WADD TECHNICAL REPORT 60-814

AUDIO WARNING SIGNALS FOR AIR FORCE WEAPON SYSTEMS

J. C. R. Licklider

Bolt Beranek and Newman Inc.
Cambridge 38, Massachusetts

MARCH 1961

Contract No. AF 33(616)-5611
Project 6190
Task 71556

BEHAVIORAL SCIENCES LABORATORY
AEROSPACE MEDICAL LABORATORY
WRIGHT AIR DEVELOPMENT DIVISION
AIR RESEARCH AND DEVELOPMENT COMMAND
UNITED STATES AIR FORCE
WRIGHT-PATTERSON AIR FORCE BASE, OHIO

1,300 – May 1961 – 29-1077

FOREWORD

This report presents the outcome of a study of problems in the design of audio warning signals for Air Force weapon systems. This study was carried out for the Engineering Psychology Branch, Behavioral Sciences Laboratory, Aerospace Medical Laboratory, under Project 6190 "Flight Display and Flight Control Integration," Task 71556 "Design Requirements for Decision Making Displays," with Dr. Dwight E. Erlick acting as Task Scientist and monitor and Lt. Colonel Elizabeth Guild of the Bioacoustics Branch, Biomedical Laboratory advising on operational aspects of the problem.

The main bases for the procedure described in the report are experience and rational analysis. Some semiformal experiments were conducted in the course of the study, but most of the decisions were based on pre-existing psychoacoustical and human factors information and on informal listening tests. In the listening tests and in the group decision process that delineated the procedure, J. I. Elkind, G. H. Flanagan, D. M. Green, K. D. Kryter, J. C. R. Licklider, T. Marill, R. W. Pew, J. A. Swets, A. Z. Weisz, C. S. Williams, and L. Zeitlin participated.

Because the signals treated in the report are acoustic, they are illustrated better by auditory than by visual displays. Accordingly, a magnetic tape recording has been prepared to supplement the printed report. It illustrates the several "degrees" of each of the dimensions that figure in the specification of an audio warning signal. A copy of the tape may be obtained on short-term loan from the Engineering Psychology Branch, Behavioral Sciences Laboratory, Aerospace Medical Laboratory, Wright Air Development Division, Wright-Patterson Air Force Base, Ohio.

ABSTRACT

This report presents technical information on the design, selection, and use of audio warning signals. It describes a procedure for specifying the acoustical characteristics of warning signals required to meet the exigencies and conditions expected for a given Air Force system, whether it be ground-based, airborne, or spaceborne. The procedure brings together the several applicable design criteria, the constraints imposed by the system, and the conditions under which the system must operate. The procedure then sets those considerations into interaction with characteristics of human sensation, perception, and reaction and from that interaction determines the acoustical specifications of the warning signals. A magnetic tape recording illustrates the procedure and provides an auditory display of the dimensional system in terms of which the warning signals are specified.

PUBLICATION REVIEW

Walter F. Grether
WALTER F. GRETHER
Technical Director
Behavioral Sciences Laboratory
Aerospace Medical Laboratory

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

INTRODUCTION

SUBJECT AND PURPOSE

This report concerns auditory warning signals. An auditory warning signal is a sound presented to an operator to alert him to an emergency, an abnormal condition, or a critical event.* Most of the warning signals currently employed in Air Force systems are visual. There is a growing tendency, however, to use auditory signals. Auditory signals are being used mainly to supplement, rather than to replace, visual signals.

The main purpose of this report is to present technical information that bears upon design, selection, and use of auditory warning signals. The broad aims are:

1. To preclude piecemeal development of a haphazard collection of signals that would be ineffective and confusing.

2. To facilitate the development of an integrated system of signals that will serve well the intended purposes and impose upon operating personnel little requirement for special training.

It is not the intention, however, to propose a ready-made ensemble of signals for adoption by all the organizations of the Air Force. The diversity of Air Force systems and of system-imposed requirements is too great for that. The intention is to provide a set of guide lines, procedures, and examples, that will permit variation within a standard format to meet the diverse needs of the various Air Force systems.

Most of the discussion in which reference is made to specific exigencies and signals will concern airborne weapon systems. However, the treatment is intended to be general enough to apply also to ground-based systems and to space systems.

FACTORS THAT SHOULD GOVERN DETERMINATION OF REQUIREMENTS AND SPECIFICATIONS

Warning signals are designed or selected for use in a weapon system. The factors that should be allowed to govern the determination

---

* It is convenient to distinguish among: (1) warning signals, (2) warning displays, and (3) warning systems. A warning display is a warning signal, together with the equipment that generates it and presents it. A warning system includes, also, the sensors that determine the clues and the decision-making equipment that decides that an exigency is imminent or exists.

of the requirements and specifications for audio warning signals are:

1. Characteristics of the weapon system and its operational environment.

2. Standardization of warning signals throughout the Air Force and throughout the Armed Services.

3. Characteristics of human hearing, perception, and reaction.

4. Engineering and economic considerations, such as weight, cost, reliability, and feasibility of instrumentation.

The main difficulty that must be overcome is that the four factors are understood by different sets of people.

On first thought, one might imagine designing one grand system of warning signals, taking into consideration all the foreseeable weapon systems and all the conceivable emergencies, abnormal conditions, and critical events. That would let us bring the several cognizant groups together to settle the matter once and for all. It would promote orderly analysis of the requirements of the individual weapon systems. It would greatly facilitate standardization. It would be efficient from the human-factors point of view, for there would be only one occasion for setting forth the relevant facts and principles and relating them to the specific problems.

It is not feasible, however, to leap suddenly from the present situation to one that features an ultimate, rigid, comprehensive system of warning signals. Moreover, although comprehensive systematization is desirable, finality and rigidity are not. Air Force systems must be designed to adapt to conditions not foreseeable at the time of conception, and flexibility and modifiability are especially important characteristics for a system of warning signals because emergencies are by their very nature difficult to foresee.

GENERAL NATURE OF THE PROCEDURE TO BE PRESENTED

The procedure to be presented for determining the requirements and specifications for audio warning signals is designed to put the decisions into the hands of the people who best know the weapon systems in which the signals will be used. In the case of an existing system, they are likely to be operating personnel and design engineers. In the case of a system that is in the process of design, they are likely to be design engineers and operators of related systems. The procedure contains within itself the relevant facts and principles of human sensation, perception, and reaction, and it provides assurance of reasonable standardization. It imposes upon the user little or no requirement for special knowledge of human factors or standardization. It does require him, however, to specify correctly what the signals must do and the conditions and constraints under which they must function.

The procedure employs:

1. A series of procedures and an engineering formula that contain the relevant facts and principles of human sensation, perception, and reaction and that effects a transformation from goals, conditions, and constraints, on the one hand, to signal specifications, on the other.

2. A catalog of recorded audio warning signals that illustrates the various signal specifications with concrete examples.

As mentioned earlier, the hope is, through the agency of the over-all procedure, to put the main decisions into the hands of the people who best know the weapon or supporting systems in which the warning signals will be used. The actual preparation of the signals and actual design and construction of the displays, however, will of course remain matters for specialists in other fields. It appears likely that the ensemble of audio warning signals to be used in each system will be recorded on magnetic tape (or belt or drum) and played back, under the control of equipment designed to sense and interpret the available clues, through earphones or loudspeakers to the operator(s). It appears likely, also, that the warning signals will include speech segments as well as sounds of inanimate origin. Specialists in magnetic recording, psychoacoustics, and electroacoustics, will therefore have important roles in the execution.

REFERENCES

Two reports are of major importance in the background of this one. Auditory Signals in Air Force Weapon Systems and Equipment(5) by H. E. Price, and H. J. Older, gives the results of a survey (1955-56) of the use of auditory signals other than speech: It describes the audio warning signals then in use or under consideration. Evaluation of Audio Warning Displays for Weapon Systems(1), by D. E. Erlick and D. P. Hunt, discusses the needs for audio warning signals, the signal formats that appear most promising, and the problem of evaluation.

PREVIEW OF PROCEDURE

The procedure to be proposed is a series of subprocedures, each intended to systematize a portion of the over-all task.

The first thing to do, in setting up a system of warning signals for a particular application, is to determine what the possible and probable exigencies will be. The second step is to decide to which of them to assign warning signals. The third step is to decide upon the nature of the warning signal to assign to each exigency. Thus, it is not until we reach the third step that we come to a problem that is clearly within the scope of this report.

The subprocedures that we shall examine have been developed to:

1. Organize a picture of what features are considered desirable for the audio warning signals.
2. Determine the conditions under which the audio warning signals will have to function.
3. Determine the constraints that limit achievement of the goals of step 1.
4. Decide whether or not to use speech in the warning signals.
5. Decide whether or not to use two-channel (dichotic) audio warning signals.
6. Arrive at specifications, in physical terms wherever possible, for the nonspeech parts of the audio warning signals.
7. If speech is to be employed, systematize the selection and recording of the speech parts of the audio warning signals and the blending together of the speech and nonspeech parts.

CHARACTERISTICS OF PRESENT AUDIO WARNING SIGNALS

We turn our attention, now, to the warning signals. What general characteristics should audio warning signals have? What rules should govern the design or selection of audio signals for specified exigencies? We may begin our examination of these questions by looking briefly at signals presently employed in aural warning.

The acoustic signals listed in the report by Price and Older(5) fall into the categories:

1. bell,
2. buzzer,
3. horn,
4. siren, and
5. tone.

There are several kinds of bell and many kinds of tone. Most current warning tones have complex spectra, but some are almost pure tones. Most are interrupted, but some are steady.

At present, warning signals and exigencies are not paired in a thoroughly definite and standard way. A bell may mean that a target has been detected, a radar track needs attention, the system has a malfunction, there are too many data, there are too few data, there has been a sudden increase in data rate, there has been a sudden decrease in data rate, the crew should prepare to ditch, the crew should prepare for a crash landing, the crew should brace for impact, there is a fire, or someone should answer the telephone or the door. True, different kinds of bell (or different temporal patterns of ringing) are usually employed for different exigencies, and one operator is unlikely to encounter all of them. Nevertheless, it is evident that, in the present usage, there are too few basically different kinds of signal for the number of exigencies to be designated.

Several of the tonal warning signals currently employed sound very much like non-warning sounds sometimes heard over the radio. The use of a continuous 400-cps tone for warning seems particularly undesirable, since the systems in which it is employed have 400-cps electrical power, and it would take only a minor equipment failure to put a spurious 400-cps tone into the headset.

Progress is being made toward more effective design and assignment of warning signals. The MA-1 warning signal generator is an example. Nevertheless, examination of the situation suggests that we should systematically list the goals, the conditions, and the constraints, and try to work out a systematic solution to the audio-warning-signal problem.

GOALS, CONDITIONS, AND CONSTRAINTS

It is convenient to distinguish among:

1. The goals -- the warning-signal features and capabilities that appear to be desirable in their own right, that one would specify if there were no constraints or restrictions. These features and capabilities might better be called "desiderata" if that were not so cumbersome a word. They express the aspirations toward which the design strives. They strongly influence the design criteria.

2. The conditions under which the warning signals will have to function.

3. The constraints imposed by relations among the goals and the characteristics of human hearing and perception.

4. The constraints imposed by system considerations.

GOALS OR DESIGN OBJECTIVES

An audio warning signal should ordinarily be --

1. Audible: easy to hear above background noise or other signals.

2. Quick-acting: capable of evoking a reaction quickly.

3. Alerting: effective in attracting the attention of the operator.

4. Discriminable: easy to differentiate from other signals.

5. Informative: capable of specifying precisely the exigency.

6. Compatible: consistent with warning signals already in use.

7. Nonmasking: not prone to interfere with other functions by drowning out other auditory signals with which it happens to coincide.

8. Nondistracting: not startling; not prone to interfere with other functions by insistently holding attention away from other signals.

9. Nonpainful: not uncomfortable or painful to the ears.

10. Nondamaging: not prone to cause irreversible damage to hearing.

Procedure: Relative Importance of Goals

We shall need quantitative data on the relative importance of those attributes. The proposed procedure for obtaining the data follows:

1. Prepare a set of 10 cards, one card for each of the first 10 goals. Type the name of the goal attribute and a brief explanation on the card.

2. Select 10 or 20 operator-advisors and have each one rank the 10 goal attributes in order of importance for the system under consideration.

3. Determine the mean rank of each attribute.

4. Rank the attributes in order of mean rank. Call the final ranks r_g, the subscript denoting the goal attribute. Save these final ranks for later use.

CONDITIONS

Some of the conditions under which the audio warning displays will be called upon to operate have a special bearing upon the design of the warning signals. Unfortunately, these conditions may vary from phase to phase of the mission, and particular exigencies are more likely to arise in one phase than another. For example, the emergency of coming in to land with wheels not down and locked is by nature restricted to the landing phase, and the background noise is likely to be low, the communication signal density high, during that phase. It is therefore necessary to consider the conditions separately for each exigency. Happily, not many conditions have to be considered. The main ones are:

1. Noise spectrum: level and shape of spectrum of background noise at operator's eardrums.

2. Communication and navigation signals: frequencies of occurrence, levels and shapes of spectra, modulation and interruption characteristics, redundancy, and criticalness of communication and navigation signals with which warning signals will have to compete.

3. Other signals: same attributes as in item 2.

4. Requirement for rapt attention to displays used in ongoing tasks: importance, and percentage of time.

5. Requirement for immediate manual reaction: importance, and amount of time needed for reaction.

Procedure: Determination of Conditions

The conditions must be specified quantitatively. At best, the data will ordinarily be rather uncertain estimates, but it will be possible to do a fairly effective job even with rough data, so long as the data are not systematically in error. It is essential to obtain the estimates from qualified experts.

1. Obtain estimates of noise power-density spectra for each phase of mission and each location within system that must be distinguished. Ordinarily, these estimates will not take into account the acoustic attenuation characteristics of headgear, earphones, etc., worn by operators. Obtain estimated attenuation characteristics of items that will be worn by operators. Subtract attenuation (in decibels, as function of frequency) from each ambient power-density spectrum (also in decibels, as function of frequency) to obtain noise power-density spectrum at eardrum.

2. List the communication and navigation signals that might occur simultaneously with each exigency (or with the associated warning signal). Use the following check list and the list of manipulatable signal characteristics (Table 1)

to systematize their description:

a. Speech

  1. Male, earphone
  2. Female, earphone
  3. Male, loudspeaker
  4. Female, loudspeaker
  5. Male, direct
  6. Female, direct

b. Code: dominant audio frequency
c. LF/MF radio range
d. Radio beacon
e. Fan marker
f. ILS marker
g. VOR-omnirange
h. TACAN
i. Radar beacon
j. Other audio communication and navigation signals

For each item, obtain an estimate of the fraction of the time (in the interval during which the exigency is most likely to occur) the specified communication or navigation signal is expected to fill. This estimation will be facilitated by reference to "time line" analyses of the missions the system is being designed to fulfill. The other characteristics are adequately determined by the designation of the communication or navigation signal and by the background noise level.

3. List the other sounds that might occur simultaneously with each warning signal. Use the following check list to organize their description:

a. Electrical interference

  1. Power-system hum (60-cps or 400-cps fundamental, 120-cps or 800-cps second harmonic)

  2. D-C motor noise

  3. Resistance and shot effect ("white") noise

  4. Radar "leak-through"

        5. Thunderstorm static

    b. Radio signals

        1. Adjacent-channel interference
        2. Carrier difference frequencies

    c. Radar Signals
    d. Countermeasures signals
    e. Air-conditioning noise
    f. Adventitious environmental sounds (e.g., hammering, sawing, drilling)

For each item, obtain estimates of the fraction of the time the specified disturbance is expected to fill, the expected level and shape of spectrum, and modulation or interruption characteristics.

4. Obtain estimates of the importance to the system or mission of extreme concentration on the part of the operator during intervals when each exigency is most likely. These should be in the form of ratings on a scale from 0 to 10. Obtain estimates of the percentage of time concentrated attention is required.

5. Obtain estimates of the duration of the interval from the beginning of the exigency to the last moment at which reaction will be effective. Obtain estimates of the importance to the system or mission of rapid reaction on the part of the operator in the event that each exigency does occur. These should be in the form of ratings on a scale from 0 to 10. Obtain estimates of the length of time in seconds required for the reaction to each exigency.

CONSTRAINTS IMPOSED BY RELATIONS AMONG GOAL ATTRIBUTES AND BY CHARACTERISTICS OF HEARING AND PERCEPTION

Audible, Discriminable, and Informative

To be clearly audible, an acoustic signal must be stronger than the background noise. The main determinants of auditory signal detection are understood quantitatively. Given the shape and level of the interfering background spectrum and the shape of the signal spectrum, we can say what the level of the signal must be for a specified probability of detection and a specified probability of "false alarm." Alternatively, we can specify the spectral shape that will yield highest detectability.

Unfortunately, to achieve high audibility against a strong background noise without making the signal prohibitively intense, it is necessary to concentrate the signal energy within a narrow band of frequencies. For a given over-all signal power and duration, the most clearly audible signal will usually be a segment of sinusoidal tone. The reasons why it is unfortunate are (1) that most listeners are not good at discriminating among simple tones unless their frequencies are widely separated, (2) that -- as follows directly from the poorness of the discrimination -- one cannot convey much information rapidly through a system of simple tonal signals, and (3) that the radio channels are full of tonal signals that might be confused with warning signals if the latter were tonal.

Strong background noise therefore creates a problem of incompatibility among goal attributes: although audibility is a prerequisite for discriminability, the requirements for highest audibility are contrary to the requirements for highest discriminability and for highest information measure.

Audible, Quick-Acting, Alerting, Nondistracting, Nonpainful, and Nondamaging

If we should attempt to provide an audio warning signal that would be clearly audible under extremely noisy conditions -- e.g., high-speed flight with open canopy, rocket launching of manned vehicle with inadequate sound attenuation -- we would run into the following problem: A signal strong enough to over-ride a background noise of 115 or 120 db re 0.0002 microbar in the listener's ears is simply too strong to listen to under ordinary circumstances. If such a signal were turned on suddenly under ordinary operating conditions, the effect on the operator would be traumatic. If an extremely strong signal is to be employed, there must be protection against rapid onset when the background noise is not strong.

One possible arrangement is to repeat the signal over and over at progressively increasing levels until the operator hears it. An essential part of such a system would be a silencing switch (employed in most current systems): the operator would turn the signal off as soon as he heard and understood it. That would ordinarily protect him from exposure to the very intense repetitions. But such a scheme has serious drawbacks. It is undesirable to require the operator to make an additional response (silencing the warning signal) during an emergency. It is undesirable to introduce delay into the warning process, under the unusually noisy conditions, by having low-level (and hence inaudible) presentations of the warning signal precede the effective ones.

A second way of solving the problem is to use noise-operated gain control. A noise-operated gain control would adjust the warning signal to a level sufficiently high for audibility in the presence of the existing background noise. At the same time, it would prevent the occurrence of unnecessarily intense signals and, thereby, avoid the difficulty of startling and distracting the operator(s). To make an effective noise-operated gain control, it will be necessary to incorporate into the electronic circuitry a number of characteristics corresponding to those of hearing. An effective device will, therefore, necessarily be somewhat complex. That fact poses problems of size and weight and, also, problems of reliability. Thus, although a noise-operated gain control would be an extremely convenient thing to have in designing a warning-signal subsystem, it is not entirely clear that it would make a sufficiently important positive contribution to warrant the necessary increment in over-all system complexity. In any event, it is beyond the scope of this report to recommend one way or the other on this question, so we may simply note that there is no satisfactory noise-operated gain control in existence at the present time, and there is no indication that there will be such a device in the near future.

The strong signals required to over-ride strong noise have several undesirable effects. At a rms sound pressure level of 115-120 db, sounds become uncomfortably loud. At 125-130 db, they sometimes give rise to a disturbing tactile sensation that might have a distracting effect. When the momentary peaks reach about 140 db, there is a "tickling" sensation in the ears. And when the rms sound pressure level reaches 140 or 145 db, actual pain is reported and irreversible damage to hearing can be produced in an exposure of only a few seconds. Fortunately, most of the amplifiers, earphones, and loudspeakers likely tc be employed in warning subsystems are incapable of developing the highest of those levels. It is worth keeping in mind, however, that acoustic conditions can arise in which there is a basic incompatibility between audibility and tolerance.

Quick-Acting and Informative

To evoke a reaction very quickly, a signal must be brief or must carry the essential information in its initial part. The total amount of information that a signal can convey tends to increase, however, in proportion to the duration of the signal. Thus there is a problem of incompatibility between the goal attributes, quick-acting and informative.

Alerting and Nondistracting

To be highly effective as an alerter, a signal must be strong, must move about in phenomenal space, and must fill up -- or come close to every part of -- the space within which the listener's attention can focus itself. Such a signal tends to hold attention after attracting it. But attracting and holding attention are almost exactly what we mean by distracting. We cannot have both attributes, alerting and nondistracting, in high degree. It is necessary to compromise. Operating personnel tend to put more weight on the attribute, "nondistracting," than a person without much operational experience would suppose. In an effort to design signals with high attention value, it is easy to create signals that operating personnel will judge prohibitively distracting.

Audible and Nonmasking

To be audible, a signal must be strong enough to over-ride the background noise. The stronger a signal, the more it masks other signals. There is therefore an incompatibility between the attributes, audible and nonmasking. This incompatibility can be obviated by delivering the warning signal to one ear and the signal that it might otherwise mask to the other ear.

CONSTRAINTS IMPOSED BY SYSTEM CONSIDERATIONS

Auditory Acuity of Air Force Personnel

Although most operating personnel of the Air Force have normal hearing, there are some active pilots whose hearing is not good. There is no medical requirement for acuity above 2000 cps. One cannot count on adequate hearing in both ears.

Those facts do not impose insurmountable obstacles, but they must be taken into account. At least a major concentration of energy in a warning signal should be below 2500 cps, and the signal should be readily identifiable on the basis of components below 2500 cps.

(There are only occasional hearing losses that are severe at 2500 cps but not at 2000 cps.) The signal may be presented to one ear at a time, but it should not be presented only to one ear. A feasible and effective procedure is to alternate the warning signal back and forth from one ear to the other at a rate of two or three alternations per second.

Availability of Separate Channels to the Two Ears

The scheme just outlined, depends of course, upon having separate channels to the two ears, so that the signal can be delivered to one and not simultaneously to the other. Other dichotic (different stimuli to the two ears) presentation systems, including those that produce stereophonic effects, pose the same requirement.

A word should be said about the advantages of two-channel dichotic presentation: It makes it possible to create warning-signal effects characterized by motion and by unusual location in phenomenal space -- effects that cannot be produced by ordinary communication and navigation signals or by an enterprising enemy making ingenious use of radio. These dichotic effects can endow warning signals with a characteristic quality which will set them apart from all other signals. Dichotic signals are markedly more effective in alerting the operator than are ordinary diotic signals. Moreover, as mentioned earlier, dichotic presentation offers a way of obviating the masking of one signal by another.

Restrictions on Size, Weight, and Power

System considerations may of course impose restrictions on the size, weight, and power of audio warning displays. These are not likely to be severe obstacles except perhaps in the most parsimonious space systems. An existing 12-double-channel tape device, including elementary logic circuitry and power amplifiers but not power sources, is 3" x 4" x 6" in size and weighs 6 lbs, and those figures can be bettered. The warning display subsystem should, however, include a source of power for use in the event of system power failure.

Size, weight, and power restrictions usually favor earphone presentation over loudspeaker or direct-from-source presentation of warning signals. In most systems in which headsets are always worn by all personnel, earphone presentation seems likely to become standard.

SYSTEMS OF AUDIO WARNING SIGNALS

FUNCTIONS OF WARNING SIGNALS

The functions of a warning signal are

1. To attract and, if necessary, hold the operator's attention.

2. To designate to him the general category within which the signalled exigency lies.

3. To identify the condition specifically.

4. In some instances, to suggest appropriate action.

We may refer to the first function as A (for attention or alert), the second as G (for general category), and the third and fourth as S (for specific condition or suggestion).

TEMPORAL PATTERNS OF WARNING SIGNALS

A priori, there are four reasonable ways to arrange signals to fulfill the stated functions.

1. (A) (G) (S) -- Three signals, one for each function.

2. (AG) (S) -- Two signals, one to attract attention and designate the general category, the other to identify a specific condition or suggestion.

3. (A) (GS) -- Two signals, one to attract attention and the other to identify the general category and also the specific condition or suggestion.

4. (AGS) -- One signal to serve all three purposes.

Essentially all the systems of warning signals currently employed are of type 4. Rational analysis based on knowledge of attention and hearing, plus experience, has led to the conclusion that the most promising arrangement for future development is type 2, with a nonspeech signal fulfilling the AG function and a speech signal fulfilling the S function.(1) The bulk of the experience gained in the present work confirms that conclusion.

DETERMINATION OF SPECIFICATIONS FOR AUDIO WARNING SIGNALS

ACOUSTIC CHARACTERISTICS OF AUDIO WARNING SIGNALS

A diotic warning signal is completely defined by its waveform or by its spectrum (amplitude and phase as functions of frequency). A dichotic warning signal is completely specified by a pair of waveforms or spectra. However, neither waveform nor spectrum presents all of the essential acoustic characteristics in usable form. We must think in terms of some time-domain characteristics of the signals, some frequency-domain characteristics of the signals, and some features of their intensity-frequency-time patterns.

Among the most important characteristics and features of the audio warning signals or of their main component parts (AG and S) are:

1. Over-all average power if signal is homogeneous over time. Running average power if signal is not homogeneous.

2. Shape of power spectrum or power-density spectrum if signal is homogeneous over time. Shape of running power-density spectrum if signal is not homogeneous. (The latter is equivalent to the normalized intensity-frequency-time pattern.)

3. Duration.

4. Diotic or dichotic. If dichotic, characteristics 1, 2, and 3 for right-ear and left-ear signals.

5. If speech, average duration of phonemes.

6. If speech, subjective urgency as judged on scale from 0 (dead monotone) through 5 (normal exposition) to 10 (extremely dramatic).

7. Separability into natural components (in analogy to separability of an orchestra sound into sounds of individual musical instruments; e.g., bells plus siren). If separable, foregoing characteristics of components.

RELATIONS BETWEEN ACOUSTIC CHARACTERISTICS AND GOAL ATTRIBUTES

We must consider, first, the acoustic characteristics that are required to endow the signals with the attributes listed earlier as desirable. Some of the required characteristics relate to individual signals; others relate to the ensemble of signals that constitute a system.

The main relations may be summarized as follows: To make the signal --

1. Audible: Use high sound pressure level. Concentrate signal energy in band in which background noise is lowest.

2. Quick-acting: Focus attention on first 0.5 sec of AG component, then upon first 2.0 sec of S component. Make sure that first 0.5 sec is discriminable from corresponding interval of each other signal that may be presented or appear. Make sure that all essential information is contained in first 2.5 sec. Use characteristics that favor audibility, discriminability, and alerting capability.

3. Alerting: Use high sound-pressure level. Use dichotic presentation, alternating signal from one ear to the other. Use glissando. Use sudden onset. Make signal spend at least 0.1 sec in each octave band from 200 cps to 3200 cps.

4. Discriminable: Use a small ensemble of signals or use speech. Build up a set of signals with large "subjective distances" between members. Use signals that have established names or associations.

5. Informative: Use large ensemble of signals. Use characteristics that favor discriminability. Use speech.

6. Compatible: Use established signals if they are acceptable on other grounds. Do not use established signal to convey a new meaning.

7. Nonmasking: Use dichotic (alternating) presentation of warning signals, diotic presentation of communication and navigation signals. Concentrate energy of warning signals in audio frequency bands unused or little used by communication and navigation signals. Do not use unnecessarily strong warning signals.

8. Nondistracting: Avoid abrupt onset. Avoid movement in phenomenal space; use diotic presentation. Do not use unnecessarily high sound-pressure levels. Avoid signals with strong affective associations.

9. Nonpainful: Avoid very high sound-pressure levels. If very high sound-pressure must be used, avoid concentration of energy in frequency. Use very low frequency band.

10. Nondamaging: Avoid very high sound-pressure levels. Avoid concentration of energy in frequency. Use signals that do not cause discomfort or "ringing in the ears" as an after-effect. Use short durations.

RELATIONS BETWEEN ACOUSTIC CHARACTERISTICS AND CONDITIONS

Approximately comparable to the relations just listed are the following relations between acoustic characteristics of warning signals and the operational conditions under which they are most likely to be used. To insure satisfactory matching of signals to conditions, take the following steps:

1. Background noise: Adjust signal level (over-all average power of speech, average power in first 0.5 sec of nonspeech signal) to yield the required value of the index d'. Use value of d' given in Sections on Ranking and Weighting Procedure and Specification of Signal Characteristics (below). The d' is an index of detectability (6).

2a. Speech communication signals: If speech communication signals are presented via loudspeaker or direct air path from talker, avoid similar mode in presenting audio warning signals. If speech communication signals are presented diotically via earphones, (1) avoid diotic earphone presentation for warning signals -- use dichotic presentation designed to give warning signals distinctive spatial ("stereo") characteristics, or (2) superpose upon the speech components a distinctive nonmasking, nonspeech component that will identify the speech as part of a warning signal, or (3) use a very distinctive voice in recording the warning signals.

2b. Code communication signals: Avoid warning signals that sound like coded radio transmissions, i.e., that consist of interrupted tones. Especially, avoid interrupted tones with frequencies near the frequency usually employed for code reception in the system under consideration. (That rules out several signals that are currently used in special applications, but it is very important not to use warning signals that tend to be confused with other signals.)

2c-j.* Navigation signals: Avoid warning signals that sound like navigation signals, i.e., that consist of modulated or interrupted tones. (This is essentially the same constraint as 2b. The prevalence of modulated and interrupted tones in communication and navigation work essentially rules out that entire class for use as warning signals.)

3a. Electrical interference: Avoid steady tones. Avoid signals that sound like random noise or static. Avoid signals consisting of trains of impulses, whether regularly or irregularly spaced in time.

3b. Radio signals: (The radio signals to which reference is made here are adventitious signals, exclusive of communication and navigation in the channel to which the receiver is intended to be tuned.) Avoid steady tones. Avoid simple glissandi of the type made by two carriers when one is being shifted in frequency (beat-frequency-oscillator effect). Avoid effects such as scrambled speech that would be confused with "monkey chatter" from adjacent channels.

3c. Radar signals: Avoid periodic impulsive signals.

3d. Countermeasures signals: Avoid signals that consist of random noise, periodic pulses, steady or frequency-modulated simple tones, and signals (e.g., "bagpipes") made by standard countermeasures generators.

3e. Air conditioning noise: Avoid signals consisting of random noise.

---

* This outline parallels the one on pages 8 and 9. This item refers to items 2c through 2j of that one.

3f. Adventitious environmental sounds: In selecting familiar sounds and sources for use in warning signals, avoid those that are likely to occur adventitiously under operational conditions. Note that this rules out musical sounds, which otherwise might be used to good effect as warning signals. Musical sounds are almost sure to enter the system if it contains a radio receiver.

4. Requirement for concentration of attention: Use signals with relatively high alerting capacity if the system imposes a strong requirement for concentration of attention.

5. Duration of critical interval: Use rapidly-acting signals if duration is short and if quick reaction is important. Avoid warning signals that persist -- particularly avoid signals that increase progressively in level -- if the action required to silence the warning signal will interfere with the reaction required to correct the exigency.

FORMULA RELATING GOALS AND CONDITIONS TO SPECIFICATIONS OF SIGNALS

At this point, we have in hand all the factors and relations that should go into the determination of the specifications of the warning signals. The next problem, therefore, is to put them all together.

FORMAL APPROACH

The problem can be approached in several ways. Probably the most elegant way is to express the characteristics of the individual audio warning signals that will be optimal as a system of logical functions of a multivariable argument, the argument representing the goals, the conditions, and the required number of audio signals. The structure of the formula would then contain the constraints and the relations that we discussed in earlier sections of this report.

We have studied and experimented with this approach. In artificially simplified examples, it works well. In application to realistic problems of warning-signal design, however, it becomes prohibitively complex. There are many members of the set of "good" examples, and each member has many attributes. We have been unable to work out a way of determining the logical structure in simplified form from the matrix of examples and attributes. Clearly, the problem faced is an important one in the field of system analysis. The present project has been concerned, however, with a substantive problem.

RANKING AND WEIGHTING PROCEDURE

A less elegant but more workable approach to specification of the aural warning signals is based on weighting the ranks assigned to the goals in the section on Relative Importance of Goals. The idea underlying the procedure is to emphasize most strongly the signal characteristics that are required to achieve the most important goals and, conversely, to give relatively little weight to the signal characteristics that would favor achievement of the least important goals.

First, prepare a table with the various manipulatable characteristics of the auditory signals as rows and the various goal attributes as columns. (See Table 1. Delete the characteristic, diotic-dichotic, if two-channel headsets cannot be used.) In each cell, place the weight that specifies the importance of the signal characteristic in realizing the goal. (The weights are shown in Table 1.) Negative weight indicates that increasing the degree of the signal characteristic reacts against achievement of the goal.

Second, assign to each goal a coefficient proportional to the inverse of its rank r_g. The coefficients are:

Rank	Coefficient
1	100
2	50
3	33
4	25
5	20
6	17
7	14
8	12
9	11
10	10

Third, multiply the coefficient associated with each column (goal attribute) of Table 1 by the weight of each cell in the column.

Fourth, add the weighted coefficients of each row and associate the sum Σ_r with the corresponding signal.

Fifth, with each sum Σ_r, enter the appropriate line chart of Fig. 1 and read out the corresponding normalized value Σ_r'.

Sixth, record the normalized value Σ_r' in the appropriate cell of Table 2.

Carry out the foregoing procedure for each audio warning signal (i.e., for each exigency for which an audio warning signal is to be prepared).

ADJUSTMENT FOR DISTRIBUTION OF CHARACTERISTICS AMONG SIGNALS

Purpose. It is important to avoid making the warning signals similar either to one another or to other signals that are likely to be encountered in the course of a mission. This means that we cannot accept as final the specifications provided by the values of Σ_r', arrived at through consideration of the several warning signals separately. We must distort the Σ_r' specifications systematically in order to take into account the requirement of dissimilarity among warning signals. We shall handle this problem in two different ways, one for the speech signals and one for the nonspeech signals. The following paragraphs refer to the latter. The signal characteristics that should be adjusted to ensure dissimilarity are (2) center frequency, (3) amount of variation of center frequency, (4) spread in frequency, (7) emphasis of low frequencies, (8) dwell time, (12) repetitiveness, and (13) rhythm. Characteristics (2), (3), (4), and (8) are particularly important.

Preliminary procedure. The procedure that must be used to ensure dissimilarity among the warning signals focuses upon the normalized values read from the line charts of Fig. 1. These values refer, of course, only to warning signals. It is therefore necessary to introduce a preliminary procedure to bring into the picture other (non-warning) signals that are expected to be encountered during typical missions.

This preliminary procedure employs the line charts of Fig. 2 and tables similar to Tables 1 and 2. Using the data obtained in the section on *Relations Between Acoustic Characteristics and Conditions*, enter the line charts with the appropriate acoustical parameters of each non-warning signal and determine a value of Σ'. (If the non-warning signal is separable into components, treat each component as a separate signal.) These values of Σ' will be of the same general nature as the values Σ'_r determined for the warning signals. We may think of the set of values of Σ' for a given signal as defining a point in a multidimensional space, the dimensions being what we have been calling the signal characteristics. The non-warning signals correspond to points in this space, just as the warning signals do.

Analogy. In order to conceptualize concretely the problem of avoiding undue similarity among signals, it is convenient to simplify matters by thinking of only three dimensions (signal characteristics) and representing each signal by a point in ordinary three-dimensional space. We may then think of the tentative specification -- points arrived at through consideration of individual signals -- as fixed points in the space, and we may even imagine them as colored black. We want the final specification-points to be not too far removed from these tentative ones, but we want the final ones to be spread out, well separated from one another. We want them separated, also, from points representing the non-warning signals that are likely to be encountered.

Now let us locate the non-warning signal points in the space, and color them red. They are fixed points, similar in every way (except for not representing warning signals) to the black points.* Next, let us introduce into the space freely movable, green points, one immediately adjacent to each black point. And let us connect each green point to its black mate by a rubber band. (All the rubber bands are equally strong.) Finally, let us charge all the red and green points electrically, all with like charge so they will repel one another. (The black points do not enter into the electrical picture; they are merely anchor points for the fixed ends of the rubber bands.) We use an amount of charge that is calculated, in relation to the stiffness of the rubber bands, to provide the desired compromise between separating the signal points one from another and keeping the warning-signal points near their tentative specifications.

---

* Any warning signals for which the specifications are fixed *a priori*, e.g., by standardization considerations, should be entered as red points, not as black ones.

The result of the interplay between the retentive force of the rubber bands and the repulsive force of the electric charges is, of course, to move the green points from their initial positions to final positions in which they are: (1) still reasonably in agreement with the specifications based on separate considerations of the individual signals, and (2) yet fairly well separated, one from another and also from the red (non-warning signal or predetermined warning signal) points. This is just what we want to have, but we want to have it not merely in the three dimensions of the analogy but in the 10 dimensions of Table 1.

Procedure. The problem of implementing accurately the compromise illustrated by the foregoing analogy is an interesting one and a difficult one. It seems worthy of consideration in its own right. A 'dynamic' computer program might be designed to effect such a compromise for arbitrary numbers of points in an arbitrary number of dimensions, with any specified division of importance between keeping the green points near their black mates and separating the members of the set of reds and greens. But such a refinement of method is beyond the present scope. What we need, here and now, is a way of making only a rough compromise, but a way of doing it simply and systematically. We shall therefore approach the problem in two steps, first looking only at the matter of too-close neighboringness on individual dimensions, and then considering the over-all separations between pairs of points in the space. Because it is simple, and because there is some evidence that "distance" between signals in human perception is a sum and not the square root of a sum of squares, we shall simply add up the separations of paired signals on the individual dimensions and make adjustments until there are no too-close proximities.

Adjustment to prevent too-close proximity in individual dimensions. The average separation of signals, one from another, on any normalized characteristic should not be less than 0.20. If there are many signals -- if the number k of warning and non-warning signals is large -- then the average separation between neighbors on one normalized characteristic cannot be larger than 1/k. We may take 1/(2k) as the danger line.

As a first step, note the pairs of signals that are separated by less than 1/(2k) on each normalized characteristic. Keep the pairs separated by characteristics. Use pairs consisting of two warning signals and pairs consisting of one warning and one non-warning signal, but not pairs consisting of two non-warning signals.

Next, determine which pairs, if any, have been noted more than once.

Then, readjust the normalized values for those pairs (i.e., for one or both of the warning signals in a pair consisting of two warning signals, but not for the non-warning signal in a mixed pair), trying not to create new close pairs in the process. Readjust until no pair of warning signals comes inside the 1/(2k) separation limit for more than one signal characteristic. Replace the values in the cells of Table 2 by the new values just found.

Adjustment to ensure adequate over-all separation. Having made sure that each pair of signals is characterized by very close proximity on no more than one dimension, we turn now to consider the over-all separations of pairs. Again, we consider only pairs that include two warning signals or one warning signal and one non-warning signal. And, again, we restrict attention to dimensions (2), (3), (4), (7), (8), (12), and (13).

First, make a table with rows corresponding to the various signal pairs and columns corresponding to the dimensions.

Second, determine from Table 2 (as readjusted in the preceding section) the distance between the signals of each pair in each dimension. The distances are, of course, the absolute values of the differences between the normalized coordinates.

Third, add the distances found in each row (i.e., sum over dimensions) to determine an index D_{ij} of over-all separation for each pair of signals i and j. ($i \neq j$.)

Fourth, letting x be the number of dimensions in the table, take x/7 as the lower limit for D, and readjust the coordinates of the warning-signal numbers of the pairs for which $D < x/7$.

Ideally, we should have an automatic procedure that would effect the readjustment for pairs with too little separation without pushing other pairs too close together. We might even go on to consider trios and larger groupings of signals, realizing that the operator could learn to handle a lone pair of signals separated by a small distance D much more easily than a large cluster with an average separation equal to D. As indicated earlier, however, it is not within the present scope to strive for such refinement. We must leave mainly to intuition, therefore, the problem of selecting the particular adjustments to make in order to separate the too-similar signals without creating other too-close similarities in the process.

Even without an automatic procedure, however, we may make sure of arriving eventually at a successful adjustment by testing to determine whether the adjustment just made has eliminated the troublesome pairs without causing difficulty in pairs that initially did not need adjustment. It is probably worthwhile, if the number of signals is large, to make such a test after the initial adjustment.

The problem of avoiding too great similarity among the speech components of the warning signals is relatively minor. It is necessary only to check through the list of speech items, looking for obvious phonetic similarities, and to correct them by substituting for one or the other member of the pair likely to be confused. It is very important, however, to make sure that the speech components of the warning signals are clearly identified as parts of warning signals, and that the other speech signals that may reach the operator's ears are not misinterpreted as parts of warning signals. Confusion can be prevented by using a very distinctive voice, by connecting a nonspeech sound very closely with the speech, or by using dichotic presentation for the warning signals but not for any others.

CONVERSION OF NORMALIZED CHARACTERISTICS INTO PHYSICAL SPECIFICATIONS

Through the procedure just described, we arrive at final values Σ'_r of the normalized signal characteristics. We now turn to the transformations that will take us from the normalized characteristics of specifications in physical terms. These transformations are summarized in Fig. 2. There are 15 signal characteristics in the table. The ones under consideration now are the first nine. The remaining six, not as susceptible to specification in physical terms, are deferred for later treatment with the aid of scaling. Fortunately, they are less important than the present nine.

The first major decision is whether or not to employ speech signals in the warning-signal subsystem. This decision may of course be forced into the negative by system constraints. It if is not, take the values of Σ'_9 from row 9 of Table 2, and locate markers at corresponding points on line chart 9 of Fig. 2. (Use line 9A if there are to be seven or fewer audio warning signals; use line 9B if there are to be eight or more.) If more than one-third of the markers fall to the right of the midpoint (0.5) of the line chart, the decision should be to employ speech signals, i.e., to use warning signals of the (AG) (S) format. If that decision is reached, use the following procedure for the (AG) signals. (If a nonspeech signal is used to label the speech parts as parts of warning signals, put that nonspeech signal in with the (AG) signals.) If the decision is made to use only nonspeech signals, then use the following procedure for all the warning signals.

The next major decision is whether or not to employ dichotic signal presentation. This decision, also, may be forced into the negative by system constraints. If it is not, take the values of Σ'_{10} from row 15 of Table 2, and locate markers at corresponding points on line chart 10 of Fig. 2. If more than one-third of the markers fall to the right of 0.3, the decision should be to use two-channel presentation. (Considerations related to equipment cost may reasonably move the cutoff point either side of 0.3 by perhaps 0.1 unit.)

Audibility. For each warning signal, refer to the background noise spectrum estimated in the section of Conditions. If it has not already been smoothed, smooth it enough to prevent abrupt variations within 'critical bands' and measure the ordinate at the center of each critical band. The boundary and center frequencies of the critical bands are given in Table 3.(7) Now do the same for the warning signal itself, the over-all signal power being set arbitrarily at P.

Next, determine the power signal-to-noise ratio S/N at the center frequency of each critical band. Then enter Table 3 with the band center frequency, and read out the corresponding value of (S/N)', which is the signal-to-noise ratio that yields d' = 1. Next, find the ratio of (S/N) to (S/N)'. Empirically(3), $d' \alpha S/N$, and the value of the ratio of (S/N) to (S/N)' may be used as the value of d' for the critical band and signal under consideration.

To find the value of d' for the over-all signal, use the formula(2)

$$d'_P = [\Sigma(d'_{ij})^2]^{1/2}.$$

This value is for the arbitrarily selected over-all power P.

Finally, to find the value of signal power required to achieve the value of d' specified by line chart 1 of Fig 2, use again the empirical relation, $d' \alpha S/N$. That is, set the over-all warning-signal power equal to

$$P' = (d'/d'_p)P.$$

That will provide the desired degree of audibility as determined in the compromise with the other relevant goals.

Center frequency, amount of variation of center frequency, spread in center frequency, and duration. Happily, these signal attributes are determined more directly. Their specifications are found simply by entering the appropriate line charts of Fig 2 with the corresponding normalized sums of Table 2. The specifications are provided directly in physical units: cycles per second for f_c, σ_c, and σ_f, and seconds for Δt.

Emphasis of initial segment. For the nonspeech signals with which we are dealing here, the degree of emphasis of the initial segment can be specified physically as of the fraction C of the total duration comprised by the interval containing the first half of the total signal energy. That fraction is read directly from line chart 6 of Fig 2. The initial-segment emphasis of a recorded warning signal can be adjusted conveniently by multiplying the electrical time function (waveform) by an exponential function of time ($e^{\Gamma t}$ or $e^{-\Gamma t}$) in an electronic analogue multiplier. If the amount of emphasis is monitored -- e.g., with a square-law circuit, integrator, and pen recorder -- the desired adjustment can be obtained by varying Γ until the integral of the squared voltage reaches half its maximum value in the desired fraction C of its total duration.

Emphasis of the low frequencies. In discussing emphasis of the initial segment, we did not consider the opposite pole, emphasis of the final segment, because there is no likely requirement for the latter. In dealing with the analogous quantity in the frequency domain, however, high-frequency emphasis is perhaps as probable a requirement as is low-frequency emphasis. Accordingly, we may let the value 0.5 represent spectral uniformity over what might be called the "warning-signal frequency range," and we may consider high-frequency emphasis to be associated with values of s_7 (and, as seen in line chart 7 of Fig. 2, also with values of the variable K) less than 0.5. Let the warning-signal frequency range extend (see line chart 2 of Fig. 2) from 250 to 2300 cps, and associate K = 0.5 with a balance of signal energy (within that range) about 1000 cps. It is natural, then to define K as the ratio between the power in the band 250-1000 cps and the power in the whole range 250-2300 cps. K may therefore be monitored conveniently with the aid of two filters and two square-law meters (e.g., Ballantine Model 320 True Root Mean Square Voltmeters), and K can be adjusted to have any desired value with the aid of conventional tone controls.

Dwell time. The dwell time τ is defined, for frequency-modulated narrow-band signals, as the length of time spent within a frequency band one critical band wide. (See Table 3 for critical bandwidths.) If the signal is not a narrow-band signal, but is nevertheless modulated in frequency through a range great compared with the bandwidth, τ may be taken to be the time spent within a critical band by the center frequency of the signal. If the signal is not modulated through such a range, dwell time is not an applicable dimension. Consider this dimension, therefore, only if σ_t (amount of variation of center frequency) is more than twice σ_f (spread in frequency).

Dwell time is difficult to measure directly with electronic measuring equipment, but can be determined readily from the spectrographic records provided by a Sound Spectrograph or (better) a Vibralyzer (Kay Electric Company). The critical band limits and time lines may be marked off on a sheet of transparent plastic, which is then used as an overlay. The dwell times can then be read off from visual inspection. There will, in general, be different dwell times for different bands or different temporal segments of the record. It is sufficient to make the average agree approximately with the specified value of τ.

Speech-nonspeech. This characteristic was considered in the first main decision of the procedure. We shall defer discussion of the speech parts of warning signals until later. The speech-nonspeech dimension does not play any further role.

Diotic-dichotic. The diotic-dichotic dimension figured in the second main decision of the procedure, the decision whether or not to employ two-channel warning equipment and signals. If that decision came out in the affirmative, the diotic-dichotic dimension enters the procedure again here. It indicates how marked or vivid the stereo effect should be.

Unfortunately, there is no simple way of reducing the markedness or vividness of a stereo effect to a recipe stated in physical terms. As mentioned earlier, this unamenableness to direct physical specification, in fact, characterizes all the remaining signal features. We are forced, therefore, to adopt an indirect procedure. In the catalogue of illustrative recordings that is associated with this report, Part II contains a series of examples for each signal characteristic illustrating various amounts of that characteristic, i.e., various locations on the corresponding line chart in Fig 2. The procedure for realizing a given value of 'dichoticness' (given value of ∂) makes use of the series for dimension 10.

This procedure may be formal or informal and may employ as many listeners (raters) as the importance of the over-all project warrants. Essentially, we need ratings by listeners of each of the illustrative examples in the recorded catalogue and of the warning signals under development. The medians of the ratings of the illustrative examples determine a scale and define its relation to ∂. The medians for the signals under development then locate those signals on the ∂ scale. The divergencies of these values of ∂ from the values specified by Table 2 indicate in which direction to modify the signals. It is necessary to test and remodify until a sufficiently close approximation to the desired values is obtained. One reiteration of the procedure will probably suffice.

In constructing dichotic signals, it is desirable to have the two parts (right ear and left ear) sufficiently similar to ensure perception of a single, fused sound image. (If the aim is to make a warning signal with two separable components, each component should be presented binaurally. Two completely different signals, one in one ear and the other in the other, are too likely to cause the listener momentary confusion.)

For small amounts of stereo effect, it is sufficient to start with one acoustic signal and produce a pair of signals from it with the aid of a pair of microphones. The magnitude of ∂ can be controlled by varying the spacing of the microphones. For larger values of ∂, use a movable source, and move it relative to the microphones while it is sounding. For very large values of ∂, it is convenient to employ an electronic switch and to present the signal first to one ear and then to the other. This technique has the great advantage of minimizing the masking produced by the warning signal -- masking that might otherwise drown out communication or navigation signals.

Established associations, affectiveness, repetitiveness, and rhythm. These signal characteristics must be handled, if they are taken into account at all, with the aid of rating procedures analogous to the one described for dichoticness. As is seen in Table 1, these characteristics do not offer much control in achieving the goals. They need not be introduced into the overt procedure unless the number of warning signals to be constructed is quite large (say, $k = 25$) or the system for which they are intended is extremely important.

If these signal characteristics are to be handled in the overt procedure, ratings should be made of the illustrative examples in Part II of the catalogue and of the warning signals under development, and adjustments should be made until the desired values of the characteristics (ϵ, α, π, and ρ) are approached.

SELECTION AND PREPARATION OF SPEECH COMPONENTS

If the warning signals are to include speech segments, the selection and preparation of the speech signals are extremely important parts of the task. The procedure to be followed must include:

a. Selection of the words or phrases,

b. Selection of an announcer or announcers,

c. Rehearsal to establish voice level, timing, and adjust tone of voice, rate of speaking, affectiveness, urgency, etc.

d. Recording,

e. Combining speech with nonspeech parts to produce over-all signal.

Given the definitions of the exigencies, the main decisions that must be made concern:

a. Whether to state what the trouble is, to tell the operator what to do about it, or both.

b. How much detail to present; how precisely to pinpoint the situation or the required response.

c. How to phrase the message -- choice of words, amount of redundancy.

d. What qualities of speech to use -- how calm or excited (relaxed or urgent), how personal or impersonal (formal or informal), etc.

e. What relation between speaker and operator to imply, which includes the "image" sex, age, and status of speaker.

For the speech signals, there is no formula to relate the goals to physical specifications. It is necessary to rely on the judgments of listeners. The judgmental procedure can be systemized, but the adjustment of the speech signal to meet the criteria must of necessity be either a trial and error process or guided by intuition. The examples given in the recorded catalogue part of this report may serve as anchor points in setting up and realizing the criteria.

Situation versus response. As mentioned earlier, a warning signal may tell the operator (a) what the situation is, (b) what action to take, or (c) both. Ordinarily, the operator is brought into the picture either because his integrative and decision-making capabilities are required -- because the equipment cannot take into account all the contingencies that may affect the decision -- or because he must be forewarned of an impending emergency action. In the first case, alternative (b) is ruled out. The warning signal may either (a) define the situation as sensed by the sensors of the warning-signal subsystem, leaving it to the operator to combine that information with other information available to him and to decide on the composite basis what action to take, or (b) define the situation as sensed and then make non-binding suggestions concerning appropriate responses.

The decision among the alternatives just mentioned can be made without a formal procedure, but it should be made deliberately and explicitly for each exigency under consideration.

Amount of detail. The amount of detail to present depends heavily upon the amount of time available. (See item 5 of the earlier discussion involving Table 2 and Fig. 2.) If the time is short, present the most essential information first. Then there are two alternatives: (a) to have the warning signal terminate itself, and (b) to let the warning signal continue, as long as it may contribute any helpful information. In the latter case, there should be a turn-off switch. The operator should not have to continue listening if he chooses not to, for the distracting effect of the signal may outweigh its informativeness if the operator has made up his mind and undertaken a course of action. The turn-off switch should not prevent the presentation of *other* warning signals. It should reset itself after an interval so that it will not preclude a second later presentation of the activated warning signal.

The decision, whether or not to include a switch for turning off warning signals, is a basic design decision. Experience indicates that such a switch is desirable unless the operator will *always* be extremely busy, and there appear to be no present systems in which that is the case. We support the statement in the Handbook of Instructions for Aircraft Designers (HIAD) which specifies inclusion of a turn-off switch.

If there is to be a turn-off switch, then the question arises, for each signal, whether or not to use a (long) signal that might lead to use of the switch. This decision should be based on two factors: (a) how busy the operator is likely to be when the signal is sounded, and (b) how helpful the extra information afforded by a long-continuing signal would be likely to be. As before, the decision should be made deliberately and explicitly for each signal.

Choice of words. We studied in detail the question of selecting words in such a way as to optimize the intelligibility of the speech parts of the warning signals. This study led us to the conclusion that intelligibility should be considered only after aptness and conciseness have been assured. This is not to say that intelligibility is unimportant. It is simply that aptness and conciseness are also important, and that they are difficult to achieve if intelligibility is allowed to come first -- whereas intelligibility suffers little if primary consideration is given to the other qualities.

The recommended procedure, therefore, is to collect, in written form, at least a dozen suggestions for the wording or phrasing of each speech signal. These should be composed by several different people. Then a group discussion should be held to determine (a) the format for the entire set of speech signals and, (b) within the constraint imposed by that format, the two or three best wordings for each signal.

It is possible, but prohibitively time-consuming, to determine through phonetic analysis rather accurately what overall set of the selected wordings will lead to the highest intelligibility, the least confusion among warning signals. We considered incorporating into a digital-computer program the phonetic and psychoacoustic discriminators necessary to make the selections automatically, but that project was rather beyond the present scope. The recommended procedure is therefore a substitute that relies on subjective judgment, but leads to reasonably good results:

Prepare cards containing all the pairs of acceptable speech signals (i.e., best two or three for each exigency). Read these in the presence of background noise to a group of listeners. Have the listeners rate each pair on a scale of 0 to 10, indicating how similar or confusable are the two members of the pair. (The rating 10 means "subjectively identical.") Assign each rating to both members of the pair. Find the sum of the ratings for each speech signal. For each exigency, use the corresponding signal with the lowest sum of ratings.

Calmness versus urgency. There are some reasons for believing that a signal that warns of a serious emergency should convey a sense of urgency. There are other reasons for believing that it should be calming and inspire a rational, collected approach to the problem it poses. But urgency and calmness occupy extreme positions on dimensions that are far from orthogonal. Although they are not precisely opposites, it is difficult to achieve them both in high degree.

Insofar as both calmness and urgency are judged to be desirable attributes for a particular warning signal, the procedure used in creating the signal should lead to as high values of both attributes as can be achieved under the existing constraints. Ordinarily, however, signals with high urgency will be desired only for "killer" exigencies or for "warning" exigencies to which the operator has not responded despite warning. A simple way to systematize the procedure for controlling the degree of urgency and the degree of calmness is the following:

a. Refer to the examples for the two dimensions that are provided in the catalogue. Select the values (interpolating if necessary) that are desired. Call them z'_c and z'_u for calmness and urgency, respectively.

b. Have the announcer(s) listen to the examples nearest the target value and then generate several speech recordings for each desired signal, attempting to hit near the target values of calmness and urgency in each instance but varying the effort to a moderate degree in whatever ways appear promising.

c. Have listeners rate the samples separately on both variables, using scales from 0 to 10. Include samples from the catalogue in the rating to anchor the judgments.

d. Average the ratings for each signal on each attribute. Let the averages be z_c for calmness and z_u for urgency.

e. Prepare a line chart for each attribute, relating the average ratings obtained to the values given in the catalogue.

f. With the aid of the line charts, convert the average ratings for the announcer's signals to catalogue-type values.

g. Decide upon a weight w_c for calmness, relative to $w_u = 1$ for urgency.

h. Select the version for which the value of $D = w_{ci}(z'_{ci} - z_{ci}) + (z'_{ui} - z_{ui})$ is minimum.

Formal-informal. Another pair of dimensions along which speech signals vary are formality-informality and personalness-impersonalness. These dimensions are closely correlated. The impression gained from contacts with operators is that, although an informal, personal approach from an automatic device is interesting and even refreshing on first experience, fair amounts of formality and impersonalness are preferred for a steady diet. Our recommendation is that the most-frequently-expected signals (i.e., status and caution signals, if those categories are employed) be quite formal and impersonal, rather highly stereotyped, leaving for the less frequent and more urgent signals the leeway provided by moderate degrees of informality and personalness.

The matter of establishing and realizing the desired degree of these attributes can ordinarily be handled informally. In situations in which an explicit procedure is required, the procedure used in connection with calmness and urgency can be employed. It is necessary only to substitute subscripts: f (formality) for c and p (personalness) for u.

Role and status. In the last section, in discussing formality and personalness, we considered part of the problem of the simulated role and status that are assumed by the warning-signal subsystem when it employs speech signals. It is important to decide, also, on the age, sex, and social or command status relative to the operator(s) that the speech parts of the warning signals should simulate.

On the basis of experience, we conclude that the voice should sound mature, that a man's voice will in the long run prove more acceptable than a woman's voice to male operators, and that it is best to keep the social or command status ambiguous. Each of those conclusions must be amplified.

So long as his voice is mature and does not show signs of advanced age, it does not provide very accurate clues to the announcer's age. Insofar as age is concerned, therefore, all that is necessary is to avoid youthful-sounding voices and faltering voices.

It has been suggested many times that a woman's voice be used for the speech parts of warning signals. This is an interesting suggestion because there are fewer women's voices than men's voices on military radio channels (and the woman's voice in the warning signal would therefore be less likely to be confused with part of an ordinary communication signal) and, also, of course, because a woman's voice may have a natural priority in commanding a man's attention. The first advantage is perhaps balanced by the consideration that women's voices, because they have fewer harmonics to define their formants, tend to be slightly less intelligible than men's. The second advantage may be genuine if the warning signals are heard infrequently, but (unless means were provided for varying the expressions from presentation to presentation) it would doubtless be attenuated by the mechanical sameness of frequent repetition. We recommend experimental use of women's voices in warning signals, but we do not think it wise at present to incorporate them into important operational systems.

If one could be sure, at the time of design, of homogeneity in rank of the operators of a system, it might be a good idea to incorporate into the voice signals some signs or suggestions of the authority that stems from superior rank. Usually, however, it is dangerous to assume that the operators will be homogeneous. It seems better to keep the authoritarian quality nonspecific, to rely on the intrinsic authoritativeness of the warning-signal subsystem.

The voice should sound, of course, as though it expects to be heeded, and it should inspire confidence to some reasonable degree. If there are many signals, these matters should be handled through the explicit procedure outlined in the section on calmness and urgency. Appropriate names for the present dimensions are 'authoritativeness' (a) and 'reassuringness' (r).

The last six dimensions were treated in three pairs of two because there will be few occasions on which all will be handled formally. If it does seem worthwhile, in any application, to handle them all formally, define weights for all six dimensions and select the version for which the sum of the weighted differences between desired and actual coordinates is least.

PROCESSING OF SPEECH

On the basis of the experience gained during the study of warning signals containing speech, we believe that only a limited amount of processing should be done in preparing the speech signals. There are not many things that can be done to speech waves to make them distinctive or attention-compelling that do not to some extent impair intelligibility or other qualities that we want to retain. Moreover, if the product of processing is a novelty effect, it is likely to become unappealing after repeated presentation. It seems best, in short, to avoid processing the speech parts of warning signals for purposes other than to increase or preserve intelligibility.

The types of speech processing that can make a definite contribution to the effectiveness of warning signals are (1) processing that increases the strength of consonant sounds (which are weak in normal speech) relative to the strengths of vowel sounds and (2) processing that produces subjective movement or other stereo effects.

Compression of the type used in telephonic "companding" circuits produces the desired result to a limited degree. A stronger effect is produced by symmetrical peak clipping. We recommend use of 6 to 8 decibels of peak clipping for warning signals that will be used in intense noise (greater than 105 db re 0.0002 microbar at the listener's eardrums).

Peak clipping. For applications in which weak or moderate noise is expected, it is best to apply the clipping only to the speech parts of the warning signals. The speech waves should be clipped before the speech and nonspeech parts are combined in the preparation of the warning signals. For applications involving intense noise, on the other hand, the clipping (in excess of 6 or 8 db) should be introduced in the playback circuit. The circuit should be so arranged that the clipper limits the extreme voltage swings, both positive and negative, of the over-all warning signal to a level of about 123 db re 0.0002 microbar. That limitation should be fixed, independent of the setting of any volume control. The clipping will then serve to protect the listener's ears against overload, and the amount of clipping will depend upon the level of the warning signal, being great only if the warning signal is played back at such a high level that the speech peaks would be uncomfortably loud if they were not clipped off.

In connection with the latter arrangements, three notes of caution should be made:

(1) It is dangerous to have a manual volume control in a warning signal subsystem. If the control is not turned up during very noisy periods, the value of the subsystem is of course nullified.

(2) Since the intense noise is presumably not expected during all phases of the mission, it is desirable to have the warning-signal volume set high only during those phases in which intense noise may occur. That arrangement, which might involve setting the warning-signal volume control as part of mode switching, would protect the operator against the trauma caused by hearing an unexpectedly loud sound under critical circumstances. In some systems, however, it is possible for intense noise to arise as part of an emergency situation. The problem posed by that possibility cannot be handled by ganging the warning-signal volume control to the mode switches. There seems to be no satisfactory alternative to the noise-operated volume control mentioned earlier in this report.

(3) No strong low-frequency nonspeech signals should be allowed to coincide with the speech signals. (This does not preclude the very small amount of overlapping recommended later.) When strong low-frequency waves and speech waves pass together through a peak clipper, intelligibility is impaired.

Symmetrical peak clipping and the advantages to be gained through use of speech clipping in noisy situations are discussed in reference (4). The speech signal delivered to the clipper should not contain accentuated low-frequency components. The speech spectrum should be either normal or tilted upward very slightly -- not more than 3 db per octave.

Stereo effects. For the production of stereo or dichotic effects, the basic techniques are those mentioned earlier:

1. Artificial introduction of phase variations into one of the two (right ear, left ear) speech channels or, dichotically, into both channels.

2. Artificial introduction of amplitude variations -- particularly, switching the speech from one ear to the other at rates between 4 and 25 cycles per second. (Faster rates of switching do not produce as distinctive effects, and they fail to yield the freedom from masking that is achieved with the lower rates.)

3. Use of two microphones and moving sources of speech and nonspeech signals.

Effects produced by these techniques are illustrated in the magnetic tape recording.

BLENDING OF SPEECH AND NONSPEECH PARTS

It is important to blend together the speech and nonspeech parts of each warning signal to produce a unified, coherent whole. This cannot be done satisfactorily by splicing magnetic tape. It requires either dubbing simultaneously from two tape machines to a third or from one tape machine and a live announcer to a second tape machine.

The second of the latter alternatives has a great advantage if the announcer is allowed to monitor the nonspeech part of the signal and if the nonspeech signal is played back over and over from a tape loop. With a little training, an announcer can achieve very precise control of timing. Ordinarily, there should be enough overlap between the (AG) and (S) parts to carry attention from the first into the second. However, this overlap should not be allowed to degrade the intelligibility of the first word of the speech part.

Dubbing from two tape machines to a third. With an arrangement of equipment similar to that shown in Fig. 3, one can blend together separately recorded speech and nonspeech parts of warning signals. The two tape playbacks are conventional two-channel machines with reels. The slanting pointers in the diagram represent gain controls. The pointers that are connected together by dotted lines are mechanically coupled ("ganged"); one adjustment varies the gains of both the right-ear and left-ear channels.

(AG) nonspeech signals are played back from tape playback 1, (S) speech signals from tape playback 2. Reference marks (e.g., patches of splicing tape affixed to the magnetic tape) are used to mark points just ahead of the beginnings of the segments to be dubbed. The two tapes are positioned and repositioned by trial and error plus iterative correction until the proper time relation between the speech and nonspeech parts is achieved. Then the pair of signals, superposed, is recorded on the tape recorder, in the right-hand side of the diagram.

The oscilloscope and VU meters are for monitoring. They may be connected, of course, to other points in the circuit than the ones indicated in the diagram.

It is desirable to blend the (AG) signal into the (S) signal to create a bond between them. We have found that that decreases the probability of confusion between the speech part of the warning signal and an adventitious, isolated segment of speech occurring at approximately the same time. Blending is readily effected by turning the ganged (AG) gain controls down and the ganged (S) gain controls up simultaneously, just as the (AG) signal is terminating and the (S) signal is coming on. The initial phoneme of the (S) signal should be sustained longer than is normal during the course of speech. That protects it against masking by the end of the (AG) signal. Moderate lengthening of the first speech sound is a technique widely employed, no doubt unconsciously, by persons attempting to break into animated group conversations.

The arrangement of Fig. 4 is the same as that of Fig. 3 except for the substitution of the tape loop for reel tape playback 1 and the live talker and microphones for reel tape playback 2. The talker listens to the (AG) signal as it is repeatedly played back from the loop. He utters the (S) signal with what seems to him to be natural timing and adjusts his speech reiteratively until he achieves the desired result. The talker should use the VU meters in monitoring his speech.

REFERENCES

1. Erlick, D. E. and Hunt, D. P. Evaluating audio warning displays for weapon systems. Technical Report 57-22, Wright Air Development Division, Wright-Patterson Air Force Base, Ohio, April 1957.

2. Green, D. M., McKey, M. J. and Licklider, J.C.R. Detection of a pulsed sinusoid in noise as a function of frequency. J. Acoust. Soc. Am., 31, 1446-1452, 1959.

3. Green, D. M., Birdsall, T. G. and Tanner, W. P., Jr. Signal detection as a function of signal intensity and duration. J. Acoust. Soc. Am., 29, 523-531, 1957.

4. Licklider, J.C.R. Effects of amplitude distortion upon the intelligibility of speech. J. Acoust. Soc. Am., 18, 429-434, 1946.

5. Price, H. E., and Older, H. J. Auditory signals in Air Force weapon systems and equipment. Tech. Report No. 56-53, Operational Applications Laboratory, Air Force Cambridge Research Center, Bolling Air Force Base, D. C., June 1956 (CONFIDENTIAL).

6. Tanner, W. P., Jr. and Birdsall, T. G. Definitions of d' and η as psychophysical measures. Tech. Report 80, Electronic Defense Group, Engineering Research Institute, Univ. of Michigan, Ann Arbor, Michigan, February 1958.

7. Zwicker, E., Flottorp, G., and Stevens, S. S. Critical band width in loudness summation. J. Acoust. Soc. Am., 29, 548-557, 1957, Table I.

Table 1

Transformation from Goals to Preliminary Signal Specifications

		Goals											
		1	2	3	4	5	6	7	8	9	10		Σr
	Manipulatable Characteristics of Signals	Audible	Rapid-acting	Alerting	Discriminable	Informative	Compatible	Nonmasking	Nondistracting	Nonpainful	Nondamaging		
1	Sound pressure level	5	3	4	3	3	1	-4	-3	-4	-4		
2	Center frequency	-3	0	0	0	0	0	3	-1	-2	-2		
3	Amount of variation in center frequency	-2	0	3	2	2	0	0	-2	1	2		
4	Spread in frequency	-5	1	1	2	3	0	-2	3	3	3		
5	Duration	2	-3	3	3	5	0	-3	-2	-2	-3		
6	Emphasis of initial segment	0	5	0	0	0	0	0	-2	-1	-1		
7	Emphasis of low frequencies	2	-1	-1	0	0	0	-2	1	2	2		
8	Dwell time	4	2	3	0	2	0	-2	-1	-2	-3		
9A*	Speech-nonspeech	-3	-3	3	-3	10	-3	-2	-2	-2	-1		
9B**	Speech-nonspeech	-3	-3	3	8	10	-3	-2	-2	-2	-1		
10	Diotic-dichotic	1	2	7	4	1	-3	8	-5	0	0		
11	Established associations	0	3	3	0	3	1	0	-2	0	0		
12	Affectiveness	0	3	3	0	0	0	0	-3	0	0		
13	Repetitiveness	0	0	2	0	0	0	0	1	0	0		
14	Rhythm	0	1	2	0	0	0	0	-1	0	0		

* Use if there are only a few signals in the ensemble, i.e., a few exigencies to be distinguished among.

** Use if there are many signals, many exigencies.

Table 2

Summary Table for Results from Table 1

	Manipulatable Characteristics of Signals		Warning signals or exigencies										
			1	2	3	4	5	6	7	8	9	10	D_{ij}
1	Sound pressure level	d'*											
2	Center frequency	f_c											
3	Amount of variation of center frequency	σ_c											
4	Spread in frequency	σ_f											
5	Duration	Δt											
6	Emphasis of initial segment	c											
7	Emphasis of low frequencies	k											
8	Dwell time	τ											
9A	Speech-nonspeech	ϕ											
9B	Speech-nonspeech	ϕ											
10	Diotic-dichotic	∂											
11	Established associations	ϵ											
12	Affectiveness	α											
13	Repetitiveness	π											
14	Rhythm	ρ											

* Actually, d' is the symbol for an intervening variable dependent upon sound pressure level.

Table 3

Boundary and Center Frequencies of The Auditory Critical Bands and Associated Values of The Signal-to-Noise Ratio (S/N)' Required to Yield the Detectability Index d' = 1

Band	Boundary	Center Frequency	(S/N)'**	Band	Boundary	Center Frequency	(S/N)'
	20			12		1580	112
1		65	316*		1700		
	110			13		1835	126
2		155	79*		1970		
	200			14		2130	145
3		250	60		2290		
	295			15		2480	170
4		345	63		2670		
	395			16		2900	204
5		450	66		3120		
	503			17		3400	257
6		560	69		3680		
	625			18		4020	339
7		690	74		4360		
	755			19		4780	490*
8		830	79		5200		
	900			20		5700	740*
9		980	85		6200		
	1060			21		6850	1260*
10		1155	91		7500		
	1250						
11		1355	100				
	1460						

---

* Estimated

** (S/N)' is the ratio of signal power to noise power density required to make d' = 1. If the signal duration Δt is between 0.5 and 1.0 second, divide the values of (S/N)' given in the table by $(\Delta t)^{1/2}$. If the $\Delta t < 0.5$ second, divide the tabled values by $(1.4)\Delta t$.

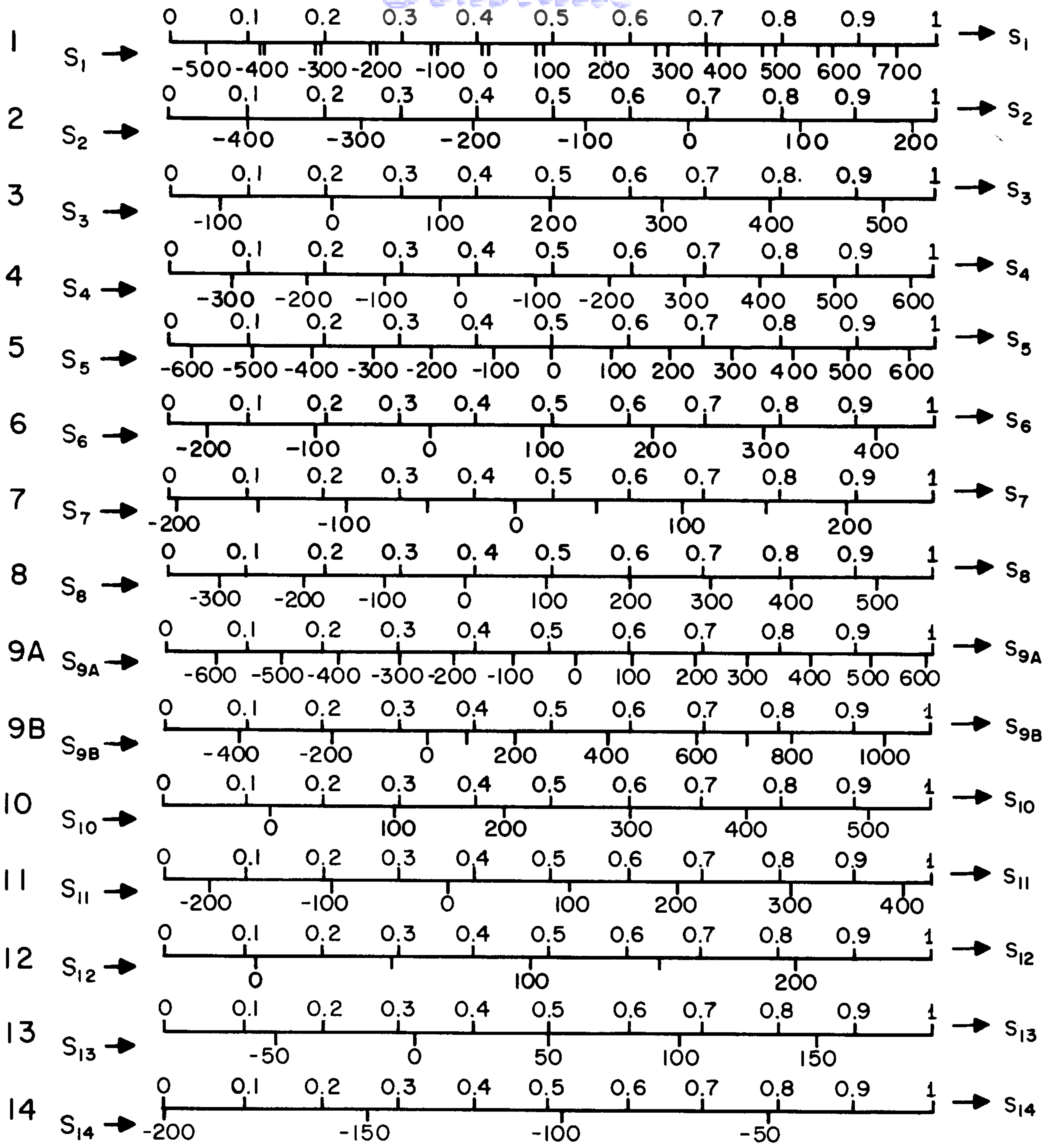

FIG. 1 LINE CHARTS FOR NORMALIZATION OF SUMS IN THE PROCEDURE INVOLVING TABLE 1

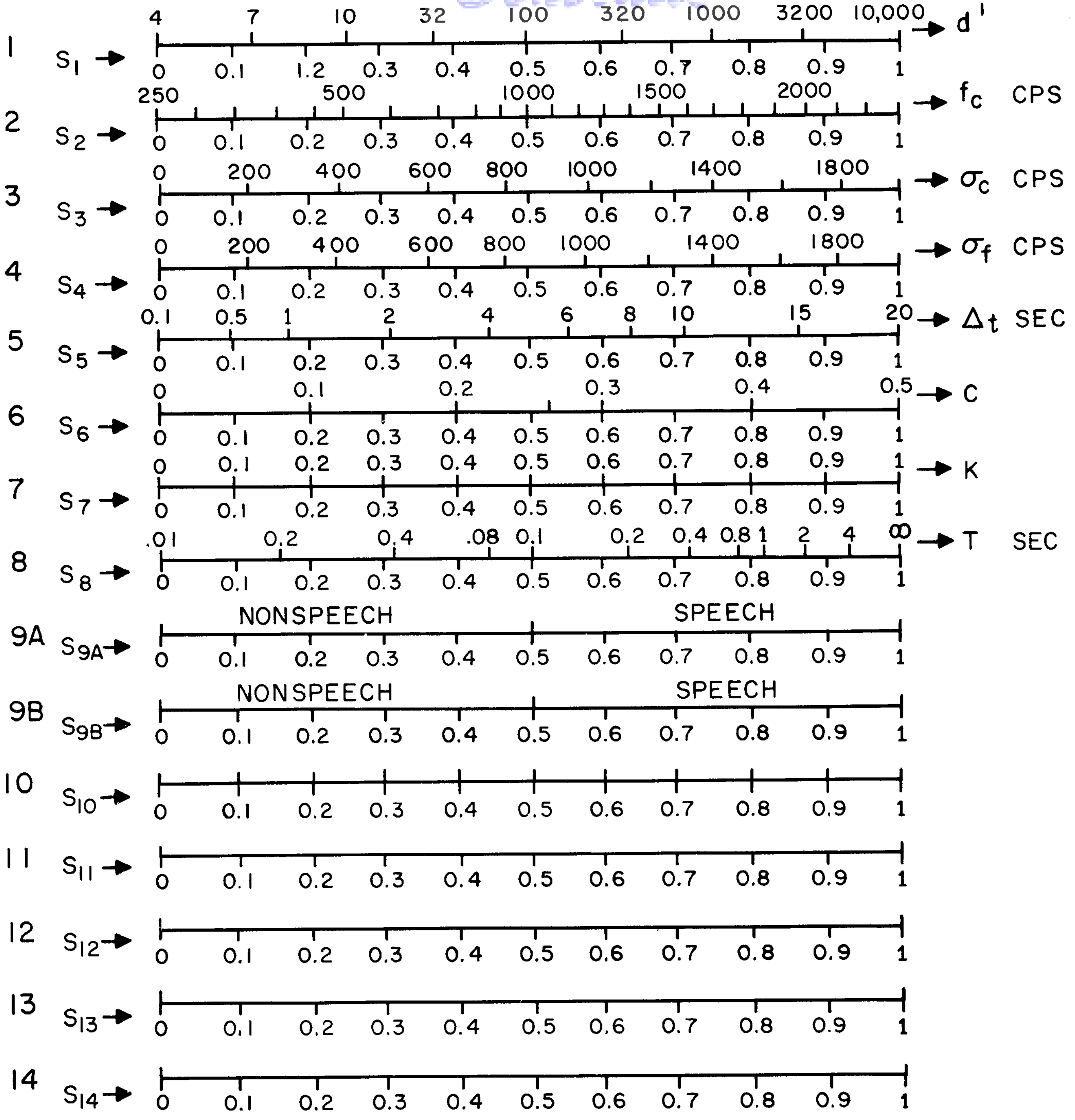

FIG. 2 LINE CHARTS FOR TRANSFORMING FROM NORMALIZED DIMENSIONS TO PHYSICAL SPECIFICATIONS

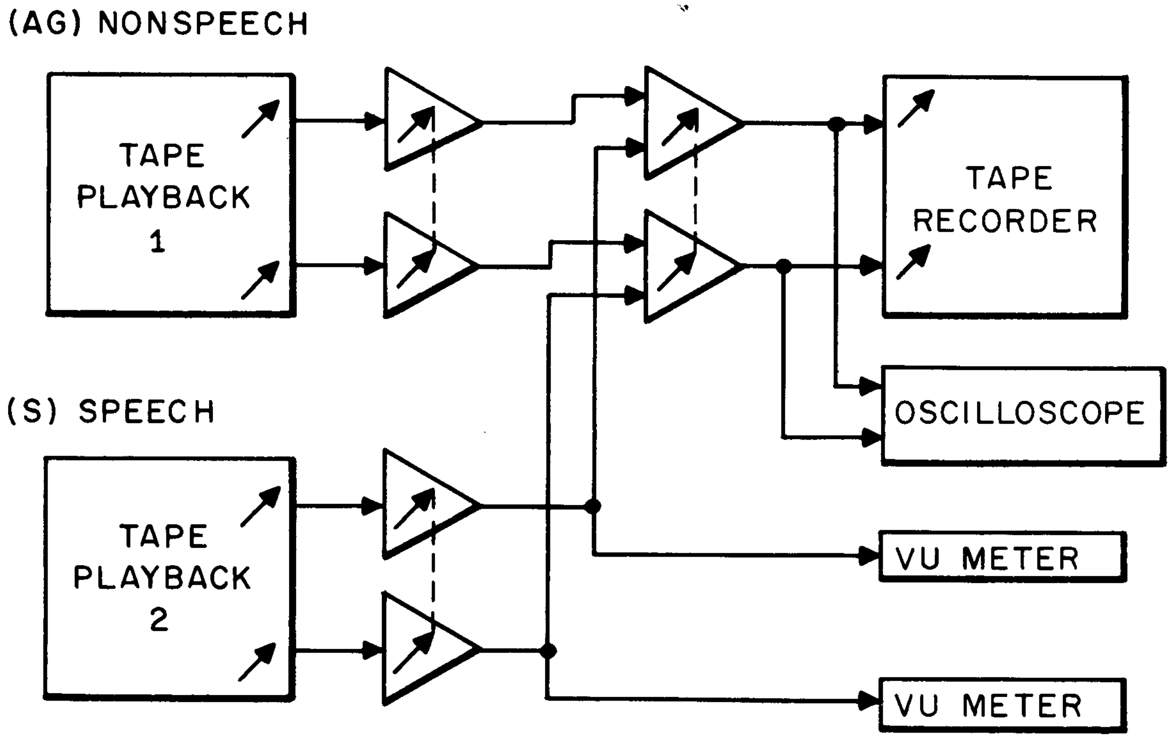

FIG. 3 EQUIPMENT SETUP FOR SUPERPOSING RECORDED SPEECH AND NONSPEECH PARTS OF WARNING SIGNALS

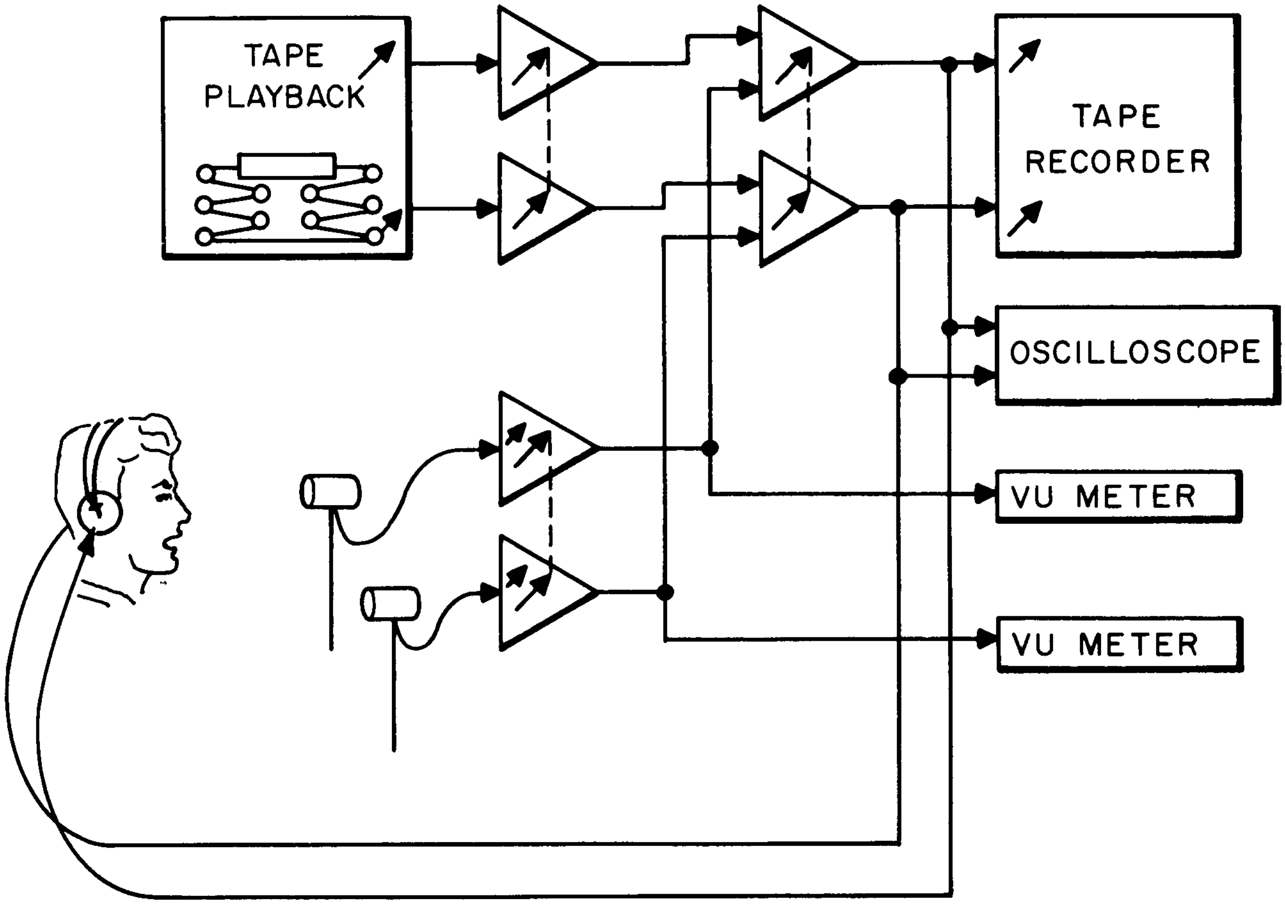

FIG. 4 EQUIPMENT SETUP FOR SUPERPOSING "LIVE" SPEECH AND RECORDED NONSPEECH PARTS OF WARNING SIGNALS

LIST OF SYMBOLS

A	alerting (abbreviation)
C	emphasis of initial segment (coefficient); center (subscript)
c	calmness (subscript); center (subscript)
d'	index of detectability
e	constant, 2.7183
f	frequency (subscript); formality (subscript)
G	general category (abbreviation)
g	goal or goal attribute
i	index (subscript)
j	index (subscript)
K	emphasis of low frequencies (coefficient)
k	index (subscript)
P	power assumed for warning signal at beginning of calculation; variable related to automatic correction
P'	power (calculated) to which warning signal should be adjusted
p	probability
R	inverse measure of the requirement for a warning signal
r	rank of exigency in which need for warning signal or of desideratum in order of importance; reassuringness (subscript)
S	inverse measure of the appropriateness of audio warning; special category (abbreviation)
S/N	signal-to-noise ratio
t	rank of exigency in need for an audio warning signal; time
Δt	duration
u	urgency
w	weight
x	final priority in line for audio warning signal; number of dimensions
z	obtained value of speech attribute
z'	desired value of speech attribute
α	affectiveness (coefficient); "is proportional to"

LIST OF SYMBOLS (Cont.)

Γ	damping (coefficient)
ϵ	established associations (coefficient)
ϕ	speech-nonspeech (coefficient)
π	repetitiveness (coefficient)
ρ	rhythm (coefficient)
Σ	sum of $100/r_{di}$ over row i
Σ'	normalized value of Σ
σ	spread; standard derivation; variation
τ	dwell time
∂	"dichoticness," the intensity of variables of the stereo effect