had a multi-harmonic spectrum with a continuously rising fundamental frequency (a burst of a siren). It is important to stress that the listeners were naive and that their rate of confusion is very high with respect to the rate that might be expected to occur on the flight-deck. The results do, however, indicate that any potential for confusion would be reduced by employing a richer variety of temporal patterns.

The experiment was designed, in part, to establish a method whereby an existing or proposed set of warnings could be tested for potential confusions. The method has been applied to the warning set proposed for a new aircraft and it appears to be quite sensitive. It revealed that the temporal pattern of two pairs of warnings should be made more distinct, and as before, there were no obvious spectral confusions.

2.2.3 Pulse spacing and perceived urgency

The procedure for identifying confusions does not, in and of itself, specify what the patterns of pulses should be. However, the data indicate that the temporal patterns of current warning sets could be improved by ensuring that the pulse rate and priority of warnings do not conflict. A burst of pulses with varying amplitude and a short inter-burst interval sounds more urgent than a continuous sound with the same average level. The horn that is often used as a configuration warning on the flight-deck has no temporal pattern and currently depends on its abrupt onset and excessive level to convey urgency. It would be better to use a burst of pulses with a high pulse rate but a lower overall level and longer rise times; it would carry the same sense of urgency but in a much more acceptable form. The waveform inside the basic pulse could be taken from the existing horn sound and in this way the pairing of horn sound and configuration fault could be preserved.

On flight-decks where there are a relatively large number of warnings, the conflict between pulse rate and urgency is often further compounded by the presence of lower priority alerts, like passenger evacuation and selective call, which have been assigned sounds with higher pulse rates, and thus a greater sense of urgency, than the immediate-action warnings. The most recent warning systems are electronic and it is possible that the pulse spacing of these warnings could be increased without replacing the entire system. In the next generation of warning systems, the lower priority alerts will not have their own individual warning sounds. Rather they will be grouped according to priority and the entire group will be signalled by an attenson. Here again the pulse spacing should reflect the priority.

3. THE SPECTRAL CHARACTERISTICS OF FLIGHT-DECK WARNINGS

The spectral characteristics of existing flight-deck warnings are quite good by comparison with the existing levels, temporal characteristics, and ergonomics. There is still considerable room for improvement, however, and as the warning level is reduced to make the system more acceptable, it will become increasingly important to optimise the spectral characteristics of the warning sounds. In this section guidelines are developed to specify

(a) the appropriate spectral region for warning-sound components,

(b) the appropriate number and spacing of warning components, and

(c) the relative amplitude of the components.

The warning horn and firebell of the Boeing 727 are then reviewed in the light of these guidelines.

3.1 Frequency Limits for Auditory Warning Components

Although the hearing of young normal listeners ranges from about 16 Hz to 16 kHz, the lower and upper portions of the range should not be used for presenting auditory warnings on the flight-deck because they are less dependable as communication channels than the mid-range, and sound power below the optimum frequency region can actually reduce the effectiveness of components in the optimum region.

3.1.1 The lower bound for warning components

The absolute threshold of hearing is about 60 dB at 60 Hz and it falls about 15 dB per octave up to 0.5 kHz. There is no appreciable loss of low-frequency hearing with age; the expected loss at 0.5 kHz and below is less than 10 dB up to about 70 years of age (Ref. 8). Thus, it would, at first, appear that the lower limit for components of auditory warnings could be set as low as 0.125 kHz, and some existing warnings do contain power in this region. In practice, however, several factors combine to reduce the utility of low-frequency components for communication purposes.

On the flight-deck, noise components below 0.5 kHz typically have as much or more power than components in higher frequency regions. For example, the spectra for the flight-deck noise on the BAC 1–11 (Fig. 1.1) show either that level decreases monotonically with frequency (the climb and descent spectra), or that the noise level is uniform up to a cutoff frequency beyond which it falls as frequency rises (the level-flight spectrum). The climb and descent phases of flight sound much quieter than the level-flight phase because the levels are lower for climb and descent throughout most of the frequency range. But in the region below 0.5 kHz they are essentially as high as in level-flight. Thus low-frequency warning components have to have relatively more power than higher components for simple acoustic reasons, and the levels cannot be reduced for warnings that are only relevant during the take-off and climb phases of flight.

The efficiency of the auditory system deteriorates as frequency decreases below 0.5 kHz; that is, the K in the general masking equation (Eq. 1.1) increases as frequency decreases. The width of the auditory filter continues to decrease with centre frequency down to about 0.125 kHz but the reduction in the noise power passed by the filter is more than offset by the loss of efficiency at low frequencies. Thus, low-frequency warning components have to have relatively more power for auditory reasons as well.

The absolute level of the low-frequency noise means that low-frequency warning components have to be in excess of 85 dB SPL to reach the minimum of the appropriate-level range, and this creates two problems. First, high-intensity, low-frequency sounds cause a disproportionate amount of masking at higher frequencies; in other words the lower skirt of the auditory filter broadens at high intensities. Second, high-level components put a disproportionate burden on the sound production system and the intercom system because they increase the probability of clipping the signal. This type of distortion gives rise to harmonics that fall in the optimum listening region and so change the character of the warning sound, typically making it harsher. If the warnings were always played at the same level, so producing the same distortion, the problem might not be too serious. But the level is bound to vary with time and across aircraft, and this would lead to warnings whose character, and in particular whose perceived urgency, would vary with aircraft.

As a guideline, then, the lower limit for the spectral components of auditory warnings is 0.5 kHz on civilian jet aircraft.

There is one further caution with respect to the lower limit for the frequency of warning components. The flight-deck noise of propeller driven aircraft and helicopters has proportionately more low-frequency energy and the absolute level of the components is much higher than that on jet airliners. As a result, on propeller driven aircraft and helicopters, warning components below 1.0 kHz are essentially useless. Thus any attempt to standardise warning systems across aircraft types would require raising the lower bound for components to around 1.0 kHz.

3.1.2 The upper bound for warning components

For a young, normal listener the absolute threshold of hearing stays below 10 dB SPL all the way up to 10 kHz and does not rise to 60 dB until beyond 15 kHz (Ref. 9). And in the region above about 4.0 kHz, the level of flight-deck noise is so low that if it prevailed throughout the spectrum there would be no need to raise one's voice at all on the flight-deck. Unfortunately, this does not mean that there is a vast high-frequency region available for auditory-warning components.

High-frequency hearing deteriorates with age; at 10 kHz, 25% of the male population will have a loss of about 55 dB at age 55 (Ref. 10). The phenomenon is referred to as presbyacusis and it exhibits considerable variability; 25% of the population will have 20 dB or less hearing loss at age 55. The absolute level of the hearing loss is not actually as much of a problem for warning systems as the variability across listeners. Since the spectrum of flight-deck noise is relatively low at high frequencies, there is sufficient space between the maximum of the appropriate range and levels that produce severe annoyance (100 dB) to boost the warning components and so compensate for the presbyacusis of the older pilots. But the solution would be specific to one level of hearing loss and the warnings tailored for pilots with this loss would still sound overly loud to pilots with normal hearing. The effects of presbyacusis are most noticeable above 5.0 kHz, and so the dominant components of auditory warnings should be below 5.0 kHz.

Prolonged exposure to noise produces an accelerated loss of hearing. The natural resonance of the mechanical part of the auditory system causes the peak of this 'noise-induced' hearing loss to appear in the region of 4 – 6 kHz. Factory and farm workers often exhibit losses of 30 – 40 dB between 4 and 6 kHz well before 50 years of age (Ref. 8). It is difficult to predict the losses that might be expected for commercial airline pilots because of the diversity of training. The older pilots are more likely to have flown with the armed forces in noisy aircraft, and it seems likely that a portion of them will have at least mild, noise-induced losses. The younger pilots will show fewer and smaller losses, but if they have a significant amount of experience in propeller aircraft they too may exhibit mild noise-induced losses. This type of loss has only a marginal effect on speech comprehension and so it does not affect the medical status of the pilot. As with presbyacusis, however, the problem is not so much the absolute level of the hearing loss as the increase in the variability of hearing in the pilot population. Given a specific noise-induced loss, the spectra of the warning sounds could be tailored to accommodate the loss; but the warnings would then be inappropriate for pilots with normal hearing. Thus, it is preferable to keep the dominant components of auditory warnings below 4.0 kHz and so avoid the complications that might otherwise be introduced by noise-induced hearing losses.

There are two technical limitations on the use of high-frequency warning components. First, the frequency response of existing intercom systems and headsets falls off fairly rapidly in the region above 4.0 kHz and there is a requirement for the warnings to be presented on the headsets as well as the flight-deck speaker. Although there is no technical difficulty in producing intercom systems and headsets with a greater frequency range, it would be costly and the potential advantage of warnings with high-frequency components would not appear to justify the expense. Second, warning components much in excess of 4.0 kHz would also increase the cost of the micro-computers envisaged for future warning systems. A digital system with a frequency limit of 4 – 5 kHz is capable of providing fairly good quality digitized speech. This frequency limit requires the system to have a sampling rate of about 10 kHz, and this rate will probably becoming an industry standard for practical, as opposed to high-fidelity, digital speech systems. If the warning system is not compatible with digital speech systems it will be considerably more expensive, which again argues for an upper bound on warning components in the region of 4.0 kHz.

Thus, auditory, acoustic and cost considerations suggest that the prominent components in auditory warnings should not fall outside the frequency region 0.5 to 5.0 kHz, and the majority of the prominent components should occur in the region 1.0 to 4.0 kHz.

3.2 The Pitch

The common view of pitch perception (the place theory) is based on the power spectrum; stated in its simplest form 'We hear a pitch if, and only if, the spectrum of the stimulus contains a peak at the corresponding frequency'. In reality, this theory only works for stimuli with one, or at most two, components. The spectra of the important pitch-producing stimuli (the vowels of speech, the notes of music, and the tones of auditory warnings) show that they are comprised of sets of many harmonically-related components, and for these stimuli we hear a 'residue' pitch corresponding to the fundamental of the harmonic series. In this subsection, the basic phenomenon of residue pitch is introduced, and then the properties of the process are reviewed as they pertain to the pitch of auditory warnings.

3.2.1 Residue pitch

Fig. 3.la presents a schematic representation of a 12-component residue-producing stimulus; it <u>does not</u> produce pitch sensations corresponding to the individual sinusoidal components at 1200, 1400, 1600, ... 3400 Hz; it <u>does</u> produce a strong 200 Hz pitch even though there is no power in the spectrum below 1200 Hz (Ref. 11). Furthermore, the phenomenon is not the result of non-linear distortion in the periphery of the auditory system. The 200 Hz pitch dominates the perception even at the lowest stimulus levels and it cannot be masked by noise in the region of 200 Hz. And when the components are frequency shifted <u>en masse</u> without altering the spacing between the components, the residue pitch shifts. For example, if the components are shifted up 60 Hz, as in Fig. 3.lb, the pitch shifts about 10 Hz. Since the component spacing has not changed, the pitch would not change if it were based on a difference- frequency distortion component. The pitch shift of the residue shows that the mechanism by which pitch is extracted from the stimulus is complex, and the models developed recently to explain residue pitch reflect this complexity (Ref. 12, 13).

Fortunately the details of the process are not essential for the design or evaluation of auditory warning systems. What is important, however, is the realisation that the place theory, or the power-spectrum model, of pitch perception is totally inadequate for auditory warnings, and that residue pitch is the 'normal' mode of pitch perception and not just a laboratory phenomenon. Two examples will serve to make this argument better than a lengthy review of the research.

Consider the pitch of speech and the transmission of that pitch by the telephone. The pitch of speech is carried by the vowels of speech and the spectra of vowels reveal them to be sets of 20 – 40 harmonics of a low-frequency fundamental – on the order of 100 Hz for men and 200 Hz for women. The fundamental is rarely the strongest harmonic and it is virtually certain that the pitch of speech is residue pitch. The frequency response of the telephone cuts off at about 300 Hz and so speech received via the telephone has no low-frequency energy. Thus, if the pitch of speech were not carried by the higher harmonics, the telephone in its present form would not work.

Consider also the pitch of music and the transmission of melody via a small radio. The spectra of musical notes show that they too are sets of harmonics. The low harmonics have proportionately more energy than the low harmonics of vowels, but as with vowels the fundamental of musical notes is often not the strongest component, particularly for notes below ‘middle C’. And when the lower brass and stringed instruments play in the lower half of their range, the fundamental is commonly absent. Thus the pitch of music is also residue pitch. ‘Middle C’, the middle of the range of musical notes, has a frequency of 261.6 Hz. A pocket radio with a small speaker transmits very little energy in the region below 300 Hz, but one has no difficulty in hearing a melody played by a trombone or sung by a baritone on such a radio. Clearly, it is the harmonics, which fall in the mid-range of hearing and are transmitted by the small speaker, that carry the pitch of music.

3.2.2 The minimum number of components for auditory warnings

Theoretically, one spectral component is sufficient to carry the information that a warning is intended to convey. There are three related reasons, however, to indicate that it is preferable to have four or more prominent components; that is, components in the appropriate-level range between 1.0 and 4.0 kHz.

The first and most obvious reason is that it is more difficult to mask a multi-component sound. Whereas the occurrence of an extraneous sound on the flight-deck might mask a single component warning, it is much less likely to mask a sound comprised of four or more components. Provided all of the components come on and go off together, the auditory system will perceive them as a group and assign them to one perceptual source.

If the warning has only a few components, the masking of one or two can markedly alter the character of the sound and thus reduce the probability of recognising the warning. As the number of components increases, the effect of masking one or two naturally decreases; indeed, if the sound has six or more components, and they are harmonically related, the effect of masking one or two is surprisingly small. Thus the second reason for multi-component sounds is that they maintain their character better under varying conditions of masking, so increasing the probability of recognising the warning. The third reason is simple logic; the more components in the warning sound, the greater the scope for making the sounds distinctive. Single-component sounds (sinusoids) all have the same sound quality or timbre; they differ only in pitch and it is difficult to remember pitch with any accuracy.

3.2.3 The regularity of component spacing

In general, if the members of a multi-component spectrum are harmonically related the sound has a precise pitch and a well defined character; if the components are inharmonic the sound has multiple pitches, or an ambiguous pitch, and a diffuse sound character. Although there is no ostensible research, since it is easier to associate harmonic sounds with imagined sources, it seems likely that they are easier to learn and remember than inharmonic sounds. And since harmonic sounds are more cohesive, they are probably more resistant to disruption by extraneous maskers. As a guideline, then, there is a preference for harmonic, or near harmonic, sounds for auditory warnings.

Strictly harmonic, multi-component stimuli sound like musical notes or chords. The lower components are better resolved by the auditory system, that is, the number of harmonics per auditory filter is lower. As a result, they contribute more to the quality of the sound. If the first five harmonics have a significant proportion of the energy, the note will sound ‘smooth’, ‘sonorous’, and ‘full’. As the energy shifts to higher harmonics the note sounds ‘sharper’ and it has more ‘edge’. As a guideline, then, warning-sound quality and warning priority will be better matched if the low priority warnings, like attensons, have relatively more energy in the first five harmonics, and high-priority warnings, like immediate-action warnings, have relatively more energy in harmonics 6 – 10. If a multi-component stimulus is essentially harmonic, but incorporates a small number of inharmonic components, it still sounds like a musical note, but with a harsher or shriller timbre. The auditory system is quite sensitive to this type of irregularity and it offers an effective means of making warning sounds distinctive. As a guideline, the incorporation of quasi-harmonics should probably be reserved for the higher priority warning sounds.

3.2.4 The appropriate range for the pitch of warning sounds

A warning sound must contain at least four harmonics if it is to produce a reliable residue pitch rather than several high pitches corresponding to individual component frequencies. Furthermore, four of the first ten harmonics should be prominent to ensure a strong pitch and to provide for sufficient distinctiveness among the members of the warning set. Sets of exclusively high harmonics produce a thin, buzzy sound with a relatively weak pitch. To avoid masking, the four prominent components should fall in the appropriate frequency region (1.0 – 4.0 kHz).

This implies that the upper bound for the <u>pitch</u> of warning sounds is 1000 Hz; this is the highest fundamental for which four harmonics (numbers 1 – 4) can be generated in the region below the upper bound for component frequencies (4.0 kHz). Similarly it implies that the lower bound for the <u>pitch</u> of warning sounds is 143 Hz; this is the lowest fundamental for which four components (numbers 7 – 10) can be generated with the lowest frequency (7 x 143 Hz) above the lower bound of the appropriate range for component frequencies (1.0 kHz).

As a guideline, then, the appropriate range for the pitch of warning sounds is 150 to 1000 Hz.

3.3 The Relative Amplitude of Warning Components

3.3.1 Relative amplitude and pitch

The pitch of a <u>harmonic</u> multi-component sound is essentially insensitive to the relative, amplitude of the components. The pitch associated with a quasi-harmonic, multi-component sound, like the set of shifted harmonics in Fig. 3.lb, is slightly altered by the introduction or removal of components. But there is very little effect of component amplitude for components more than 15 dB above threshold, confirming the basic stability of residue pitch.

The pitch of <u>inharmonic</u> multi-component sounds does depend on the relative amplitude of the components. When the set of components is not fused into a residue pitch, one hears pitches corresponding to one or more of the individual components and the more intense components tend to dominate the sound. Thus when a change in relative amplitude changes the dominance pattern it changes the pitch.

There is considerable variation in the level and shape of the spectrum of flight-deck noise as an aircraft proceeds through its succession of phases of flight, and changes in flight-deck noise alter the signal-to-noise ratio of warning components. Some of the flight-deck warnings, like the firebell, are appropriate to more than one phase of flight, and so the insensitivity of harmonic and quasi-harmonic sounds to component amplitude is a particular advantage for these multi-phase warnings.

3.3.2 Relative amplitude and timbre

The timbre of a harmonic sound is affected by the relative amplitude of the components; in general, as the distribution of energy shifts towards the higher harmonics the sound becomes less musical and more buzzy, as if from a smaller, or more distant, version of the same source. But it seems unlikely that timbre changes would reduce the probability of recognising an auditory warning significantly. This type of change must be fairly common in the natural environment, and it is possible that the auditory system uses its knowledge of the existing noise conditions to renormalise the components internally and thus minimise the effects of the background.

3.3.3 Relative amplitude and warning-sound consistency

In order to minimise the probability of warning confusion, auditory warnings ought to sound the same throughout all phases of flight, including the preflight check on the ground. The guidelines set out in this section are designed to promote this kind of consistency of perception.

The warning horn and the firebell on the Boeing 727 provide convenient examples of warnings that do and do not have the balance of factors required to support consistency of perception. Both sounds are good in that they are quasi-harmonic, multi-component stimuli with fundamentals in the central position of the appropriate-pitch range, and they have numerous low-order harmonics in the optimum range for component frequencies. The relative amplitude of the components is far from optimum, however, and in the case of the firebell it effects the consistency of the sound.

All of the harmonics of the horn are well above threshold, even in the noisiest phase of flight (Fig. 1.3), and so the warning horn always has the same character and pitch. Unfortunately, this consistency is the direct result of the warning's excessive level. If the warning level is reduced to bring the prominent components into the appropriate-level range, and so make the warning less aversive, the relative amplitude of the components between 1.0 and 2.0 kHz should be increased.

The situation with the firebell is much less acceptable: During the preflight check and on takeoff, all of the firebell components are well above threshold (Fig. 1.4), and so the firebell will have its normal rich character if it occurs. But when the noise level rises during climb, many of the components would fall to just over threshold, and in level-flight the majority of the components would be well below threshold, leaving only two inharmonic components to carry the information. Thus the sound of the firebell becomes thinner and its pitch probably changes during the course of flight. The remedy, of course, is to increase the relative amplitude of the mid-range components. The bell-like character of the sound will be preserved so long as the high inharmonic components are in the appropriate-level range, and the irregular temporal envelope is maintained.

Similar analyses apply to the horn and firebell on the BAC 1-11 (Figs. 1.5, 1.6).

3.4 Frequency and Amplitude Modulation in Warning Sounds

It was recommended in Section 2.2 that warning sounds should be bursts of brief pulses to minimise the disruption and masking that they might cause on the flight-deck. This largely eliminates the possibility of using slow amplitude and/or frequency modulation to make warning sounds more distinctive; that is, it eliminates effects like tremolo and vibrato.

It is still possible, however, to incorporate a uni-directional frequency glide into a 100-ms pulse. A rapid sweep of the fundamental of a set of harmonics provides an effective means of drawing attention. It imparts a sense of urgency and so should be reserved for high-priority warning sounds. Note that it is important to ensure with rapid frequency sweeps that four or more mid-range components remain in the appropriate-level range and the appropriate-frequency range for the full 100 ms.

4 ERGONOMICS

In a sense, the ergonomics of auditory warnings on the flight-deck is the topic of this entire document, and many of the ergonomic issues have already been discussed. The method developed in Section 1 to predict the appropriate-level for warning sounds attempts to specify not only the minimal acoustic power required to make the warnings effective, but also the ergonomic maximum beyond which they will be aversive. In Section 2, the attempt to minimise the potential for startle by rounding the pulses and starting at moderate levels, and the attempt to minimise the disruption the warnings cause by minimising the total on-time, have the additional ergonomic aim of producing warnings that the flight crew do not immediately cancel. The concern for matching perceived urgency and warning priority, in both Sections 2 and 3, reflects an ergonomic attempt to make the warnings state their priority level in their sound character, so that it is immediately obvious and does not require further cognitive processing. The trade-off between voice-warning length and warning priority is reviewed in the next and final section. This section, then, is confined to two remaining ergonomic issues: manual and automatic volume control, and the maximum number of auditory warnings on the flight-deck.

4.1 Manual and Automatic Volume Control

It has been suggested by flight crew and others, that either the crew should be given a volume control knob to set their own warning levels, or the system should include automatic volume control so that warnings are not presented at their full volume during the quieter phases of flight.

The problem with the first suggestion, manual volume control, is that the crew would probably want to set the volume during the preflight check on the ground when the flight-deck noise is minimal, and this would often leave the warning level below the minimum of the appropriate range for level-flight. The upper limit of the appropriate range is primarily a matter of annoyance and it is conceivable that an occasional crew might actually prefer a slightly higher level. This suggests that the optimum solution might be a volume control with a minimum that still leaves the warnings loud enough for the level-flight condition, and a maximum as much as 10 dB above the maximum of the appropriate-level range. On the other hand, it seems likely that the desire for a volume control stemmed primarily from a desire to attenuate existing warnings, and that the next generation of warnings with lower overall levels and automatic attenuation beyond the voice warning component would render a volume control knob largely superfluous.

Automatic volume-control systems, for presenting sound in vehicles rather than recording sound in studios, have never achieved the level of acceptance that might have been expected. Efforts to develop a measure of vehicle noise that predicts preferred listening level have not been particularly successful. Part of the problem is the enormous intensity range of hearing and the fact that listeners prefer lower signal-to-noise ratios at higher intensities. Since the appropriate range is limited in both frequency and level on the flight-deck, it is possible that an automatic volume control operating on the noise level in the octave about 2.0 kHz, and with a limited range of 10 to 15 dB, would be successful in this environment. And since the warning levels appropriate for level-flight will leave the prominent components close to the annoyance level in most other phases of flight, some limited, automatic volume control might improve the acceptability of the warning system sufficiently to make it worthwhile.

4.2 The Maximum Number of Auditory Warnings on the Flight-Deck

The fact that some pilots flying aircraft with many auditory warnings made even a small number of confusions when asked to identify tape recordings of their warnings (Ref. 1) suggests that the number of auditory warnings on the flight-deck should be limited to a value well below the 13 used on some current aircraft. The surveys presented in Refs 2 and 3 show that the pilots are of this opinion, and ARINC Characteristic 726 recommends a maximum of four individual warning sounds for immediate-action warnings, plus two attensons, one for the immediate-awareness warnings and one for the advisory alerts.

The investigation of the learning and retention of warning sounds discussed in Section 2 and Appendix B, shows that naive listeners learn about seven warnings quickly, but thereafter the rate of learning slows considerably. Over the course of an hour, all but one of the 20 listeners learned all ten of the warning sounds in the test set, and a week later a few minutes' practice brought them back to perfect performance. Thus there is no inherent difficulty in learning warning sounds, but beyond the first seven it does require appreciably more effort. And it seems reasonable to assume that larger warning sets would also require more maintenance training to keep performance at a high level over a number of years.

The results of the learning and retention study cannot be directly applied to the situation on the flight-deck. They do, however, reinforce the growing belief that aircraft with sets of more than 10 warnings have too many. On the other hand, the ease with which even naive listeners learn up to seven arbitrary warnings suggests that the limit of four immediate-action warnings in ARINC Characteristic 726 is probably overly conservative. A set of up to six immediate-action warnings plus two attensons should prove entirely reliable if

(a) the warning sounds have distinctive temporal as well as spectral patterns,

(b) the perceived urgency of the warnings matches their priority, and

(c) the warning sounds are reinforced by key-word voice warnings with good speech quality.

5. VOICE WARNINGS ON THE FLIGHT-DECK

Speech is the original high-speed communication system. The messages are transmitted via a versatile, redundant code, language, and the information-bearing units (words) require no new learning. The technology associated with speech synthesisers and digital-speech systems is developing rapidly and air-worthy systems with limited vocabularies and acceptable speech quality ought to be available soon. It should, therefore, be possible to incorporate the advantages offered by voice-warning and verbal-message systems into the next generation of flight-deck warning systems. In the first part of this section, the advantages and disadvantages of speech as a warning are briefly discussed in relation to warning priority. In the second section, the acoustic characteristics of speech are introduced and compared with the appropriate-level range and the appropriate-frequency range on the flight-deck.

5.1 Speech and Warning Priority

In ARINC Characteristic 726 (Ref. 14), the speech system is envisaged as having two roles: The top-priority warnings, which require the flight-crew's immediate action, would incorporate a brief voice warning to add redundancy to the warning sounds; an example of this role occurs in the prototype warning described in Section 0.3 (Fig. 0.2). The second-priority warnings, which signal an abnormal condition and require the crew's immediate awareness, would be verbal messages, and the entire set would be announced by one specific sound played before the verbal message to draw the crew's attention — an 'attenson'.

The reason for the two separate roles follows from a brief consideration of speech as a flight-deck warning. The advantages of speech are that it is versatile and reliable. The disadvantages are:

(a) There is already a lot of speech on the flight-deck, including synthetic speech; thus a verbal warning might go unattended for some short period because of a lack of perceptual contrast.

(b) It is difficult to communicate in the presence of a recorded message.

(c) Speech occupies the entire auditory communication channel, and it does so for a relatively long period of time.

The disadvantages are most noticeable in the case of the immediate-action warnings where time and the minimisation of disruption are important. Since the number of immediate-action warnings is small, the versatility that speech offers is not a particular advantage in this case. Thus, the verbal component in these warnings should probably be restricted to brief, key-word messages to increase reliability, and the message should probably not be repeated in the background version of the warning.

The disadvantages of speech are somewhat less important in the case of abnormal-condition warnings where time and disruption are a little less critical. And since the number of alternatives is larger, the versatility of speech is a more important advantage in this case. Thus it would probably be best to use full-format messages and the natural redundancy of language (Ref. 15) in this case, and to repeat the complete warning after a suitable pause.

5.2 The Sounds of Speech and Their Generation

Phonemes are the basic acoustic elements of speech. We use about 40 phonemes in English and all the words, phrases, and sentences we speak are combinations of these 40 sounds. The phonemes fall into three basic groups, the vowels, the voiced consonants, and the unvoiced consonants. The vowels often have 1000 times the energy of the unvoiced consonants, and so speech requires a fairly large dynamic range. The noise on the flight-deck limits the available dynamic range, and so, broadly speaking, the problem of tailoring speech for flight-deck warnings is one of minimising its dynamic range and adjusting its overall level to position as much of the energy as possible in the appropriate-level range.

5.2.1 Speech production

Speech sounds are perhaps best understood in terms of their production. Speech is a highly specialised version of controlled breathing which occurs on the exhalation portion of the breathing cycle. The diaphragm pushes upwards and in so doing pushes air from the lungs through the throat and out the mouth or nose. If the muscles are relaxed and the vocal tract is open the result is simply breathing. Speech sounds occur when some portion of the vocal tract is constricted and air is forced through the constriction producing turbulence and the rapid changes in air pressure that we perceive as sound. The method of constriction distinguishes the three classes of speech sounds.

5.2.2 Vowels

In the throat there is a structure composed of cartilage, muscles, and ligaments known as the larynx, or voice box. If the ligaments, or vocal cords, are pulled tight by the muscles and air is forced through the throat by the diaphragm, the vocal cords vibrate to produce a ‘voiced’ sound. The air is broken into a stream of regular, short, puffs that occur at the rate of 75-100 pulses per second for men and about 150-200 pulses per second for women. The spectrum of the pulse train is a set of harmonics of the basic pulse rate; the amplitude of the harmonics falls slowly with frequency above about the fifth. The cavities of the throat, mouth and nose operate together to produce resonances that amplify or attenuate the individual harmonics, producing local maxima in the spectrum that are referred to as formants. There are typically three or four formants spread across the region 0.5 to 5.0 kHz.

If the lips and teeth are open and the tongue is out of the way so that the only constriction in the vocal tract is the vocal cords, the result is a vowel. The specific vowel is determined by the position of the formants relative to each other and relative to the fundamental frequency. The entire set of components is perceived as a unit with a residue pitch corresponding to the fundamental of the harmonic series. As noted in Section 3.2.1, there may be little or no energy at the fundamental itself.

5.2.3 Unvoiced consonants

If the vocal cords are open but the flow of air through the vocal tract is constricted by some combination of the articulators (the lips, teeth, and tongue), the result is an unvoiced consonant. The articulators 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 resonances of the vocal tract and so it is not a white noise; but since the source is not a regular vibration, the sound is a noise rather than a set of harmonics. Examples of unvoiced consonants are the sounds symbolised by s, sh, th, and h when used at the start of a word. In the case of unvoiced consonants the important formants are the third and fourth, which are fairly broad and not particularly well defined. The unvoiced consonants have far less energy than vowels in that part of the spectrum up to and including the second formant (about 2.0 kHz). In the region above 2.0 kHz the unvoiced consonants have roughly the same power as the vowels but since many of the unvoiced consonants have short durations (like the sound of p in pot) their total energy is often remarkably small.

5.2.4 Voiced consonants

Constriction of the vocal cords and the articulators 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; in this case the vocal cords are vibrating, the lips are closed and the air is forced out through the nose resulting in a nasal resonance. Another group of voiced consonants are the plosives associated with the letters b, d, and g; the identifying characteristic of these consonants is the transition from the wideband noise associated with the puff of air as the lips open, to the harmonic spectrum of the vowel that follows. The spectra of the voiced consonants are highly variable, ranging from the vowel-like spectra associated with m and n to the initially broadband spectra associated with the plosives, but basically they fall in between the vowels and consonants. Similarly, the power of the voiced consonants is greater than that of the unvoiced consonants but less than that of the vowels.

5.3 The Spectral Distribution of Speech Power and Speech Intelligibility

Although speech power occurs at frequencies from 50 Hz to 12 kHz, the distribution of speech power is far from uniform. About 80% of the energy in the long-term spectrum of speech occurs in the three octave bands about 0.5 kHz, the region of the first formant. The second formant, in the region of 1.5 kHz, contains most of the remaining energy with only a small proportion occurring in the third and fourth formants around 2.5 and 3.5 kHz respectively. The different regions of the spectrum do not, however, contribute to the comprehension of speech in direct proportion to their power. Whereas the majority of the power is below 1.0 kHz, it is the region 1.0 to 3.0 kHz that is most important for comprehension. In general terms, the reason that the -comprehension distribution peaks at a higher frequency than the power distribution, is because the consonants which appear at higher frequencies carry a far larger proportion of the speech information than they do of the speech power. Vowels often have 30 dB more energy than the consonants that surround them in a word, but the consonants are at least as important for comprehension.

5.3.1 Speech and the appropriate-frequency range on the flight-deck

In English, the voiced and unvoiced consonants v and f are distinguished from the voiced and unvoiced consonants z and s, primarily by the presence of more energy in the region above 5.0 kHz in the latter cases. Communication systems that cut off at 5.0 kHz or lower make these distinctions more difficult. But it is not considered a serious problem, and with these exceptions, the upper limit of the appropriate-frequency region for warning components is also appropriate for speech, in the sense that it does not reduce comprehension appreciably.

There is considerable speech energy below the lower bound of the appropriate-frequency region for warning components (0.5 kHz), and the region of the spectrum between 0.2 and 0.5 kHz does contribute to comprehension, particularly for vowels with a strong first formant. There is a substantial amount of flight-deck noise in the region below 0.5 kHz, which might at first suggest that speech components should be differentially amplified in the region below 0.5 kHz. There are two related reasons for not doing this, however. First, the information in the region below 0.5 kHz is duplicated in the region just above 0.5 kHz where the signal-to-noise ratio is better, so it is not essential to transmit the low-frequency energy for good comprehension. Second, since speech requires a fairly large dynamic range to begin with, and since flight-deck noise eliminates the use of low to moderate levels, the more intense speech has to be presented at rather high levels on the flight-deck already, and further amplification of the low-frequencies would be undesirable. Indeed, it would be preferable to reduce rather than enlarge the dynamic range of flight-deck speech, since loud, low-frequency components increase the probability of masking disproportionately. In addition, they increase the probability of clipping the speech and so introducing distortion components into the mid-frequency region of the spectrum.

5.3.2 Speech and the appropriate-level range on the flight-deck

The appropriate level for voice warnings and messages is established in the same general way as for warning sounds; the overall level is adjusted to position the prominent components of the average speech spectrum in the appropriate-level range as determined in Section 1.

In practice, since the majority of the speech energy is in the first two formants, and since the first formant is stronger than the second in the long-term spectrum, the adjustment usually involves positioning the peak of the first formant near the maximum of the appropriate range in order to include as many of the components of the second formant as possible in the appropriate-level range. The adjustment is most critical for the level-flight phase where the minimum of the appropriate range is often as high for the second formant as it is for the first; this is the case on the BAC 1–11 and the Boeing 727 (Figs 1.1 and 1.2). In the other phases of flight the spectrum of the flight-deck noise is typically lower in the region of the second formant and the adjustment has more tolerance.

Treating the components in the first and second formants as individuals and adjusting the overall level on this basis is not strictly correct; as frequency increases, two or three harmonics may fall within the passband of the auditory filter and thereby increase the detectability of the speech in that region by up to 5 decibels. However, since this occurs more in the region of the second formant, and since the second formant is the weaker in the average speech spectrum, it simply means that the problem of fitting the components of both the first and second formants into the appropriate-level range is not as critical as it might at first have appeared.

A similar analysis applies to the upper half of the speech spectrum. Since the power in the first two formants is 30 – 40 dB greater than that in the higher formants, it might at first seem advantageous to reduce the dynamic range of the voice warnings by differentially amplifying higher frequencies — a process referred to as whitening since it makes the spectrum more like that of a white noise. Indeed 9.0 dB of whitening across the region 0.5 to 4.0 kHz (3 dB/octave) would probably prove advantageous. More whitening is not necessary, however, for two reasons: The flight-deck noise is much lower (20 – 30 dB) in just that region of the spectrum (3.0 – 4.0 kHz) where the softest, unvoiced consonants occur. And the integration of signal energy by the auditory filter, which is ignored in determining the appropriate-level range, increases the signal-to-noise ratio for broadband signals, like unvoiced consonants even more than it does for the upper formants of vowels (up to 15 dB). Excessive whitening should be avoided as it reduces comprehension because it reduces the prominence of the formants and accentuates the noisy aspect of the consonants.

As a guideline, then, it is important to adjust the level of speech so that the peak of the first formant in the average speech spectrum falls near the maximum of the appropriate-level range. A small amount of whitening (3 dB/octave) is probably advantageous in the region above 0.5 kHz, but in the region below 0.5 kHz attenuation is preferable to amplification.

The conclusions of this and previous sections are presented, for the readers' convenience, as an appendix at the very end of the document (Appendix C).

ACKNOWLEDGEMENTS

The author would like to thank Mr John Rye and Mr Arthur Thorning of the Civil Aviation Authority for their continual support of this project and for their comments on an earlier draft of the report. The author is also indebted to Ms Nel Evers for her assistance in preparing the report.

REFERENCES

1. DuRoss, S.H. (1978) ‘Civil Aircraft Warning Systems – A survey of pilot opinion within British Airways,’ Royal Aircraft Establishment, Technical Report No. 78056.
2. Cooper, G.E. (1977) ‘A survey of the status and philosophies relating to cockpit warning systems,’ NASA Ames Contract Report No. NAS2-9117.
3. Lauber, J. (1976) ‘Airline warning systems operational data received in response to IATA Letter 4007-F,’ NASA Ames International Report, unpublished.
4. Potter, R.C., Fidell, S.A., Myles, M.M., and Keast, D.N. (1977) ‘Effectiveness of audible warning devices on emergency vehicles,’ Report No. DOT-TSC-77-38, U.S. Department of Transportation, Washington DC.
5. Wilkins, P.A. (1981) ‘Assessing the Effectiveness of Auditory Warnings,’ British Journal of Audiology, 15, 263–274.
6. Patterson, R.D. and Milroy, R. (1979) ‘Existing and Recommended Levels for Auditory Warnings on Civil Aircraft,’ Civil Aviation Authority Contract Report 7D/S/0142 R2.
7. Patterson, R.D. and Milroy, R. (1980) ‘Auditory Warnings on Civil Aircraft: the learning and retention of warnings,’ Civil Aviation Authority Contract Report 7D/S/0142 R3.
8. Green, D.M. (1976) An introduction to Hearing. New Jersey: Lawrence Erlbaum.
9. Fletcher, H. (1953) Speech and Hearing in Communication. New York: van Nostrand.
10. Robinson, D.W. and Sutton, G.L. (1979) ‘Age effect in hearing – A comparative analysis of published threshold data.’ Audiology 18, 320–334.
11. Patterson, R.D. (1973) ‘The effects of relative phase and the number of components on residue pitch.’ J. Acoust. Soc. Am. 53, 1565–1572.
12. Plomp, R. (1976) Aspects of tone sensation. London: Academic Press.
13. Moore, B.C.J. (1977) Introduction to the psychology of hearing. London: Macmillan Press.
14. ARINC Characteristic 726 (1979) ‘Flight warning computer systems,’ Aeronautical Radio, Annapolis, Maryland.
15. Simpson, C.A. and Hart, S.G. (1977) ‘Required attention for synthesized speech perception for two levels of linguistic redundancy,’ J. Acoust. Soc. Am. 61, S7.
16. Carpenter, A. (1971) ‘Voice Communications,’ in Human Factors for Designers of Naval Equipment. Medical Research Council, Royal Navy Personnel Research Committee.

FIGURES

0.1 Component patterns for an advanced auditory warning sound. The basic pattern is four regular-spaced pulses of sound followed by two irregularly spaced pulses. The different rows show varying levels of urgency.

0.2 The time course of a complete auditory warning. Each trapezium represents a burst of pulses, the rectangle represents a voice warning. The heights indicate the relative levels of the component sounds.

1.0 Calculation of the threshold imposed by a hypothetical flight-deck noise (solid line with dots) at multiples of 0.5 kHz. Auditory filters with centre frequencies of 1.0 and 4.0 kHz, and their rectangular equivalents, are shown below the noise curve. The resulting threshold values are indicated by the row of dots in the upper portion of the figure.

1.1 The range of appropriate levels for auditory warnings on the flight-deck of the BAC 1–11 (vertical-line shading). The minimum of the appropriate-level range is approximately 20 dB above auditory threshold (the broken line) which is calculated from the spectrum of the level-flight noise (the solid line) using a power-spectrum model of auditory masking (Ref. Al0). The faint dashed and dotted lines show the spectra of the flight-deck noise during steady climb and shallow descent.

1.2 The range of appropriate levels for auditory warnings on the flight-deck of the Boeing 727 (vertical-line shading). The minimum of the appropriate level range is approximately 20 dB above auditory threshold (the broken line) which is calculated from the spectrum of the level-flight noise (the solid line) using a power-spectrum mode1 of auditory masking (Ref. Al0). The faint dashed and dotted lines show the spectra of the flight-deck noise during steady climb and steep descent.

1.3 The principal spectral components of the take-off warning (interrupted vertical lines) and the undercarriage warning (solid vertical lines) on the flight-deck of the Boeing 727. The components are superimposed on Fig. 1.2 which presents the appropriate-level range, auditory threshold, and the noise backgrounds for the aircraft. The figure shows that the warnings are about 20 dB too loud.

1.4 The principal spectral components of the firebell (solid vertical lines) on the flight-deck of the Boeing 727. The components are superimposed on Fig. 1.2 which presents the appropriate-level range, auditory threshold, and the noise backgrounds for the aircraft. The figure shows that the firebell is roughly the correct level.

1.5 The principal spectral components of the take-off warning (interrupted vertical lines) and the undercarriage warning (solid vertical lines) on the flight-deck of the BAC 1–11. The components are superimposed on Fig. 1.1 which presents the appropriate-level range, auditory threshold, and the noise backgrounds for the aircraft. The figure shows that the warnings are about 15 dB too loud.

1.6 The principal components of the firebell (solid vertical lines) on the flight-deck of the BAC 1–11. The components are superimposed on Fig. 1.1 which presents the appropriate-level range, auditory threshold, and the noise backgrounds for the aircraft. The figure shows that the firebell is about 15 dB too loud.

3.1 Schematic representation of the power spectra of two, residue-producing, 12-component stimuli with frequency shifts of (a) 0 and (b) 60 Hz.

Figure 0.1

COMPONENT PATTERNS FOR AN ADVANCED AUDITORY WARNING SOUND

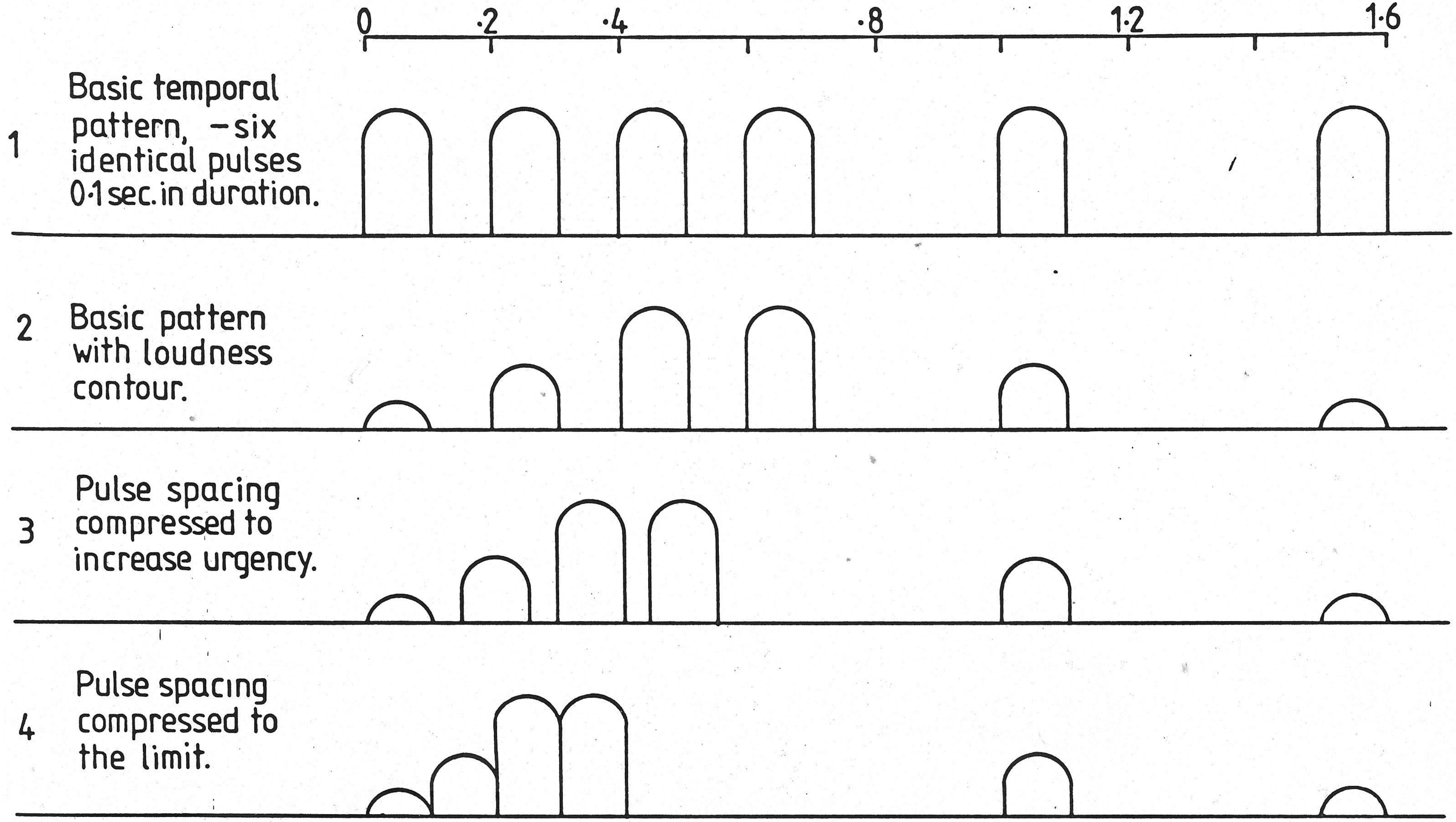


Figure 0.2

TIME COURSE OF COMPLETE WARNING

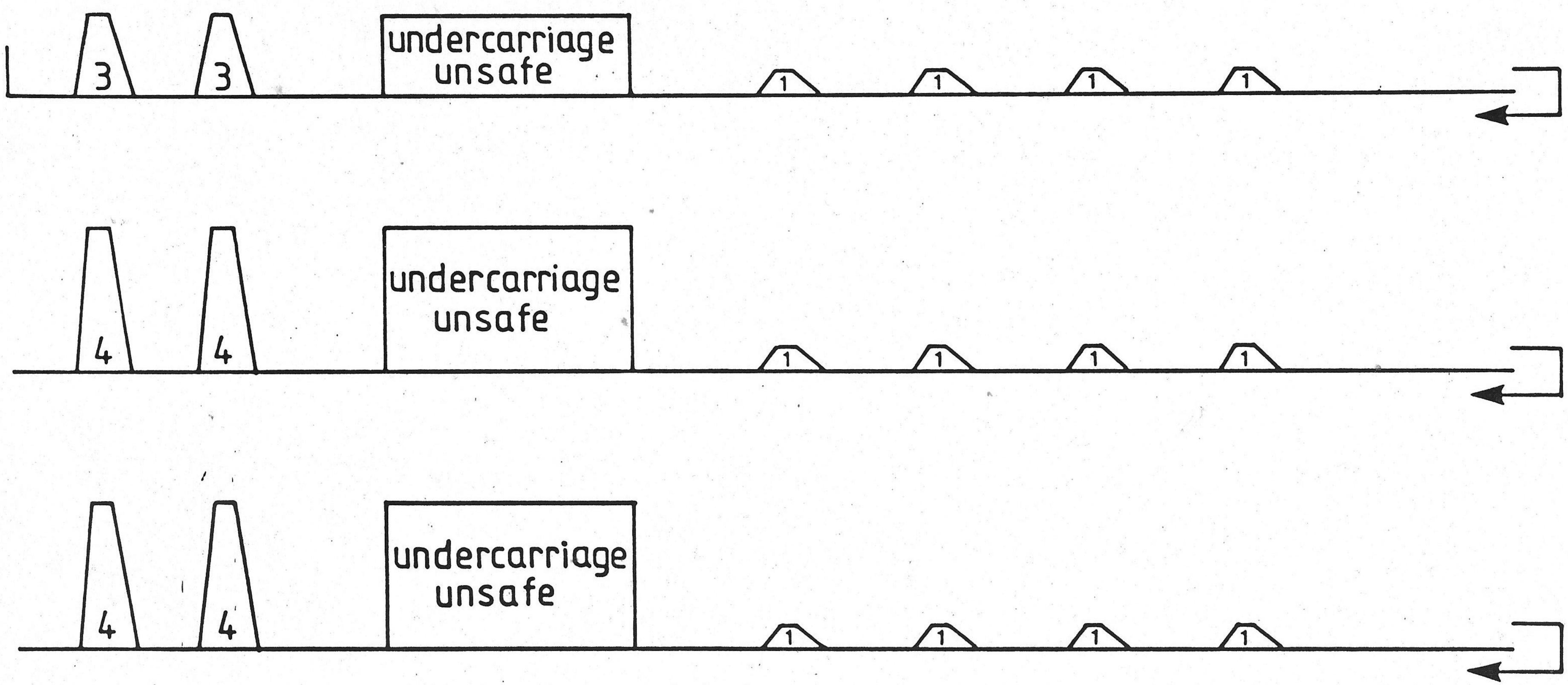
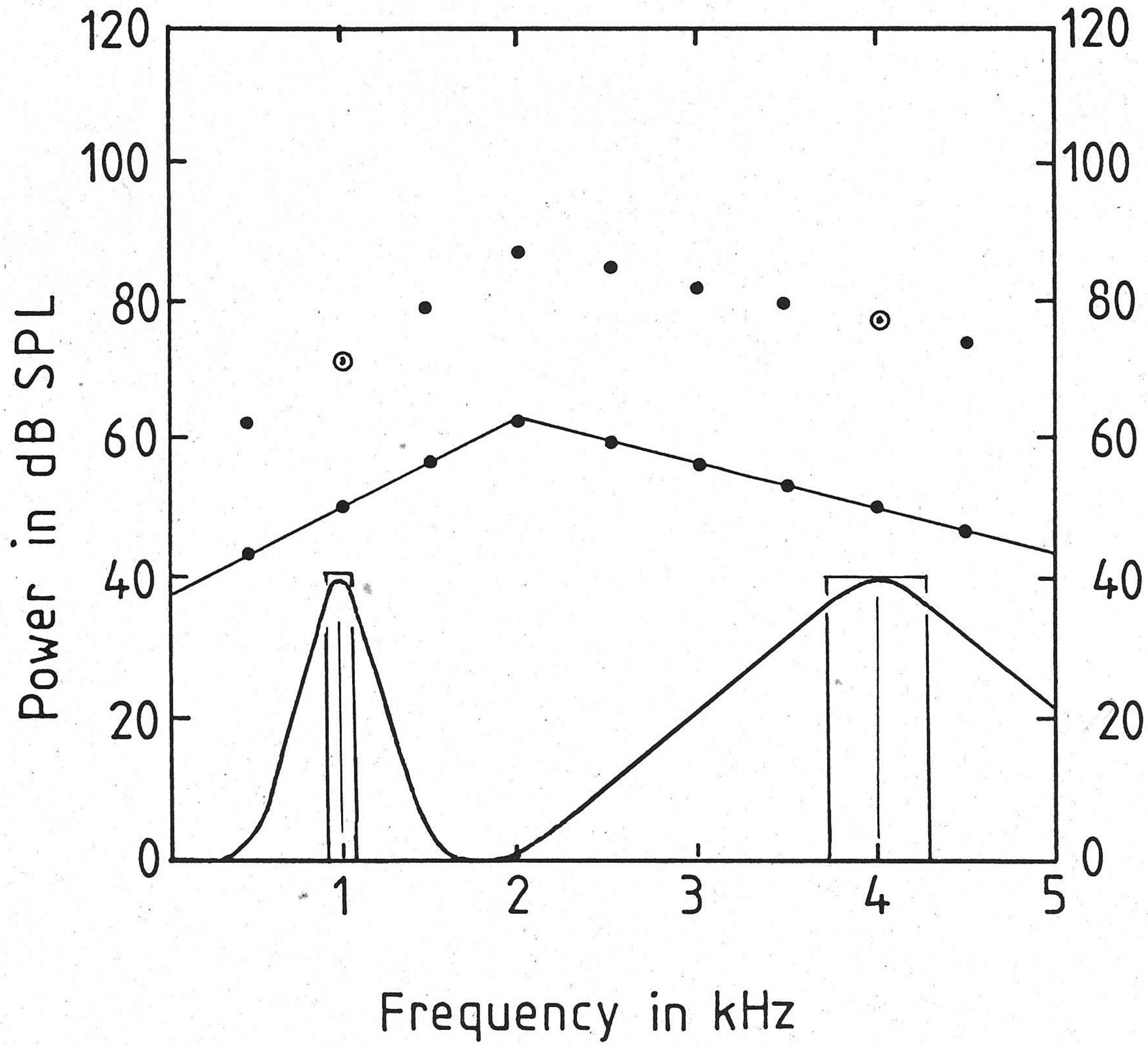
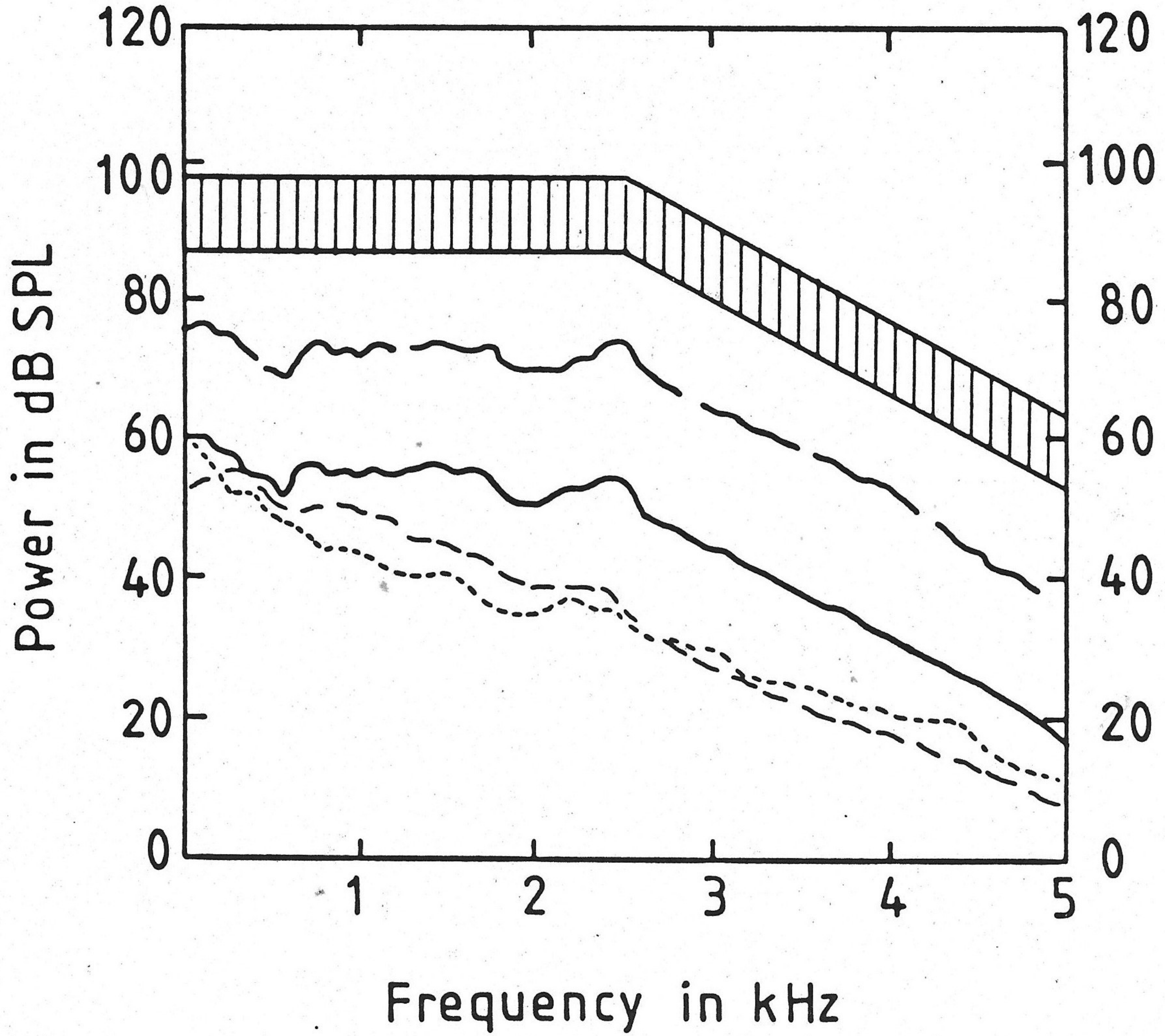


Figure 1.0



120
100
80
60
40
20
0
120
100
80
60
40
20
0
1
2
3
4
5
Power in dB SPL
Frequency in kHz

Figure 1.1



120
100
80
60
40
20
0
120
100
80
60
40
20
0
Power in dB SPL
1
2
3
4
5
Frequency in kHz

Figure 1.2

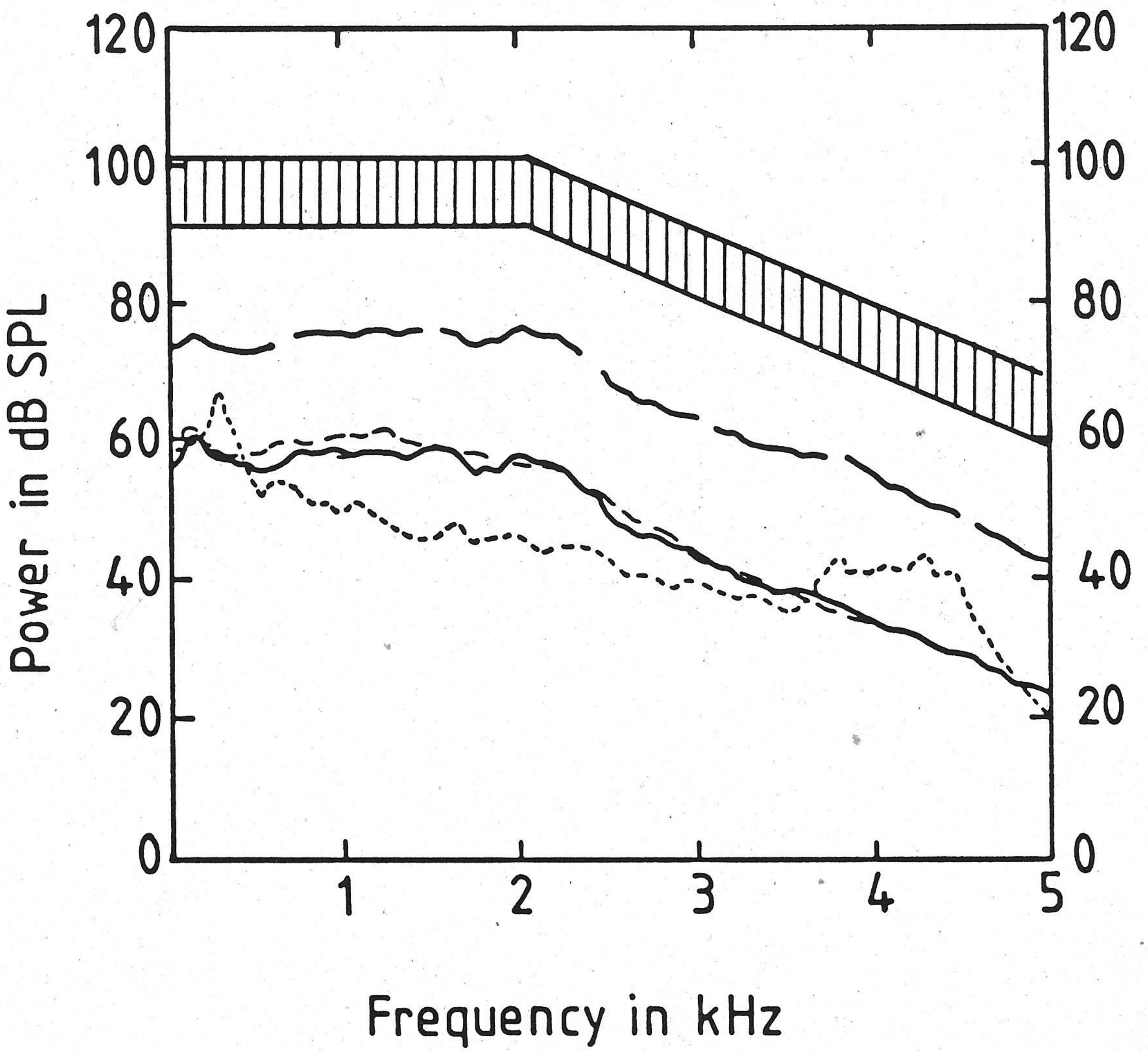


Figure 1.3

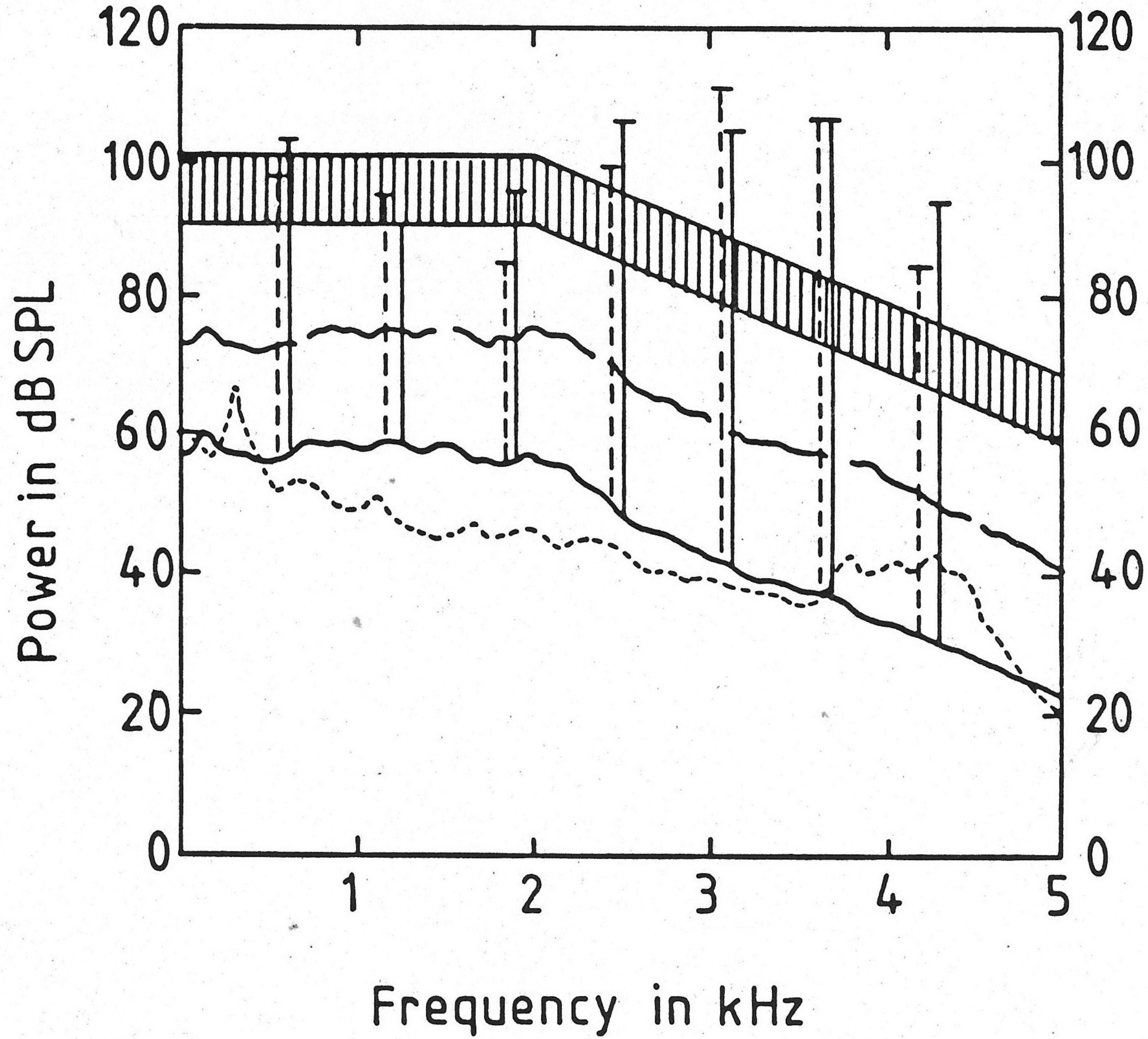
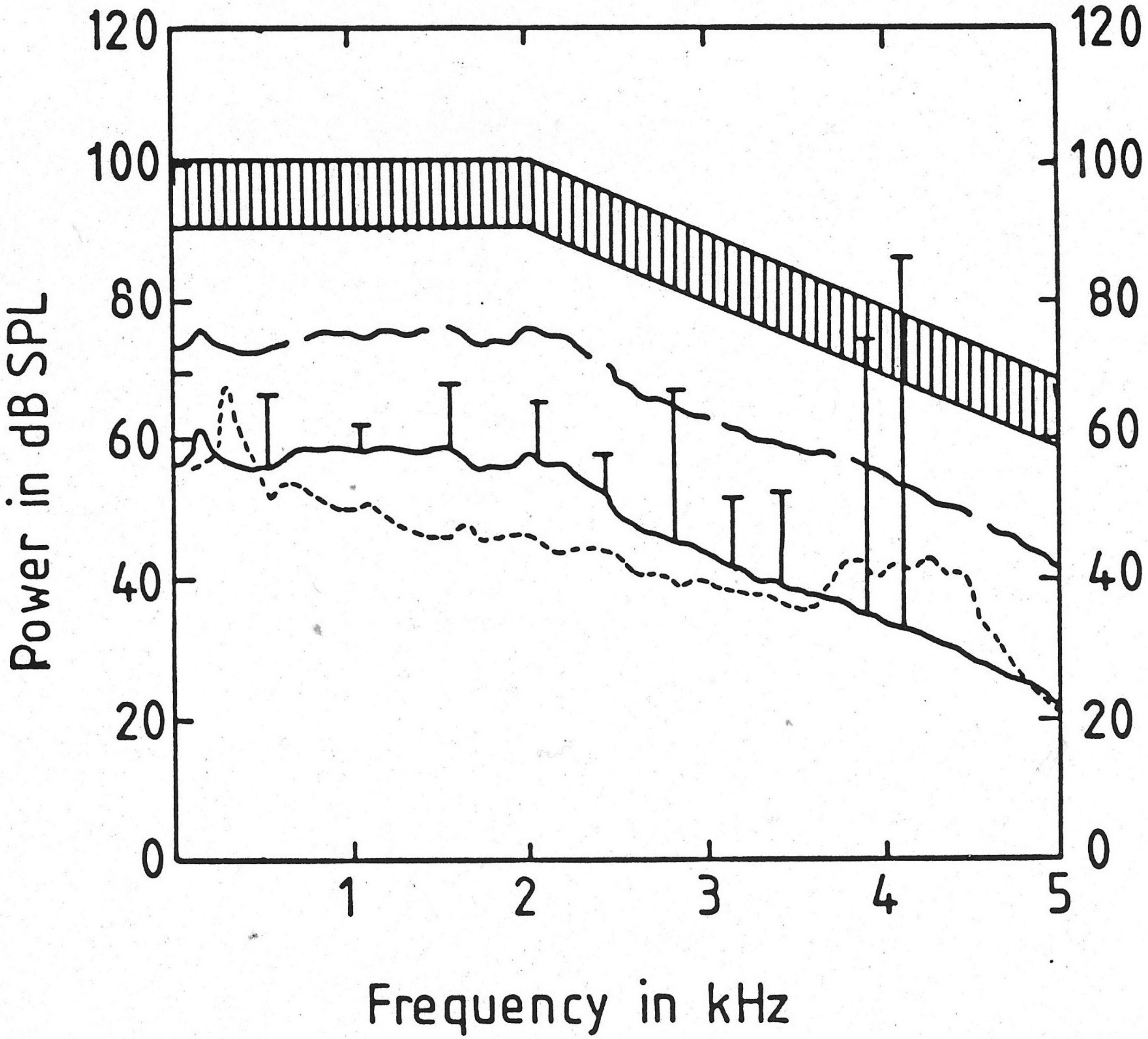
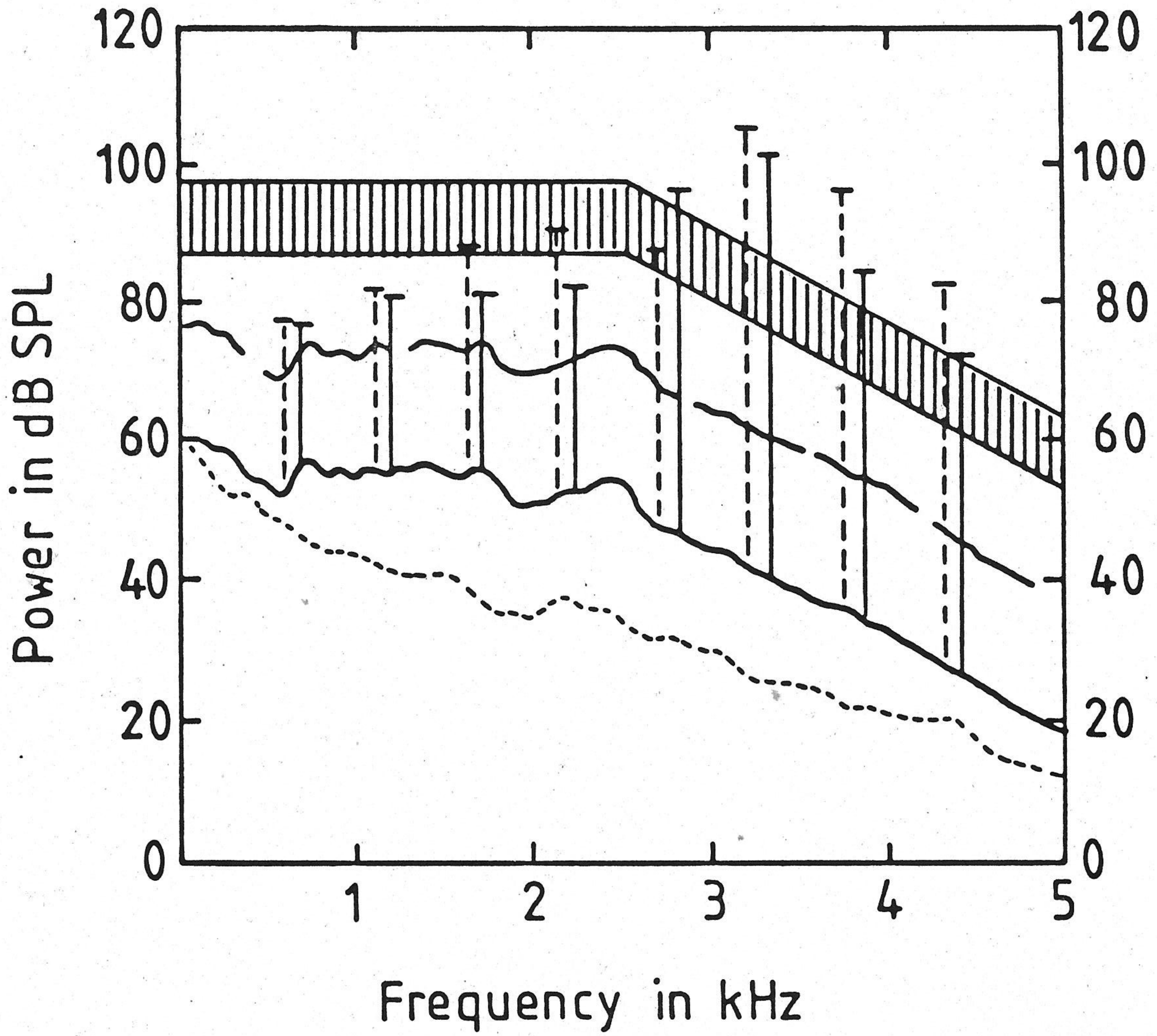


Figure 1.4



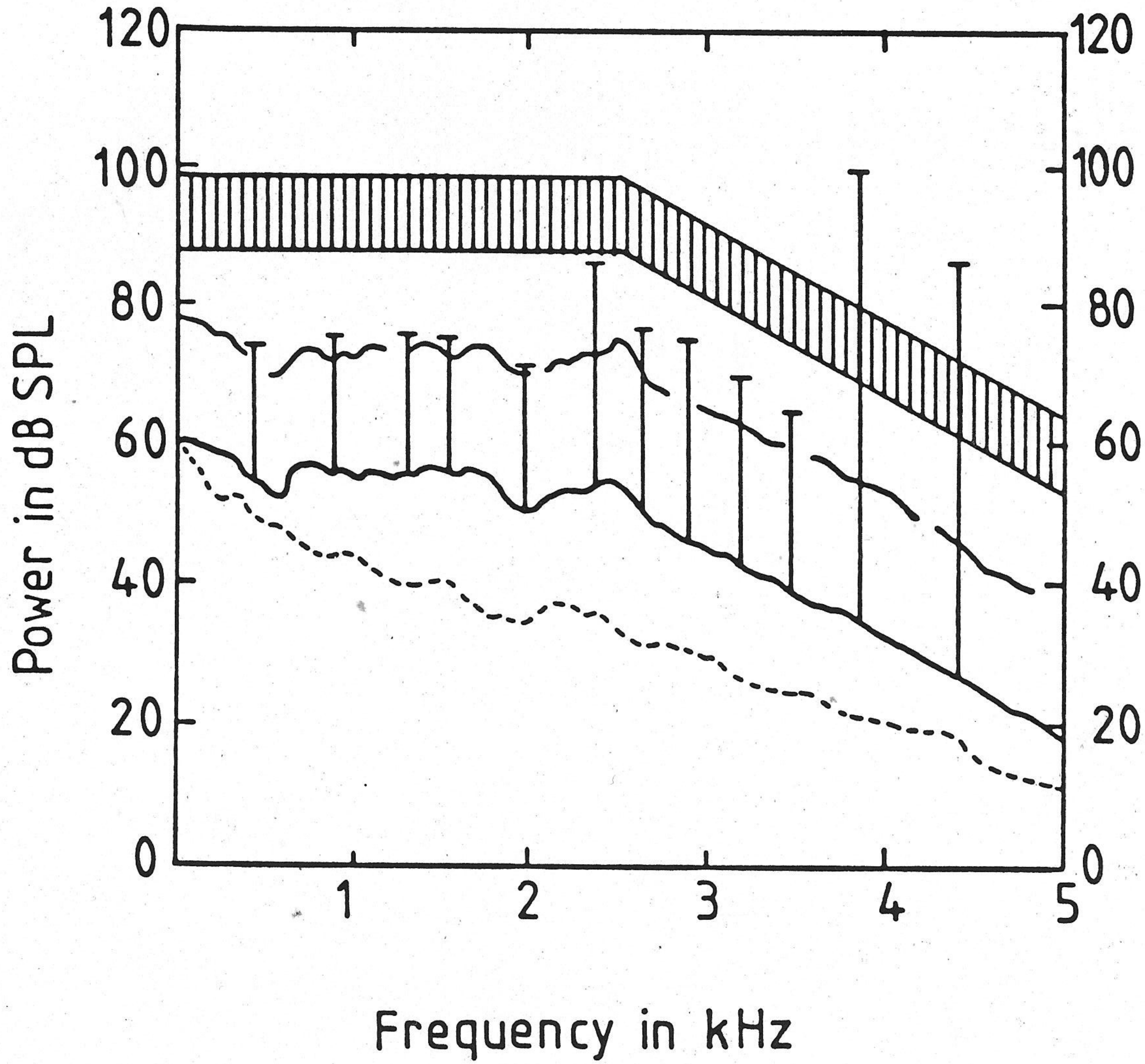
120
100
80
60
40
20
0
120
100
80
60
40
20
0
1
2
3
4
5
Power in dB SPL
Frequency in kHz

Figure 1.5



120
100
80
60
40
20
0
120
100
80
60
40
20
0
1
2
3
4
5
Power in dB SPL
Frequency in kHz

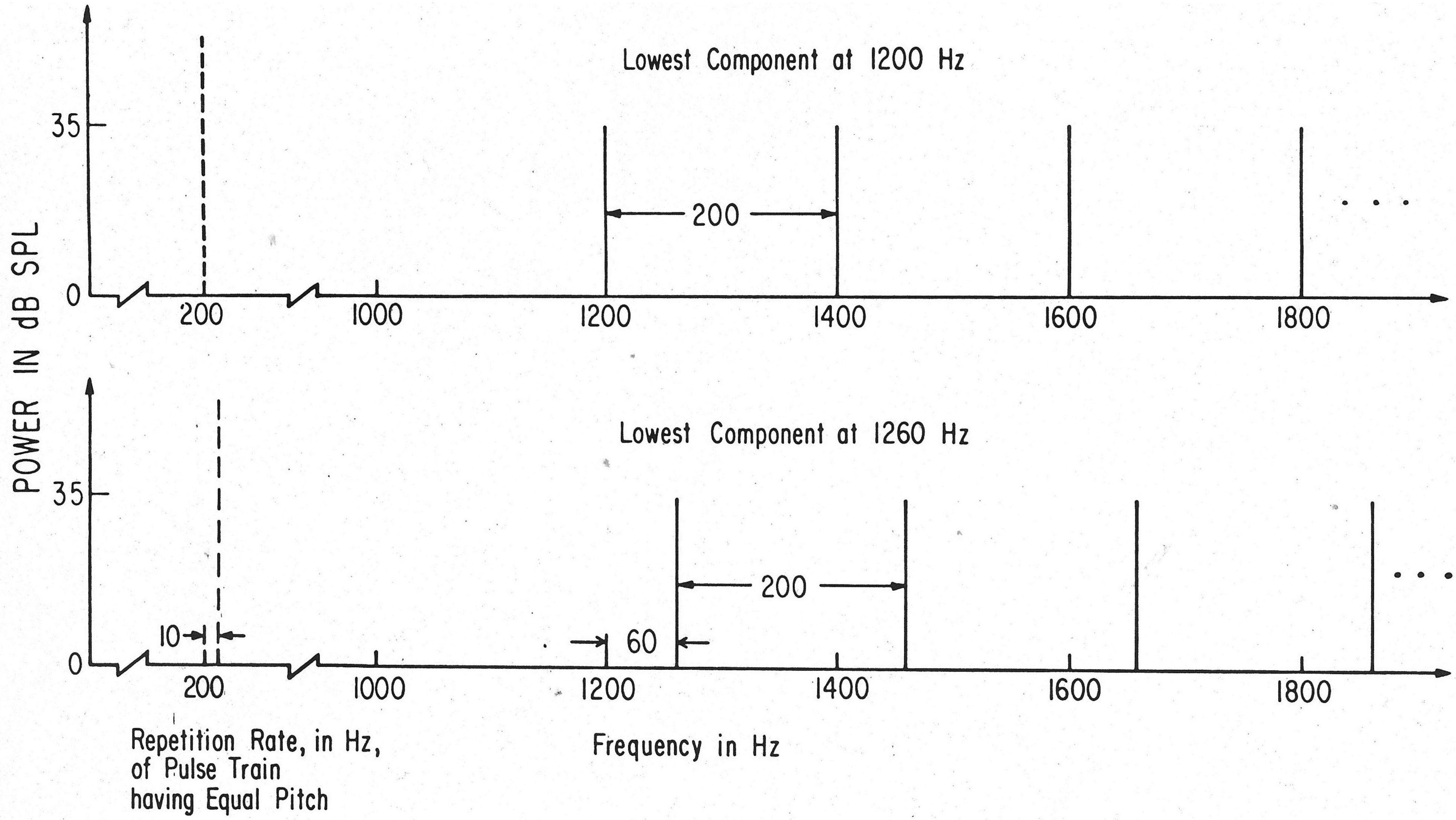
Figure 1.6



120
100
80
60
40
20
0
120
100
80
60
40
20
0
1
2
3
4
5
Power in dB SPL
Frequency in kHz

Figure 3.1

POWER SPECTRA OF RESIDUE PRODUCING STIMULI



APPENDIX A Determining the Appropriate Levels for Auditory Warnings

This appendix presents a three-step procedure for determining the appropriate sound level for flight-deck warnings. The spectrum of the background noise on the flight-deck is used in conjunction with a recent model of auditory masking to predict auditory threshold, as a function of frequency, on the flight-deck. Then, this general threshold curve is used to establish the range of appropriate levels for the spectral components of warning sounds. The spectrum of individual warnings can then be compared with the band of appropriate levels to determine which spectral components dominate the sound: the correct level for the warning is produced by adjusting the overall level so that these crucial components fall within the range of appropriate levels.

1. THE RANGE OF APPROPRIATE LEVELS FOR FLIGHT-DECK WARNINGS

In this section it is argued that a warning should be 15 dB above masked threshold to ensure that it will be noticed, and not more than 25 dB above threshold or it will be disruptive.

1.1 The Lower Limit — Threshold + 15 dB

In the auditory system, as in most physical systems, the detection of a signal presented over a noise background is a two-step process; the incoming stimulus is first subjected to a spectral analysis, and then a detection mechanism operates on the results of that analysis. The spectral analysis performed by the ear is outlined in the next subsection; the current topic is the definition of threshold and its role in determining the minimum of the appropriate range for flight-deck warnings.

The statistical nature of noise means that threshold for a signal in noise does not occur at one discrete signal-to-noise ratio; rather the probability of detecting the signal in an arbitrary sample of noise rises from chance to 100% over some range of signal levels. Typically, then, signal threshold is defined with respect to the probability of signal detection and the same applies to the auditory system.

The standard tools for measuring threshold in hearing are the ‘psychometric function’ in combination with the ‘two-interval, forced-choice’ procedure. The procedure and function are illustrated in the upper and lower sections of Fig. Al. The listener is presented with two sound samples one after the other (Fig. Ala); both have masking noise, but one of the samples, chosen at random, also has the signal. The listener's task is to report at the end of the trial, which interval had the signal, and in the event that he does not know, he must guess. The signal level is varied between the trials of the experiment, and over the course of many trials the percent correct associated with a range of signal levels is determined. The function relating percent correct to signal intensity is the psychometric function, as illustrated in Fig. A1b. Since chance in this experiment is 50%, this is the asymptote at which the function begins. The signal proceeds from inaudible to clearly audible over a range of some 20 dB. The central portion of the function is essentially linear, and typically, the function rises about 5% per decibel in this region. By convention, threshold is taken to be the signal level required to support 75% correct identification of the signal interval. When the signal is 10 dB above threshold it is easy to detect and by the time it is 15 dB above threshold it is difficult to miss. Thus an important signal like an auditory warning should be 15 dB above masked threshold. An excellent review of signal detection by humans is presented in Ref. Al.

The conclusion that the interference produced by a background noise is minimal once the signal is 15 dB above threshold applies to auditory processing quite generally. For example, the ability to discriminate a small frequency difference varies with signal-to-noise ratio near threshold. Just above threshold it is difficult to discriminate a relatively large frequency difference, but the difference limen, as it is termed, decreases rapidly as the ratio increases, and once the signal is 15 dB above threshold the frequency limen is essentially as small as it would be in the absence of noise (Ref. A2). Similarly, when a signal is just above masked threshold, its loudness is reduced by the presence of the noise. But as the level of the signal is raised, its loudness grows out of proportion to the increase in level, and once it is 15 dB above threshold the effect of the background is essentially gone (Ref. A3). The ability to localise a sound shows a corresponding effect; near threshold localisation is poor, but by 15 dB above masked threshold it has recovered.

Two vigilance experiments performed with sirens and machine sounds as signals confirmed that the psychometric function has the same form when environmental sounds are substituted for tones (Ref. A4, A5). The experiments also show that the basic shape of the function is not disturbed by the tracking tasks used to occupy the observers; as before, detection performance rises from chance to near perfect over a 20 dB range of signal levels, and once the signal exceeds threshold by 15 dB performance is indistinguishable from its asymptotic level.

1.2 The Upper Limit – Threshold + 25 dB

The upper limit for the appropriate range for warnings is not as well defined as the lower limit because it is difficult to convert the annoyance and interference caused by excessively loud warnings into a quantitative limit in decibels. However, the effects of annoyance and interference are real enough, and the lack of quantitative accuracy does not excuse the excessive levels of some existing warnings.

With regard to annoyance, the situation is fairly straightforward. Although the flight-deck noise of commercial, civilian, jet aircraft is low with respect to that of propeller driven aircraft or military jets, it is still sufficiently high to necessitate rather intense warnings. Sounds in the mid-frequency range that exceed 90 dB(A) are generally agreed to be annoying despite the variability of the measure, and their loudness will not be reduced by the background noise since they will be better than 15 dB above threshold. The total power of future warnings will not be far below 90 dB(A). Thus it seems likely that to avoid annoyance, the warnings should not be any louder than other factors require.

One factor that could dictate high levels is safety: it has been suggested that the louder a sound is, the better the chance that it will draw the attention of a person engrossed in an important task such as landing an airplane. Presumably it is this kind of argument that prompted the manufacturers to set some of the warnings to such high levels. Although the argument seems reasonable, the vigilance experiments with sirens and machinery noise (Ref. A4 and A5) show that there is a ceiling effect; once a sound is well above masked threshold it already has a very good chance of drawing your attention, and further increases in level do not improve detection performance measurably.

The inordinate reliability of supra-threshold signals was considered a paradox in signal detection theory until recently, when advances in the statistics of signal detection processes offered a solution to the paradox (Ref. A6). It is perhaps worth pointing out that threshold for a signal in a broadband noise background is amazingly consistent across listeners – for any particular signal and background combination the vast majority of listeners would fall within a 3-dB range.

The vigilance experiments provide one further insight. When observers are occupied in an engrossing task, and the rate of occurrence of signals is low, they will occasionally fail to recognise a signal no matter what its level. That is, asymptotic performance in these experiments is near but still below 100%. Thus, although flight-crew may fail to process the occasional loud warning correctly, it is not the case that increasing the sound power would overcome this problem. Once performance has reached its asymptotic level, further increases produce very little, if any, improvement. Whereas increasing warning levels from loud to extremely loud offers very little prospect for improving flight-deck safety, it does increase the probability of disrupting thought and communication, and thereby producing a safety hazard. Some of the existing warnings are actually loud enough to mask everything but a loud shout. It is not possible, at this point, to evaluate precisely the tradeoff between (a) having signals loud enough to ensure that they will interrupt an engrossed flight-crew, and (b) having them not so loud as to disrupt communication more than necessary. However, it seems likely that a cost-benefit analysis would reject at least the loudest of existing warnings because, whereas detection performance reaches asymptotic levels soon after threshold, the costs associated with increasing signal level continue to rise well beyond this point. In other words, the loudest of existing warnings could probably be reduced to levels that are much less disruptive and much less aversive without reducing their reliability.

2. THE PREDICTION OF MASKED THRESHOLD

2.1 The Power-Spectrum Model of Auditory Masking

A noise can have a million times the acoustic power of a signal, yet the signal will remain perfectly audible, provided the signal power occurs in a different portion of the spectrum from the noise power. On the other hand, if the noise power is concentrated in the region of the signal, it need have scarcely more power than the signal to mask it. Thus the occurrence of auditory masking depends not so much on the total power of the signal and noise, but rather on the distribution of the power across frequency. Indeed, to a first approximation, the prediction of auditory masking is really quite simple; a noise will mask a signal if the spectra of the stimuli show that the noise has more power than the signal at every point in the spectrum.

The accuracy of this generalisation depends primarily on the width of the filter used to perform the spectral analysis of the stimuli; whereas a sound level meter with octave-band filters will overestimate the masking produced by the noise, a spectrum analyser set to use a 1.0-Hz fi1ter will underestimate the masking of the noise. This observation laid the foundation for recent power-spectrum models of masking, because it showed that the spectrum analyser could provide a useful analogy for the filtering mechanism of the ear if the attenuation characteristic, or shape, of the auditory filter could be determined with sufficient accuracy.

To be more specific, it is assumed that when an observer is asked to detect a signal presented over a noise, he centres an auditory filter at a local peak of the signal spectrum and listens for the signal through that filter to improve the signal-to-noise ratio, and thus his chances of hearing the signal. An example is presented in Fig. A2 where the signal is a tone and the masker a low-pass noise, that is, a broadband noise that has been low-pass filtered. The filter passes the signal and progressively attenuates the noise components as their distance from the signal increases. A portion of the noise leaks under the skirt of the filter, however, and it is this noise that does the masking

If the edge of the noise is moved closer to the tone, more noise leaks under the filter skirt and, as expected, threshold rises; if the noise cutoff is lowered, threshold falls.

If the power of the signal at threshold is P_s, the long-term power spectrum of the noise is N(f), and the auditory filter shape is W(f), then the general equation for the power-spectrum model of masking is

$$P_s = K \int_{-\infty}^{\infty} N(f) W(f) \, df \qquad \text{(A1)}$$

That is, the power of the signal at threshold is some constant, K, times the integral of the noise spectrum times the filter function. The model is referred to as a ‘power-spectrum model’ because the fluctuations of the noise are ignored and the impulse and phase responses of the filter are not known. Thus, the stimuli are represented by their long-term power spectra and, mathematically, the filter shape is just a weighting function applied to the spectrum.

The general masking equation is used twice in the prediction of masked threshold: To begin with, it is used in conjunction with laboratory experiments to determine the filter function; a noise with a spectrum that simplifies the integral in the general equation is used to mask a tone, and in this way tone threshold can be employed to measure the filter shape. The remainder of this subsection is devoted to a brief description of filter-shape measurement. Then the filter shape is substituted into the general equation (Eq. Al), and the equation is used in the reverse direction to predict the threshold that will be imposed by a given environmental noise. This is the procedure used in section 1.3 to predict threshold on the flight-deck.

2.2 Auditory Filter Shape

The first attempt to measure the filter shape directly (Ref. A7) was based on the experiment suggested in the diagram of Fig. A2. The noise spectrum in that experiment has a constant value, N_o, up to the cutoff frequency F, and beyond this point the noise spectrum is zero. In this case the general masking equation reduces to

$$P_s = KN_o \int_{-\infty}^{F} W(f)\ df \qquad \text{(A2)}$$

The equation shows, that for this particular form of noise, the tone threshold imposed by the noise actually provides a direct estimate of the integral of the auditory filter up to the cutoff frequency, F. Thus, when the experiment is expanded, and tone threshold is determined repeatedly as the edge of the noise is moved through the region of the filter, the set of rising threshold values provides a measure of the function that defines the integral of the auditory filter. The filter shape is then obtained by taking the derivative of the threshold curve with respect to F.

Unfortunately, the situation is not quite as simple as suggested so far. The mathematical derivation of the auditory filter requires the filter to be centred on the tone; but when the masker is a low-pass noise and the signal is a tone, the filter will not be centred precisely on the tone. For, if the skirts of the filter are steep and the passband of the filter has a fairly flat top, as is typical of most physical filters, the signal-to-noise ratio at the output of the filter can be improved by shifting the filter up a little in frequency; the shift reduces the noise passed by the filter substantially, but it only reduces the signal a little. It is possible to derive the filter shape using the assumption that the filter is centred, not on the signal, but at the point where it yields the maximum signal-to-noise ratio at the output of the filter. The mathematics is more complex, however, and it requires the use of experiments in which a second masker is inserted to control the position of the filter. The experiments and their analyses are presented in Refs. A8 and A9; it is sufficient at this point to note that this approach makes it possible to measure the shape of the filter with considerable accuracy over the octave about its centre frequency.

The measurements of the auditory filter revealed one surprising finding; providing the stimulus level was not too high, the passband of the filter is close to symmetric when plotted on a linear frequency scale. Aside from this, the filter shape is fairly typical of those associated with other, well-tuned, physical systems. The filter has a passband with sides, or skirts, that fall at a rate of roughly 100 dB per octave. The dynamic range of the passband is almost always over 40 dB. Outside the passband the slope of the filter drops rapidly. For listeners in their early twenties, the equivalent rectangular bandwidth of the filter, BW_{ER}, decreases from about 14 to 11% of the centre frequency as the centre frequency rises from 0.5 to 4.0 kHz; the bandwidth increases about 2% per decade beyond age twenty (Ref. A10). Since the rate of change along these functions is slow the bandwidth data can be summarised for practical purposes in a single figure, and if it is assumed that the average age of flight-crew is around 40 years, then a reasonable guideline for BW_{ER} is 15% of the centre frequency of the filter.

On a plot of filter attenuation versus linear frequency, the sides of the passband are essentially straight, indicating that this part of the filter function can be approximated by a pair of back-to-back exponential functions. Since the filter is roughly symmetric, only one exponential parameter is required, and so a first approximation to the filter shape is provided by

$$W(g) = e^{-pg}, \tag{A3}$$

where g is the normalised separation from the centre of the filter, to the evaluation point f; that is, $g = |f - f_c|/f_c$. The parameter p determines the width of the passband of the filter. Since the auditory filter has a rounded, rather than a peaked top, and limited, rather than endlessly descending skirts, the approximation is considerably improved by introducing a rounding factor $(1 + pg)$ and a dynamic range restriction, r. Thus, the filter approximation becomes

$$W(g) = (1 - r)(1 + pg)\, e^{-pg} + r. \tag{A4}$$

The factor (1 – r) is introduced to ensure that the value of the filter remains unity at its maximum point of sensitivity. A diagram of this <u>ro</u>unded-<u>ex</u>ponential filter appears in the lower part of Fig. A3; for convenience it is referred to as the Roex(p,r) filter.

2.3 The Calculation of Masked Threshold

The expression for the filter shape can now be substituted into the general masking equation (Eq. Al) to provide an expression for calculating threshold for an arbitrary noise spectrum. The experiments used to determine the filter shape also provide an estimate of the proportionality constant, K, in the general masking equation. It varies slightly with the centre frequency of the filter being a little lower at 2.0 kHz than at 0.5 or 4.0 kHz; it does not, however, vary with age. For practical purposes, it can be assumed to have a value of 1.0 across the range, provided the warning sound contains more than four pulses and the pulse spectrum has more than four components in the appropriate range. In this case, the expression for threshold is

$$P_s = f_c \int_0^{0.8} N(g) \left[(1 - r)(1 + pg)\, e^{-pg} + r\right] dg \qquad \text{(A5)}$$

The constant, f_c, converts the integral value into physical power from the relative frequency domain. Since the filter range restriction is implemented with a constant, r, the integration is restricted in frequency to 0.8.

The indefinite integral of the Roex(p,r) filter has the form

$$-(1 - r)\, p^{-1} (2 + pg)\, e^{-pg} + rg \qquad \text{(A6)}$$

The tail integral of the filter, that is the integral from g to 0.8 provides a convenient means of calculating threshold. The tail integral is

$$(1 - r)\, p^{-1} \left[-(2 + 0.8p)\, e^{-0.8p} + (2 + pg)\, e^{-pg}\right] + r(0.8 - g), \qquad \text{(A7)}$$

and it is shown in the upper part of Fig. A3. Since the filter shape is roughly exponential, the filter and its integral have similar shapes, and so they are parallel in the centre portion of the range. The differences occur at the ends of the functions; the integral is slightly steeper to start with and it reflects the range limitation somewhat earlier.

When the noise spectrum does not change by more than 30 dB across the range of the filter the dynamic-range restriction can be ignored and the tail integral reduces to

$$-\,p^{-1}(2+0.8p)\,e^{-0.8p}+p^{-1}(2+pg)\,e^{-pg} \qquad \text{(A8)}$$

Provided the noise spectrum is not dominated by pure tone components, it can be approximated by a step function. If the width of the step is restricted so that the noise spectrum does not diverge from the step value by more than +3 or -6 dB, then threshold at a particular frequency, f_0, can be calculated by summing terms of the form

$$NL\,p^{-1}(2+pC)\,e^{-pC}-NL\,p^{-1}(2+pF)\,e^{-pF} \quad , \qquad \text{(A9)}$$

one for each step in the range of the filter centred at NL is the average noise level in the step in (dynes/cm^2)/Hz, and C and F are the <u>c</u>loser and <u>f</u>arther edges of the step in relative frequency terms. Threshold is the sum of the contributions from the individual steps times f_c. When the noise level is given in dB SPL, (NLDB) then the value NL should be replaced by $10^{NLDB/10}$ in Eq. A9. Note, that if a noise step crosses the centre frequency of the filter, it must be divided into two steps at f_c and the two contributions must be calculated separately.

REFERENCES

A1 Green, D.M., and Swets, J.A. (1966) 'Signal Detection Theory and Psychophysics,' (J.Wiley and Sons, New York).

A2 Henning, G.B. (1967) 'Frequency discrimination in noise,' J. Acoust. Soc. Am. 41, 774-777.

A3 Scharf, B. (1964) 'Partial Masking,' Acoustica, 14, 16-23.

A4 Potter, R,C., Fidell, S.A., Myles, M.M., and Keast, D.N. (1977) 'Effectiveness of audible warning devices on emergency vehicles,' Report No. DOT-TSC-77-38, U.S. Department of Transportation, Washington, DC.

A5 Wilkins, P.A. (1981) 'Assessing the Effectiveness of Auditory Warnings,' British Journal of Audiology, 15, 263-274.

A6 Green, D.M., and Birdsall, T.G. (1978) 'Detection and Recognition,' Psychological Review, 85, 192-206.

A7 Patterson, R.D. (1974) 'Auditory Filter Shape,' J. Acoust. Soc. Am. 55, 802-809.

A8 Patterson, R.D. (1976) 'Auditory Filter Shapes Derived with Noise Stimuli,' J. Acoust. Soc. Am. 59, 640-654.

A9 Patterson, R.D., and Nimmo-Smith, I. (1980) 'Off-frequency listening and auditory-filter asymmetry,' J. Acoust. Soc. Am. 67, 229-245.

A10 Patterson, R.D., Nimmo-Smith, M.I., Weber, D.L., and Milroy, R. (1982) 'Deterioration of Frequency Selectivity and Speech Intelligibility with Age,' J. Acoust. Soc. Am. (manuscript submitted).

FIGURES

A1 The Measurement of Auditory Threshold:

(a) The two-interval, forced-choice procedure (2IFC). On each trial, the listener is presented with two observation intervals one after the other. They are preceded by a warning light and followed by a response interval. The masker (vertical lines) is presented in both observation intervals, the signal (the sinusoid) is present in one interval or the other at random. The listener's task is to indicate the interval containing the signal.

(b) The psychometric function presents a summary of the data from a 2IFC experiment in which the level of the signal is varied between trials. It is a plot of the percentage of correct responses attained at each signal level; threshold is defined as the signal level that supports 75% correct identification of the signal interval.

A2 Schematic representation of a tone, a lowpass-filtered broadband noise, and the hypothetical auditory filter. The shaded area where the noise and filter overlap represents the noise that is effective in masking the tone.

A3 The Roex(p,r) auditory filter. The lower and upper solid lines show the righthand half of the attenuation characteristic of the auditory filter, and the filter integral, respectively. The dashed lines show where the predominance of the exponential term gives way to the dynamic-range limit.

Figure A1

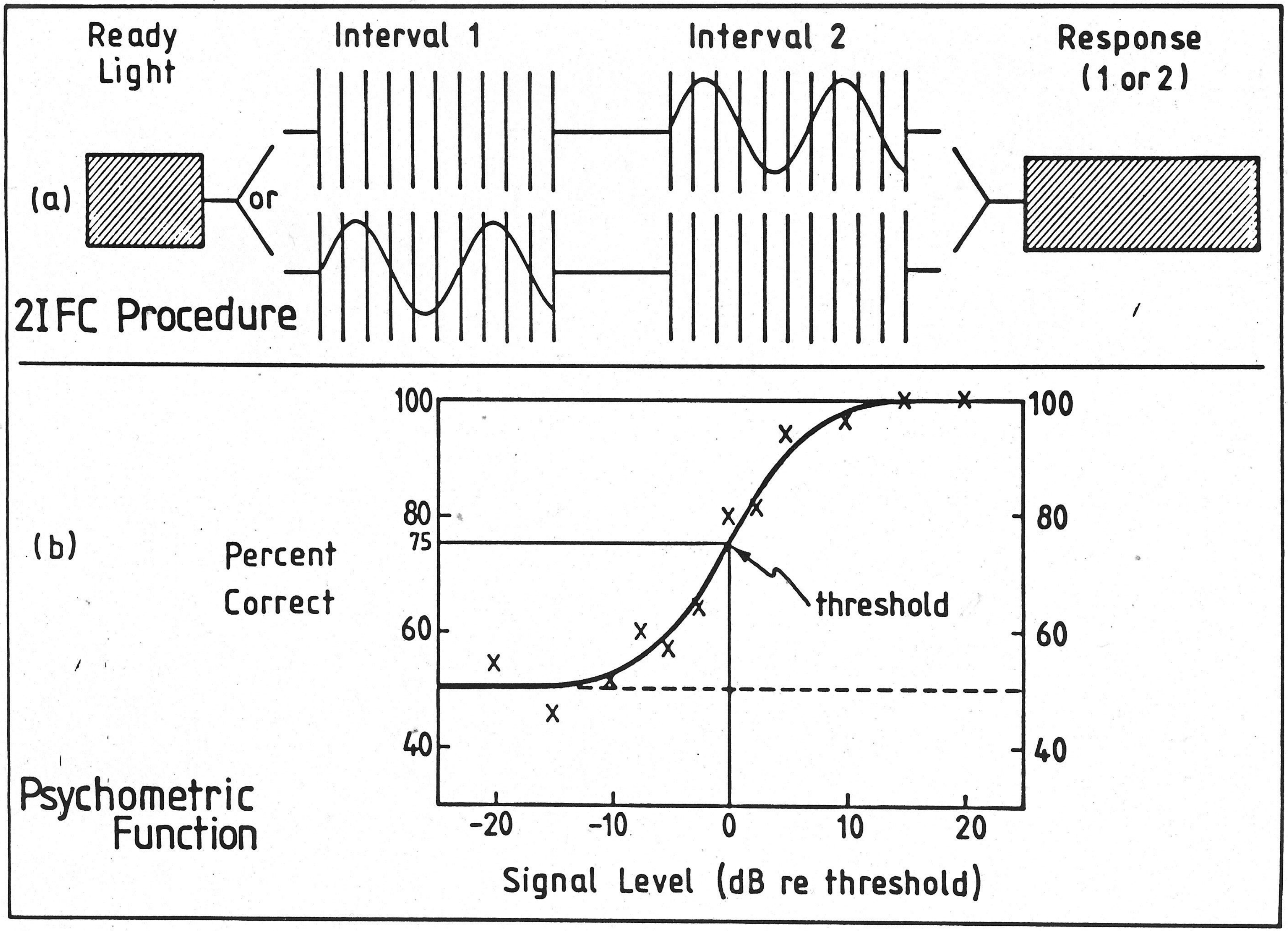

Ready Light
Interval 1
Interval 2
Response (1 or 2)
(a)
or
2IFC Procedure
(b)
Percent Correct
100
80
75
60
40
100
80
60
40
threshold
-20
-10
0
10
20
Signal Level (dB re threshold)
Psychometric Function

Figure A2

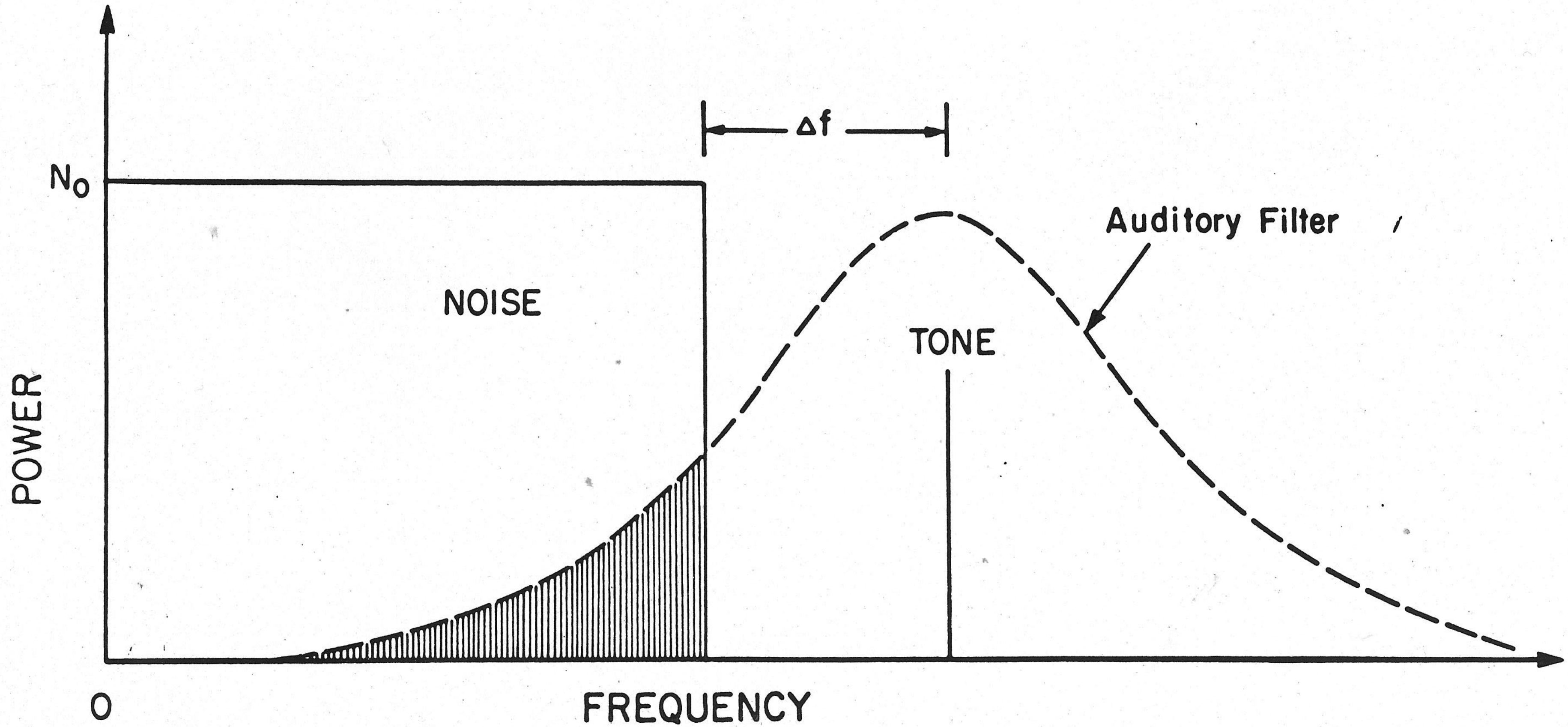
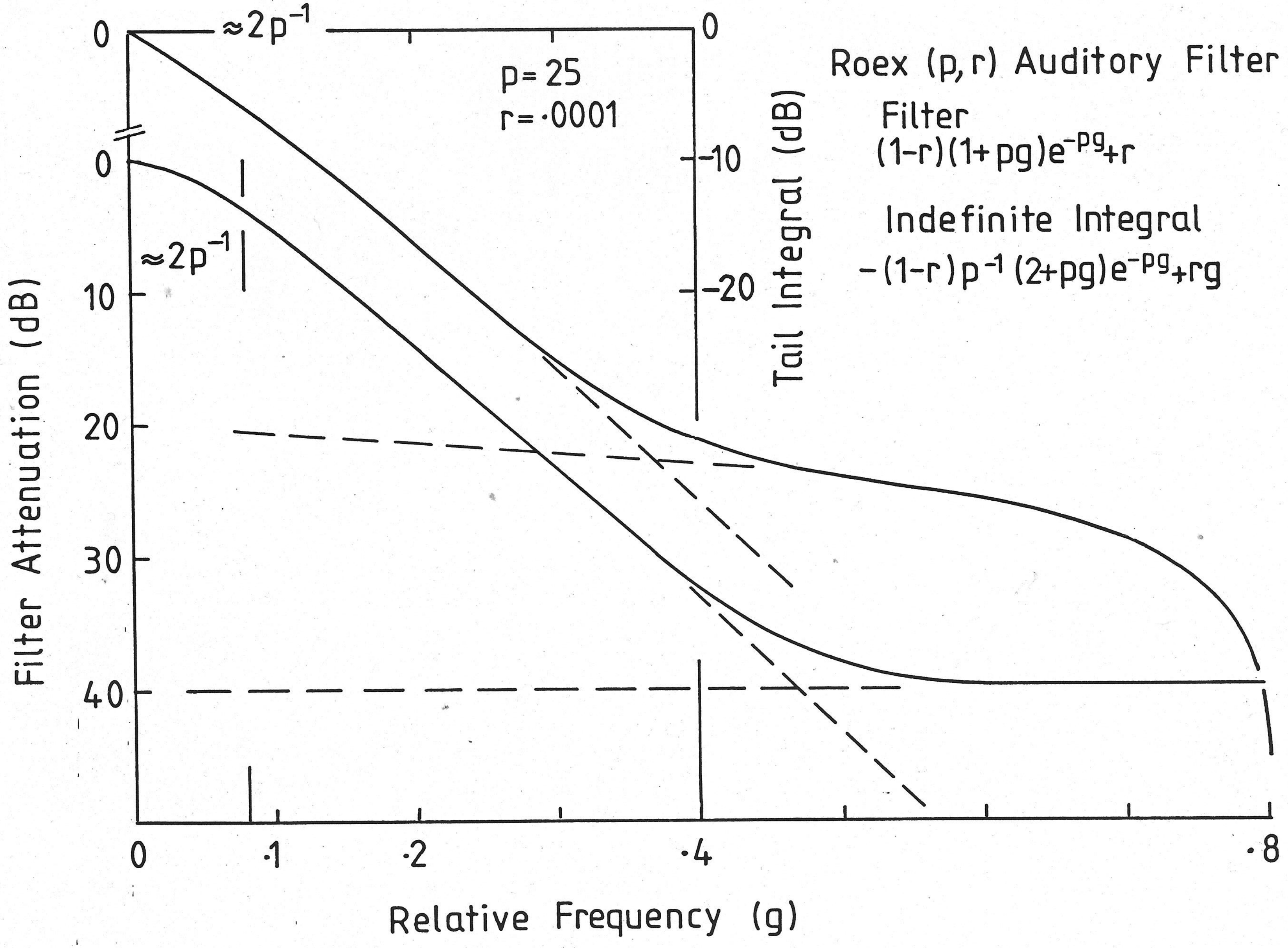


Figure A3



≈2p⁻¹
p = 25
r = ·0001
Roex (p, r) Auditory Filter
Filter
(1-r)(1+pg)e^-pg+r
Indefinite Integral
-(1-r)p^-1(2+pg)e^-pg+rg
≈2p⁻¹
Filter Attenuation (dB)
Tail Integral (dB)
0
10
20
30
40
0
-10
-20
0
·1
·2
·4
·8
Relative Frequency (g)

APPENDIX B The Learning, Retention, and Confusion of Auditory Warnings

1. INTRODUCTION

This appendix presents a precis of a report on the learning and retention of auditory warnings (Ref. B1). The original report contains a method for assessing an existing or proposed warning system to determine the potential for confusion amongst the members of the set, and the precis concentrates on this aspect of the study.

1.1 Background

The flight crew of commercial aircraft have repeatedly complained that the auditory warnings used on the flight-deck are confusing as well as too loud. And it has been demonstrated that operational pilots flying civil aircraft with a relatively large number of warnings make some errors when asked to identify tape recordings of the auditory warnings used on their own aircraft (Ref. B2). Although the error rate was not high it seemed important to determine whether auditory warnings are intrinsically difficult to learn and remember, or whether the observed confusions were generated by ergonomic errors that could be corrected.

No systematic research on the learning and retention of auditory warnings like those found on the flight-decks of civil aircraft was found. There were, however, several studies to show that it takes much longer to learn a set of arbitrary sounds than to learn a list of words. The problem is that the listener has to learn the sounds themselves as well as learning that these items are members of the set to be remembered. This general conclusion seems likely to prove true for sets of auditory warnings as well. But to extrapolate beyond the generality does not seem reasonable because those results were obtained with organised sets of synthetic, laboratory stimuli. As a result, an experiment was performed to determine how quickly observers could learn, and subsequently, how well they could identify a set of ten aircraft warnings.

2. METHOD

2.1 The Warning Sounds

A total of 54 auditory warnings were recorded from seven civil aircraft using a Nagra tape recorder and a Knowles miniature microphone positioned about 25 cm to the right of the first officer at seated ear height. For four of the aircraft (DC1O, 707, 727 and

BAC 1–11) the recordings were made in flight; for the remaining three aircraft (747, L1011, and Trident) the recordings were made from the appropriate simulator. A subset of 10 warnings was selected for the experiment in such a way as to offer maximum acoustic distinctiveness within the set, while at the same time preserving the combination of sound and name most commonly found on this type of aircraft. For example, five of the seven aircraft employed an intermittent horn as the take-off warning sound, so this pairing of sound and name was used in the experiment. The warnings selected, and the aircraft from which they originate, are given in the first two columns of Table B1.

Samples from the tape-recorded warning sounds were digitized using a small computer. The sampling rate was 8.5 kHz and the digitized versions of the warning sounds were stored on floppy diskettes for subsequent replay. The warning sounds were bandpass filtered on input at 0.05 and 4.0 kHz and lowpass filtered at 4.0 kHz when replayed. The duration of the digitized samples was 1.47 sec. The warnings were edited with the computer (a) to cut out background noise on the original recordings, (b) to reduce long-term amplitude fluctuations, and (c) to increase distinctiveness. For example, one cycle of the glide-slope clucker was separated from the background noise and copied repeatedly to produce a non-fluctuating warning based on one clean cycle of the original. The warnings were attenuated or amplified to prevent differences in loudness being used as a cue in the experiment.

2.2 Procedure

Two groups of listeners learned the warnings under serial-learning or cumulative-learning conditions. The results from the two conditions were quite similar and so the description in this precis is limited to the serial-learning condition which produced a slightly higher error rate.

The experiment was performed in four stages: In the first stage, the listeners learned the set of 10 warnings. At the end of this session they were given no further information about the experiment but simply asked to return one week later. In the second stage the listeners were given a recognition test to measure their retention of the warnings, and then after a short rest they were retrained (Stage 3) using the same procedure and warnings as in Stage 1. At this point the listener's audiogram was measured and then in the final stage the retention of the warnings was remeasured.

The basis of the serial-learning method was the trial which consisted of (a) a presentation session in which the 10 warning sounds were played in a random order with the warning name appearing concurrently on a visual display unit (VDU), and (b) a test session in which the warnings were replayed in random order without the printed names, and after each, the listener was required to enter the initial letter of the warning's name on the keyboard of the VDU. The procedure was repeated until perfect identification was achieved on a test session.

The listener was given a brief explanation of the experiment and a short description of the function of each warning sound as the sound was presented over a loudspeaker. The experiment was controlled by a small computer and the trials were self-paced. The sounds were presented over headphones at a moderate level. A list of the warning names was attached to the VDU and the initial letter of each name was used to identify the warning during testing.

3. RESULTS

3.1 Learning

The acquisition of the auditory warnings by the listeners is shown in Fig. B1 for both the initial learning session (Stage 1) and the retraining session (Stage 3). The figure shows the mean number of warnings correctly identified by the ten observers as a function of the cumulative average time required to complete a trial. On the first trial of Stage 1 just under four warnings were identified correctly on average, and over the next three trials about three more warnings were acquired. Thereafter, however, the rate of acquisition slowed dramatically and it took the next nine trials to acquire two more warnings. Although the figure does not show it, all listeners went on to learn the complete set of warnings. When the listeners returned a week later, they began with a retention test in which they heard each sound once and tried to name it. After this limited refresh and no feedback they commenced the retraining session, where the average number correct on the first trial was nine out of ten. Only two listeners did not achieve perfect performance in their first four trials. A full two minutes of each trial was taken up by the presentation of the sounds and sequencing instructions; the average time per trial was 2.83 and 2.67 minutes in Stages 1 and 3 and so the listeners were taking only 2.75 s on average to respond.

The mean number of errors on each trial for the two stages of the experiment are shown in Fig. B2. The vertical lines with horizontal dashes show one standard deviation above the mean; the corresponding standard deviation below the mean was not plotted to avoid clutter. As would be expected from the acquisition data, the error rate shows a steady decline with increasing trial number.

3.2 Retention

When the listeners returned for Stage 2 of the experiment after a one week absence they were able to identify, on average, 7.1 of the warnings learned in the first stage. After the retraining session and an intervening task that took 45 minutes the listeners were tested for their retention of the warnings once again (Stage 4) and, on average, were able to correctly identify 8.9 of the warnings. The distribution of listeners for Stage 4 of the experiment showed that even the worst listeners would eventually learn and retain the warnings.

3.3 Warning Confusions

The errors made by the listeners were broken down by warning type to try to identify any consistent confusions. In brief, the analysis consists of pooling all of the responses and making a table that shows the distribution of responses made to each warning by the listeners (a confusion matrix). The data appear in Table B2. The left-most column shows the warnings that were presented and the top row shows the letters that the listeners entered on the VDU keyboard to indicate their response. The warnings and responses are in the same order; the cell entries are percentages. Correct responses appear on the negative diagonal (marked by the broken line) and the majority of the responses appear here (63.6%); the errors appear in the other cells broken down according to warning presented and response used. The fire warning shows the typical pattern for a non-confusable item; fully 98% of the responses were correct and the warning was misidentified as the altitude alert or the passenger-evacuation warning on only 1% of the presentations. The next warning, the take-off warning, shows the typical pattern of a confusable item; it was correctly identified only 59% of the time and it was misidentified as the disconnected autopilot and the selective call 13% and 11% of the time respectively. In addition, all of the other responses are used at some time in response to presentations of this warning. Not only do the listeners use many different responses when presented with this warning sound, they also use the take-off response (T) when unsure about other warning sounds. Whereas all of the T responses ought to fall in the second cell of the second column, there are many entries in the remaining cells of the column. The column to the left showing the use of the fire response (F) is empty by comparison.

The statistical significance of the observed confusions can be assessed by 1) scaling the proportion correct for each warning in accordance with the listeners' response biases (the column marginals), 2) predicting all of the individual confusion scores using the scaled percent-correct values and the assumption that the errors are randomly distributed with respect to these values, and 3) comparing these predicted error rates with those observed. The result is a matrix of standard normal scores whose significance can be assessed in the usual way. There are ten significant confusions which are underlined in Table B2, that is, ten confusions that deviate from the predicted chance level with a probability of occurrence less than 0.01. Of the ten, six occur in pairs: The take-off warning often elicits the incorrect response D, and the disconnected-autopilot warning often elicits the incorrect response T. Similarly the undercarriage warning is confused with the altitude alert and the overspeed warning is confused with the glide-slope warning. The take-off warning also elicits the S response more often than would be expected by chance, but the reverse is not true. The remaining underlined values occur in the lower, righthand section of the matrix and, although there are only three significant values, the pattern of errors in this region is interesting. One of the values indicates a significant lack of confusion rather than a significant confusion; by chance the selective-call warning should have elicited the G response several times since the listeners have a strong bias for using this response, but the G response was never given to a presentation of the selective-call warning. In fact, the relative frequency of errors involving the glide-slope warning and the other three warnings in this subsection of the table is consistently low. The other warnings – selective call, passenger evacuation, and cabin pressure – are confused. The confusions involving the passenger-evacuation warning and the S response, and the cabin-press-e warning and the P response are significant at the 0.01 level. The confusions involving the selective-call warning and the P response and the selective-call warning and the C response are significant at the 0.05 level.

There were not sufficient confusions in Stage 3 to support a proper confusion analysis. There was some evidence that some listeners were still having difficulty identifying the take-off warning, but the most noticeable effect was that the number of errors was markedly reduced, indicating that the observed confusions are not intransigent, and that with a little more training they would probably disappear.

3.4 Acoustic Characteristics Underlying the Confusions

A simple listening test, in which the members of confusable pairs of warnings were played alternately, immediately revealed that warnings with similar repetition rates are likely to be confused, and that this occurs despite large spectral differences between the warnings. The temporal and spectral characteristics of the warning sounds are listed in columns four and five of Table B1; a verbal description of the sound is provided in column six. The prominent confusions identified earlier are marked by vertical lines in column three of the table.

The take-off warning and the disconnected-autopilot warning both have 2.5-Hz repetition rates and on/off ratios of about 3/4, and they are reliably confused. Their spectra, however, are very different; whereas the take-off spectrum is static with two main peaks at 3.1 and 3.6 kHz, the disconnected-autopilot spectrum has many more components and they glide up in frequency throughout the burst of sound. The warnings are highly discriminable when played one after another, but not in isolation, presumably because the listeners are attending more to the temporal than the spectral characteristics of the sound. The undercarriage warning and the altitude alert are both continuous sounds, and temporal similarity would appear to be the basis of confusion here too since the pitch of the undercarriage warning is almost a minor seventh above that of the altitude alert and the relative prominence of the harmonics is completely different.

The spectrum of the passenger-evacuation warning is a set of lines, while that of the cabin pressure warning is like a broadband noise with one low-frequency component superimposed on it. But the warning sounds are temporally similar insofar as they are trains of pulses and they have about the same repetition rate, 6 Hz, and this is presumably why they are confused. The confusions between these two warnings and the selective-call warning would at first appear to contradict the repetition-rate hypothesis since, strictly speaking, the repetition rate of the selective-call warning is 3 Hz. However, the selective call is composed of two, equal-amplitude tones that alternate with no intervening gap and it gives the impression of having a 6-Hz repetition rate. The tonal quality of the selective call is quite unlike the shrill whistle of the passenger-evacuation warning and the sharp thump and swish of the cabin-pressure warning. Together, then, these confusions indicate that warnings with similar temporal characteristics are prone to confusion.

There is one remaining confusion – that between the overspeed and glide-slope warnings. It is not based on repetition-rate similarity since the repetition rate of the overspeed warning is fully three times that of the glide-slope warning. There is some similarity in the spectra of these warnings in that they are not line spectra, like most of the other warning spectra, and they have broad humps in approximately the same portion of the spectrum. Thus, spectral similarity probably contributes to confusion in this case. At the same time, temporal characteristics probably also play a role here for, despite the difference in repetition rate, the two sounds are both trains of brief pulses separated by silence. The common names for these sounds are 'clacker' and 'clucker', which suggests a perceptual similarity based on both spectral and temporal features.

A complementary analysis was performed to identify clustering amongst the warnings as a result of the perceptual confusions, and to illustrate the grouping of the warnings graphically. A dissimilarity coefficient was obtained for each pair of warnings by averaging the probabilities that one of the warnings would not be given as the response to a presentation of the other warning. On the basis of this set of scores a tree diagram was constructed which shows the level of dissimilarity to which one must rise in order to establish a chain of links between any two warnings. The tree diagram for the Stage-1 data is presented in Fig. B3. It shows that there are close links between a) the undercarriage warning and the altitude alert, b) the take-off and disconnected-autopilot warnings, c) the selective-call, passenger-evacuation, and cabin-pressure warnings, and d) the glide-slope and overspeed warnings. It also shows that one must rise to a high level of dissimilarity to establish a link between the fire warning and any other. Finally, it is interesting to note that, with one minor exception (cabin pressure and glide-slope), when the warnings are ordered according to repetition rate the tree is free of crosses, i.e. places where horizontal links cross verticals.

4. CONCLUSIONS

The learning curves from Stage 1 of the experiment show that:

a) Listeners acquired from four to six auditory warnings quickly – in a few trials;

b) Thereafter, learning slowed dramatically – each additional warning required, on average, an extra five minutes of training; and

c) despite the slow rate of progress, learning continued steadily, and all listeners eventually learned the entire set of warnings.

The retraining curves from Stage 3 of the experiment show that listeners returned to near perfect performance after a week in only a few trials. Thus although it takes a considerable amount of time to learn the warnings, once learned, they are remembered. The retention distributions from Stages 2 and 4 show that the warnings are retained well.

An analysis of the errors that occurred during learning revealed that there were significant confusions. Listening tests and a clustering analysis of the confusions revealed that it is similarity of temporal characteristics, and in particular similarity of repetition rate, that leads to confusion. Given temporal similarity, significant confusions can arise even when the spectral characteristics of the sounds make them highly discriminable.

REFERENCES

B1 Patterson, R.D. and Milroy, R. (1980) 'Auditory Warnings on Civil Aircraft: the learning and retention of warnings,' Civil Aviation Authority Contract Report 7D/S/0 142 R3.

B2 DuRoss, S.H. (1978) 'Civil Aircraft Warning Systems – A survey of pilot opinion within British Airways,' Royal Aircraft Establishment Technical Report No. 78056.

TABLES

B1 The function (column 1) and aircraft of origin (column 2) for the ten auditory warnings used in the learning and retention experiment. The temporal characteristics, spectral characteristics, and a verbal description of the warning sounds are presented in columns 4, 5, and 6. The confusions observed during learning are noted by vertical lines in column 3.

B2 The confusion matrix associated with the initial learning stage. The cell entries along a row show the distribution of responses made to that particular warning (percentage values). The negative diagonal, marked by the broken line, shows the correct responses. The underlined cell values indicate statistically significant confusions.

TABLE B1

Warning	Aircraft	Confusions	Temporal characteristics	Spectral characteristics	Verbal description
1 Fire	BAC 1 – 11		85-Hz repetition rate. 8.5-ms pulses in an irregular train.	2 dominant components near 4.0 kHz	Ringing bell
2 Take-off	BAC 1 – 11		2.5-Hz repetition rate. 300-ms bursts separated by 100-ms gaps.	2 dominant components at 3.1 and 3.6 kHz	Intermittent horn
3 Overspeed	BAC 1 – 11		17-Hz repetition rate. 35-ms triangular pulses separated by 25-ms gaps.	Broad spectrum with broad hump near 2.4 kHz.	Clacker
4 Under-carriage	L1011		Continuous. No modulation.	Harmonics of 290 Hz with those between 2.0 and 3.0 kHz dominating.	Hollow resonator (horn)
5 Altitude	L1011		Continuous. Just noticeable, 6-Hz, amplitude modulation.	Broad line-spectrum. Harmonics of a fundamental near 150 Hz.	Buzzy, hollow, musical note.
6 Disconnected autopilot	747		2.5-Hz repetition rate. 260-ms bursts separated by 140-ms gaps.	Broad line-spectrum with rising frequency.	Siren being started repeatedly.
7 Selective call	747		3.0-Hz repetition rate. 170-ms bursts of two alternating tones.	Sets of harmonics of 0.51 and 0.89 kHz.	Rapidly alternating tones about a minor 7th apart.
8 Glide-slope	DC1O		5.1-Hz repetition rate. 15-ms pulses separated by 185-ms gaps.	Broad spectrum with humps at 1.7, 2.2, and 2.7 kHz.	Chicken clucking.
9 Passenger evacuation	747		5.0-Hz repetition rate. 120-ms pulses separated by 80-ms gaps.	Complex tone with peaks kHz.	Pulses of a shrill bell-whistle
10 Cabin pressure	L1011		6.1-Hz repetition rate. 35-ms pulses over 160-ms noise background.	Broad spectrum with peak at 0.6 kHz and a broad hump at 2.5 kHz.	Train of ‘bonks’ over a swishing background.

TABLE B2

Warning	F	T	O	U	A	D	S	G	P	C
1. Fire	98				1				1	
2. Take-off	1	59	2	2	4	13	11	5	2	2
3. Overspeed	2	5	63	1	2	4	1	21	1	1
4. Undercarriage		3	2	71	9	8			4	3
5. Altitude		3	1	15	50	8	4	5	5	9
6. Disconnected autopilot		11	2	4	7	62	2	4	5	4
7. Selective call	2	5	3	2	6	5	50	_	14	13
8. Glide-slope		3	6	2			2	86		1
9. Passenger-evacuation		9	2	2	5	3	13	5	50	11
10. Cabin pressure	1	4	4	1	7	3	9	3	21	47
Total	105	102	84	98	91	105	91	129	103	91

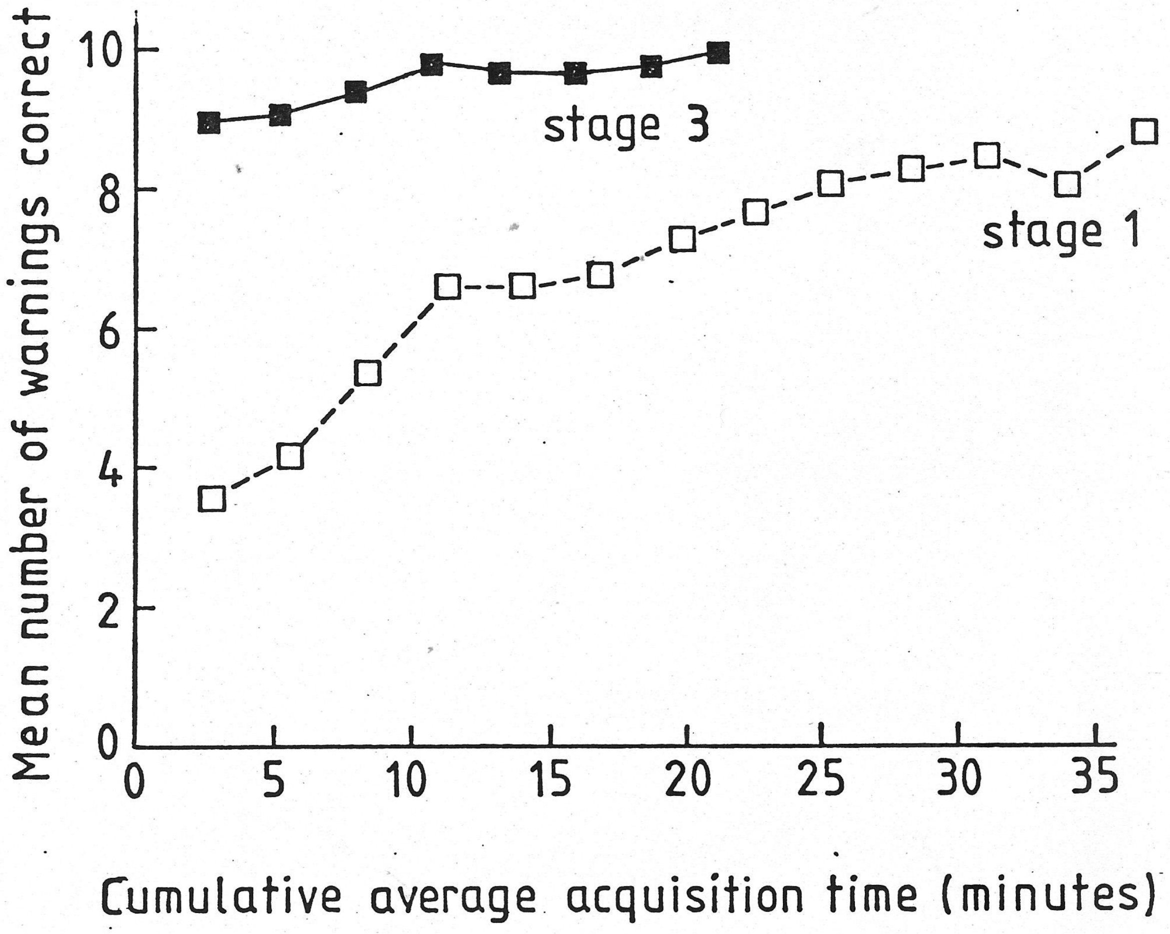
FIGURES

B1 Learning curves for Stages 1 and 3 of the experiment. The data show the average correct for the 10 listeners plotted as a function of the cumulative, average time per trial.

B2 Mean number of errors for the 10 listeners on each trial in Stages 1 and 3. The error bars show one standard deviation above the mean.

B3 A tree diagram illustrating the clustering among the warning sounds indicated by the confusion data. The ordinate shows the relative level of dissimilarity to which one needs to rise to establish links between two warnings. The abscissa values are in Hz.

Figure B1



Mean number of warnings correct
10
8
6
4
2
0
0
5
10
15
20
25
30
35
stage 3
stage 1
Cumulative average acquisition time (minutes)

Figure B2

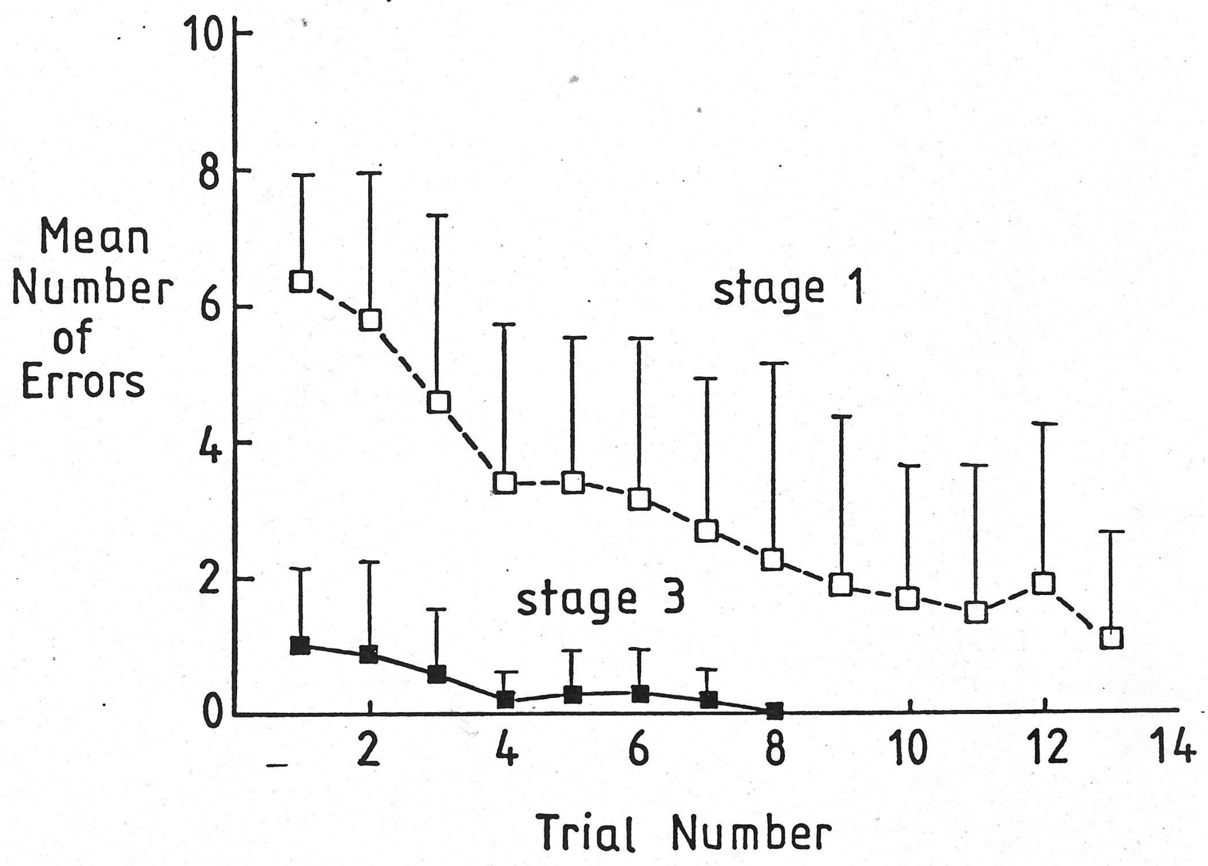
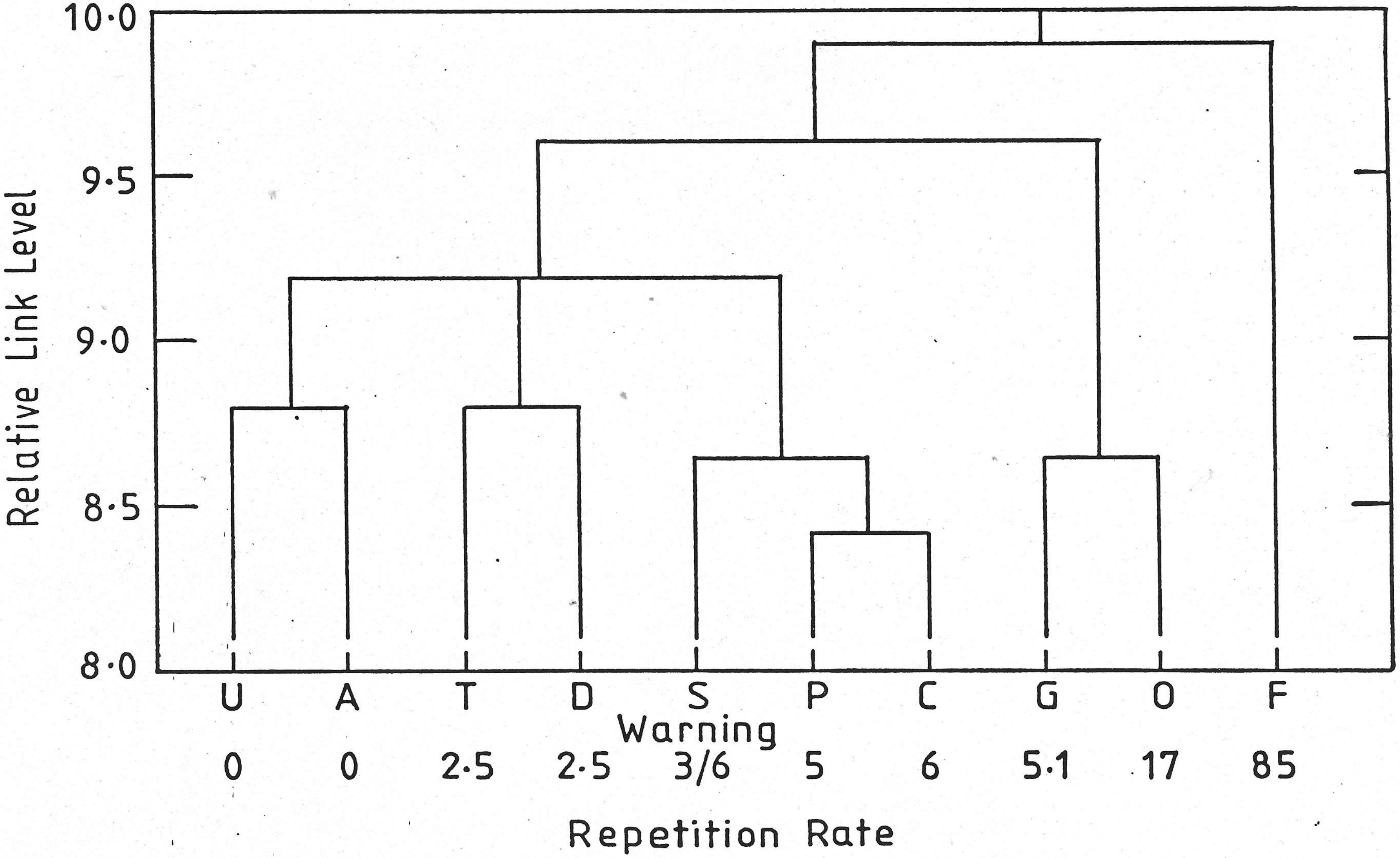

10
8
6
4
2
0
Mean Number of Errors
stage 1
stage 3
2
4
6
8
10
12
14
Trial Number

Figure B3



10·0
9·5
9·0
8·5
8·0
Relative Link Level
U A T D S P C G O F
Warning
0 0 2·5 2·5 3/6 5 6 5·1 17 85
Repetition Rate

APPENDIX C Summary of Auditory Warning Guidelines

The guidelines developed in the main body of the document are listed by section in this appendix.

1. THE OVERALL LEVEL FOR FLIGHT-DECK WARNINGS

The lower limit for the range of levels appropriate for the prominent spectral components of auditory warning sounds is 15 dB above the threshold imposed by the background noise on the flight-deck.

The upper limit for warning-sound components is 25 dB above threshold since the levels imposed by the noise in level-flight are already rather high.

For many civil jet aircraft threshold on the flight-deck, P_s, can be calculated as a function of filter centre-frequency, f_c, using the equation

$P_s = 0.15 f_c NL$,

where NL is the average spectrum level of the background noise in the region about f_c. The level-flight phase of flight is usually the loudest. Note, NL is in (dynes/cm^2)/Hz in this equation.

Many existing flight-deck warnings contain components well over the maximum of the appropriate-level range.

2. THE TEMPORAL CHARACTERISTICS OF FLIGHT-DECK WARNINGS

The pulses of sound used to build a warning sound should have onsets and offsets that are 20 – 30 ms in duration. The gating function should be rounded and concave down.

The sound pulses should be 100 - 150 ms in duration.

For urgent warning sounds the inter-pulse interval should be less than 150 ms. For non-urgent warnings the interval should be over 300 ms.

The warning sound should be composed of 5 or more pulses in a distinctive temporal pattern to minimise the probability of confusion among the members of the warning set.

3. THE SPECTRAL CHARACTERISTICS OF FLIGHT-DECK WARNINGS

The appropriate-frequency region for the spectral components of flight-deck warnings is 0.5 – 5.0 kHz.

The warning sounds should contain more than four components and the components should be harmonically related so that they fuse into a concise sound.

The fundamental of the harmonics should be in the range 150 – 1000 Hz, and at least four of the prominent components should fall in the range 1.0 – 4.0 kHz.

For immediate-action warnings the sounds might contain a few quasi-harmonic components and/or a brief frequency glide to increase the perceived urgency of the sounds.

4. ERGONOMICS

Manual volume control should be avoided. Automatic volume control should be restricted to a range of 10 – 15 dB and used primarily to reduce the volume when the aircraft is on the ground or in the climb or approach phases of flight.

There should be no more than six immediate-action warning sounds and up to three attensons.

5. VOICE WARNINGS ON THE FLIGHT-DECK

The voice warnings incorporated into the immediate-action warnings should be brief and use a key-word format. They should not be repeated in the background version of the warning.

The voice warnings used as immediate-awareness warnings should use a full-phrase format and be repeated after a short pause.

The frequency range appropriate for warning-sound components is also appropriate for speech (0.5 – 5.0 kHz).

The appropriate level for voice warnings can be achieved by positioning the maximum of the average speech spectrum (typically the components of the first formant) near the maximum of the appropriate-level range for warning components.

In the region 0.5 to 5.0 kHz, a progressive amplification of about 3 3 dB per octave will improve the speech intelligibility.